Surface Enhancement Using Real-time Photometric Stereo and Reflectance Transformation

Tom Malzbender[1], Bennett Wilburn[2], Dan Gelb[1], Bill Ambrisco[3]

[1] Hewlett-Packard Laboratories, USA
[2] Microsoft Research Asia, China
[3] Foxhollow Technologies, USA

Abstract

Photometric stereo recovers per-pixel estimates of surface orientation from images of a surface under varying lighting conditions. Transforming reflectance based on recovered normal directions is useful for enhancing the appearance of subtle surface detail. We present the first system that achieves real-time photometric stereo and reflectance transformation. A high-speed video camera, computer controlled light sources and fast GPU implementations of the algorithms enable both methods. We also present novel GPU-accelerated normal transformations before relighting that "amplify" shape detail. By applying standard image processing methods to our computed normal image, we can selectively enhance surface detail at different frequencies. Our system allows users in fields such as forensics, archeology and dermatology to investigate objects and surfaces by simply holding them in front of the camera. Real-time analysis of surface roughness for metrology can also be performed from the extracted normal field.

Catagories and Subject Descriptors: I.3.7 [Computer Graphics]: Three Dimensional Graphics and Realism: Color, shading, shadowing, and texture, I.3.2 [Computer Graphics]: Graphics Systems, I.4.3 [Image Processing and Computer Vision]: Enhancement

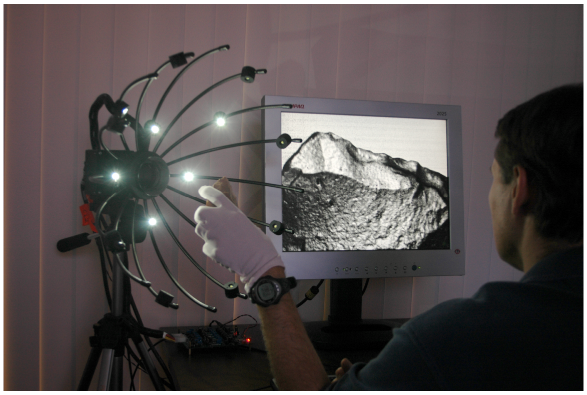

Figure 1. *Our real-time system for inspecting objects with transformed surface reflection properties that are more conducive to perceiving shape detail.*

1. Introduction

The perception of shape and form is crucial in the interpretation of surfaces and objects in a wide variety of disciplines. Examples include such diverse fields as the study of ancient artifacts and remains, forensics, diagnostic medicine, industrial inspection and manufacturing (quality control). The field of image enhancement [Rus02] provides a variety of methods that assist in making surface detail more apparent. Contrast enhancement methods in particular, such as histogram equalization, unsharp masking, and high-pass filtering are useful for this purpose. These methods are inherently two dimensional and do not recover shape information.

By collecting multiple images under varying lighting conditions, several authors have extracted 3D information in the form of surface normals and used that information for enhancing surface detail [DHT*00], [WGT*05], [MGW01], [HBM*02], [McC01], [BSM05]. Similar to our work is that of [WGT*05], which also uses computer controlled Luxeon V L.E.D.s to achieve reflection transformation for theatrical performances. Acquisition is done in real-time in their system, but processing is done offline due to complexity. None of these past efforts provide real-time interactivity.

We achieve real-time reflection transformation end-to-end, which allows unskilled users to inspect objects as they would unaided, but with an enhanced rendering of shape detail. In addition, we introduce simple normal transformations that "amplify" surface detail. Real-time analysis of the extracted normal field can also be performed to measure roughness of the surface patches being imaged, which appears to be useful for dermatology and cosmetics applications.

2. Overview and Background

Our system triggers up to 16 lights, one at a time, synchronized with a high-speed video camera to capture views of a surface under various lighting directions. We use photometric stereo to extract surface normals, and reflectance transformations to relight objects by introducing specular highlights. In this section, we review the computations necessary for photometric stereo and reflectance transformation. Our system performs these operations on graphics hardware to achieve real-time operation.

