Performance Relighting and Reflectance Transformation with Time-Multiplexed Illumination

Andreas Wenger Andrew Gardner Chris Tchou Jonas Unger[†] Tim Hawkins Paul Debevec

University of Southern California Institute for Creative Technologies[1]

Figure 1: A live-action performance illuminated several different ways in postproduction.

Abstract

We present a technique for capturing an actor's live-action performance in such a way that the lighting and reflectance of the actor can be designed and modified in postproduction. Our approach is to illuminate the subject with a sequence of time-multiplexed basis lighting conditions, and to record these conditions with a high-speed video camera so that many conditions are recorded in the span of the desired output frame interval. We investigate several lighting bases for representing the sphere of incident illumination using a set of discrete LED light sources, and we estimate and compensate for subject motion using optical flow and image warping based on a set of tracking frames inserted into the lighting basis. To composite the illuminated performance into a new background, we include a time-multiplexed matte within the basis. We also show that the acquired data enables time-varying surface normals, albedo, and ambient occlusion to be estimated, which can be used to transform the actor's reflectance to produce both subtle and stylistic effects.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture I.4.8 [Image Processing and Computer Vision]: Digitization and Image Capture—Reflectance I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Tracking

Keywords: relighting, compositing, environmental illumination, image-based rendering, reflectance models

[1]USC ICT, 13274 Fiji Way, Marina del Rey, CA, 90292
Email: wenger@ict.usc.edu, gardner@ict.usc.edu, tchouster@gmail.com, jonun@itn.liu.se, timh@ict.usc.edu, debevec@ict.usc.edu.

[†]Supported by the Swedish Research Council and receiving supervisory support from the Swedish National Graduate School in Computer Science.

Figure 2: The lighting apparatus used for capturing a performance under time-multiplexed illumination. Behind the actor is the gray background matte surface, and at left is the high-speed camera.

1 Introduction

In motion pictures, lighting is not only used to help actors and sets look their best, but as an integral part of storytelling to set mood, direct attention, and underscore performance. This importance is reflected in the high proportion of time and expense spent on lighting: by many estimates, one half or more [Trumbull 2000] of the valuable time spent on a set is involved in setting up the lighting.

Several aspects of film production can be performed after principal photography, such as editing, sound effects, scoring, color timing, and visual effects. In each case, the fact that the process can be performed as part of postproduction allows results to be progressively improved and revised by the filmmakers after principal photography. Lighting, in contrast, must in large part be finalized at the time each scene is filmed. This requirement adds complication and cost to principal photography and provides limited options for modification and improvement during postproduction. A situation where this is a particularly difficult constraint is when shooting actors in front of a green screen. In this case, the lighting on the actor often must be chosen before the virtual backgrounds are finalized, posing difficulties for achieving consistent illumination between the actors and the background.

Interactive techniques for lighting computer-generated characters and scenes (e.g. [Marks et al. 1997; Gershbein and Hanrahan 2000; Calahan 2000]) provide a glimpse of the increased convenience and creative control achievable when lighting can be designed and refined during the postproduction process. Inspired by the artistic control available for lighting computer-generated scenes, the goal of this work is to develop a technique for creating and changing the lighting of live-action photography as a postproduction process, and to demonstrate this technique on several close-up shots of live performances.

Our Approach In our approach, we illuminate the performance with a rapid series of basis lighting conditions which we record with a high-speed camera. We then recombine these images to create novel illumination conditions as has been seen for computer-rendered imagery [Nimeroff et al. 1994; Marks et al. 1997] and still photography [Haeberli 1992; Debevec et al. 2000]. To apply this process to a moving subject, we develop a motion compensation technique based on optical flow to temporally re-align the frames and subsequently reintroduce an appropriate amount of motion blur. We finally show how the basis illumination data can be used to modify the reflectance properties of the performance for both subtle and stylistic effects.

