Using a Raster Display for Photometric Stereo

Nathan Funk
Singular Systems
Edmonton, Canada
nathan.funk@singularsys.com

Yee-Hong Yang
Computing Science
University of Alberta
Edmonton, Canada
yang@cs.ualberta.ca

Abstract

This paper presents a new controlled lighting apparatus which uses a raster display device as a light source. The setup has the advantage over other alternatives in that it is relatively inexpensive and uses commonly available components. The apparatus is studied through application to shape recovery using photometric stereo. Experiments on synthetic and real images demonstrate how the depth map of an object can be recovered using only a camera and a computer monitor.

1. Introduction

The ability to control the lighting of a scene to known conditions is generally beneficial for shape recovery. Without any knowledge of the lighting, it is challenging to determine other properties of the scene such as the geometry of objects or their surface properties. Lighting estimation methods have been proposed to counteract this problem, but they often rely on limiting assumptions about the scene.

Controlled lighting can be accomplished in many ways. One way is by positioning a large set of light sources in a manner such that a wide range of different lighting conditions are achievable [6]. Or one can use a smaller set of light sources and devise a method of accurately moving them to desired locations [8]. Two key factors are the *accuracy* of the positioning and the *ease of control* of the sources. This paper presents a method which meets these requirements through the use of a raster display device. Such devices include CRT monitors, LCD screens, and LCD projectors in conjunction with a projection screen. All these devices provide a dense grid of accurately positioned light emitting cells. They can be easily controlled with the use of a computer. The wide availability of raster display devices is also a definite advantage over constructing a rig specialised for controlled lighting. At least two published works [13, 17] have used raster display devices as illumination devices in the past, however not in the same manner presented in this paper.

The constructed apparatus could be used for a range of applications in computer vision. This paper however solely discusses the application of the setup for shape recovery using photometric stereo [16]. The display is used to generate a set of six different lighting configurations while a camera captures images of the scene for each configuration. The first published proposal for the use of displays for shape recovery was made by Clark in a recent paper [4]. The work however only offers a theoretical analysis without an implementation. This paper uses a slightly different approach which can be practically implemented with common display devices.

The traditional photometric stereo approach was implemented [16] due to its simplicity in comparison to more recent approaches [1, 2, 9, 10, 11, 14, 15]. It is based on a Lambertian reflectance model and assumes that the light sources are distant point sources. The Lambertian reflectance assumption is kept in this paper, yet modifications are made to handle near point sources. Perspective projection is also considered. So in comparison to other implementations which assume orthographic projection [10, 11], a long focal length lens is not required.

In Woodham's approach [16], the surface normal $\hat{n}$ and albedo ρ are determined locally from at least three surface radiance measurements R_i and the corresponding light source vectors $\vec{L_i}$. The magnitudes $|\vec{L_i}|$ correspond to the local surface irradiance. Lambertian reflection at non-shadowed locations can be expressed as a dot product of the light vector with the scaled normal vector $\vec{N} = \rho\hat{n}$. This allows multiple radiance values to be expressed through matrix multiplication of the combined light vectors $\mathbf{L}$ with the scaled normal vector, resulting in $\vec{R} = \mathbf{L}\vec{N}$.

There is a unique solution $\vec{N} = \mathbf{L}^{-1}\vec{R}$ when $\mathbf{L}$ is invertible. If more than three light sources are used, the system may become over-constrained, making a least-squares solution applicable. Finally, the albedo and surface normal can be extracted as $\rho = |\vec{N}|$ and $\hat{n} = \vec{N}/\rho$.

2. Performing Photometric Stereo with a Raster Display

The setup we propose consists of three main components: the *raster display device* which serves as the light source, the *scene* which reflects the light from the display and a *camera* to capture images of the scene. We assume that the light entering the camera was reflected only once from the scene after being emitted by the display. Our model for the light emission from a raster display device considers the effects caused by the near placement of the display from the scene. It also considers the dependence of the display brightness on the viewing direction. We refer to this effect, which is especially significant for LCD monitors, as *directionality* of the display.

2.1. Model

We propose an empirical model which defines the irradiance at a point in the scene from a point on the display. This model is used to determine the magnitude of the light source vectors for photometric stereo. We consider three factors influencing the irradiance of the scene: (1) the unattenuated radiance emitted from a white pixel (2) the effect of directionality on the emitted radiance and (3) the inverse square law as it applies to point sources.

