Acquiring Input for Rendering at Appropriate Levels of Detail: Digitizing a Pietà

Holly Rushmeier, Fausto Bernardini, Joshua Mittleman and Gabriel Taubin

IBM TJ Watson Research Center, Yorktown Heights, NY 10598 USA

Abstract: We describe the design of a system to augment a light striping camera for three dimensional scanning with a photometric system to capture bump maps and approximate reflectances. In contrast with scanning an object with very high spatial resolution, this allows the relatively efficient and inexpensive acquistion of input for high quality rendering. This system is being used in a project to digitize a Michelangelo Pietà in Florence, Italy.

1 Introduction

All of the input required to accurately render an object could be derived from a highly detailed description of the object's geometry and material properties. A key to efficient rendering is not to work directly with such a detailed description, but with appropriate representations for various length scales – geometric models, surface maps and reflectances [6]. One approach is to derive the appropriate representations for the various scales from a highly detailed description. In this paper we suggest a step towards acquiring input at appropriate levels of detail directly, to increase the efficiency of acquistion and the subsequent processing required.

The motivation for this work is an ongoing project to create a digital model of the Michelangelo Pietà located in the Museum of the Opera del Duomo, in Florence, Italy (Fig.1a) This work was initiated by art historian Jack Wasserman, who is preparing a book [1] on this sculpture. A geometric model, on a relatively coarse (approximately 5 mm resolution) scale is needed to examine questions about the composition of the figures, to determine the shape left when Michelangelo removed several pieces that have subsequently been reattached, and to speculate on the shape of a piece that is still missing. Some of these issues will be examined with calculations of distance and volume. However, much of the study will be performed by rendering the sculpture in configurations and views specified by the art historian. These rendering requirements dictate the accuracy of the base model.

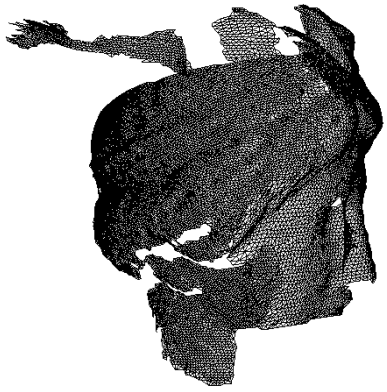
Finer resolution (on the order of 1 mm or less) descriptions are needed to render the appearance of the work for a CD-ROM to accompany the book, and to examine issues such as what type of tools were used based on fine tool marks. While photographs, panoramic views or light fields would be efficient means for rendering the sculpture as it now appears in the museum, this type of acquisition and rendering would not be adequate for this application. The goal is not to simulate a visit to the museum, but to render the sculpture in detail in alternative locations, such as a reconstruction of the tomb where the work was

[1] To be published by Princeton University Press

originally destined to stand. The sculpture is also to be rendered with altered geometry to see it as Michelangelo saw it with parts removed. To perform these renderings we wish to acquire the fine level geometry and reflectance. Surface maps, such as bump or height maps, are needed to render small details such as tool marks. Reflectance variations are required to show where there are cracks that have been repaired, and to show the extent of a yellow patina that was applied to a portion of the sculpture.

There are a number of challenges in this project. One is the sheer size of the sculpture, which stands 2.25 m high. Based on our initial scans, attempting to scan this in detail at even a spatial resolution of 0.5 mm would require on the order of 100 million points. Furthermore, the sculpture has a complicated topology with cavities formed by the overlapping limbs of the figures. No large piece of equipment that simply rotates about the sculpture would be adequate to capture the interior of these cavities. Finally, obviously samples can not be removed to measure reflectance in a separate laboratory. All measurements must be taken in the museum with minimal physical contact. Based on a review of previous projects to scan three dimensional works of art (e.g. [3],[9]) we decided to design a system around an existing multiview camera for three dimensional scanning.

