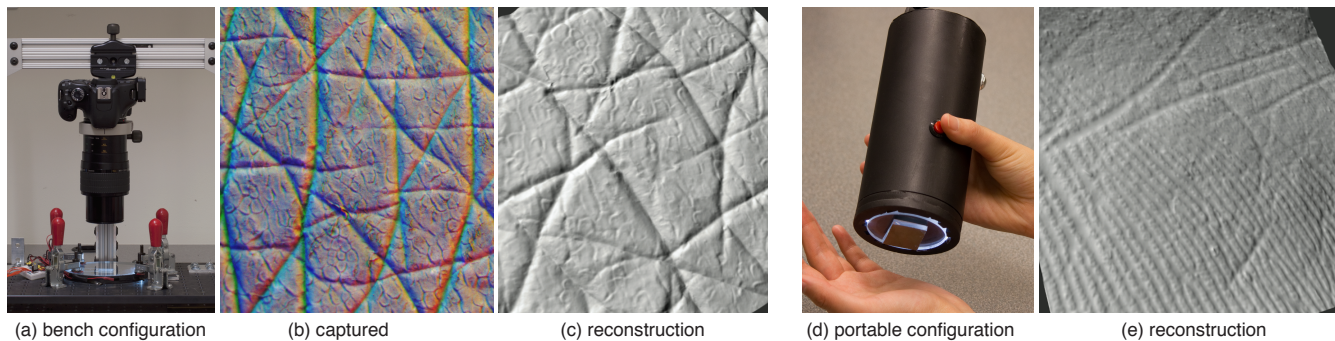# Microgeometry Capture using an Elastomeric Sensor

Micah K. Johnson[*]   Forrester Cole[†]   Alvin Raj[‡]   Edward H. Adelson[§]
Massachusetts Institute of Technology

Figure 1: *Our microgeometry capture system consists of an elastomeric sensor and a high-magnification camera (a). The retrographic sensor replaces the BRDF of the subject with its own (b), allowing microscopic geometry (in this case, human skin) to be accurately captured (c). The same principles can be applied to a portable system (d) that can measure surface detail rapidly and easily; again human skin (e).*

Abstract

We describe a system for capturing microscopic surface geometry. The system extends the retrographic sensor [Johnson and Adelson 2009] to the microscopic domain, demonstrating spatial resolution as small as 2 microns. In contrast to existing microgeometry capture techniques, the system is not affected by the optical characteristics of the surface being measured—it captures the same geometry whether the object is matte, glossy, or transparent. In addition, the hardware design allows for a variety of form factors, including a hand-held device that can be used to capture high-resolution surface geometry in the field. We achieve these results with a combination of improved sensor materials, illumination design, and reconstruction algorithm, as compared to the original sensor of Johnson and Adelson [2009].

CR Categories: I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Geometry; I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

Keywords: geometry, texture, material, microstructure

Links: ◆DL 🗎PDF 🌐WEB

[*]e-mail: kimo@csail.mit.edu
[†]e-mail: fcole@csail.mit.edu
[‡]e-mail: alvin@csail.mit.edu
[§]e-mail: adelson@csail.mit.edu

1 Introduction

This paper presents a new system for capturing microscopic surface geometry of a wide range of materials, including translucent materials such as human skin. Our system adopts the retrographic sensor approach of Johnson and Adelson [2009] and extends it to allow fast, accurate capture of surface detail with spatial resolution as fine as 2 microns.

Current systems for capturing fine scale surface detail based on active light scanning [Levoy et al. 2000; Alexander et al. 2009] or photometric stereo [Woodham 1980; Tagare and de Figueiredo 1991; Hernández et al. 2007] can capture detail at sub-millimeter resolution. Systems based on shape-from-focus [Nayar and Nakagawa 1994] can resolve microscopic surface detail under certain conditions.

Systems based on passive or active scanning, however, are often confounded by the optical properties of surfaces at the microscopic scale. Most scanning systems based on active light, for example, assume an opaque, diffuse subject material. While this assumption often holds at the macro scale, it generally does not hold at the micro scale. For example, paper appears matte at the macro scale, but when viewed at a micro scale the individual cellulose fibers are transparent and specular. Diffuse paint can sometimes be applied to the subject to alleviate these issues, but paint has many disadvantages. It is inconvenient or impossible to paint many surfaces, and it is difficult to attain a coating that preserves the detail of a microstructured surface.

To circumvent these difficulties, commercial instruments for estimating depth at the micron scale or below use sophisticated techniques such as white light interferometry or scanning focal microscopy. These laboratory-based devices tend to be large, slow, and expensive ($100,000 or more).

The retrographic sensor proposed by Johnson and Adelson is immune to the problems posed by transparent or specular surfaces, because the sensor skin imposes a known BRDF. However, limitations in the sensor material, lighting design, and reconstruction algorithm prevented the original retrographic sensor from achieving the fidelity possible with our system. This paper introduces a new sensor material, an accompanying new lighting design, and a

new near-field photometric stereo algorithm that handles spatially varying illumination and cast shadows. These advances allow for more accurate measurement than the previous system, and increase the spatial resolution by an order of magnitude in both the x and y directions, yielding images with as many as 1 million pixels per mm^2 (i.e., imaging at 1 micron per pixel). These same advances also simplify and make practical the construction of a compact device for capturing surface geometry in the field. We demonstrate a bench configuration (Figure 1a) and portable version (Figure 1d), both of which are constructed from low cost, easily available components.

2   Related work

Our system builds on the work of Johnson and Adelson [2009]. In their paper, they describe a method for measuring the shape and surface texture of many materials with a single image, by using a block of clear elastomer with a thin layer of elastic paint on one side (a *retrographic sensor*). When an object is pressed into the elastomer, it is non-destructively "painted" by the reflective paint, removing the influence of its underlying optical properties on the shape estimation. As a result, a standard photometric stereo algorithm can be used to estimate the surface normals of the object.

