A Perceptually Validated Model for Surface Depth Hallucination

Mashhuda Glencross[*]
The University of Manchester

Gregory J. Ward[†]
Dolby Canada

Caroline Jay[‡]
The University of Manchester

Jun Liu[§]
The University of Manchester

Francho Melendez[¶]
The University of Manchester

Roger Hubbold[‖]
The University of Manchester

Figure 1: *The left image is a photograph of a Mayan carving under diffuse lighting, which was combined with a similar flash photo to derive a height field and albedo map for this surface. The middle image uses the derived model to render the same view with novel lighting. The rendering on the right shows an oblique close-up with a second novel lighting condition and added specularity.*

Abstract

Capturing detailed surface geometry currently requires specialized equipment such as laser range scanners, which despite their high accuracy, leave gaps in the surfaces that must be reconciled with photographic capture for relighting applications. Using only a standard digital camera and a single view, we present a method for recovering models of predominantly diffuse textured surfaces that can be plausibly relit and viewed from any angle under any illumination. Our multiscale shape-from-shading technique uses diffuse-lit/flash-lit image pairs to produce an albedo map and textured height field. Using two lighting conditions enables us to subtract one from the other to estimate albedo. In the absence of a flash-lit image of a surface for which we already have a similar exemplar pair, we approximate both albedo and diffuse shading images using histogram matching. Our depth estimation is based on local visibility. Unlike other depth-from-shading approaches, all operations are performed on the diffuse shading image in image space, and we impose no constant albedo restrictions. An experimental validation shows our method works for a broad range of textured surfaces, and viewers are frequently unable to identify our results as synthetic in a randomized presentation. Furthermore, in side-by-side comparisons, subjects found a rendering of our depth map equally plausible to one generated from a laser range scan. We see this method as a significant advance in acquiring surface detail for texturing using a standard digital camera, with applications in architecture, archaeological reconstruction, games and special effects.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture;

Keywords: relighting, shape-from-shading, albedo estimation, textured surfaces, perception, computational photography

[*]e-mail: mashhuda@manchester.ac.uk

[†]e-mail: gward@dolby.com

[‡]e-mail: caroline.jay@manchester.ac.uk

[§]e-mail: jun.liu@manchester.ac.uk

[¶]e-mail: francho.melendez@manchester.ac.uk

[‖]e-mail: roger.hubbold@manchester.ac.uk

1 Introduction

Textured surfaces such as brick, stone, wood and many other building materials have local variations in their surface *meso-structure*. Shading variations due to self-shadowing provide important perceptual cues necessary to convey a correct impression of shape. An interesting question, however, is how accurate does surface meso-structure need to be for shape and corresponding shading to appear plausible? This is an important question as our objective is to produce synthetically relit results that are perceptually difficult to distinguish from photographs. (See Figure 1.) In this paper we show that an approximate representation of the real surface (depth + albedo map) may be used to relight predominantly diffuse textured surfaces in a visually plausible manner. To this aim, we introduce a practical method to recover approximate surface texture information from a single viewpoint. From a 2D picture, we infer surface depth where it is not fully divulged in the image. We call this *depth hallucination*.

Representing surface detail is useful to increase the visual realism in a range of application areas, especially architectural reconstructions. In particular, accurately assessing the effect of new buildings on lighting requires modeling of gross 3D geometry, meso-structure, and *albedo* (equivalent to diffuse reflectance), so that simulations of appearance at different times of the day are possible. Our method is aimed primarily at the materials recovery part of such architectural reconstructions. We aim to acquire surface meso-structure so that it may be combined with gross 3D geometry (obtained using another method) to convey the appearance of

visually realistic surface detail. However, this technique is equally applicable to recovering and representing surface detail for use in graphically rich games and movies. Currently, our method is being used to recover depth maps at Chichén Itzá (Mexico), for the production of a dome-projected movie.

Our main contribution is a novel, experimentally validated shape-from-shading method, which takes diffuse-lit/flash-lit image pairs and produces a plausible textured height field that can be viewed from any angle under any lighting. In the absence of a flash-lit image, we apply histogram matching against a visually similar texture for which we have recovered a model from captured pairs. This practical optimization simplifies the capture requirements for large surfaces composed of the same material but containing significant meso-structure variation. Since our goal is to recover enough surface detail for plausible relighting, accuracy requirements are purely perceptual and are evaluated based on the final imagery. To date, no published method for recovering and relighting textured height fields has been validated against equivalent photographs. Our experimental studies demonstrate that participants cannot reliably identify our relit images as synthetic, and more importantly that they believe these to be as plausible as geometrically correct laser-scanned reconstructions.

2 Previous Work

Creating 3D models directly from photographs is appealing since it offers the potential of economical acquisition for photorealistic visualization. The landmark method of Debevec et al. [1996] produces visually pleasing results for architectural applications. However, detailed surface meso-structure of building materials is rarely considered in such models. To correctly relight different materials requires separation of the way surfaces scatter light and the actual light striking the surface. Although solutions to separate these under specific constraints have been proposed [Narasimhan et al. 2003], the problem is not generally solvable. To fill in the missing information, humans use tacit knowledge gained from experience of real world illumination to estimate material properties [Fleming et al. 2003]. A number of meso-structure recovery methods capture normal and texture maps with multiple sources [Rushmeier and Bernardini 1999; Lensch et al. 2003]. An accurate but data intensive approach is to capture and encode the appearance of textured surfaces with a gantry under a large number of lighting and viewing conditions [Dana et al. 1999]. Other methods to recover albedo and meso-structure exist, but require sets of images and/or specialized equipment [Yu et al. 1999; Li et al. 2006; Ngan and Durand 2006; Paterson et al. 2005].