2.1 Photometric Stereo

The recovery of per-pixel surface normals from multiple images of a static scene under varying lighting is termed photometric stereo. It was introduced in 1980 by Bob Woodham [Woo80], and has been extensively built on [BMM*02],[NRD05], [CJ82], [TDe91], [Geo03]. An overview of early work on photometric stereo is found in [Hor86]. Particularly relevant to our work is that of [Woo94], where photometric stereo running at 15 frames/second is achieved. This is done without switching light sources by using 3 spectrally distinct light sources from varying orientations. The response to each light source is inferred separately from each channel of an RGB color camera. However, this approach not only requires the subject to have uniform albedo, it requires calibration from an object of known shape with the same albedo, limiting its practical application.

Although a minimum of only three images with non-coplanar light sources are needed for photometric stereo, more images can improve results in several ways. First, removing the lightest and darkest sets of samples per image location is a useful technique for avoiding the use of shadowed and specular samples [EYB96], [RTG97], [WGT*05], [WVM*05]. Second, using larger numbers of lighting direction samples can lead to increased normal-noise averaging and thus higher quality normals [MGW01], [HBM*02].

A diffuse surface with orientation $n = (n_x, n_y, n_z)$ at pixel (i,j), illuminated from direction $l=(l_x,l_y,l_z)$, will appear with luminance

$$e = a n \cdot l \qquad (1)$$

where a is the albedo of the surface.

M images illuminated with lighting directions given by matrix

$$L = (l_1^T, l_2^T, .. l_m^T)^T \qquad (2)$$

result in a column vector of measured per-pixel luminances for each lighting condition,

$$E = (e_1, e_2, ... e_m)^T \qquad (3)$$

The surface orientation, n, for this pixel can be solved for by computing

$$n' = L^+ E \qquad (4)$$

where L^+ is the Moore-Penrose generalized pseudoinverse of L. Normalizing n' yields the surface albedo and surface normal for every pixel:

$$a = \|n'\| \qquad (5) \qquad\qquad n = \frac{n'}{a} \qquad (6)$$

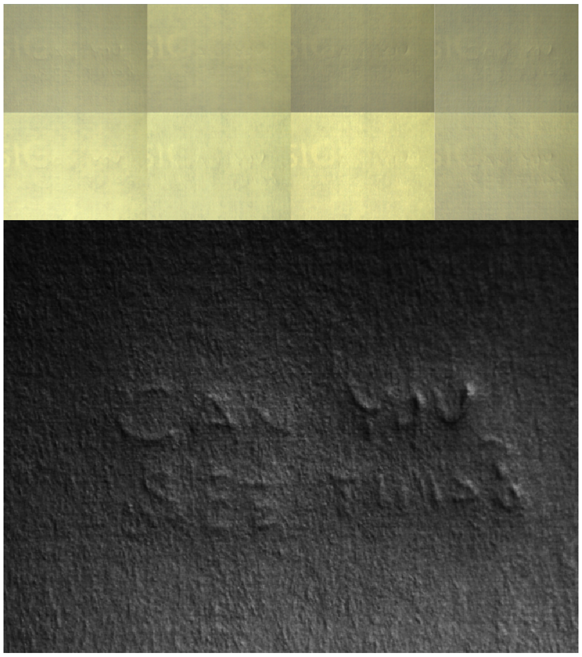

Figure 2. *A real-time specular enhancement example of indented writing on a post-it note. Top: Eight input camera images taken at varying lighting directions. Bottom: Relit computing synthetic specular highlights from surface normals extracted with photometric stereo.*

2.2 Reflectance Transformation

Reflectance transformation has been shown to be useful for relighting human faces, especially actors, in [DHT*00], [WGT*05], and for the enhancement of surface detail in [MGW01], [HBM*02], [McC03], [Rus02]. Surface normals recovered by photometric stereo are used to transform a measured reflectance function in some manner, keeping the estimate of the surface normal constant. This changes the effective material properties of a surface without changing its underlying geometry. Specular enhancement is one such method where synthetic specular highlights are added to an image of a predominantly diffuse surface. This change in the appearance of the surface can make surface shape detail more visible, as demonstrated in figure 2.