In previous work, specialized reconstruction systems were developed to exploit the properties of certain materials, such as shininess, to capture detail. Specular surfaces allow for detailed geometry capture because specular highlights are highly localized, are not influenced by scattering within the material and the physics of specular reflection is simple and well understood [Nayar et al. 1990]. Wang and Dana [2006] measured mesostructure of specular surfaces at a spatial resolution of 0.1 mm using an imaging device with a concave parabolic mirror to view multiple angles in a single image. Their approach estimated surface orientation by detecting the position of the specularity in an image for a given light position. Chen et al. [2006] proposed a similar system using off-the-shelf hardware: a camera, a hand-held light, and a calibration target. Francken et al. [2008] further improved the performance of this approach using coded illumination. While these techniques show promising results, they are limited to glossy surfaces.

Shape-from-focus is one approach for capturing microgeometry of textured surfaces [Nayar and Nakagawa 1994]. The method progressively varies the distance between the camera and the specimen, for each pixel estimating a depth based on the distance that provides sharpest focus. The accuracy of shape-from-focus depends on the ability to make an effective estimate of focus. Smooth or translucent materials may confound the focus estimate and produce inaccurate results.

Effort has also been placed in hybrid approaches that simultaneously infer shape and reflectance from photographs [Wang and Dana 2004; Goldman et al. 2005; Paterson et al. 2005; Ruiters and Klein 2009]. With sufficient input these methods can produce good results, but must make some assumptions (such as moderately diffuse BRDFs [Paterson et al. 2005]) because the simultaneous reflectance and geometry problem is fundamentally underconstrained. In contrast, our method removes the reflectance estimation problem through the use of the retrographic sensor, leaving us the much easier problem of estimating geometry given known reflectance.

The Johnson and Adelson system was able to capture fine details, such as the ridges of a fingerprint, or the raised ink on a twenty-dollar bill. However, the thickness of the reflective skin, the size of the metal flake particles, and the random orientations of the flakes all reduce the possibility of resolving surface microstructure (Figure 2). In addition, the illumination and reconstruction algorithms

(a) Metal-flake pigment   (b) Silver powder
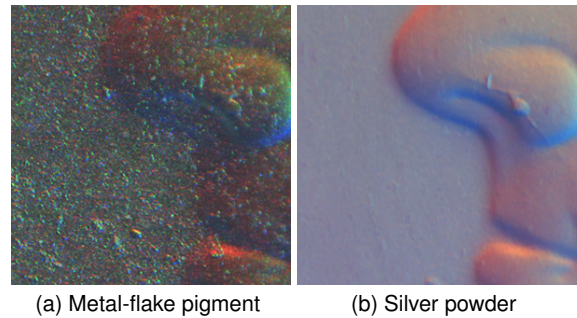
Figure 2: *The size and shape of the reflective pigments impose a limit on the resolution of the system. Both images show Washington's nose on a US quarter. (a) Using metal-flake pigment, the size of the flakes and their random orientations are visible in the image. These effects will cause noise in the estimated surface. (b) With silver powder, the average particle size is below 1 micron. The near-spherical shape reduces noise due to random particle orientation.*

were optimized for single-frame capture to allow for real-time reconstruction, at the expense of accuracy. By improving the pigments, illumination, and the surface estimation algorithm, we are able to achieve results at resolutions an order of magnitude greater than before with improved accuracy.

3   Methods

The retrographic sensor is a slab of clear elastomer with a reflective skin. When an object is pressed into the sensor, the reflective skin conforms to the shape of the object's surface. Viewed through the elastomer, the surface appears to be painted by the reflective skin. This property allows for simple computer vision algorithms (e.g., photometric stereo) to be used to estimate the shape of the surface.

While simple in concept, the performance of the retrographic sensor depends heavily on the construction of the sensor skin, the elastomer used for the sensor itself, the illumination design used for acquisition, and the reconstruction algorithm. We describe here the improvements to hardware and software that enable both microgeometry capture in a laboratory setting and fine scale capture in a portable device.

3.1   Acquisition Hardware

The original work of Johnson and Adelson used a sensor skin constructed with a metal-flake pigment. The metal flakes are shiny and the reflected intensity changes rapidly for even small deflections in the surface normal, making shallow surface relief easily visible. However, metal-flake pigment has two important drawbacks. First, when viewed at high resolution, the flakes themselves are visible and introduce noise (Figure 2a). Second, the narrow reflection lobe of the shiny pigment limits the range of surface angles for which a usefully bright image can be recovered (Figure 3a).

To allow for high-fidelity capture at microscopic scales, we instead construct the skin from pigment containing near-spherical particles of silver less than 1 micron across.[1] This extremely fine powder appears as a smooth coating even at high resolution (Figure 2b). We chose a metallic pigment because it is opaque even in a thin layer. The thickness of the skin affects the ability of the sensor to

---

[1] 47MR-21S, Inframat Advanced Materials (www.advancedmaterials.us)
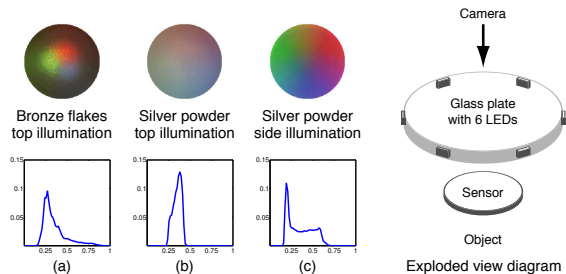
Figure 3: *Illumination design for different pigments. (a) A metallic pigment provides good contrast for many lighting conditions, but loses intensity at the steep sides of the sphere. (b) A diffuse pigment under the same illumination as (a) has limited contrast. (c) Grazing illumination improves contrast across the sphere. Luminance histograms are shown below each calibration sphere. Grazing illumination is achieved by placing surface mount LEDs along the edge of a glass plate, as shown in the exploded view diagram.*

resolve fine surface detail. A thick skin acts as a mechanical low-pass filter, attenuating fine detail. For a high-resolution sensor, the skin should be at most a few microns thick. Dielectric pigments such as titanium dioxide are not sufficiently opaque in such thin layers. Metallic pigments, however, are reasonably opaque even in thin layers. We even apply the silver pigment without a binder to further limit thickness.