(a)



(b)

Figure 1a: A photograph of the Pietà in the Museum of the Opera del Duomo,
1b: An example of the geometry captured in the intial scan of the Pietà on site in Florence. Two merged patches are shown, in wire frame.

We have performed one scan of the geometry of the sculpture on site in Florence. Typical results for part of the sculpture are shown in Fig. 1b. A Virtuoso shape camera [14] was used to capture the geometry in approximately 620 overlapping patches, referred to as "shape photographs." The shape camera uses light striping and 6 black and white cameras to obtain three dimensional points, and one color camera to acquire a texture. The Virtuoso software processes the points into a triangle mesh with *uv* indices to a color image to be used as a

texture map. The results shown in Fig. 1b are the result of merging two patches. Figure 2 (see color plate) shows the texture maps acquired for the two patches that were merged in Fig. 1b. The red dots are from lasers projected on the sculpture to facilitate the alignment of patches. The differences in overall lightness and hue in the texture maps are due to the automatic exposure control and color balance of the color camera used in the Virtuoso. The images clearly show highlights and shadows that are unique to the particular lighting conditions. Even without the red laser spots, these textures are not suitable for realistically rerendering the sculpture. The geometry is more than adequate for the questions of form and composition. We only need finer geometric detail for rendering the appearance at a smaller scale, and a method to capture a better approximation of the spectral reflectance of the surface.

2 Previous Work

Recently, more high-precision three dimensional scanners have become available [12]. Many scanners can produce very dense point clouds representing the object's geometry. The problem of efficiently finding a triangle mesh or other surface approximation to fit these points, and then reducing the number of vertices needed to define the surface is a very active area of research [1]. A reduced number of vertices is necessary for efficient storage and rendering.

Some researchers have considered the problem of retaining detail from the point cloud while simplifying the base geometric model. Krishnamurthy and Levoy [7] describe a method for deriving B-Spline patches with an accompanying height field or bump map from a dense triangle mesh. Although this provides primitives useful for rendering, it still requires an high precision initial scan, fitting a triangle mesh to a dense cloud, and significant processing.

Most scanning systems, like the Virtuoso, acquire color texture data by taking color images that are precisely aligned with the geometry. As illustrated by Fig. 2, this can produce poor results because the maps include the effects of the specific lighting conditions and camera characteristics. One approach to improving this has been developed by Sato et al. [11]. In their method, multiple color images are used to estimate the BRDF (bidirectional reflectance distribution function) of an object which has been digitized by a light striping range finder. Their method eliminates many problems with texture artifacts, although they do not discuss the issues of how to find absolute (rather than relative) reflectance, or how to correct for the alteration in color due to the light source and camera transfer function.

An alternative, more expensive, technology, for acquiring accurate maps of object color is to scan with three color lasers. The result is a dense, unorganized cloud of three dimensional points, each with a color attached. Soucy et al. [13] describe a method for deriving simplified meshes and texture maps from these dense three dimensional clouds.

In [10], the use of a version of photometric stereo is described for obtaining bump maps and maps of the approximate relative values of diffuse reflectance. Photometric stereo can produce much denser descriptions of an object for a given camera resolution than binocular (or multiview) stereo because an estimate can be made at each pixel, rather than at object edges or the edges of projected stripes. Photometric stereo is good for estimating surface normals,

but poor for reconstructing the three dimensional surfaces. In the field of computer vision various researchers, e.g. Ikeuchi [5] and Wolff [15] have developed hybrid techniques combining photometric stereo and multiple camera views. In these methods multiple cameras take images for multiple lighting conditions in order to reconstruct a denser range image than could be obtained with multiple cameras alone.

In our application, we do not need additional three dimensional points. The Virtuoso output is robust, and is more than adequate for our geometric model. In our system we use photometric stereo to compute normals only, which will be aligned with the three dimensional output and used as bump maps.