Classic shape-from-shading solutions aim to acquire 3D depth information from a single image [Koenderink and van Doorn 1983; Horn 1989; Malik and Maydan 1989; Haddon and Forsyth 1998a; Haddon and Forsyth 1998b; Prados and Faugeras 2005]. This is an under-constrained problem. Numerous shapes, surface reflectances, and lighting conditions can give rise to the same shading pattern [Belhumeur et al. 1999], and associated ambiguities in shape perception [Ramachandran 1988]. However, shape-from-shading approaches are attractive for our application as they do not require special equipment or lengthy data-capture processes. Khan et al. [2006] successfully demonstrated how, under certain circumstances, limitations in our ability to correctly interpret depth and lighting [Ostrovsky et al. 2005] can be exploited to create plausible synthetic images from a dark-is-deep approximation [Langer and Bülthoff 2000]. Our depth hallucination approach is inspired by their ideas.

A large body of literature on the topic of shape-from-shading exists, and we refer to published surveys for a review of existing meth-
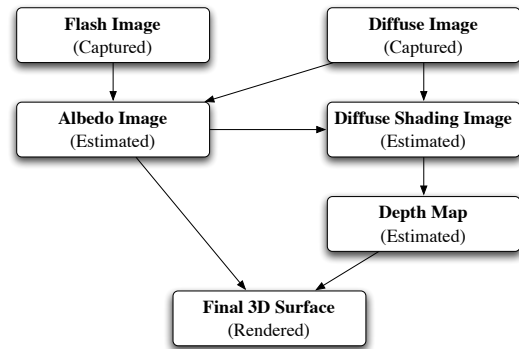
Figure 2: *Flow chart showing the steps in our process.*

ods [Zhang et al. 1999]. Broadly, our approach performs irradiance estimation and is similar in spirit to the iterative technique of Langer and Zucker [1994]. Langer and Zucker's model is specifically designed for recovering shape-from-shading on a cloudy day. They observe that under diffuse lighting, surface luminance depends primarily on a local aperture function defined as the solid angle subtended by the visible sky at each surface point. They formulate a set of constraints, applying a robust numerical approach to solve for depth. However, there are a few practical limitations to their method. First, the model assumes uniform albedo, which is a problem for a wide range of textured surfaces. Second, their approach suffers from quantization errors since they perform discretized sampling of light source directions over a hemisphere at each point on a hypothetical surface. Third, their implementation is based on an iterative ray-tracing scheme, which is computationally expensive. Since our goal is to recover sufficient depth for plausibly relighting textured surfaces, we develop a simpler, deterministic image-space solution that approximates their results.

3 Depth Hallucination Method

We assume our surface can be plausibly represented as a height field, whose underlying material matrix is approximately Lambertian and opaque, with average reflectance, $2\% < \rho < 70\%$ or so. Our overall process is illustrated in Figure 2. The individual steps are image capture, albedo and shading estimation, depth estimation, and relighting the surface. Specifics of how we estimate albedo and shading depend on whether the input to our process is a diffuse-lit/flash-lit image pair [Eisemann and Durand 2004], or a single diffuse-lit image. Subtracting the diffuse-lit image from the flash-lit image gives a reasonable estimate of albedo, and a comparison of our diffuse-lit image and albedo provides a usable estimate of diffuse shading for depth estimation. We discuss this in further detail in Section 3.2. Our depth estimation method is described in 3.3, and rendering of our final images is described in 3.4. Throughout these sections, we illustrate the steps in our process with a case study of a brick path and show the output of each intermediate step.

3.1 Image Capture

To capture our input images, we employ a standard digital SLR camera mounted on a tripod, and an attached strobe. Our method requires that we capture a sample of the textured surface without global curvature, as might be found on a wall or floor. If the textured surface contains significant specularities, cross-polarization (i.e., the polarizer on the flash is perpendicular to the polarizer on the lens) can be used to minimize highlights [Hershberger 2008].

First we capture a RAW format image[1] under indirect illumination (i.e., overcast skies or shadow). We call this the diffuse-lit condition. A second photo is taken from the same point with the flash fired at full power. The camera is set to its maximum flash synchronization speed, while position, aperture, and focus are fixed to ensure good pixel registration between the diffuse-lit and flash-lit conditions. Ideally, the flash should be mounted as close to the camera lens as possible in order to minimize shadows, though the images shown in this paper were all taken with a standard flash mount. See Figure 3 for an example input image pair.

3.2 Albedo Map and Shading Image

The first stage in our method requires estimation of albedo and diffuse shading. We begin by calibrating our RAW image captures to one another based on their aperture A (f-stop), ISO I, and shutter speed T_s and convert to linear, floating-point pixel values using the following exposure correction factor C_e:

$$C_e = \frac{A^2}{(T_s I)} \qquad (1)$$

If absolute values were required, there would be an additional conversion factor, which is unnecessary for relative measurements such as ours.