When illuminated, the powder appears diffuse and rather dark. Counterintuitively, a dark sensor is desirable as it reduces inter-reflection between surface features. The diffuse surface reflects illumination at the full range of surface orientations, but can present problems for capture if the illumination is not designed to create sufficient contrast. Figure 3b shows a calibration sphere pressed into the silver powder skin with illumination similar to the original retrographic sensor; the corresponding intensity histogram is below. The lack of contrast reduces the discriminability of surface orientations and increases the sensitivity to noise.

For a diffuse surface, contrast is maximized under grazing illumination. We introduce a new lighting configuration to create grazing illumination across the sensor. Six surface-mount LEDs are spaced equally around the edge of a glass disc. The glass disc is used as a mounting plate for the elastomeric sensor and the light from the LEDs propagates within the disc by total internal reflection. Each LED provides a different lighting condition for photometric stereo. This illumination design dramatically increases contrast, as shown in Figure 3c.

The resolution of the sensor is also dependent on the rigidity of the elastomer. Following Johnson and Adelson, we use 3M VHB mounting tape for our high-resolution results for hard surfaces. For skin and other soft surfaces, we use a thick slab of very soft gel elastomer, to avoid applying too much pressure. We typically use thermoplastic elastomers like those in the Versaflex series from GLS Corp[2], which can be made even softer by the addition of plasticizer oil.

We have created two embodiments of our system, one for microgeometry capture in a laboratory environment, and one for fine-scale capture in the field. Both share the same sensor and lighting design, but have different optics.

---

[2] http://www.glscorp.com

Bench configuration The bench configuration (Figure 1a) consists of a 18-megapixel Canon EOS Rebel T2i camera with a Canon MP-E 65 mm macro lens mounted vertically over an optical bench. The elastomeric sensor is mounted on a 0.5-inch thick, 5.5-inch diameter glass plate with six LEDs evenly spaced around the perimeter. The glass plate and sensor are secured to the subject material with toggle clamps.

Portable configuration The portable configuration (Figure 1d) is constructed from an acrylic tube with 3-inch outside diameter. The tube is approximately 8 inches long. The sensor is mounted on the exterior of a 0.25-inch thick, 2.25-inch diameter glass plate at one end of the tube. Our grazing illumination configuration leaves the interior of the tube free for a 0.8-megapixel Point Grey Flea2 firewire camera (1032×776 pixels). When an exterior button is pressed, the system rapidly captures the six lighting conditions.

3.2 Reconstruction Algorithm

The grazing illumination achieved by mounting LEDs along the edge of a glass plate is effective at creating contrast, but it no longer fits the assumptions of traditional photometric stereo techniques. In particular, the illumination varies spatially across the image and shadows are possible if the object has large depth discontinuities. The lookup-table approach, used for the original retrographic sensor, assumes distant illumination and does not model cast shadows. To improve the accuracy of our surface normal estimation, we developed a new photometric stereo method that accounts for both spatially-varying illumination and cast shadows.

In order to model the spatially-varying illumination in our system, we constructed a dense calibration target, shown in Figure 4. The target consists of an array of spheres (radius 1.5mm) evenly spaced in a grid. Our system automatically calibrates from two images of the target (6 lighting conditions stored as color channels). The automatic calibration process takes advantage of the even spacing of the spheres and the sharp boundary caused by the grazing illumination to accurately locate the spheres in the image.

At each sphere in the image, we fit two illumination models: linear and quadratic. The linear model is a good initial approximation of the illumination, but we find that it is not sufficiently accurate for our choice of pigment and illumination. The quadratic model accounts for the fact that our pigment is not perfectly Lambertian and that our illumination is not well-approximated by a point light source due to internal reflections within the glass disc.

The linear model is similar to the model used in other photometric stereo methods [Barsky and Petrou 2003; Basri et al. 2007]. In particular, the model assumes Lambertian reflectance, constant albedo and a single point light source positioned at infinity. For surface normals facing the light direction, we model the intensity in channel k as:

$$s_k(\vec{N}) = \rho_k \vec{L}_k^T \vec{N}(\vec{p}) , \qquad (1)$$

where $\vec{N}(\vec{p})$ is the surface normal at position $\vec{p}$, $\vec{L}_k$ is the light direction and ρ_k is the albedo for channel k. Since the reflective skin has constant albedo, the factor ρ_k is absorbed into the vector $\vec{L}_k$ as its length. Thus, the linear model has three components which are estimated from the calibration sphere using least squares.

The quadratic model is derived from a truncated spherical-harmonic shading model, which has been shown to be a good approximation for Lambertian reflectance under arbitrary illumination conditions [Basri and Jacobs 2003; Ramamoorthi and Hanrahan 2001]:
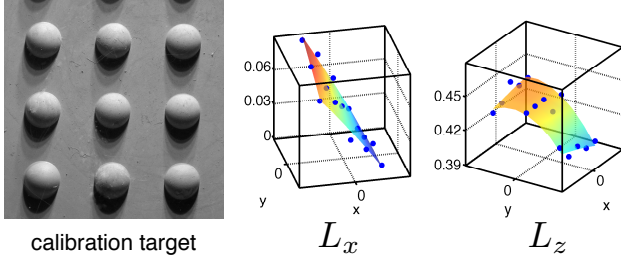
Figure 4: *Modeling spatially-varying illumination. The relative positions of the lights varies across the image, as shown by the different shadows cast by each sphere on the calibration target. A linear model with three coefficients (L_x, L_y, and L_z) is fit independently to each sphere. The spatial variation of these coefficients is modeled with a quadratic surface (shown colored).*

$$s_k(\vec{N}) = \sum_{n=0}^{2} \sum_{m=-n}^{n} l_k^{n,m} Y^{n,m}(\vec{N}) \,, \qquad (2)$$

where $Y^{n,m}$ is the m-th spherical harmonic basis function of order n and $l_k^{n,m}$ is the associated illumination coefficient for channel k. The quadratic model has nine coefficients which are estimated from the calibration sphere using least squares. We find that this model can fit the appearance of our diffuse (but non-Lambertian) pigment under our particular illumination design.