3 Proposed Method

Because of the need for a relatively small and mobile system to capture areas in the cavities of the sculpture, and because of the constraints of time and budget, we are building our acquisition system around the Virtuoso. To augment the output, we are adding a photometric system, similar to that described in [10], to obtain bump and texture maps at a finer level of detail.The texture maps will be calibrated by spot measurements using a Colorton II Digital Color Ruler [8] to produce maps of approximate spectral diffuse reflectance.

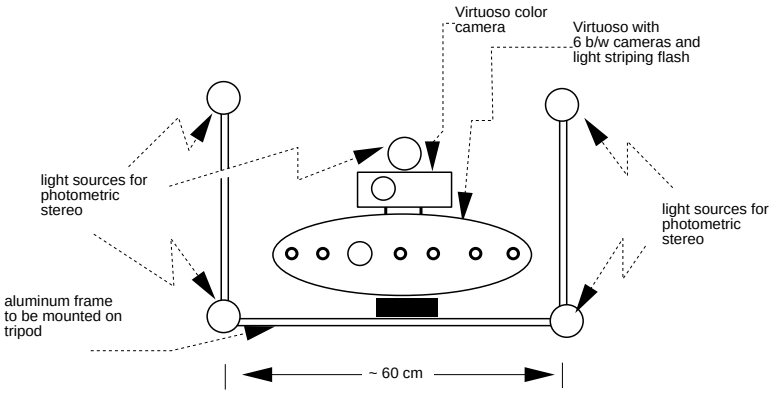


Figure 3: A diagram showing the Virtuoso system augmented by a photometric stereo system.

A diagram of the system currently being tested is shown in Fig. 3. Five lights are positioned around the Virtuoso. The Virtuoso color camera is used to capture images with each of the lights turned on in sequence. The light source positions must be known with respect to the Virtuoso's internal coordinate system. This is required so that the coordinate system of the normals computed from the photometric images is aligned with the coordinate system of the three dimensional points.

A Faro digitizing arm [2] is used to measure the light locations. In a one time only measurement, the Virtuoso coordinates of specified points on a simple target geometry were found by taking a shape photograph of the target. These

coordinates were then entered as data as the Faro arm digitized the target points, thus aligning the Faro coordinate system with the Virtuoso system. With the Faro arm realigned, the coordinates of three points on the Virtuoso camera body were found in the Virtuoso's own coordinate system. The reference points on the camera body are being used in subsequent experiments to align the Faro arm to the Virtuoso coordinate system for measuring light source positions.

The format of the shape photograph produced by the Virtuoso development software is an Open Inventor mesh of 3-D points with uv texture coordinates, and a color image trimmed to just the rectangular area covered by the captured geometry. The photometric software takes as input the 5 colored images for the 5 light positions, after trimming. Normals for each pixel in these images are computed using the photometric stereo method described in [10].

The issues to consider in successfully applying photometric stereo to non-Lambertian, non-convex surfaces are discussed in [10]. Briefly summarizing the basic calculations, let ρ and $\hat{n}$ be the diffuse reflectance and surface normal for a point P on the object. The lights used in the photometric system each have radiance L_o and subtend solid angle $\Delta\omega$ from point P. The radiance $L_{r,j}$ is reflected from P when it is illuminated by light j. The vector $\hat{l}_j$ points from P in the direction of light j. Grey level values are computed for each color image by combining the red, green and blue values for each pixel. The highest and lowest of the 5 grey level values for each pixel are discarded to minimize the effects of specular highlights and shadows. In the image captured by the color camera, the 0 to 255 grey scale value of the pixel in which P is visible is equal to $\alpha L_{r,j}$, where α is the constant of proportionality accounting for the linear response of the CCD camera. With the remaining three grey level values the following equation is solved for $\rho L_o \Delta\omega \hat{n}/\alpha\pi$:

$$\rho L_o \Delta\omega/\pi \begin{bmatrix} l_{1,1} & l_{1,2} & l_{1,3} \\ l_{2,1} & l_{2,2} & l_{2,3} \\ l_{3,1} & l_{3,2} & l_{3,3} \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \begin{bmatrix} \alpha L_{r,1} \\ \alpha L_{r,2} \\ \alpha L_{r,3} \end{bmatrix} \tag{1}$$