We use a simple Phong/Blinn lighting model [Bli78], [Pho75], to compute these highlights from the specular coefficient, k_s, half (h) and normal (n) vectors.

$$i_s = k_s h \cdot n \qquad (7)$$

3.0 Normal Transformation

In our system, we do not compute an explicit model of the reflectance function, only per-pixel surface normals and albedo. This prohibits the use of diffuse gain [MGW01], a useful reflectance transformation method. However, a similar effect can be achieved using simple normal transformations that make subtle shape detail more visible.

Cignoni et al. [CST05] also propose a method for amplifying high-frequency variations in surface normals. They require a geometric model of the object and modify the mesh as a preprocessing step, both of which are not feasible for a live, interactive system. Instead, we use image-based methods and avoid explicitly deducing structure by working directly on computed normals. Applying image-processing techniques to images of normals guarantees bounded compute times independent of the object surface complexity.

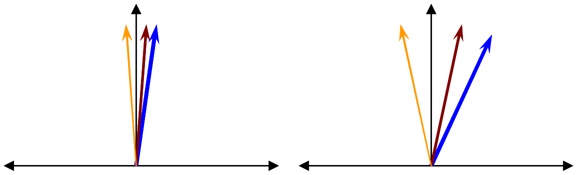

Figure 3. *Normal Transformation. Left: surface normals for three pixels extracted using photometric stereo. Right: transformed normals.*

3.1 Global Reference Normal Transformation

One normal transformation that we have found useful is exaggerating the component of the surface normal that points away from the camera optical axis. In particular, the transformation that maps normal n into $n*$, given by:

$$n* = (gn_x, gn_{y,} \sqrt{1-(gn_x)^2-(gn_y)^2}) \qquad (8)$$

can be used to map small surface normal perturbations into larger changes for gain values, g, under user control. Figure 3 shows this mapping graphically. This method is similar to that independently employed and demonstrated by [WVM*05] called *slope exaggeration*. These transformed normals, clamped to the unit hemisphere, are evaluated with Phong/Blinn shading to introduce synthetic specular highlights, as with specular enhancement. An example of camera input and rendered output frames computing eq. 8 in real-time is shown in figure 4.

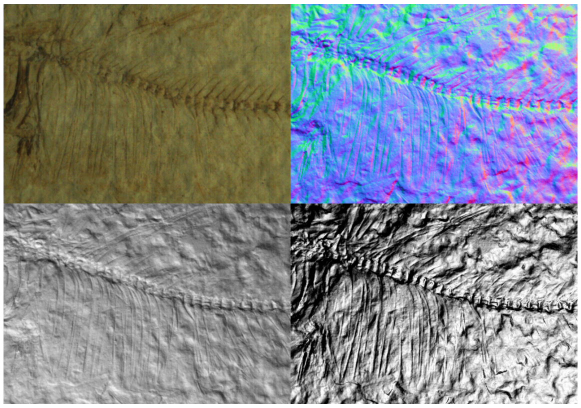

Figure 4. *Effects of first transforming the surface normal before adding specular highlights. Upper Left: One of eight camera input images of a fish fossil. Upper Right, surface normals extracted with photometric stereo, transformed by eq. 8. and rendered using pseudocolor. Lower Left: Specular Enhancement. Lower Right: Specular enhancement with transformed normals.*

3.2 Unsharp Masking Normals

The normal transformation described by equation 8 is performed globally relative to the optical axis of the camera. With our system we are also free to apply standard image processing methods to the computed surface normals. As one example, we implemented a high-frequency emphasis filter for surface normals. At each pixel, we compute a local reference normal which is the average of normal vectors over a local patch, m, (9x9 pixels in our case), then amplify the difference between the computed normal and this reference vector:

$$n* = n + k\left(n - normalize\left(\sum_{j=1}^{m} n_j\right)\right) \qquad (9)$$

Here, k is a user-selectable scale factor determining how much to amplify the normal variations. We set the z-component of n* to zero if it is negative to ensure that no normals face away from the camera, then normalize to produce the final transformed normal.