We observe that the coefficients of both the linear and quadratic illumination models vary smoothly over the image. For example, consider the image of a calibration target in Figure 4. A simple visual inspection reveals that the light direction is more grazing for the bottom spheres in the grid (L_z is smaller). The light direction is also more vertical for the right spheres in the image (L_x is smaller). To model the spatial variation in the coefficients, we fit a quadratic surface using the center of each sphere as the spatial position of the coefficient. On the right of Figure 4, we show the measured coefficients for the linear model at each sphere and the quadratic surface. Note that the shape of each surface matches our intuition of the lighting variation.

3.2.1 Linear estimate

Given a set of m observed intensities, I_1 to I_m, at a pixel $\vec{p}$, our goal is to estimate the surface normal that produced those intensities under the known illumination model. We begin with the linear model to obtain an initial surface normal estimate $\vec{N}$. We then refine the estimate using the more accurate quadratic illumination model.

At each pixel $\vec{p}$ in the image we obtain an local light direction estimate for channel k from the spatially-varying linear illumination model. We denote this light direction $\vec{L}_k$. The surface normal at the position $\vec{p}$ is estimated via least squares:

$$
\begin{aligned}
\vec{N} &= \begin{bmatrix} \vec{L}_1^T \\ \vdots \\ \vec{L}_m^T \end{bmatrix}^+ \begin{bmatrix} I_1(\vec{p}) \\ \vdots \\ I_m(\vec{p}) \end{bmatrix} , \\
&= L^+ \vec{I}(\vec{p}) \,,
\end{aligned}
\qquad (3)
$$

where the $+$ denotes pseudoinverse.

In the presence of cast shadows, however, the estimate will be biased. To reduce the influence of shadows, we define a binary diagonal weight matrix W where the k-th entry is zero if the intensity

in channel k is below a predefined threshold τ (e.g., $\tau = 5/255$). The weighted version of Eqn. 3 is:

$$\vec{N} = (WL)^+ W \vec{I}(\vec{p}) \,. \qquad (4)$$

If more than three channels at pixel p are in shadow, we set all weights to 1. Pixels that are mostly in shadow will be processed separately during the nonlinear refinement stage.

3.2.2 Nonlinear refinement

While many photometric stereo approaches are based on the linear model in Eqn. 3, we find that the model is not sufficiently accurate in our setup (Figure 6b). The nonlinear shading model, Eqn. 2, accounts for the fact that our pigments are not perfectly Lambertian and our illumination is not well-approximated by a point light source due to internal reflections within the glass disc. Starting with the solution of the linear model, we improve the surface normal estimates using a nonlinear refinement step.

The nonlinear refinement step uses the quadratic shading model in Eqn. 2. As with the linear model, the coefficients vary spatially and the variation is modeled during calibration. At position $\vec{p}$, the nine coefficients are interpolated to approximate the local illumination. To simplify notation, we rewrite the spherical-harmonic shading model as a quadratic function of the surface normal using the derivations provided in [Ramamoorthi and Hanrahan 2001]. Let the matrix A_k, vector $\vec{b}_k$, and scalar c_k represent the local shading function for channel k:

$$s_k(\vec{N}) = \vec{N}^T A_k \vec{N} + \vec{b}_k^T \vec{N} + c_k \,, \qquad (5)$$

We define an error function E_1, with a diagonal weight matrix W to account for shadows:

$$E_1(\vec{N}) = \|\sqrt{W} \vec{f}(\vec{N})\|^2 = \|\sqrt{W} \left(\vec{s}(\vec{N}) - \vec{I}(\vec{p}) \right) \|^2 \,, \qquad (6)$$

where the vector $\vec{I}$ represents the color values across all channels at pixel $\vec{p}$.

To optimize Eqn. 6, we use an iterative scheme, such as a Gauss-Newton iteration. At a current surface normal estimate $\vec{N}_i$, the Gauss-Newton update rule is:

$$J(\vec{N}_i)^T W J(\vec{N}_i)\vec{h} = -J(\vec{N}_i)^T W \vec{f}(\vec{N}_i) \,, \qquad (7)$$

where $J(\vec{N}_i)$ is the Jacobian matrix (i.e., the matrix of partial derivatives) of the function $\vec{f}$ at the current estimate $\vec{N}_i$. The update vector $\vec{h}$ satisfying Eqn. 7 is added to the current estimate: $\vec{N}_{i+1} = \vec{N}_i + \vec{h}$.

Surface normals are not general vectors in $\mathcal{R}^3$ and the update vector $\vec{h}$ can move the estimate off the surface of the unit sphere, slowing convergence. To address this issue, we parameterize the surface normal using 2D coordinates u and v as follows:

$$\vec{N}(u,v) = R_0 \begin{bmatrix} u & v & r \end{bmatrix}^T \,, \qquad (8)$$

where $r = \sqrt{1 - u^2 - v^2}$ and R_0 is a rotation matrix such that R_0^{-1} maps the initial estimate $\vec{n}_0$ to $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$.