Since $\hat{n}$ has a magnitude of 1 we can obtain $\hat{n}$ and $\rho L_o \Delta\omega/\alpha\pi$ separately. Since the source radiance, solid angle subtended, and camera response is the same for all pixels, $\rho L_o \Delta\omega/\alpha\pi$ is equal to ρ_{rel}, the reflectance of the object point viewed through this pixel, relative to the points viewed through the other pixels in the image. As discussed in [10], approximate relative reflectances for red, green and blue (RGB) channels can be obtained by adjusting the ratio of red to green to blue found in the images outside of the shadow and highlight regions. For pixel n, the adjusted values of $R'G'B'$ then are:

$$[R', G', B']_n = \rho_{rel,n}[R, G, B]_n / greylevel([R, G, B]_n) \tag{2}$$

Equation 2 gives a crude relative estimate of reflectance for the imprecisely defined red, green and blue channels. To obtain more precise, absolute measurements, we make spot measurements of absolute diffuse reflectance using the Colortron II. The Colortron II measures absolute spectral reflectances in 32 wavelength bands in the visible spectrum. To use the measurement to adjust the relative reflectances, we select values at three wavelengths, 700, 580, and 450 nm to correspond roughly with the R, G and B channels. For the location of the spot measurement c, we have reflectance measurements $\rho_{700,c}$, $\rho_{580,c}$ and $\rho_{450,c}$, and

we can obtain the adjusted $R'G'B'$ values computed for c from the processed photometric images. Estimates of reflectance at the three selected wavelengths are then made for all of the image pixels n using:

$$[\rho_{700}, \rho_{580}, \rho_{450}]_n = [\rho_{700,c} R'_n / R'_c, \rho_{580,c} G'_n / G'_c, \rho_{450,c} B'_n / B'_c)] \quad (3)$$

Because we only obtain three values to represent color in our images, we only obtain three values estimating the spectral reflectance at each point. A series of spot measurements of the spectral reflectance on the actual Pietà showed that the spectrum for the marble, like many other common materials, is quite smooth. Using three values gives a reasonable representation for our application. Analogous to the bump map recording small variations between more sparsely measured three dimensional points, the processed color images are used to record small variations in spectral reflectance between spatially sparse diffuse reflectance measurements.

Using the uv coordinates from the Open Inventor model, the resultant map of surface normals computed using Eq.1 and map of reflectances computed using Eqs. 2 and 3 can be rendered with the three dimensional geometry.

Using this plan for scanning, we can adjust our system to acquire data at the resolution needed for rendering. The resolution of the geometric model needed determines the distance at which we want to focus the cameras for the light stripe image, since for the same stripe pattern the spacing between the points increases with the distance between the camera and target. We have some control over the relative resolution acquired by the photometric portion of the process by using a color imaging camera with the appropriate resolution relative to the light stripe frequency. Along with this specification for the color camera we require manual control of exposure, so that the exposure is the same for all of the images used for the photometric solution.

The scan of one patch of the sculpture proceeds as follows:

- Turn on laser dots for geometric alignment.
- Turn on ambient light for adequate illumination for stripes
- Take light stripe pictures
- Turn off laser dots and ambient light
- Take five photometric images

This scan procedure will be under software control, with most of the elapsed time required for each scan being the time to download data from the camera to a removable storage card. We are currently estimating the precise resolution needed for the geometry and normals maps based on our inital scan of the actual piece and photometric tests in our laboratory.