To calculate albedo $I_a(j)$ we perform the operation expressed below at each pixel j:

$$I_a(j) = \frac{I_f(j) - I_d(j)}{I_c(j)} \qquad (2)$$

Pixel values in the diffuse-lit image I_d are subtracted from our flash-lit capture I_f, and we divide the result by pixel values in the flash calibration image I_c taken of a white Lambertian surface at a similar distance and aperture. This yields approximate reflectance values at each pixel, simultaneously correcting for vignetting, fall-off, and the global cosine factor. Since the cosine factor also depends on the local surface normal, we may underestimate albedo in steeply sloped areas. In the overall method, such errors will manifest as slight edge shifts, which are very difficult to detect visually. We apply a daylight white balance that provides a good match to the flash, therefore image subtraction results in a good color balance in our albedo image, as shown in Figure 4(a). In cases where flash shadows are present, we also apply a simple thresholding and neighbor-filling technique that copies detail from the flash-lit areas [Petschnigg et al. 2004].

To compute the diffuse shading image, we take the ratio of the diffuse-lit condition over the albedo at each pixel. This can result

---

[1] We use RAW format to simplify calibrating the images to each other, however our technique also works with linearized JPEGs.
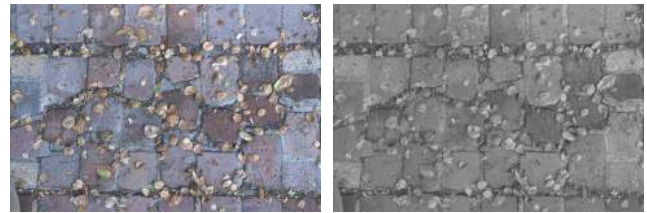

(a) Photograph of a brick path taken in shadowed daylight conditions. (b) Flash-lit photograph of the brick path.

Figure 3: *An example input photograph pair.*
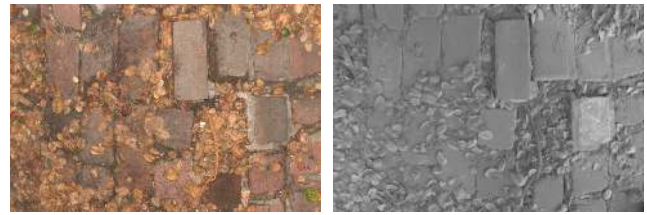
(a) Derived albedo map. (b) Diffuse shading image.

Figure 4: *Example albedo map and shading image generated from the photographs in Figure 3 of the brick path.*

(a) Alternate photograph of a brick path taken in shadow. (b) Histogram-matched diffuse shading image.

Figure 5: *Example input image for histogram matching and generated shading image.*

in a color cast due to skylight or cloudy illumination, but our depth estimation method uses only the luminance channel. A computed grayscale shading image for our brick path is shown in Figure 4. The depth estimation method described in the following section assigns a height of 0 to a pixel intensity of 0.5, so we normalize our shading image to this mean value.

In cases where there are significant differences in meso-structure but similar material properties to a previously captured surface, we can use a diffuse-lit image in conjunction with an existing diffuse-lit/flash-lit pair. We transfer the statistics of the diffuse-lit image to the albedo and the diffuse shading image of the existing exemplar using histogram matching [Heeger and Bergen 1995]. Figure 5(a) shows an example diffuse-lit capture taken close to the location shown in Figure 3(a). Figure 5(b) shows a synthesized diffuse shading image computed by applying histogram matching of Figure 5(a) to Figure 4(b). The histogram matching method is especially useful in architectural applications, where it is impractical to take flash-lit images of every portion of a large structure, but sample areas with similar appearance and statistics may be readily found.

3.3 Depth Estimation

The Langer and Zucker [1994] method is designed to recover shape from shading on a cloudy day, which is precisely what we capture in our technique. Applying their relaxation method entails iteratively ray-tracing a discretely sampled hemisphere of light source directions at every surface point. Instead we develop an approximate solution that works entirely in image space and yields a direct estimate of depth at each pixel. A conservative model basis ensures that we do not exaggerate depth variations, and a final, user-specified scale factor achieves the desired roughness.

Surface meso-structure can be approximated as a terrain with hills and valleys. The orientation of the surface to the sky (cosine factor) dominates on the hills, while the visible aperture effect dominates in the valleys, where the sides are at least partly in shadow. We therefore begin by developing two local models to approximate

these different types of relationships between meso-structure depth and shading. The scope of each model is shown on a hypothetical textured surface in Figure 6.
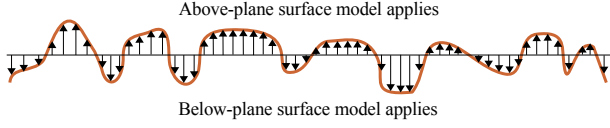

Figure 6: *Example of a profile of a textured surface and the separation between the above-plane and below-plane surface models.*

These models are derived such that an above-plane linear model is matched to a below-plane quadratic model at a tangent point, creating the smooth piecewise function plotted in Figure 7.