With this parameterization, the Jacobian can be expressed in terms of coordinates u and v through the chain rule:

$$\begin{bmatrix} \dfrac{\partial J}{\partial u} & \dfrac{\partial J}{\partial v} \end{bmatrix} = \dfrac{\partial \vec{f}}{\partial \vec{n}} R_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -u/r & -v/r \end{bmatrix} . \qquad (9)$$

Figure 5: *Shadow handling via iteratively reweighted least squares. At the end of the optimization, the weights for each image primarily correspond to shadow regions in the image.*

Since this parameterization is only defined for the hemisphere about the original surface normal estimate $\vec{n}_0$, the Jacobian will be undefined if $u^2 + v^2 > 1$. Therefore, we recompute the rotation matrix based on the current surface normal estimate when $u^2 + v^2 \geq 1/2$.

3.2.3 Coarse-to-fine propagation

The nonlinear update rule is applied iteratively to each pixel in a coarse-to-fine fashion. We use a propagation strategy and curl constraint to maintain smoothness and integrability. At position $\vec{p}$, we define a local error function:

$$E(\vec{p}) = E_1(\vec{N}(\vec{p})) + \omega E_2(\vec{N}(\vec{p}), \vec{N}(\vec{p}_l), \vec{N}(\vec{p}_u)) , \quad (10)$$

where ω is a scalar weight and E_2 is a constraint on the curl:

$$E_2(\vec{N}, \vec{U}, \vec{V}) = \|V_x/V_z - N_x/N_z - U_y/U_z + N_y/N_z\| . \quad (11)$$

The subscripts l and u indicate adjacent pixels to the current position (i.e., left and up).

The propagation strategy refines the local surface normal estimate and then refines the minimum of the adjacent estimates using the error function for the current position, Eqn. 6. The minimum of these two surface normal estimates using the local error function, Eqn. 10, is kept as the current estimate. The weight matrix is then updated for the next iteration as follows:

$$w_k = \begin{cases} 0 & \text{if } I_k < \tau, \\ \epsilon/\max\left(|s_k(\vec{N}) - I_k|, \epsilon\right) & \text{otherwise} . \end{cases} \quad (12)$$

Note that this optimization strategy is essentially an iteratively reweighted least squares approach (IRLS) [Daubechies et al. 2010].

3.2.4 Shadow pixels

The iteratively reweighted least-squares approach is effective at discounting shadow pixels when at least two of the channels are not in shadow. For example, consider the input image and weight map after optimization shown in Figure 5. The black pixels have a weight of 0 and note that these pixels primarily correspond to the shadows in the input image.

However, if at pixel $\vec{p}$, $m - 1$ channels are in shadow we use an alternative optimization strategy. We use the same parameterization of the surface normal, Eqn. 8, but update the estimate as follows:

$$\begin{aligned} u_{i,j} &= (u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1})/4 , \\ v_{i,j} &= (v_{i-1,j} + v_{i+1,j} + v_{i,j-1} + v_{i,j+1})/4 . \end{aligned} \quad (13)$$
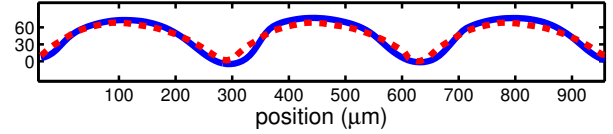
This update rule smoothly propagates surface normal estimates in regions of complete shadow.
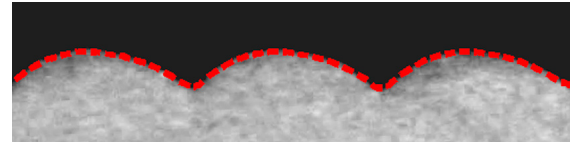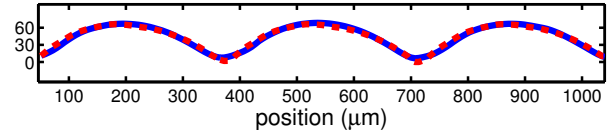

(a) Image of lenticular array with curve fit to profile
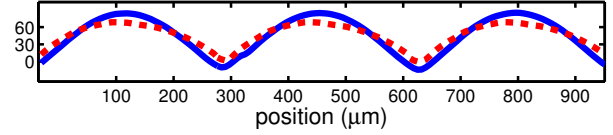
(b) Reconstructed depth profile using the linear estimate

(c) Reconstructed depth profile using a lookup table

(d) Optimized depth profile using our method

Figure 6: *Accuracy validation against a lenticular microlens array. (a) Height profile measured from an image of a lenticular array on its side. (b) Reconstructed depth profile (blue, solid) compared to measured profile (red, dashed) using the linear approach, Eqn. 4; the RMS error is 12.4 microns. (c) Reconstructed depth profile using a lookup table approach, similar to the method in [Johnson and Adelson 2009]; the RMS error is 8.1 microns. (d) Reconstructed depth profile using our method; the RMS error is 3.0 microns.*

3.2.5 Surface reconstruction

Once we have estimated surface normals at every pixel in the image, we reconstruct depth by integrating the normals. This process is common to many shape-from-X and gradient-domain methods, so we only briefly review our approach. We define an error function P on the depth z:

$$P(z) = \left\| \begin{bmatrix} D_x \\ D_y \end{bmatrix} \vec{z} - \begin{bmatrix} \vec{p} \\ \vec{q} \end{bmatrix} \right\|^\rho \quad (14)$$

where D_x and D_y are matrices representing the x and y derivative operators on the vectorized image $\vec{z}$ and the vectors $\vec{p}$ and $\vec{q}$ are $-N_x/N_z$ and $\vec{q}$ is $-N_y/N_z$ at every pixel. To reduce the influence of noisy surface normal estimates on the reconstructed shape, we use an approximate L_1 minimization ($\rho = 1$) through IRLS. Let D be the differentiation matrix and $\vec{r} = \begin{bmatrix} \vec{p} & \vec{q} \end{bmatrix}^T$. Using IRLS, the depth image estimate is updated as follows:

$$\vec{z}_{i+1} = (D^T W_i D)^{-1} D^T W_i \vec{r} , \quad (15)$$

$$W_i = \text{diag}\begin{bmatrix} \epsilon/\max\left(|D_x \vec{z}_i - \vec{p}|, \epsilon\right) \\ \epsilon/\max\left(|D_y \vec{z}_i - \vec{q}|, \epsilon\right) \end{bmatrix} . \quad (16)$$

We stop the IRLS procedure after ten iterations.