This is analogous to the 2D image processing method of unsharp masking [Rus02], except that two vector normalizations are required per pixel, computing both the reference vector and the resulting pixel normal vector. These operations all run on the GPU, and the resulting images accentuate high frequency perturbations, while leaving low frequency structure unaffected. This appears to be especially useful in the analysis of indented writing as a means of filtering out low frequency paper undulations, while preserving higher frequency indentations that are of interest [McC03]. A comparison of these methods is shown in figure 5.

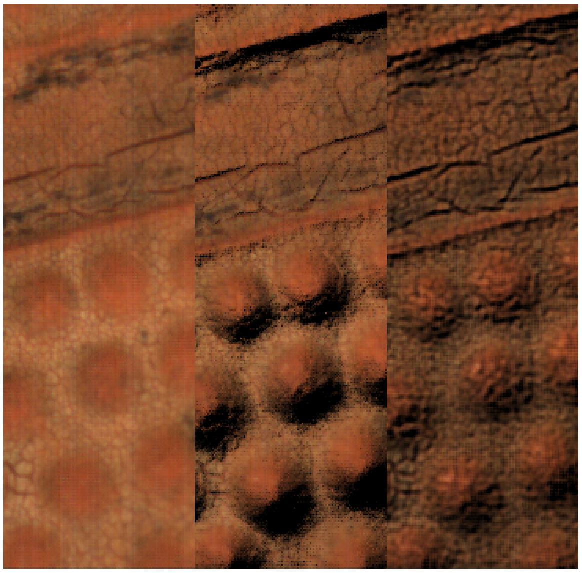

Figure 5. *Normal transformation variants. Left: Source image of basketball surface. Middle: Amplifying the component of the normal that points away from the optical axis emphasizes larger surface details. Right: Normal transformations relative to a local 9x9 normal average amplify only the high frequency detail. These images employ normals transformations without adding any specular enhancement.*

4 Quantitative Surface Roughness

Up to this point, we have emphasized the benefits of our system for visualization. Our photometric stereo system can also perform measurements of surface roughness in real-time. The loading of the CPUs on our host workstation is negligible since photometric stereo and relighting is implemented on the GPU. We exploit this to read surface normals back into main memory and compute a measure of surface roughness of a patch of m pixels in real time. In particular, the variance of surface normal coefficients given in equation 9 can be computed at greater than 30 frames/second for patches the size of the entire image.

$$s^2 = \frac{\sum_{j=1}^{m}((n_{jx}-\overline{n_x})^2+(n_{jy}-\overline{n_y})^2+(n_{jz}-\overline{n_z})^2)}{m} \tag{10}$$

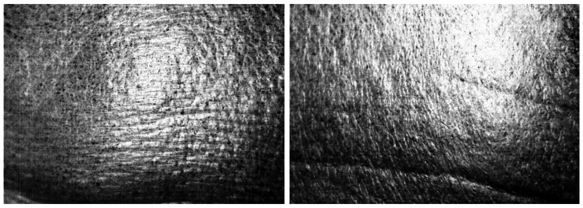

Figure 6. *Skin surface roughness. Left: variance 813, Right: variance 507.*

We have found surface normal variance to be a useful indicator of roughness for surface metrology. Specifically, its application to dermatology appears promising as an augmentation to assessing skin roughness by tactile means.

5.0. System Design and Performance

The front end to our system is a Basler high-speed video camera (A504kc) attached to a Datacube MaxResolution frame grabber. A trigger signal from the Basler controls a simple, custom CPLD board that triggers one of 16 light sources synchronized with the camera. The light sources are bright-white Luxeon V Star L.E.D.'s, pulsed at 1 amp each, and are mounted on a lightweight assembly that screws onto the lens of the camera. In practice, we drive eight lights instead of sixteen, keeping the GPU implementation simpler. This also permits us to double the exposure time, which allows either an increased signal-to-noise or smaller optical aperture, yielding increased depth of field.