Below-Plane Model
We derive our below-plane shadowing model by approximating pits in the surface as cylinders with an aperture $2a$ and depth d, as shown in Figure 8(a). In order to arrive at a simple formula, we chose to ignore interreflections, which we found affect the scale but not the character of the depth estimates. We calculate an illumination factor E_c by integrating the cosine weighting over the solid angle subtended by the visible sky:

$$E_c = 2\pi \int_0^\theta \cos\theta' \sin\theta' d\theta' = \pi \sin^2\theta \qquad (3)$$

To arrive at the shading factor S we divide E_c by the illumination factor for the full sky, E_h which can be shown to be π. Through simple trigonometry the integrated shading factor becomes:

$$S = \frac{E_c}{E_h} = \frac{\pi \sin^2\theta}{\pi} = \frac{a^2}{a^2+d^2} \qquad (4)$$

Pit depth can therefore be estimated by solving Equation (4) for d as:

$$d = a\sqrt{\frac{1}{S} - 1} \qquad (5)$$

Above-Plane Model
For the above-plane model, we approximate surface protrusions as hemispheres. Shading of these is a function of the visible portion of the hemisphere h_v subtended by the angle ψ (Figure 8(b)), and added to the remaining reflected portion of the hemisphere h_r outside this angle.

Depth can thus be estimated by a simple linear model derived as follows, where ρ is the effective surrounding surface reflectance. (While we will assume $\rho = 0$ in both models, we include it in the above-plane derivation for completeness.)

$$h_v = \frac{\pi}{2}(1+\cos\psi) \qquad (6)$$

$$h_r = \rho\frac{\pi}{2}(1-\cos\psi) \qquad (7)$$

Consequently our above-plane shading factor is calculated as the ratio of these quantities and π:

$$S = \frac{\frac{\pi}{2}(1+\cos\psi) + \rho\frac{\pi}{2}(1-\cos\psi)}{\pi} \qquad (8)$$
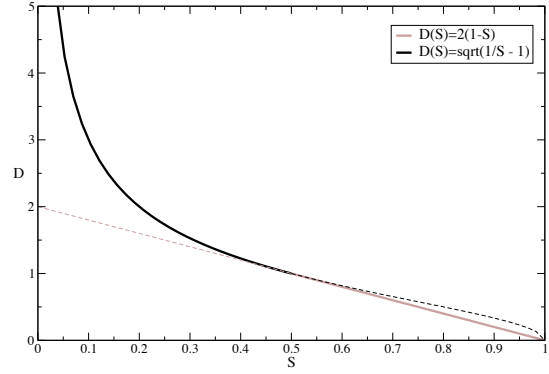
Figure 7: *The relationship between aperture and shading factor in our model. The dashed line shows the unused extensions of each model.*

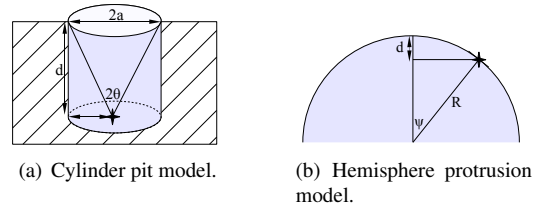(a) Cylinder pit model.   (b) Hemisphere protrusion model.

Figure 8: *Model to approximate shading of pits and surface protrusions.*

This can be simplified and solved for $\cos\psi$ to give:

$$\cos\psi = \frac{2S - (1+\rho)}{1-\rho} \qquad (9)$$

From Figure 8(b):
$$d = R - R\cos\psi \qquad (10)$$

Substituting $\cos\psi$ gives the linear model:

$$d = 2R\frac{1-S}{1-\rho} \qquad (11)$$

where R is the radius of the hemispherical hill.

Combined Model
These two models, expressed in Equations (5) and (11) can be conveniently combined at a double root solution to their intersection, by substituting $S = 1/2$ and equating the corresponding values of d:

$$a = \frac{R}{(1-\rho)} \qquad (12)$$

Recall, we assume the surrounding surface reflectance $\rho = 0$, yielding depth, d, from the diffuse shading, S, at each scale, a, in the combined aperture formula:

$$D(S) = d/a = \begin{cases} \sqrt{1/S-1} & \text{for } S \leq 1/2 \\ 2(1-S) & \text{for } S > 1/2 \end{cases} \qquad (13)$$

Multiscale Formulation
A shading change over a large region generally corresponds to a greater depth difference than the same shading change over a small
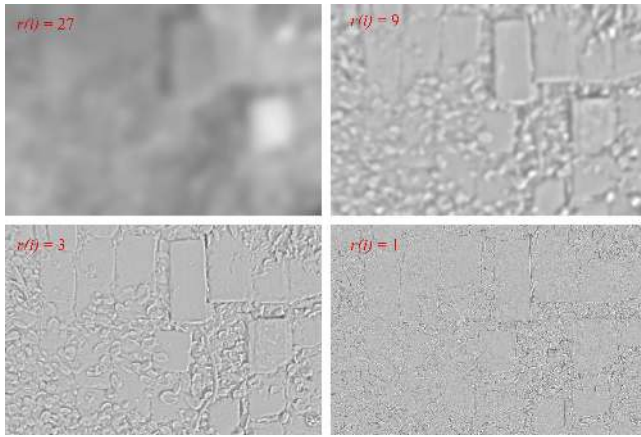
Figure 9: *The effect of different levels of Gaussian blur on the normalized shading image for the brick path example.*

Figure 10: *A comparison between a simple dark-is-deep approximation (left) and our multiscale model (right).*

(a) Rock wall image.   (b) Depth map obtained using our model.