Modeling the directionality of displays is necessary to accurately determine the radiance from the display. For example, comparing the appearance of a CRT monitor, LCD monitor and a projector screen from different angles, it is noticeable that the LCD brightness varies greatly with the viewing direction. A dependence on the viewing direction is also present for CRT monitors and projector screens although it is not as noticeable.

Since the screen radiance at each pixel is assumed to vary with the direction, it is modeled as a function over a hemisphere $R_P(\theta, \phi)$ and is expressed as

$$R_P(\theta, \phi) = R_U f(\theta, \phi), \tag{1}$$

where R_U is the unattenuated radiance and $f(\theta, \phi)$ is the *directionality function*. The directionality function is determined through a calibration procedure discussed in Section 3.4.

By considering the inverse square law, the irradiance on a scene point from a single pixel can be modeled as

$$I_S = \frac{R_P(\theta, \phi)}{|\vec{P}_i - \vec{S}|^2} = \frac{R_U f(\theta, \phi)}{|\vec{P}_i - \vec{S}|^2}, \tag{2}$$

where S is a scene point and P_C is the center of the pixel. The angles θ and ϕ depend on the location of the scene point and can be calculated from the screen orientation.

2.2. Photometric Stereo

Traditional photometric stereo [16] assumes that the light source illuminating each image is an infinitely distant point light source. With common raster displays, it is not possible to achieve the same light field as a distant point source. So instead, a close approximation is used. Point sources are simulated by displaying small white squares on the screen. A single pixel most closely approximates a finite point source, however it is not bright enough to allow accurate measurements of its effects on the image. For this reason, a group of pixels are required to act as a single light source. Increasing the size of the square acting as light source increases the accuracy of the measured image irradiance contribution, but it also further deviates from the assumption that the light source is a distant point source.

To recover the surface normal, recall $\vec{R} = \mathbf{L}\vec{N}$ where the scene radiance is expressed through matrix multiplication. Here the values in $\vec{R}$ are the surface radiance values inferred from the captured images, $\mathbf{L}$ is composed of light vectors $\vec{L}$ as rows, and $\vec{N}$ is the scaled surface normal to be recovered. The magnitude of each $\vec{L}$ vector is the scene irradiance from the associated light source. This irradiance is the sum of all the pixel irradiance contributions from the square light source. We make the assumption that they can be grouped together as one light source, where $\vec{L}$ is directed at the center of the square as an approximation of the combined effect of all pixels.

The magnitude of each $\vec{L}$ is calculated as the I_S contribution from a point at the center of the square region. Here it should be noted that only the relative irradiance values are of interest. So the number of pixels in the region and the constant R_U can be disregarded. With these considerations the relative irradiance measure can be written as

$$|\vec{L}_i| = \frac{f(\theta, \phi)}{|\vec{P}_i - \vec{S}|^2}, \tag{3}$$

where i indexes the light sources. Then, each light vector $\vec{L}_i$ can be calculated with

$$\vec{L}_i = f(\theta, \phi)\frac{\vec{P}_i - \vec{S}}{|\vec{P}_i - \vec{S}|^3}. \tag{4}$$

Since the scene points are unknown, the exact light vectors can not initially be obtained. We adopt a simple method for estimating the light vectors by making an *initial estimate* of the position of $\vec{S}$. This causes the accuracy of the normals to depend on the accuracy of the position estimate. All points are initially assumed to lie on a plane perpendicular to the camera's optical axis. The *depth estimate* parameter determines the distance of this plane from the camera's focal point. Then, for each point, the light source vectors are calculated using equation 4. With the light source vectors

known, the surface normals are determined by applying a least-squares approach at every pixel location.

To improve the accuracy in the calculations of the surface normals, an iterative approach is adopted. After obtaining the depth map from the initial calculation of the normals, this depth map is used in the calculation of the light vectors for a second iteration rather than assuming that all scene points lie on a plane. This process is repeated until the depth map converges.

Shadows are handled by thresholding the surface radiance values and omitting light sources that cause a radiance below the selected threshold value. A *shadow threshold* parameter sets the pixel value below which a radiance measurement is discarded. For each pixel, the number of radiance measurements above the threshold is counted. If the number is more than or equal to three, the photometric stereo algorithm is applied. Otherwise the normal is not recovered for the pixel in question.