4 Results

Figures 4 to 9 show some results from initial tests of the combined Virtuoso/photometric system that we performed in our laboratory. As a test object we used a 1/10 scale model of the Pietà, which is an approximate reproduction sold at the museum gift shop. The detail on this scale model, such as the fingers on the hand of the rightmost figure, are the same absolute size, 1 mm, as the features we wish to capture on the full scale model.

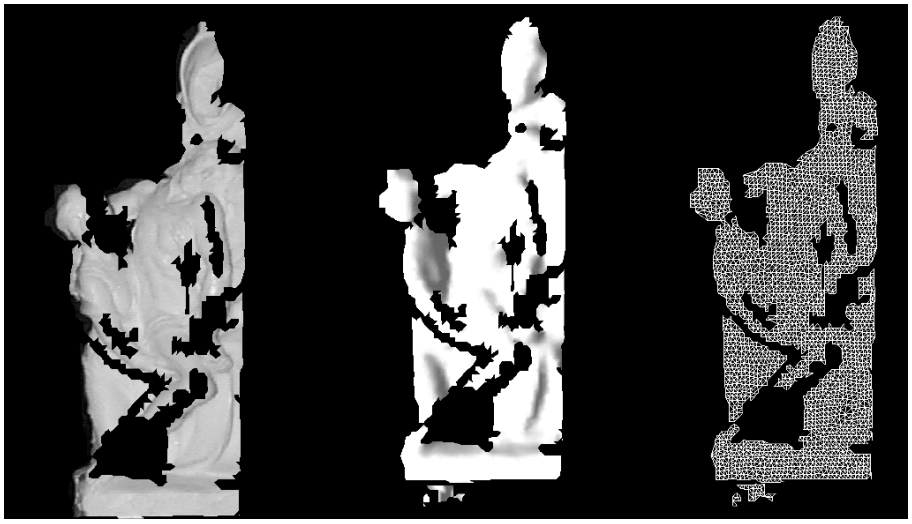


Figure 4: Results directly from the Virtuoso system for a small model of the Pietà. Results are shown (a) with texture map on the left, (b) without texture map in the center, and (c) in wire frame on the right.

Figure 4 shows the output from the Virtouso camera as purchased. Figure 4a shows the texture mapped geometry from the direction that it was acquired. The lighting in this view looks correct since it is just a display of the image as captured. There are holes in the geometry because points can only be computed for object points that are visible unshadowed from the six black and white cameras used to image the light stripes. A full model is assembled by merging together geometry from multiple views. The captured geometry is shown shaded without texture map and in wire frame in Figs. 4b and 4c . From a distance of approximately 1 m, the Virtuoso has digitized points on the 22.5 cm high statue at a spatial frequency of approximately 2 mm. This level of detail is inadequate to capture features such as the fingers on the right figure. However, it is more than enough to capture the overall coarse scale geometry. To simulate a coarser sampling, we ran the model through a simplification program. The results of using the simplification method described in [4] to reduce the number of vertices from 2927 to 1086 are shown shaded and in wire-frame in Figures 5a and 5b. The algorithm used a tolerance of 1% of the diagonal of the bounding box to determine the allowable change in the shape due to the simplification. Little detail has been lost relative to the original captured geometry.

Figure 6 shows the result of the photometric calculations. Figures 6a and 6b were rendered with the photometric results under two different novel lighting conditions (i.e. light positions that were not used in computing the normals.) The spacing of the normals on the map is approximately 0.4 mm. The spatial density of the normals is higher than the three dimensional points in part because the color camera has a higher resolution and in part because it takes a width greater than one pixel to determine the edge of the light stripes.