Figure 11: *Depth map recovered from a single image of a rock wall obtained from the Web.*

region. Since our aperture model estimates depth from diffuse shading relative to a specific feature size, *a*, we must consider each scale in our captured diffuse shading image separately. Separating our diffuse shading image into scale layers, we efficiently convert our aperture estimates into depth estimates as required for a geometric model.

Starting from a normalized version of our diffuse shading image, shown in Figure 4(b), we compute several Gaussian blurred images using kernel radii *r* increasing by powers of three up to a maximum detail size based on image content, which may be specified by the user. At each level, the image is divided by the image at the next largest kernel radius (up to the largest) and multiplied by 1/2 for normalization, effectively yielding a Laplacian pyramid of equal resolution images [Burt and Adelson 1983]. These blurred images are referred to as $\ell(i)$. We transform these images using the depth function in Equation (13) by solving for d, where a is replaced by the blur radius at each level to accumulate the depth value at pixel j. Here, depth units correspond to the width of a pixel, and must be scaled accordingly.

$$d_j = \sum_{i=1}^{N} r(i) \left[D(\ell_j(i)) - 1 \right] \qquad (14)$$

Figure 9 shows our progressively blurred shading images for the brick path example. We subtract 1 from our computed depths at each level since this is the normal value for $D(S)$ at the average image intensity of 0.5, and we want our average surface displacement to be zero.

As noted earlier, our depth estimates are conservative. First, we ignored albedo to simplify our analysis. Second, we approximated indentations in the surface as pits, where a crevice model might be more appropriate in some cases. We therefore apply a user-selected, uniform scaling factor to each depth map to compensate for this and achieve an acceptable visual match to the original surface appearance. For all our test scenes, this scaling factor was between 0.75 and 1.5. Our unoptimized implementation takes 15 seconds to generate a 900x600 resolution depth map on a single core 2.5 GHz desktop computer.

In Figure 10, we show the difference between our depth hallucination method and a global, linear, dark-is-deep approximation [Khan et al. 2006] applied to the same diffuse shading image. Notice that our model is less sensitive to noise and better approximates the upper surface as well as the crevices.
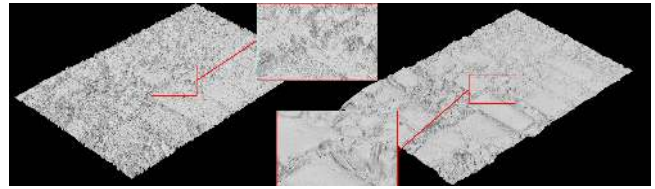
To demonstrate the flexibility of our approach, we downloaded a photograph of a textured rock wall from an online texture resource[2] taken with an unknown camera, and recovered a depth map using histogram matching to a roughly similar surface to obtain a diffuse shading image. The original image and our hallucinated depth map are shown in Figure 11.

3.4 Relighting the Hallucinated Surface

Once we have an albedo map and a depth map for our surface, virtually any rendering algorithm may be applied. We use the *Radiance* physically-based renderer [Ward 1994] with a suitable sky model that includes both direct (solar) and indirect (sky) components, choosing a low angle of solar illumination to make our depth variations more visible. A directly lit surface will have a warmer color cast, and we incorporate this in our model. Figure 12 shows the results for the brick path example with a solar altitude of $30°$. Specularity is not specifically addressed in our method, but may be added trivially to the material model by assuming a uniform value, as might be encountered on a wet day. (See right-hand image in Figure 1.) Our validation described in the following section addresses the visual plausibility of our rendered results.

4 Experimental Validation

We aim to answer two questions through two experiments. First, can our rendered images be reliably identified as synthetically generated? Second, do renderings generated using hallucinated depth maps appear plausible when compared with renderings using laser-scanned data? If users cannot reliably identify synthetically relit images created using our method while focused on assessing them, then we can conclude our method recovers sufficient detail to allow us to plausibly relight textured surfaces.

Participants with normal vision were seated in front of a standard LCD display. The experimenter ran an application that presented high resolution images to each participant. Depending on the experiment, for each stimulus, participants were asked to press a key
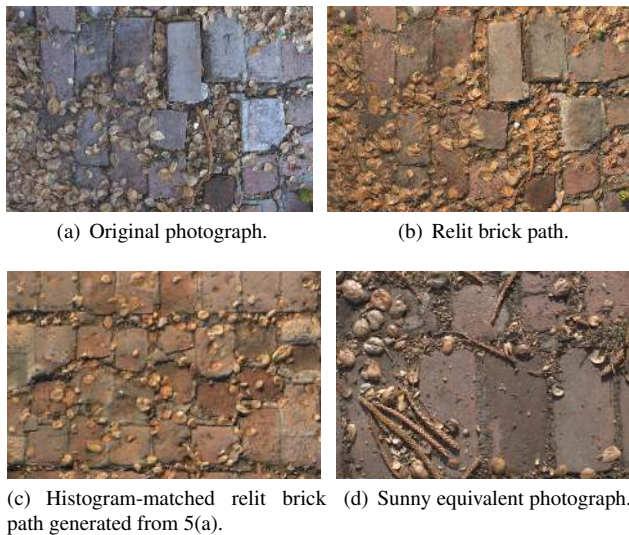
---

[2]Source of image: http://www.texturewarehouse.com

(a) Original photograph.