3.3 Median noise reduction

There are some unwanted effects in the measurement process that we treat as noise. In particular, there are sometimes random imperfections in the reflective skin, as well as dust or other debris that may become attached to the skin. If we capture multiple images with the sensor in different positions (but with the object fixed with respect to the camera), the position of the debris and imperfections will be randomized. By taking the median across multiple scans, we avoid the influence of large outliers. Note that the median approach is only appropriate if the object being measured is rigid.

Computing the median across multiple images requires image alignment. Although our capture setup is rigid, the act of pressing the sensor into the sample can cause shifts in the image of several pixels. We implemented a simple image alignment algorithm that that is reasonably efficient for this task, even on large images (15 to 18 megapixels). The algorithm performs a hierarchical coarse-to-fine grid search over a predetermined range of spatial translations. As an error criterion, we use the sum of absolute differences, after correcting for multiplicative exposure shifts. To improve efficiency, we perform the computation on a fixed number of windows and pool the estimates. These techniques are fairly common for image alignment, and a good overview can be found in a survey by Szeliski [2006]. Image alignment is also necessary for the raw images from the portable device to reduce the influence of camera shake during the capture sequence.

4 Results

Capturing geometry with our system is straightforward. We captured a range of materials using our system and implemented a basic interactive rendering system to demonstrate the results. One of our subject microgeometries has a known shape, which we use as a basis for validating the accuracy of the system (Section 4.1). We then show the flexibility and practicality of our system through samples from 8 different materials, some of which are very shiny or translucent (Section 4.3).

4.1 Accuracy validation

We validate the accuracy of our method by capturing the geometry of a lenticular sheet, a type of microlens array constructed from parallel cylindrical sections. The profile of the sheet is made up of periodic arcs (Figure 6). To measure the ground-truth profile, we cut a thin slice from the sheet and photographed it on its side at a spatial resolution of approximately 2 microns per pixel. The image of three arcs from the profile, after thresholding, is shown in Figure 6a. We extracted points along the boundary of eight arcs and fit a cubic spline to the extracted points using least-squares. The best-fit spline is shown as a dashed line superimposed on the image.

We captured images from five different sensor positions and computed the median across the images. Using our method, the RMS error in depth between the estimated and actual profiles is 3.0 microns. The initial estimate according to the linear model, Eqn. 4 achieves an RMS error of 12.1 microns. A lookup table approach similar to the method of Johnson and Adelson [2009] achieves an RMS error of 8.1 microns.

4.2 Comparison with previous approach

To demonstrate our improvements over the original retrographic sensor, we show how our methods improve both the resolution and accuracy of the system, especially in regions containing shadows.
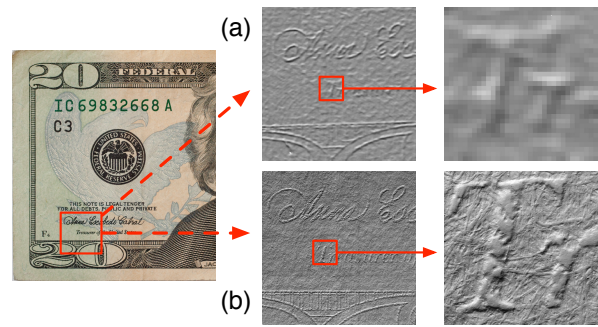
Figure 7: *Comparison with the high-resolution result from the original retrographic sensor. (a) Rendering of the high-resolution $20 bill example from the original retrographic sensor with a close-up view. (b) Rendering of the captured geometry using our method.*


(a) lookup table   (b) our method
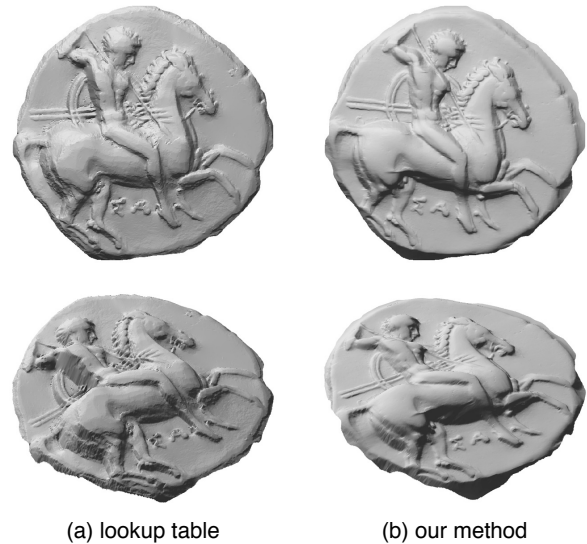
Figure 8: *The effect of shadows in photometric stereo. A Greek coin from the third century BCE, scanned using our system. The depth variation across the coin creates shadows in the input data. (a) A lookup-table based approach does not model shadows and artifacts are visible in the reconstructed surface. (b) Our optimization minimizes the effect of shadows on the reconstruction.*

For a side-by-side comparison of resolution, we acquired the high-resolution scan of the $20 bill that was captured with the original retrographic sensor. We also scanned a region of a $20 bill with our bench setup. In Figure 7 we show both results. Our method can resolve the individual fibers in the bill.