Figure 7 contrasts the results using the Virtuoso output alone with the photometric output mapped onto a simplified geometry. In Fig. 7, the light source direction is different from the conditions under which any of the Virtuoso or pho-

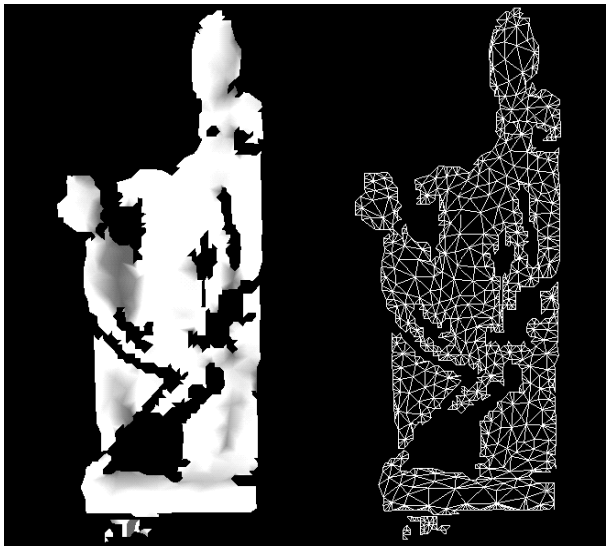


Figure 5: The geometry from Fig. 4 after simplification, (a) shown shaded on the left and (b) in wire frame on the right.

(a) (b)

Figure 6: Images rendered using photometric results under two new lighting conditions.

tometric images were acquired. Figures 7a and 7c show the result from Virtouso under the new lighting conditions. Figure 7a shows the original view, and Fig. 7c shows a view from the right side. The Virtuoso results have the obvious advantage that correct silhouettes are obtained when a new view is selected, which cannot be done with just the photometric results. The texture on the Virtuoso looks very odd in 7a and 7c – particularly noticeable are the shadow lines from the original image that are not consistent with the new lighting.

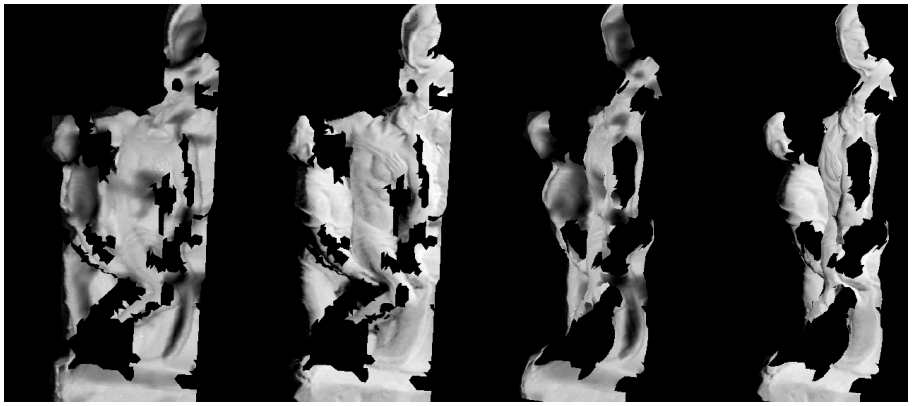
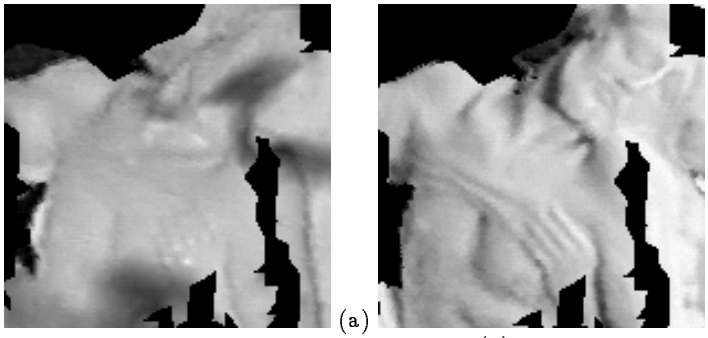


Figure 7: Comparison of original Virtuoso geometry and texture map (in 7a and 7c, the leftmost image and image third from the left) with simplified geometry and photometric results (7b and 7d).