(b) Relit brick path.

(c) Histogram-matched relit brick path generated from 5(a).

(d) Sunny equivalent photograph.

Figure 12: *Results of relighting our brick path examples.*

(a) Rendering of hallucinated depth map.

(b) Rendering of laser-scanned depth map.

Figure 13: *Matched lighting frames of the Venus North Platform.*

to either rank the image or to choose between an image pair. All collected key-presses were logged. We determined the duration for which each image (or pair) was displayed via a pilot study involving 20 participants in which stimuli were presented for 1, 3 and 5 seconds. We found no apparent differences in people's ratings between images shown for given time intervals. Study data was collected from new sets of participants who were shown each stimulus for 3 seconds. A total of 40 participants (20 in each) took part in two studies.

4.1 Experiment One

The goal of experiment one was to assess whether people can reliably identify images created using our depth hallucination approach. Single images depicting a variety of textured surfaces, consisting of both real photographs and synthetically relit images, were presented in a randomized order. (See Figure 1 and Figure 12 for examples.) A total of 27 images were presented to each person in this part of the study. This set contained 9 day-lit photographs, 9 synthetically relit images and 9 synthetically relit histogram-matched images. Due to the difficulty in acquiring photographs with natural sunny lighting conditions at exactly the same location, the set of equivalent day-lit photographs were not necessarily taken from an identical view point to the images used to recover texture hallucinations. Participants were asked to rank each image from 1 to 5, corresponding to their certainty that the image they were viewing was an untouched photograph. On this scale, we define 1 as definitely synthetically generated, 5 as definitely an untouched

photograph, and 3 as undecided.

4.2 Experiment Two

Our second study was a two-alternative forced-choice experiment in which the aim was to evaluate the visual plausibility of our estimated depth maps relative to ground-truth data. Twelve pairs of still image frames from an animation depicting changing solar position over the scene with a fixed viewpoint (but novel to the captured one) were used. Each image pair contained an image frame created using ground truth geometry acquired through a laser scanning process, and an equivalent image frame generated using our technique for estimating the depth map from photographs. The same albedo map was registered to both the laser-scan and hallucinated depth maps, and the same physically-based rendering method was used for relighting both sequences. No in-filling techniques were applied to the laser-scan. The lighting frame was matched for specific image pairs (similar to those shown in Figure 13), but similarly varied between frames in the image sequences. Within each image pair, the laser-scanned and hallucinated surfaces were presented in a randomized order. Participants were asked to choose the image they believed to be most likely to be the real surface. The image pairs shown contained clear visual differences due to the different depth capture processes, but we aimed to answer if they are equally plausible visually.

5 Results and Data Analysis

In the first experiment, where participants rated how *real* the images looked, a repeated measures analysis of variance (ANOVA) showed a slight preference for the photographs ($F_{2,38} = 21.61$, $p<0.001$). This difference was statistically significant. An important result however is that on a scale of 1 to 5: photos received a mean rating of 3.97, relit images scored 3.22 for models derived with diffuse-lit/flash-lit pairs, and 2.98 for histogram-matched versions. The difference between both classes of synthetic images was not found to be statistically significant. On our rating scale a value of above 3 suggests the image is more likely to be a photograph than synthetic.

Relit images were rarely dismissed as artificial, and equivalent photographs were not always recognized as real. For 4 out of our 9 test scenes, the mean scores in Figure 14(a) show that our synthetic images were virtually indistinguishable from equivalent photographs. In the remaining scenes, relit images were still not rejected outright. Importantly Figure 14(b) shows that around 15 out of 20 participants gave our synthetic images average ratings above 3, leading us to conclude that our renderings compare very well with photographs. This is further supported by participants commenting in post study de-briefing, on the difficulty in determining which images were synthetic.

In the forced-choice experiment, a paired-sample t-test showed no significant difference between hallucinated depth and the laser-scan. This leads us to conclude that participants could not tell which of the two looked most plausible to them. Mean scores for each choice were 54% for the hallucinated depth, and 46% for the laser-scan. Nine participants out of 20 showed a preference for the hallucinated depth (see Figure 15), while 6 show a preference for the renderings based on the laser-scanned data. The remaining 5 seem undecided. The viewpoint for each data set was kept identical across all stimuli and close to the captured view to avoid bias for or against either depth map. Within each image pair, the only variable was the depth map used to generate the image. If the view played a significant role in users' assessments then people's choices would have been highly consistent, however only 2 participants chose a particular depth map in every comparison. Between each image
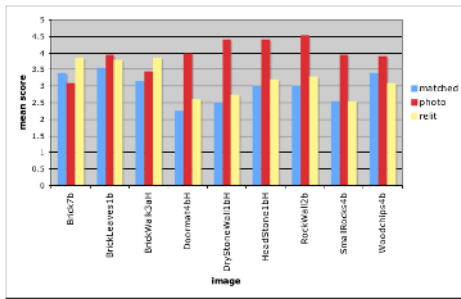
(a) Average ratings from 20 participants, for each specific scene.
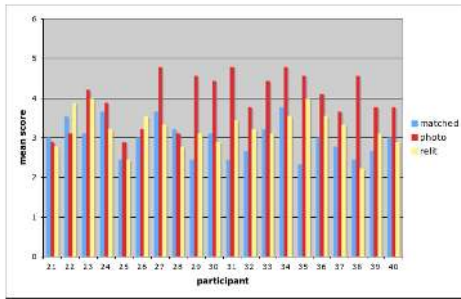
(b) Average ratings per participant, for each class of stimulus.