2.3. Depth from Surface Normals

We employ a method for determining the depth from surface normals which considers the perspective projection of the camera and also handles depth discontinuities. In our experiments it achieves better results and converges faster than Horn's method [12], and also provides better results than Frankot & Chellappa's technique [7].

Our method defines constraints relating the depth of adjacent pixels based on the direction of their normals. Each constraint is expressed as a row in a large sparse linear system which is solved with sparse matrix methods. Two rays passing through pixels adjacent in the x-direction intersect the scene at points $S_1 = (x_1, y, z_1)$ and $S_2 = (x_2, y, z_2)$. The rays are defined by the normalized coordinates (x_{n1}, y_{n1}) and (x_{n2}, y_{n2}) which constrain points on the rays such that $x_1 = x_{n1}z_1$ and $x_2 = x_{n2}z_2$. By assuming that the slope of the line connecting S_1 and S_2 is known, a relationship between z_1 and z_2 can be established. The slope is

$$\frac{z_2 - z_1}{x_2 - x_1} = \frac{-\hat{n}_{ax}}{\hat{n}_{az}}, \tag{5}$$

where $\hat{n}_{az}$ and $\hat{n}_{ax}$ are the components of the normal $\hat{n}_a$ to the line passing through S_1 and S_2. The true normal is unknown, but it can be approximated by averaging the surface normals at the intersection points. Using $\hat{n}_a \approx (\hat{n}_1 + \hat{n}_2)/|\hat{n}_1 + \hat{n}_2|$ as an approximation, equation 5 can be transformed to constrain z_1 and z_2 with the linear equation

$$(-n_{az} - n_{ax}x_{n1})z_1 + (n_{az} + n_{ax}x_{n2})z_2 = 0. \tag{6}$$

Naturally, an equivalent relationship holds for the y-direction. For pixels adjacent in the y-direction, the constraint equation becomes $(-n_{az} - n_{ay}y_{nj})z_1 + (n_{az} + n_{ay}y_{nk})z_2 = 0$.

A linear system of equations is constructed by listing the equations associated with all adjacent pixel pairs present in the image. For each equation, the z coefficients are entered in a large sparse coefficient matrix $\mathbf{H}$, such that the system can be expressed in matrix notation as the homogeneous system $\mathbf{H}\vec{z} = \vec{0}$ where $\vec{z}$ is a column vector containing all depth values.

Depth discontinuities are considered by removing constraints from the system corresponding to rapid changes in the normal map. This results in a segmentation of the image into multiple depth-independent regions. Finally, one z value per independent region is constrained to the depth estimate value before determining the least-squares solution for $\vec{z}$.

3. Experimental Setup

Figure 1 shows the experimental setup. The camera used is a Canon EOS 300D, with an EF-S 18-55mm zoom lens. Experiments were performed with a CRT monitor (19" Sony Multiscan G400) and an LCD display (17" NEC Multisync LCD 1760NX) which is shown in Figure 1.

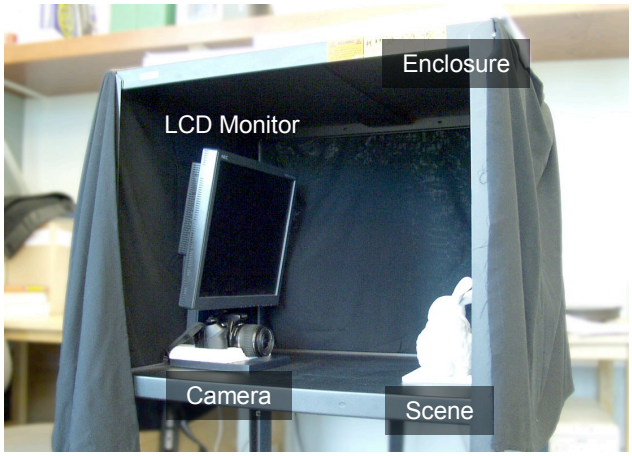

Figure 1. The experimental setup. Visible in this picture are the enclosure, the screen, the camera (below the screen) and the captured object.

The apparatus is enclosed by black cloth to reduce the amount of ambient light entering the scene. This is beneficial for the accuracy of the results although it is not required. We compensate for the environmental lighting and the radiance from the screen when it is set to black by subtracting a *dark image* from all other images. This dark image is captured with an entirely black screen. The enclosure

ensures that during the capture process the environmental lighting does not change significantly since otherwise the results would be affected.