Contributions In this work we make the following contributions:

1. We present a novel technique for designing and modifying the illumination of a live-action performance in postproduction using time-multiplexed illumination and high-speed photography.
2. We discuss and evaluate several illumination bases chosen in a manner that is informed by the system's light level requirements and the characteristics of our camera system.
3. We describe a robust process for compensating for subject motion across the basis frames by computing optical flow between specially inserted tracking frames in the basis. We show that this allows the basis sequences to be run more slowly than the target frame rate of the sequence, decreasing storage requirements and increasing exposure levels. We use the same flow fields to introduce realistic motion blur into the re-illuminated sequences.
4. We present a novel time-multiplexed matting process for compositing the performance in front of a new background.
5. We show that the data created by our system can be used to digitally alter the reflectance of the actor for both subtle and stylistic effects.

2 Background and Related Work

Performance Capture and Resynthesis Several techniques for capturing live-action performances in order to subsequently change the viewpoint and/or lighting have been proposed. [Guenter et al. July 1998] record an actor's face with five video cameras, and fits a 3D model to facial tracking markers to render the performance from new viewpoints. In contrast to our work, their technique allows the viewpoint to be changed rather than the illumination. Using different approaches, [Nishino and Nayar 2004] and [Georghiades et al. 1999] use photometric stereo [Woodham 1980] to compute surface normals, albedo, and geometry of a face from a sparse set of lighting directions to render it from novel viewpoints and directional lighting, but unlike our work they assume a Lambertian reflectance model, do not consider hair, and limit their capture to static faces rather than a performance. [Georghiades 2003] generalized the photometric stereo problem to the Torrance-Sparrow reflectance model, but assumed specularity to be constant over the face and also did not consider dynamic performances.

Other work looks more closely at simulating facial reflectance accurately. [Marschner et al. 2000] estimate spatially-varying facial albedo using different lighting and polarization conditions and uses a measured facial BRDF model to extrapolate non-Lambertian reflectance across the face. They use this to texture-map a scanned 3D face model and animate it using facial motion capture. Our technique does not provide control over the animation of the performance and we do not explore changing the viewpoint, but we record spatially-varying facial reflectance and interreflection properties and we do not need to build an explicit face model. [Debevec et al. 2000] illuminate faces from a dense set of lighting directions and recombines the basis images according to light in captured lighting environments, reproducing diffuse, specular, and translucent reflection of the face and hair. However, their significant capture time and lack of a motion compensation algorithm limits their investigation to still images of the face in static poses. [Hawkins et al. 2004] use several such illuminated datasets in different expressions and from different viewpoints to create a morphable and relightable 3D face model, but like [Marschner et al. 2000] requires explicit 3D geometry and in contrast to our work would be complicated to extend to hands, bodies, and clothing. [Debevec et al. 2002] use a sphere of RGB light sources to illuminate performances with captured lighting environments, but does not allow for the lighting on the performance to be changed in postproduction. Work that is complementary to ours [Zhang et al. 2004] present a real-time face scanning technique using stereo and rapidly projected video patterns, but it does not focus on reproducing the face's non-Lambertian reflectance properties.

Motion Compensation Our work requires temporally aligning differently illuminated frames in a high-speed video sequence. [Kang et al. 2003] described a system for temporally aligning differently exposed frames in a video sequence through optical flow and image warping to create a high dynamic range video sequence. We must also align images whose appearance differs significantly from frame to frame, but our problem is different since the pixel values of our images are not related by a single scale factor, even for a subset of the pixel brightness range. [Hager and Belhumeur 1996] described a technique for tracking a face under unknown variable illumination, but did not track motion where the lighting variation, rotation, or occlusion is as significant as what our sequences exhibit. In our technique, we leverage our control over the illumination conditions and our high frame rate to facilitate robust tracking and warping using classic optical flow techniques from [Black and Anandan 1993].

Matting Our system includes a time-multiplexed process for obtaining an alpha channel matte [Porter and Duff 1984] to composite the actor's performance over a new background. Current matting processes use blue or green screens behind the actor, or use light with specially chosen spectral properties [Fielding 1985; Debevec et al. 2002]. [Smith and Blinn 1996] use multiple backgrounds to obtained improved mattes for transparent materials, and more advanced *environment matting* [Zongker et al. 1999] techniques can acquire real-time refractive properties of foreground objects. Our time-multiplexed matte does not capture refractive properties, but obtains the same transparency performance as [Smith and Blinn 1996] since the light on the subject can be suppressed while the matte is filmed. Also, our solution is straightforward to implement in the context of our time-multiplexed lighting system.