We also demonstrate the importance of handling shadows in photometric stereo. We scanned an ancient Greek coin from the city-state of Taranto. This coin has significant depth variation that create shadows in the input data (Figure 5). A lookup table approach does not account for shadows and artifacts are visible in the reconstructed surface (Figure 8a). Our optimization strategy minimizes the effect of shadows on the reconstruction (Figure 8b).

4.3 Capture examples

To demonstrate the practicality of the system, we captured the surface structure of a range of common materials (Figure 9). Each
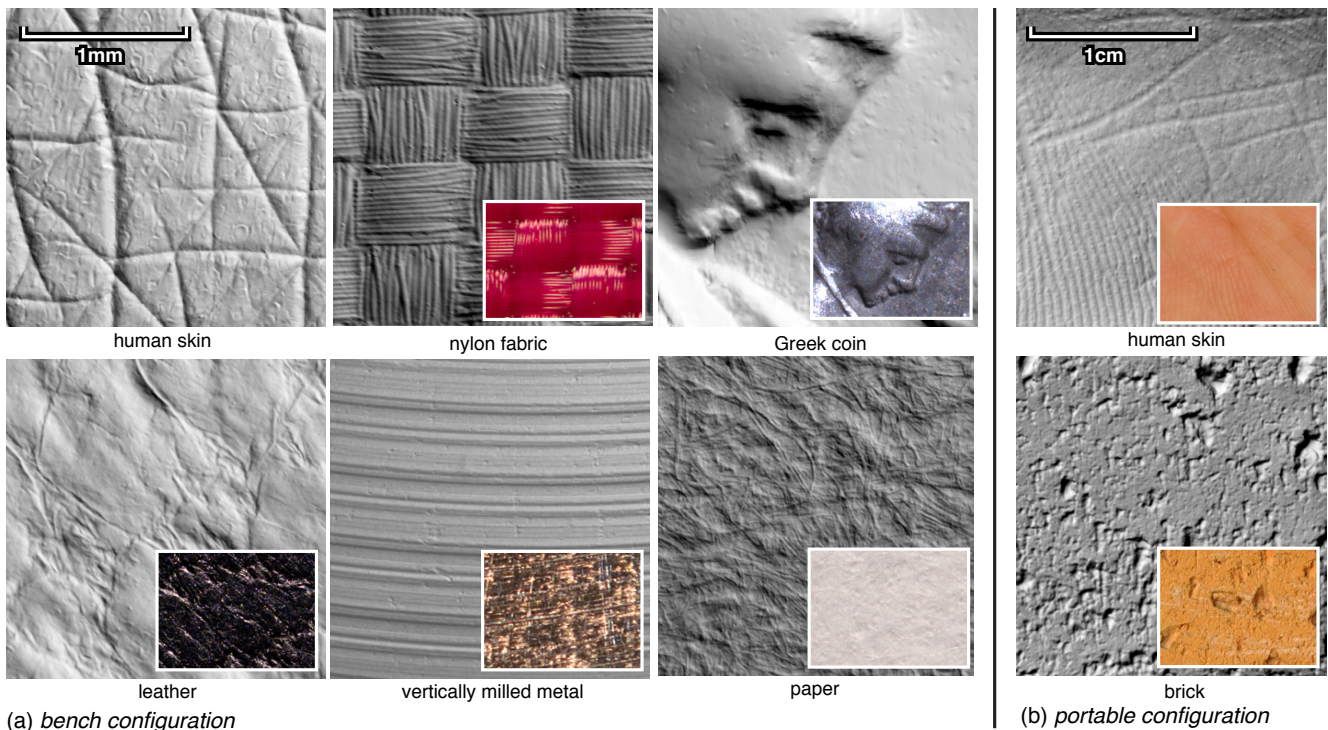
Figure 9: *Example geometry measured with the bench and portable configurations. Outer image: rendering under direct lighting. Inset: macro photograph of original sample. Scale shown in upper left. Color images are shown for context and are to similar, but not exact scale.*
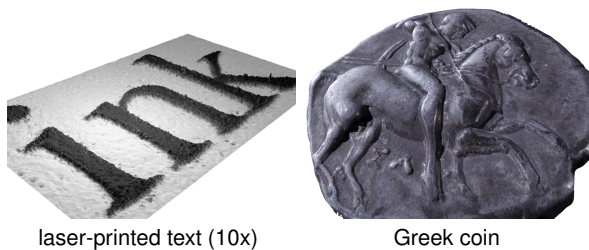
Figure 10: *Geometry rendered with a texture map. left: measured geometry of laser-printed text textured with a color photograph (height scaled by a factor of 10). right: measured geometry of a Greek coin textured with a color photograph.*

patch was reconstructed from a set of photographs, captured using our bench setup. The number of photographs captured depended on the material sample. For rigid materials, we captured 5 images and computed a median across the images to reduce noise. For materials with delicate structure, such as the fabrics, we used a single impression since the structure would be disturbed across scans.

Figure 9a shows the results from our bench configuration along with color photographs for context. The samples include difficult materials like shiny metal and fabric. In many cases, the shape of the sample is hardly recognizable from the color photograph (e.g., vertically milled metal, nylon fabric) illustrating the difficulty of capturing the surface without the retrographic sensor. The bench configuration can capture an additional color photograph to be used as a texture, prior to using the sensor (Figure 10).

To demonstrate our portable device, we captured the surface geometry of human skin and brick (captured outside). These results

are shown in Figure 9b. While currently lower resolution that our bench setup, the portable device enables texture capture in the field.

One challenge we found when preparing our results was conveying the scale of the geometry we captured. As a demonstration of scale, we printed this paragraph of text on clay-coated paper and literally captured the ink on the page.[3] Figure 10 shows the rendered geometry with the height scaled by a factor of 10. A color image of the page was used as a texture map. Note that the accumulation of toner particles is visible.