Figures 7b and 7d show the results of mapping the normals from the photometric process onto the simplified geometry obtained from the original Virtuoso mesh. In this initial application the lighting is computed using the photometric normals and a diffuse reflectance in software, and the resulting texture map is displayed on the simplified mesh using hardware texture mapping. Even though the model has fewer triangles than the original Virtuoso mesh, a correct silhouette is obtained when the model is viewed from the right. The bump map allows the display of the detailed hand of the right figure placed on the central figure. As shown in the detailed images in Fig. 8, the appearance of the geometric resolution has been increased by a factor of 5, without any explicit acquisition and processing of additional three dimensional points.

Figure 9 shows the results for a larger colored object – a ceramic cookie jar with a somewhat glossy finish similar to parts of the Pietà. The Virtuoso-only results with and without the acquired texture map are shown in 9a and 9b respectively. Figure 9b illustrates another problem common in scanning systems. Under new lighting conditions (from above and the right) slight diagonal artifacts appear on the squirrel's head and arm. The number of ridges on the squirrel's tail has been increased. The artifacts occur because even small sub-millimeter errors in computing the coordinates can result in noticeable artifacts in the slopes on the object surface. As shown in Fig. 10 even a 0.25 mm error (which is quite small by the standards of current non-contact digitizers) in samples spaced 2 mm apart can result in an angle of more than 28 degrees between normals that should be the same.



(a) (b)

Figure 8: Details from Fig. 7, showing the rendering (a) with the original texture map, and (b) with photometric results.

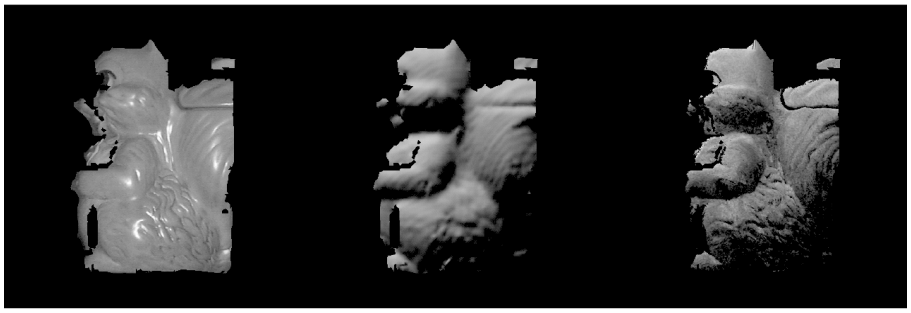


Figure 9: Results of scanning a squirrel-shaped cookie jar. The results directly from the Virtuoso are shown (a) texture mapped on the left and (b) without texture map in the center. The results of the photometric normals calculation are shown on the right (c) with the same lighting as used in (b.)

A typical remedy for small slope errors is to perform additional scans of the same area, producing even a denser cloud of points and averaging the results. However, without taking more points we already have a good description of the base geometry in one scan. Instead of taking more points, we use our photometric results, shown in Fig. 9c to replace the normals computed from the triangle mesh and eliminate the scanning artifacts and to more faithfully reproduce the portion of the jar representing the squirrel's fur. The results in Fig. 9c, while a better representation of the pattern on the tail and the fur, are somewhat noisy. The useful data for the photometric calculations is in the brown, darker areas in the images. We do not have a full 8 bits of image data for the pixels we use to compute the normals. Because our test camera has automatic exposure control, we were not able to adjust the exposure and use the full 8 bit range to eliminate the noise in this test.

The cookie jar is basically brown, but has subtle variations in the color of the diffuse portion of reflectance to enhance the impression of fur on the squirrel's body. The relative reflectance maps were computed from the photometric images, and then adjusted by a Colortron II spot measurement on the squirrel's arm.