Bases for Image-Based Relighting Most image-based relighting work has acquired images under single-source lighting conditions, although frequently such data is projected onto different bases for efficient storage and/or rendering (e.g., [Masselus et al.

2004]). [Schechner et al. 2003] suggested using *spatially* multiplexed illumination where more than one light source is turned on in each pattern. In our work we implement three different bases ranging from single lights to spatially multiplexed illumination.

[Raskar et al. 2004] acquire sequences of people and scenes using time-multiplexed flashes placed around the lens of a video camera. From the resulting shadows they extracted depth edges to create a variety of non-photorealistic rendering effects. We also film the subject under rapidly changing lighting conditions, but we use higher frame rates and many more lighting directions to capture detailed reflectance information for photorealistic relighting.

Lighting and Reflectance Transformation [Malzbender et al. 2001] used a hemispherical lighting apparatus similar to our spherical device for capturing images of textures and artifacts under a variety of lighting directions. They fit parabolic polynomials to the acquired reflectance functions, which yielded estimates of the surface normals. From the normals, they presented several forms of reflectance transformation including *specular enhancement*, where a specular lobe is added to the reflectance, and *diffuse gain*, where the width of the diffuse component is narrowed to accentuate geometric surface variation. In our work, we show that similar forms of processing can be applied to human performances photographed using our techniques for both subtle and stylized reflectance transformations. Extending this previous work, we use a complete sphere of incident illumination, and we use a surface normal estimation technique based on photometric stereo that is designed to be robust to self-shadowing and specularities. We also combine the reflectance transformation process with reflection mapping [Miller and Hoffman 1984; Greene 1986] to simulate environmental reflection, and we derive ambient occlusion maps [Landis 2002] for the reflectance functions to approximate self-shadowing in the reflection mapping process.

Some techniques modify image illumination using a single photograph: [Williams 1991] and [Petrović et al. 2000] use surface normals derived from an "inflation" operation to add novel shading to images, and [Wen et al. 2003] relight single images of a face with novel environmental illumination using a fitted generic face model. However, these papers assume simplified reflectance models that have not been shown to yield photoreal results for extreme changes in incident illumination, for example, taking an input image under harsh lighting and transforming it to diffuse illumination, or vice-versa.

3 Apparatus

Our light stage is a 2m diameter once-subdivided icosahedron, with the lowest five faces left open to accommodate the actor. A light source is placed on each edge and vertex of the stage yielding 156 light sources an average of 18° apart. Each light source is built from three Luxeon V white LEDs which together produce 360 lumens. Each light is focussed toward the subject using a Fraen wide beam tri-lens optic, yielding 420 lux at 1 meter distance. The light is even to within 20% over the area of the actor's face and shoulders.

A 60 × 40 cm 32% gray board is placed in the back of the structure to obtain the time-multiplexed matte of the actor. To light the board, six additional LED lights are attached to flexible metal arms connected to the edges of the board. With just these lights turned on, the actor appears in silhouette against an evenly-illuminated neutral background.

We use a Vision Research Phantom v7.1 high-speed digital camera capable of capturing up to 4800 frames per second at 800x600 resolution. The CMOS sensor is a 12-bit-per-channel single-chip Bayer-pattern sensor. The camera records directly to 4GB of internal RAM allowing for 4.3 seconds of capture at a 640×480 cropped resolution at 2160fps. While expensive, the camera is comparable in cost to current digital motion picture filming equipment.

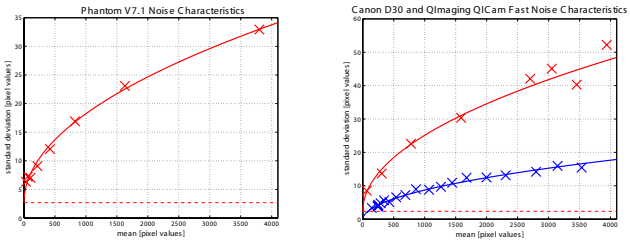
A Z-World Rabbit 2000 microcontroller drives the lights in arbitrary pattern sequences and triggers the camera's shutter in sync with each pattern. Custom driver boards distributed throughout the stage translate TTL signals from a latch board connected to the microcontroller into the requisite 0.7A current-regulated power needed to illuminate the lights.