Figure 14: *Results from experiment one.*

pair, lighting was varied. If lighting played a significant role in biasing the results, we would not have seen overall strong differences in preference between participants.

6 Limitations

Naturally, there are situations where our assumptions do not hold, and these may produce unexpected or undesired results. We examine three such cases, which we encountered while acquiring test scenes for our experiments.

The first case is shown in Figure 16(a), where ivy vines are physically separated from the stone surface below. The separation is small, but it violates one of our basic assumptions, which is that our surface may be plausibly represented as a height field. Even if our mathematical model held in this case, which it does not, our height field representation would still fail us. The vines appear to be protruding from the wall rather than next to it.
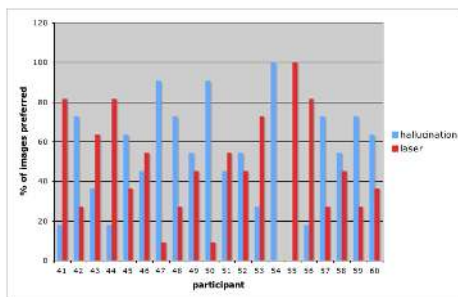

Figure 15: *Experiment two: percentage of preferred class of stimulus (hallucinated or laser-scanned) per participant.*

The second case is shown in Figure 16(b), where our surface is a reasonable match to our geometry assumptions but the daylight illumination is not. In this area, the light comes primarily from one side, as it is nearby a dark structure and only a portion of the sky is visible on the cobbled ground. This results in a bias in the shading image, which our technique translates into a bias in the geometry, making the stones appear to lean towards the original sky direction. While this problem might be overcome with large bounce cards, in a practical setting such biases may be unavoidable and would have to be corrected in a geometry post-processing step.

The third case is shown in Figure 16(c) and 16(d). The highly reflective and slightly translucent rock material violated our opaque reflectance assumption, resulting in a rather flat and unnatural appearance, though some surface structure is still obtained.
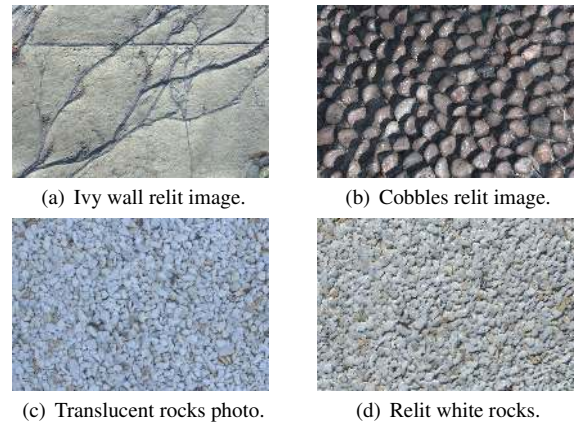

(a) Ivy wall relit image.

(b) Cobbles relit image.

(c) Translucent rocks photo.

(d) Relit white rocks.

Figure 16: *Example failure cases that violate fundamental assumptions of our algorithm.*

7 Conclusions

Ultimately, our goal is to combine models of materials approximated using hallucinated depth maps with wide-baseline 3D reconstructions of buildings. This is likely to pose further challenges, such as surfaces that are unevenly or directly lit, or where neighboring images are taken under differing illumination, or at oblique angles, and must be stitched together seamlessly. Our method complements image-based reconstruction processes by supplying surface detail.

In this work, our objective is to convincingly render the altered appearance of textured surfaces under differing lighting and viewing conditions, while requiring only simple and practical data capture procedures. Starting from diffuse-lit/flash-lit photographs, we generate both an albedo map and textured height field, which can be relit and viewed from any angle under any lighting. Our model applies a surface aperture function but, in contrast to previous methods, works entirely in image space. If only a diffuse-lit image is available, we apply histogram matching with a similar exemplar pair to lift the flash requirement, further simplifying data capture. Compared to alternatives, such as laser scanning, our depth estimation method does not require additional data registration, since both albedo and depth are acquired from perfectly aligned captures.

Histogram matching permits us to hallucinate local height variations from other diffusely lit imagery, and gaps in our captured model may be filled in using texture synthesis [Efros and Freeman 2001]. Processing from image to model is also sufficiently simple that depth and albedo maps could be generated on the fly from captured or synthetic texture data on consumer-level graphics cards.

Experimental evaluation of this new approach yielded two important observations. First, when presented with relit images, 75% of participants rated them as more like photographs. Second, participants were unable to decide whether hallucinated depth renderings or those generated using ground truth acquired by laser scanning looked most realistic. Since depth is never fully divulged by shading, our estimates may fall short of absolute accuracy. Our experimental results show that in many practical situations this is unimportant, because the hallucinated depth method reproduces surface appearance that is perceptually tantamount to photographs.