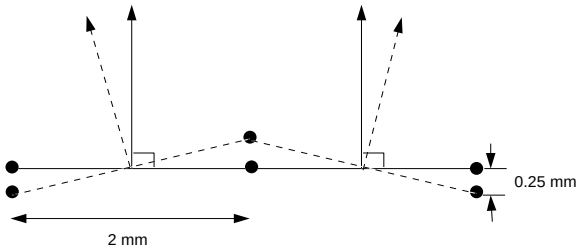


Figure 10: The solid line shows the actual solid surface. The dotted line shows the results of a scan with a small error that results in noticeable errors in the surface normals.

The resulting map of reflectances is compared with the original texture map in shown in Fig. 11 (see color plate), with the R, G and B channels used to display $\rho_{700}, \rho_{580}, \rho_{450}$.

5 Future Work

A number of issues need to be addressed to obtain high quality results from this system. For example, it is clear from the cookie jar example that a method is needed to estimate the specular component of reflectance. We can not measure the absolute value of the specular component with the limited dynamic range of our camera, however we can estimate the width of the specular peak using a modified form of Sato et al.'s method. This may need to be supplemented by spot measurements.

With our planned system, we can scan the statue at a coarser resolution than our original 600 patch scan. However, the result will still be over 100 overlapping maps of normals and reflectances. While careful calibration can ensure that these maps are consistent with one another, having numerous overlapping maps is inefficient for storage and rendering. A method for merging and reducing the number of surface maps without loss of resolution is needed.

In this paper we have presented a step forward in acquiring input at different levels of input directly, rather than starting with a very dense cloud of points and colors. Refining this idea by combining various modes of point and image acquisition has great potential for acquiring data for rendering complex objects efficiently and economically.

References

1. Bernardini, F., Bajaj, C., Shen, J., and Schikore, D. Automatic reconstruction of 3D CAD models from digital scans. *International Journal of Computational Geometry and Applications* (to appear).
2. FARO Technologies, Inc. The FaroArm. *http://www.faro.com/*.
3. Fraunhofer Institute for Computer Graphics. SCULPTOR. *http://www.igd.fhg.de/www/igd-a7/Projects/model/examples.html*.
4. Guéziec, A. Surface simplification inside a tolerance volume. *IBM Research Report*, RC20440 (1997).

5. Ikeuchi, K. Determining a depth map using dual photometric stereo. *International Journal of Robotics Research 6*, 1 (September 1987), 15–37.
6. Kajiya, J. Anisotropic reflection models. *Computer Graphics (SIGGRAPH '85 Proceedings) 19*, 3 (July 1985), 15–22.
7. Krishnamurthy, V., and Levoy, M. Fitting smooth surfaces to dense polygon meshes. *Computer Graphics (SIGGRAPH '96 Proceedings)* (August 1996), 313–324.
8. Light Source, Inc. Colortron. *http://www.ls.com/colortron.html.*
9. National Research Council of Canada. Visual information technology. *http://www.vit.iit.nrc.ca/image/masque.html.*
10. Rushmeier, H., Taubin, G., and Guéziec, A. Applying shape from lighting variation to bump map capture. *Proceedings of the Eighth Eurographics Rendering Workshop* (June 1997), 35–44.
11. Sato, Y., Wheeler, M., and Ikeuchi, K. Object shape and reflectance modeling from observation. *Computer Graphics (SIGGRAPH '97 Proceedings)* (August 1997), 379–388.
12. Schechter, J. 3D digitizers make their mark. *Computer Graphics World* (March 1998), 79–84.
13. Soucy, M., Godin, G., Baribeau, R., Balis, F., and Rioux, M. Sensors and algorithms for the construction of digital 3-D colour models of real objects. *Proceedings of the International Conference on Image Processing* (September 1996), 409–412.
14. Visual Interface, Inc. Virtuoso. *http://www.visint.com/.*
15. Wolff, L., and Angelopoulou, E. 3D stereo using photometric ratios. *Journal of the Optical Society of America – A*, 11 (November 1994), 3069–3078.