4 Camera and Light Source Calibration

We calibrated the intensity response curve of the high-speed camera by photographing a diffuse white card illuminated by the LED lights. The LEDs were pulsed at successively longer intervals for 100 exposures covering the working range of the sensor. We fit a spline to the response curve, which was close to linear.

To inform the choice of our lighting bases, we calibrated the noise characteristics of the Phantom V7.1 camera. Our goal was to measure the mean intensity μ and resulting noise (i.e., standard deviation) σ for the working range the CMOS sensor. We used the DC-powered, current-regulated LED lights to illuminate a grayscale chart, and acquired 100-frame sequences at several lens f/stops. For several pixels across the chart, we computed μ and σ for the pixel's value over the sequence. Figure 3(a) shows the graph of σ versus μ across the sensitivity of the Phantom's sensor. We found that the noise was well-modeled as $\sigma = 2.68 + 0.492\sqrt{\mu}$, suggesting that the noise is due to additive dark current noise plus photon noise proportional to the square root of the signal [Hewlett-Packard Components Group 1998]. By these measurements, the photon noise dominates the additive noise for values above 0.73% of the maximum signal. These characteristics have implications for the quality of different lighting bases described in Section 5.

For validation, we also measured the noise response characteristics of a Canon D30 still camera and a cooled QImaging QICam (Figure 3(b)), both typical of cameras used in recent image-based rendering research. As with the high-speed camera, photon noise was dominant for over 98% of each camera's range.



(a) (b)

Figure 3: (a) Measured Phantom V7.1 noise characteristics, plotting standard deviation of the signal σ versus mean pixel values μ. The additive noise level is indicated by the dotted red line. (b) Noise characteristics for two other cameras, a QImaging QICam (top, red) and a Canon D30 (blue).

Since our white LEDs appear somewhat bluish to our camera, we placed a pale orange Tiffen 85C filter on the camera lens to help balance the color response. We also performed a first-order correction for both camera vignetting and light beam falloff by photographing a large white card lit by all the lights in the center of the device and dividing subsequent images by this flat field response.

5 Basis Selection

Our microcontroller program can turn on and off any set of light sources for each frame to produce arbitrary binary patterns. In theory, any linearly independent set of patterns equal to the number of lights allows the recovery of the individual lighting directions by inverting a linear system. To design and evaluate bases, we considered three general factors. These included: 1) Subject Perception: whether the lights were too bright or too stroboscopic, 2) Image Quality: if the recovered lighting resolution is sharp, the images exhibit a good signal-to-noise ratio, and if the basis accurately images the dynamic range of diffuse, specular, and shadowed regions, 3) Motion Robustness: that the basis is robust to subject motion during the basis. With these criteria in mind we designed three bases: single lights, triangles of lights, and Hadamard patterns (Figures 4 and 6). An exhaustive comparison of the bases is not the focus of this paper, but in this section we describe these bases in the context of our light stage and discuss the advantages and disadvantages that we experienced with each.

5.1 Single Light Basis Each of the 156 light sources yields 420 lux at the middle of the stage. At our commonly used frame rate of 2160fps (exposing for 412 of the 462 available microseconds) and at an aperture of f/1.2, this produced pixel values of 25% of the saturation level for a 98.5% white reflectance standard and pixel values of 4-8% for diffuse facial reflectance depending on the subject's skin tone. At these levels, image noise was somewhat apparent (Figures 6(a,d)), but acceptable relighting results could be obtained when many such images were added together.

Figure 4: Six elements of each of the three lighting bases used in this work. Top Row: Single lights. Middle Row: Triangles. Bottom Row: Hadamard Patterns

5.2 Triangle Basis To increase the illumination, we designed a basis using triples of triangularly adjacent lights. For a subdivided icosahedron having a 5-vertex at the top, triangularly adjacent light groups point toward either the north or the south pole. Taking all triangles pointing toward the north pole yields a basis isomorphic to the single-light basis. For our configuration, this produced a 156-light basis that yielded 1250 lux at the center of the stage. With this basis, we used the triples of lights as if they were single individual lighting directions. While this limited the resolution of the lighting environments that could be reproduced, we did not find the effect to be particularly noticeable.