Acknowledgments

References

BELHUMEUR, P. N., KRIEGMAN, D. J., AND YUILLE, A. L. 1999. The bas-relief ambiguity. *International Journal of Computer Vision 35*, 1, 33–44.

BURT, P., AND ADELSON, E. 1983. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications 31*, 4, 532–540.

DANA, K. J., VAN GINNEKEN, B., NAYAR, S. K., AND KOENDERINK, J. J. 1999. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics (TOG) 18*, 1, 1–34.

DEBEVEC, P. E., TAYLOR, C. J., AND MALIK, J. 1996. Modeling and rendering architecture from photographs: a hybrid geometry and image-based approach. In *SIGGRAPH*, ACM, 11–20.

EFROS, A. A., AND FREEMAN, W. T. 2001. Image quilting for texture synthesis and transfer. In *SIGGRAPH*, ACM, 341–346.

EISEMANN, E., AND DURAND, F. 2004. Flash photography enhancement via intrinsic relighting. In *SIGGRAPH*, ACM, 673–678.

FLEMING, R. W., DROR, R. O., AND ADELSON, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision 3*, 5, 347–368.

HADDON, J. A., AND FORSYTH, D. A. 1998. Shading primitives: Finding folds and shallow grooves. In *ICCV*, 236–241.

HADDON, J. A., AND FORSYTH, D. A. 1998. Shape representations from shading primitives. In *ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume II*, Springer-Verlag, London, UK, 415–431.

HEEGER, D. J., AND BERGEN, J. R. 1995. Pyramid-based texture analysis/synthesis. In *SIGGRAPH*, ACM, 229–238.

HERSHBERGER, W., 2008. Taming those annoying highlights: cross-polarization flash macro photography. Online article. http://www.naturescapes.net/042004/wh0404.htm.

HORN, B. K. P. 1989. Obtaining shape from shading information. In *Series of Artificial Intelligence: Shape from Shading*. Mit Press, Cambridge, MA, 123–171.

KHAN, E. A., REINHARD, E., FLEMING, R. W., AND BÜLTHOFF, H. H. 2006. Image-based material editing. In *SIGGRAPH*, ACM, 654–663.

KOENDERINK, J., AND VAN DOORN, A. 1983. Geometrical modes as a general method to treat diffuse interreflections in radiometry. *J. Opt. Soc. Am. 73*, 6 (June), 843–850.

LANGER, M. S., AND BÜLTHOFF, H. H. 2000. Depth discrimination from shading under diffuse lighting. *Perception 29*, 6, 649–660.

LANGER, M. S., AND ZUCKER, S. W. 1994. Shape-from-shading on a cloudy day. *Journal of the Optical Society of America 11*, 2, 467–478.

LENSCH, H. P. A., KAUTZ, J., GOESELE, M., HEIDRICH, W., AND SEIDEL, H.-P. 2003. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics (TOG) 22*, 2, 234–257.

LI, H., FOO, S.-C., TORRANCE, K. E., AND WESTIN, S. H. 2006. Automated three-axis gonioreflectometer for computer graphics applications. *Optical Engineering 45*, 4, 1–11.

MALIK, J., AND MAYDAN, D. 1989. Recovering three-dimensional shape from a single image of curved objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*, 6.

NARASIMHAN, S. G., VISVANATHAN, R., AND NAYAR, S. K. 2003. A class of photometric invariants: separating material from shape and illumination. In *ICCV*, IEEE, 1387–1394.

NGAN, A., AND DURAND, F. 2006. Statistical acquisition of texture appearance. In *Eurographics Symposium on Rendering*, T. Akenine-Möller and W. Heidrich, Eds. The Eurographics Association, 31–40.

OSTROVSKY, Y., CAVANAGH, P., AND SINHA, P. 2005. Perceiving illumination inconsistencies in scenes. *Perception 34*, 11, 1301–1314.

PATERSON, J. A., CLAUS, D., AND FITZGIBBON, A. W. 2005. BRDF and geometry capture from extended inhomogeneous samples using flash photography. In *Computer Graphics Forum*, vol. 24, Eurographics, 383–391.

PETSCHNIGG, G., SZELISKI, R., AGRAWALA, M., COHEN, M., HOPPE, H., AND TOYAMA, K. 2004. Digital photography with flash and no-flash image pairs. In *SIGGRAPH*, ACM, 664–672.

PRADOS, E., AND FAUGERAS, O. 2005. Shape from shading: a well-posed problem? In *Computer Vision and Pattern Recognition (CVPR)*, IEEE, 870–877.

RAMACHANDRAN, V. S. 1988. Perception of shape from shading. *Nature*, 331, 163–166.

RUSHMEIER, H., AND BERNARDINI, F. 1999. Computing consistent normals and colors from photometric data. In *Second Conference on 3-D Imaging and Modeling 3DIM*, IEEE, 99–108.

WARD, G. J. 1994. The RADIANCE lighting simulation and rendering system. In *SIGGRAPH*, ACM, 459–72.

YU, Y., DEBEVEC, P., MALIK, J., AND HAWKINS, T. 1999. Inverse global illumination: recovering reflectance models of real scenes from photographs. In *SIGGRAPH*, ACM, 215–224.

ZHANG, R., TSAI, P.-S., CRYER, J. E., AND SHAH, M. 1999. Shape from shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 8, 690–706.