5 Conclusion

Our system allows for efficient, accurate capture of shallow relief geometry, regardless of the light scattering behavior of the material. While our system builds directly on the system of Johnson and Adelson [2009], we provide important improvements to the sensor construction, lighting design, and reconstruction algorithm. Together, these advances allow us to extend the capabilities of the retrographic sensor in two directions: extremely high resolution, and a portable form factor.

The spatial resolution of our high-resolution, bench configuration allows us to capture microgeometry that has previously been difficult or impossible to measure, such as the narrow grooves of brushed metal, the spacing of threads in fine fabric, or minute and intricate folds of skin. The microgeometry captured by our system is directly useful as source material for digital artists, and with minimal processing can be used as normal or displacement maps. For context, commercial databases of captured texture data (e.g., [XYZRGB 2010]) advertise 100-micron resolution, nearly

---

[3]Without clay-coated paper, we would also capture the geometry of the fibers in the paper.

two orders of magnitude more coarse than our measurements. The portable configuration, while not providing the same level of resolution as the bench system, allows for capturing surface geometry in seconds, in-place, and without harming the subject.

A limitation of our system is that it can only capture shallow relief geometry. While the sensor can be applied to any material, surfaces with holes will be captured with smoothed areas where the sensor fails to contact the subject. A related limitation is that fuzzy or hairy surfaces will not be captured faithfully, as the sensor tends to flatten the material. Softer elastomers can reduce this limitation to some extent, but it remains a fundamental restriction.

Our system is complementary to current systems for capturing geometry such as laser range and structured light scanning. The major tradeoff we make to achieve high-resolution is a small working area for capture. While future systems could expand the working area or increase the rate of capture, we believe a more appropriate approach is to combine small patches of microgeometry with millimeter-scale geometry captured through conventional systems. Example-based super-resolution [Freeman et al. 2002] or multi-scale texture synthesis [Han et al. 2008] could be used to extrapolate from the millimeter-scale geometry based on samples of microgeometry. Using such a combination of methods, we could construct a full model of geometry from the macro to the micro scale.

Acknowledgements

References

ALEXANDER, O., ROGERS, M., LAMBETH, W., CHIANG, M., AND DEBEVEC, P. 2009. Creating a photoreal digital actor: the digital Emily project. *IEEE European Conference on Visual Media Production* (Jun), 1–12.

BARSKY, S., AND PETROU, M. 2003. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 10, 1239–1252.

BASRI, R., AND JACOBS, D. W. 2003. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 2, 218–233.

BASRI, R., JACOBS, D., AND KEMELMACHER, I. 2007. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision 72*, 3, 239–257.

CHEN, T., GOESELE, M., AND SEIDEL, H.-P. 2006. Mesostructure from specularity. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

DAUBECHIES, I., DEVORE, R., FORNASIER, M., AND GÜNTÜRK, C. S. 2010. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics 63*, 1, 1–38.

FRANCKEN, Y., CUYPERS, T., MERTENS, T., GIELIS, J., AND BEKAERT, P. 2008. High quality mesostructure acquisition using specularities. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

FREEMAN, W., JONES, T., AND PASZTOR, E. 2002. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 56–65.

GOLDMAN, D., CURLESS, B., HERTZMANN, A., AND SEITZ, S. 2005. Shape and spatially-varying brdfs from photometric stereo. *Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005 1*.

HAN, C., RISSER, E., RAMAMOORTHI, R., AND GRINSPUN, E. 2008. Multiscale texture synthesis. *SIGGRAPH '08: SIGGRAPH 2008 papers*.

HERNÁNDEZ, C., VOGIATZIS, G., BROSTOW, G. J., STENGER, B., AND CIPOLLA, R. 2007. Non-rigid photometric stereo with colored lights. In *Proc. of the 11th IEEE Intl. Conf. on Comp. Vision (ICCV)*.

JOHNSON, M. K., AND ADELSON, E. H. 2009. Retrographic sensing for the measurement of surface texture and shape. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1070–1077.

LEVOY, M., PULLI, K., CURLESS, B., RUSINKIEWICZ, S., KOLLER, D., PEREIRA, L., GINZTON, M., ANDERSON, S., DAVIS, J., GINSBERG, J., SHADE, J., AND FULK, D. 2000. The digital Michelangelo project: 3D scanning of large statues. *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (Jul).

NAYAR, S. K., AND NAKAGAWA, Y. 1994. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*, 8, 824–831.

NAYAR, S. K., SANDERSON, A. C., WEISS, L. E., AND SIMON, D. A. 1990. Specular surface inspection using structured highlight and Gaussian images. *IEEE Transactions on Robotics and Automation 6*, 2, 208–218.

PATERSON, J., CLAUS, D., AND FITZGIBBON, A. 2005. BRDF and geometry capture from extended inhomogeneous samples using flash photography. *Computer Graphics Forum 24*, 3, 383–392.

RAMAMOORTHI, R., AND HANRAHAN, P. 2001. An efficient representation for irradiance environment maps. In *Proc. of SIGGRAPH*, 497–500.

RUITERS, R., AND KLEIN, R. 2009. Heightfield and spatially varying BRDF reconstruction for materials with interreflections. *Computer Graphics Forum 28*, 2, 513–522.

SZELISKI, R. 2006. Image alignment and stitching: a tutorial. *Foundations and Trends in Computer Graphics and Vision 2*, 1.

TAGARE, H., AND DE FIGUEIREDO, R. J. 1991. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence 13*, 2, 133–152.

WANG, J., AND DANA, K. J. 2004. Hybrid textons: Modeling surfaces with reflectance and geometry. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 372–378.

WANG, J., AND DANA, K. J. 2006. Relief texture from specularities. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*, 3, 446–457.

WOODHAM, R. J. 1980. Photometric method for determining surface orientation from multiple images. *Optical Engineering 19*, 1, 139–144.

XYZRGB. 2010. *Reptile DVD: Real World Materials*. XYZ RGB Incorporated.