The capture parameters depend primarily on the brightness of the display and the sensitivity of the camera. Images were exposed at the maximum exposure time of 30 seconds. Acceptable noise levels could also be achieved with exposure times around 10 seconds. The aperture was opened to its maximum. All images were downsampled by a factor ranging from 5 to 10. These parameters result in nearly unnoticeable image noise in the downsampled images.

3.1. Images Displayed on Screen

A number of factors were considered for deciding on the configuration of the images shown on the display. A square size of 50x50 pixels was selected as it was the minimum size that provided sufficient light. With a point source, the source is either fully visible or occluded from a scene point. By using an area on the display as a light source it is noteworthy that the light sources can be partially occluded through self-occlusion. Considering the partial occlusion of sources would add significant complexity to the shape recovery method. It is not explicitly considered in the method presented and should therefore be noted as a source of error in the results.

The implementation uses a total of 6 squares as light sources. The number of sources and their positions were chosen while considering several factors. One consideration is that photometric stereo requires at least three sources contributing irradiance at a point. Also, to maximize the number of surface normals recoverable by photometric stereo, the number of points on the object lit by at least three sources must be maximized. Another aspect to consider is that to recover the surface normal at points lit by only three sources, the light vectors must be non-coplanar. So any three sources on the screen must be non-collinear for photometric stereo to be performed using only their emitted light. Also, increasing the angular separation between sources increases the numerical stability of the method if the angle is below 90°. And finally, decreasing the number of sources decreases the total capture time.

Figure 2 shows an example set of images captured under associated lighting conditions. One can note that the first two images are darker than the others. This is an effect of both the inverse square law and the directionality of the screen.

3.2. Camera Calibration

The photometric stereo implementation and calibration procedures for screen position and directionality require that the camera is calibrated geometrically and radiometrically. Geometric calibration of the camera is performed using Bouguet's MATLAB toolbox [3] to obtain the intrinsic camera parameters. The HDR shop software [5] was used to ensure a linear relationship between the sensor irradiance and the pixel values of the raw images.

3.3. Screen Position Calibration

Our screen position calibration method calculates a 4×4 matrix that relates the pixel coordinates to 3D coordinates in the camera reference frame. If the screen was directly visible from the camera, the calibration could be accomplished by displaying a calibration pattern on the screen at a known position. But since the screen is not visible, we use a mirror to reflect the image of the screen into the camera's view. One calibration pattern is attached to the surface of the mirror and a second pattern is displayed on the screen. We use a first surface mirror rather than a standard back silvered mirror. Since it only reflects the light on a single plane, a clearer reflection of the screen is obtained allowing a more accurate calibration. An example of a calibration image is provided in Figure 3.

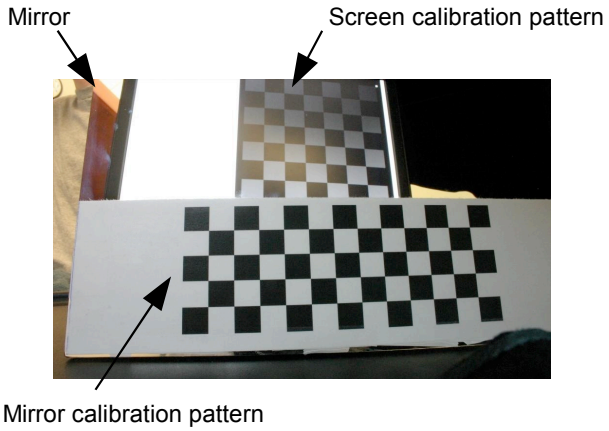

Figure 3. A screen position calibration image. With the use of this image, the screen position can be determined with high accuracy. Note that the mirror fills nearly the entire image except the left and bottom edges.