The high-speed video camera and frame grabber are capable of capturing 640 x 480 images and transferring them via DMA into the workstation's (dual processor AMD Opteron 248) system memory at 500 F/sec with negligible CPU loading. Using eight lights, this implies a maximum rate of 62 Hz. to capture each eight-frame cycle. We nominally operate up to 60 frames/second for computed imagery. Although we were initially concerned that subject motion during frame acquisition would cause visible normal estimation artifacts, we found that with a 2 msec. interval between 8 acquired frames we did not need motion estimation/image stabilization methods. However, at slower frame rates, these artifacts do become visible. Since an early, optimized CPU-only implementation achieved only 7.1 F/sec we adapted our methods to a GPU to achieve real-time interactivity.

The images from the camera in Bayer format are downloaded directly to the graphics card and demosaiced on the Nvidia Geforce 7800 GTX GPU. A dark image average captured with no L.E.D. illumination is subtracted from each of the eight light images to remove fixed pattern noise from the camera. Before photometric stereo computations can be performed, the RGB pixel quantities are converted into luminance values. Since each L.E.D. can potentially have a different brightness, we calibrate the system using a diffuse gray card. This yields a scale factor for each image which is multiplied with the per pixel luminance values.

Photometric stereo computations are next performed per pixel using the eight luminance values. If the surface being imaged has specular characteristics, or exhibits self-shadowing, then it can be useful to discard the brightest and darkest of the eight luminance values to obtain a more accurate normal estimate. We have implemented this ranking and selection in real-time, but find that switching which light sources are used to estimate the normals on a per-pixel basis introduces spatial noise. Adding more lights to our system would mitigate this problem, but make the system more complex.

The luminance values are then multiplied by L^+, as shown in equation 4, to solve for the surface n'. If the brightest and darkest luminance values are discarded, then a specific L^+ matrix must be selected based on which directions were used. The computed n' is then normalized to generate the estimated surface normal. Normal transformation can be applied based on a user specified gain value. A variety of rendering modes are useful, including rendering specular highlights from the normals, combining the specular highlights with the original captured images, visualizing the normals directly, and displaying the computed albedo values. The computed surface normals can be read back to the CPU for analysis or additional processing.

One rendering pass suffices to compute normals using photometric stereo, transform the normals according to equation 8, and shade the result. Applying the unsharp mask or other filters to the normals, however, requires two rendering passes. The first pass produces images containing the normal and albedo for each pixel. The second pass operates on these images to compute the local reference normal, amplify the difference between the pixel normal and the reference, and apply the desired shading technique. All of these operations run in real-time.

6. Conclusions

We have demonstrated what we believe to be the first real-time, photometric stereo system. Rather than capture images with multiplexed illumination for off-line processing, we exploit commodity graphics hardware to compute surface normals and display enhanced video images in a live setting. We also introduce several means of using normal transformations in conjunction with specular enhancement to highlight surface detail.

One weakness of our system is that it does not account for specularities and shadows when computing photometric stereo. This is a drawback for applications that require metric shape information, but does not severely impact our application of enhancing images for human viewing. Now that our systems operates in real-time, we are interested in improving it using recent advances in photometric stereo. This may prove difficult for some methods. For example, Mallick et al. [MZK*05] propose a rotation in color space that aligns the light source color with one of the principle axes in the color space. Working with the two remaining axes effectively removes specular highlights at each pixel. Unfortunately, for our purpose this would require a high-dynamic range camera, which is not available at the speeds we require.

Currently, the CPUs on our rendering PC are mostly idle. We showed one example of using that dormant processing power to compute a measure of surface roughness from extracted normals, but we anticipate many other applications that would benefit from real-time imagery and surface normal information.

Acknowledgements
We thank Augusto Roman and Mark Levoy of Stanford University for the use of their high speed video camera. Thanks to Mark Mudge and Carla Schroer of Cultural Heritage Imaging for providing various archeological and cultural artifacts, including those used in the video submission and shown in figure 1.

References

[BSM*05] Bertesaghi, A., Sapiro, G., Malzbender, T., Gelb, D., “Three-dimensional Shape Rendering from Muliple Images”, Graphics Models 67 (2005), pp. 332-346.

[BMM*02] Bernardini, F., Martin, I., Mittleman, J., Rushmeier, H., Taubin, G., “Building a Digital Model of Michelangelo’s Florentine Pieta”, *IEEE Computer Graphics and Applications*, 22(1):59-67.