3.4. Screen directionality calibration

This calibration step aims to determine the dependency of the display radiance on the viewing direction. In Section 2.1 this dependency is modeled as the directionality function over a sphere $f(\phi, \theta)$. A basic assumption made is that the directionality function is spatial invariant on the

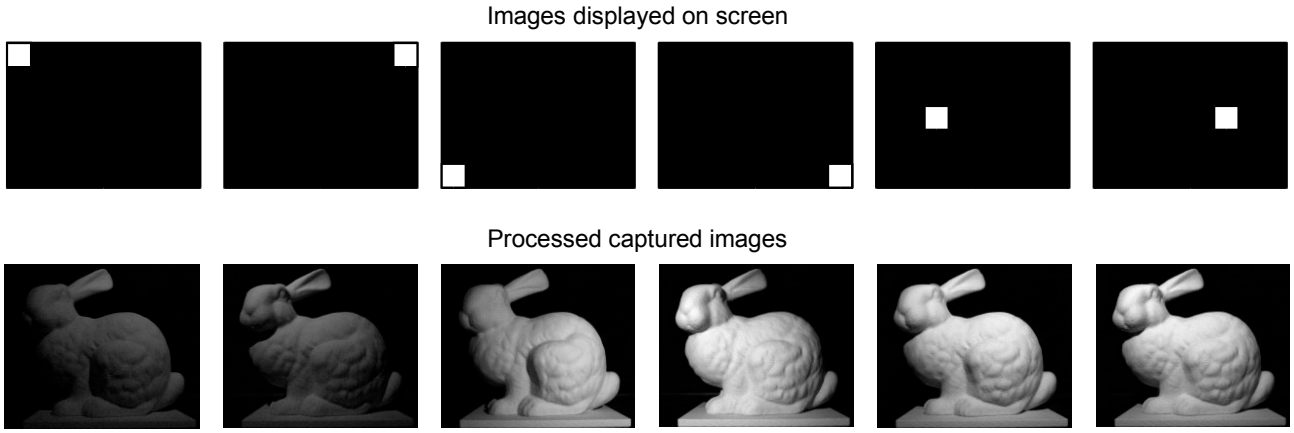

Figure 2. An example of the images displayed on the screen with their associated processed images captured by the camera. The size of the light sources is exaggerated for visibility.

display. The calibration procedure also assumes that the directionality function varies only with the angle from the screen normal ϕ (the colatitude). This assumption introduces some error into the results since there is in fact some variation of the radiance along the longitude as well. Under

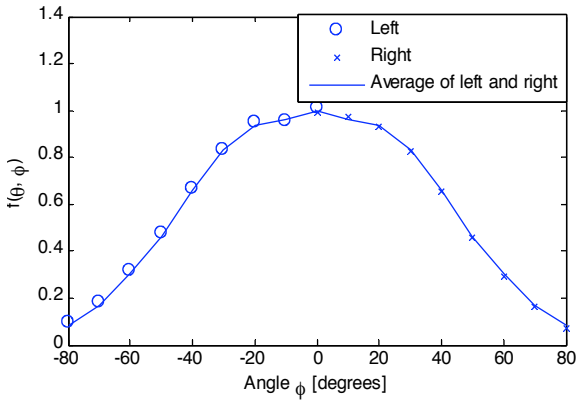

Figure 4. The directionality function for the calibrated LCD screen. Since the radiance from the screen drops considerably with increasing viewing angle, it is important to consider this property in the analysis. The left and right measurements are nearly symetric, so an average can be used for the function.

these assumptions the radiance can be measured by taking images of the screen from different angles. To accomplish this, the camera is directed at the center of the screen, which is placed on a turntable. The axis of the turntable is in the same plane as the screen, and passes through the center of the screen. A set of images is taken with the camera in full manual mode, to ensure that each of the images is exposed equally. The screen is rotated such that one image is taken at each 10° increment from $\phi = -80^\circ$ to $\phi = 80^\circ$. An average is calculated from the measurements on the left and right sides. Linear interpolation is used between the sampled orientations. We only perform a calibration for LCD monitor while the CRT is assumed to radiate equally in all directions. Figure 4 shows the calibration results for the LCD monitor.

4. Experiments

We present a quantitative evaluation of the shape recovery on synthetic and real images. The synthetic images are rendered using the intrinsic camera parameters and the calibrated screen directionality function of the LCD screen.

Two objects are examined for which both a digital model as well as a physical object were available. The first object is a sphere. In synthetic experiments a 14mm diameter model was used while the real experiments were performed using a 40mm table tennis ball painted with flat white spray paint. The second object is the Stanford bunny model which was scaled to a size of approximately 180mm from the nose to the tail. A physical model with a waxed plaster surface was printed using a 3D printer. The surfaces of both physical objects are nearly diffuse reflecting with no visible specular highlights.

To analyze the performance of the photometric stereo

method, the intermediate results of the normal map are evaluated as well as the final depth map. For the synthetic images, a 3D model is available with true orientations of surface normals as well as the true depth map. To allow a ground truth comparison for the real images, the 3D model is aligned to the images by manually selecting feature points on the object. Using the intrinsic camera parameters, the pose of the physical object is matched allowing comparison of the normals and depth to their true values. While the accuracy is limited by factors such as the accuracy of the camera calibration and manual feature point selection, the error introduced is expected to be minor in comparison to the errors in the estimated normal and depth results.

The evaluation is performed by analysing the accuracy of the surface normals and depth using two error measures. The *average normal error* is calculated as the average angle between the recovered and the true normals. It is measured in degrees. The *average absolute depth error*, measured in millimeters, is calculated by averaging the absolute difference between the recovered and true depth at each pixel. The error measures are calculated over the region of the image for which surface normals were recovered.

4.1. Synthetic Experiments

A sphere with a radius of 7mm, centered at the point $(0, 0, 300)$ is rendered to an image of size 151×151. The depth estimate corresponds to the true depth at the center pixel (293mm). The estimated depth map is scaled such that the center depth equals the depth estimate. The depth map converges within 0.001mm after four iterations, with an average normal error of 0.003 degrees and an average absolute depth error of 0.011mm.

The Stanford bunny model was rendered with an image size of 168×200. The depth estimate is set to 283mm corresponding to the depth of the center point. Similar to the synthetic sphere experiment, the accuracy of the results is improved by repeating the photometric stereo step after performing the initial shape recovery. After four iterations an average normal error of 0.42 degrees and an average absolute depth error of 0.9mm is achieved.

4.2. Real Experiments

This section presents the results for the sphere and Stanford bunny model using CRT and LCD monitors. For all experiments the depth estimate is set to the true depth of the center pixel. The shadow threshold is adjusted to minimize the error in the recovered surface normals. Only a single iteration is performed since subsequent iterations do not improve the result. A summary of the results is presented in Table 1.

Table 1. Results of the real experiments

	Image width (pixels)	Avg. normal error (degrees)	Avg. abs. depth error (mm)
Sphere / CRT	100	10	2
Sphere / LCD	100	21	5
Bunny / CRT	200	16	13
Bunny / LCD	200	21	10

Figure 5 shows the detailed results of the CRT experiments. The re-lit images show that for some regions such as the front part of the bunny's neck the surface normals could not be recovered due to shadowing. One can also note minor discrepancies in the albedo compared to the true albedo which is uniform. A slight distortion of the depth values is noticeable for both the sphere and Stanford bunny models.

5. Conclusions and Future Work

We have presented an apparatus together with a photometric stereo method for performing shape recovery using a raster display device. Our results show the potential for using display devices for controlled lighting to assist in computer vision tasks such as shape recovery.

The accuracy of the results is expected to be similar for other objects. The main source of error is expected to be the surface BRDF which, against our assumption, is likely not perfectly diffuse. The lower error in the normals recovered using the CRT versus using the LCD screen are best explained by the complexity added by the strong directionality of the LCD screen. Differences between the calibrated directionality function and the true directionality are likely the cause for the higher error values.

Future work is expected to increase the accuracy of recovery and expand the range of applicable surface types. Both of these goals might be achieved by adopting a newer photometric stereo method for handling non-Lambertian reflection [1, 9, 10, 11, 14, 15]. Another promising idea is the display of more sophisticated patterns on the screen together with a photometric stereo method which considers more complex lighting conditions [2, 9]. Increasing the amount of light from the screen would allow shorter exposure times and possibly remove the necessity for an enclosure.

We believe that our work is a significant step towards low-cost shape recovery and an increased use of display devices for computer vision techniques. For example, this technique may assist in face recognition at bank ATMs or provide easy 3D shape recovery for computer users.