[Bli78] Blinn, J.F., “Computer Display of Curved Surfaces”, *Ph.D. Thesis, University of Utah*, 1978.

[CJ82] Coleman, E., Jain, R., “Obtaining 3-Dimensional Shape of Textured and Specular Surfaces Using Four-Source Photometry”, *Computer Graphics and Image Processing*, 18, 2002, pp. 309-328

[CST05] Cignoni, P., Scopigno, R., Tarini, M., “A Simple Normal Enhancement Technique for Interactive Non-photorealistic Renderings”, Computers and Graphics 29, pp. 125-133.

[DHT*00] Debevec, P., Hawkins, T., Tchou, C., Duiker H., Sarokin, W., “Acquiring the Reflectance Field of a Human Face”, *Proceedings of Siggraph 2000*, pp. 145-156.

[EYB96] Epstein, R., Yuille, A., Belhumeur, P., “Learning Object Representations from Lighting Variation”, *ECCV*, pp. 179-199.

[Geo03] Georgehiades, A., “Recovering 3-D Shape and Reflectance From a Small Number of Photographs”, *Eurographics Symposium on Rendering*, June 2003, pp. 230-240.

[HBM*02] Hammer, O., Bengston, S., Malzbender, T., Gelb, D., ”Imaging Fossils Using Reflectance Transformation and Interactive Manipulation of Virtual Light Sources”, *Palaeontologia Electronica*, Aug. 23, 2002.

[Hor86] Horn, B, “Robot Vision”, New York, *McGraw-Hill*, 1996, ISBN 0-262-08159-8.

[MGW01] Malzbender, T., Gelb, D., Wolters, H., ”Polynomial Texture Maps”, *Proceedings of Siggraph 2001*, pp. 519-528.

[McC01] McGunnicle, G., Chantler, M., “Recovery of Fingerprints using Photometric Stereo”, *Irish Machine Vision and Image Processing Conference,* Sept. 5-7, 2001, pp. 192-199.

[McC03] McGunnicle, G., Chantler, M., “Recovering Handwriting from Background Printing using Photometric Stereo”, *Pattern Recognition* (36), 2003, pp. 1869 – 1879.

[MZK*05] Mallick, S., Zickler, T., Kriegman, D., Belhumeur, P., ”Beyond Lambert: Reconstructing Specular Surfaces Using Color”, Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2005. pp. 619-626.

[NRD05] Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R., “Efficiently Combining Positions and Normals for Precise 3D Geometry”, *Proceedings of Siggraph 2005*, pp. 536-543.

[Pho75] Phong, B., “Illumination for Computer Generated Images”, *Communications of the ACM* 18, 6, June 1975, pp. 311-317.

[RTG97] Rushmeier, H., Taubin, G., Gueziec, A., “Applying Shape from Lighting Variation to Bump Map Capture”, *Eighth Eurographics Rendering Workshop*, June 1997.

[Rus02], Russ, J., “The Image Processing Handbook, 4th Edition”, *CRC Press*, 2002., ISBN 0-8493-1142-X, pp.674-679.

[TDe91] Tagare, H., DeFigueiredo, R., “A Theory of Photometric Stereo for a Class of Diffuse Non-Lambertian Surfaces”, *PAMI* 13(2), pp.133-152.

[WGT*05] Wegner, A., Gardner, A., Tschou, C., Unger, J., Hawkins, T. Debevec, P., ”Performance Relighting and Reflectance Transformation with Time Multiplexed Illumination”, *Proceedings of Siggraph 2005*, pp. 756-764.

[WVM*05] Willems, G., Verbiest, F., Moreau, W., Hameeuw, H., Van Lerbeghe, K, Van Gool, L., “Easy and Cost-Effective Cuneiform Digitizing”, 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2005).

[Woo80] Woodham, R., “Photometric method for determining surface orientation from multiple images”, *Optical Engineering*, 19(1), pp. 139-144.

[Woo94] Woodham, R., “Gradient and Curvature from Photometric Stereo Including Local Confidence Estimation”, *Journal of the Optical Society of America, A,* vol. 11, no. 11, pp. 3050-3068, Nov. 1994.