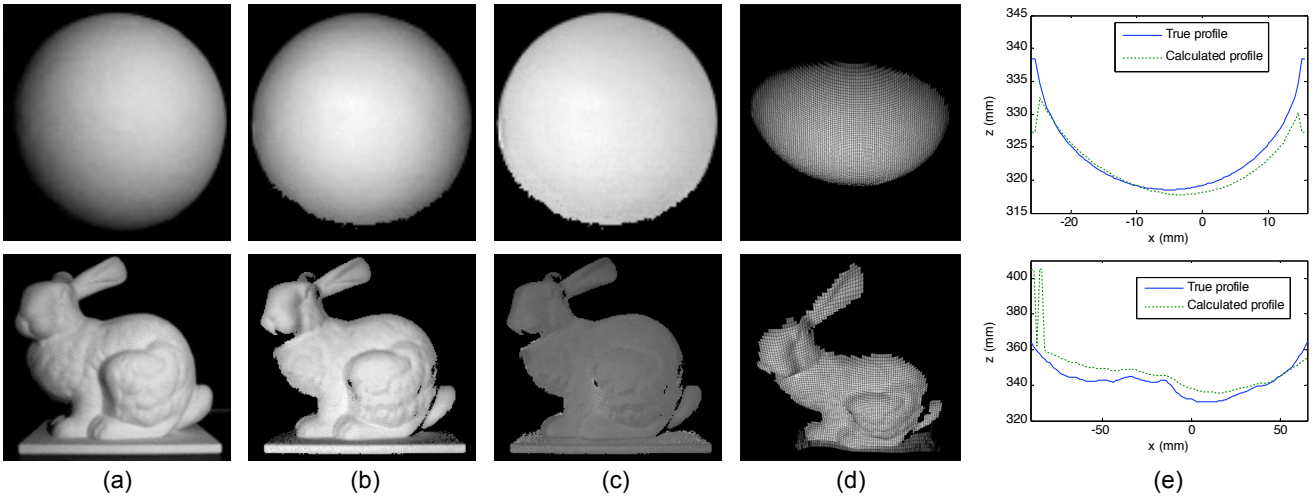

Figure 5. The real sphere and Stanford bunny results of experiments using the CRT monitor. (a) Example captured images, (b) the recovered normals lit by a distant light source from (0,0,-1), (c) the normalized recovered albedo, (d) a novel view of the re-lit recovered depth maps, and (e) a comparison of the true and recovered profiles along the vertical center.

6. Acknowledgements

We would like to express our sincere thanks for the financial support provided by NSERC and iCORE. We are also immensely grateful for the support from friends, faculty and staff at the University of Alberta. Finally, we would like to thank the anonymous reviewers for their helpful comments.

References

[1] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *PAMI*, 25(10):1239–1252, 2003.

[2] R. Basri and D. Jacobs. Photometric stereo with general, unknown lighting. In *CVPR*, volume 2, pages 374–381, 2001.

[3] J.-Y. Bouguet. MATLAB camera calibration toolbox. http://www.vision.caltech.edu/bouguetj/calib_doc/. Viewed on May 26, 2005.

[4] J. Clark. Photometric stereo with nearby planar distributed illuminants. In *CRV*, page 16. IEEE Computer Society, 2006.

[5] P. Debevec. HDR Shop. http://www.hdrshop.com/. Viewed June 8, 2004.

[6] P. Debevec, A. Wenger, C. Tchou, A. Gardner, J. Waese, and T. Hawkins. A lighting reproduction approach to live-action compositing. In *Proc. SIGGRAPH*, pages 547–556. ACM Press, 2002.

[7] R. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *PAMI*, 10:439–451, 1988.

[8] R. Furukawa, H. Kawasaki, K. Ikeuchi, and M. Sakauchi. Appearance based object modeling using texture database: acquisition, compression and rendering. In *Proc. Eurographics workshop on Rendering*, pages 257–266, Aire-la-Ville, Switzerland, 2002.

[9] A. Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *ICCV*, pages 816–825, 2003.

[10] D. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *ICCV*, volume 1, pages 341–348, 2005.

[11] A. Hertzmann and S. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *PAMI*, 27(8):1254–1264, August 2005.

[12] B. Horn. *Robot vision*. MIT Press, 1986.

[13] Y. Schechner, S. Nayar, and P. Belhumeur. A Theory of Multiplexed Illumination. In *ICCV*, volume 2, pages 808–815, Oct 2003.

[14] F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *PAMI*, 18(4):449–454, 1996.

[15] H. Tagare and R. de Figueiredo. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *PAMI*, 13(2):133–152, February 1991.

[16] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, January 1980.

[17] D. Zongker, D. Werner, B. Curless, and D. Salesin. Environment matting and compositing. In *Proc. SIGGRAPH*, pages 205–214. ACM Press/Addison-Wesley Publishing Co., 1999.