On average, each light source is used in three of the triangle patterns. We ordered the triangle sequence so that the times each light would turn on were distributed generally evenly in the sequence (Figure 5.) This significantly increased the apparent strobe rate of each light.

The triangle basis produced 75% the maximum pixel value when lighting the white reflectance standard. For diffuse facial reflectance, the pixel values were approximately 12-24% of this maximum. We found this light level to be nearly optimal, since it provided room in the response for brighter areas such as the teeth and whites of the eye, and it allowed most of the specular reflections to be imaged within the range of the sensor.

5.3 Hadamard Basis [Schechner et al. 2003] suggest using patterns based on *Hadamard matrices* to construct an illumination basis for image-based relighting of static scenes. A related process is commonly used in spectroscopy [Harwit and Sloane 1979] to increase the signal to noise ratios of spectral measurements. In this technique, there are k basis patterns, where k is an odd number. In each pattern, just over half of the lights are on, and across the basis each light is on in precisely $(k+1)/2$ of the patterns. The patterns are constructed so that one can recover an image of the subject illuminated by a single light, i.e. invert the basis, by adding the images taken when the light is on and subtracting the images taken when the light is off; all other light contributions cancel. In our work we used a 156×156 Hadamard matrix [Sloan 1999] which is based on the Williamson construction [Hall 1998].

If a sensor obeys an additive noise model, [Schechner et al. 2003] note that there will be an advantage of increased signal to noise ratio in the recovered images. For example, if there is additive noise with standard deviation σ in each image and k basis images are taken, the recovery process adds and subtracts k images yielding a total noise of $\sigma\sqrt{k}$. If a pixel appears illuminated with brightness L in each image, then its brightness in the recovered sum will be $L(k+1)/2$. Thus, the signal to noise ratio increases from L/σ to $L(k+1)/(2\sigma\sqrt{k})$, an increase of approximately $\frac{\sqrt{k}}{2}$.

Unfortunately, the geometry of our apparatus and the noise response of our camera are different than those assumed in [Schechner et al. 2003]. Since our lights are distributed on a complete sphere, half of the lights typically make little or no contribution to a given pixel's intensity. For those that do, their irradiance is generally attenuated by Lambertian cosine falloff, which averages to a factor of $\frac{1}{2}$ over the hemisphere. Thus, our expected signal to noise ratio decreases by a factor of four to $\frac{1}{4}L(k+1)/(2\sigma\sqrt{k})$, an SNR gain of $\frac{1}{8}\sqrt{k}$. For our particular case of $k = 156$, this would yield only a modest 56% SNR increase.

Another impediment to an SNR improvement is that our camera's noise is dominated by photon noise over 98% of its exposure range as seen in Figure 3(a). In this region, the photon noise proportional to the square root of the signal counteracts the $O(\sqrt{k})$ SNR gain of the additive noise model. Since we have sufficient light to operate in this region with single light sources, we would not expect an SNR gain from using the Hadamard basis.

We did notice that the Hadamard patterns significantly increased the average light output of the device. At an aperture of f/1.2, it was necessary to reduce the exposure interval to 100 μs per pattern in order to avoid saturating the sensor. We chose to reduce the exposure interval rather than narrow the aperture since this reduced the light on the performer to a more comfortable level, from 6600 to 1600 lux (A bright cloudy sky is approximately 5000 lux, so 6600 lux can seem bright when projected indoors.) We could alternately have narrowed the aperture to f/2.8, an iris setting commonly used for filming motion pictures, to increase the depth of field.

5.4 Basis Comparison Figure 6 shows a comparison the individual lighting directions derived from the three bases.

Subject Perception The performers were screened for any medical history of adverse reactions to blinking lights and asked to notify the operators immediately of any discomfort. None of the performers or members of our laboratory reported an adverse reaction to any of the bases. The Hadamard patterns produced the least perceptible strobing, but performers who had adapted to dim interior lighting noted that the basis was relatively bright. The single lights were the most stroboscopic, but relatively dim. The triangle basis was moderately stroboscopic and moderately bright.

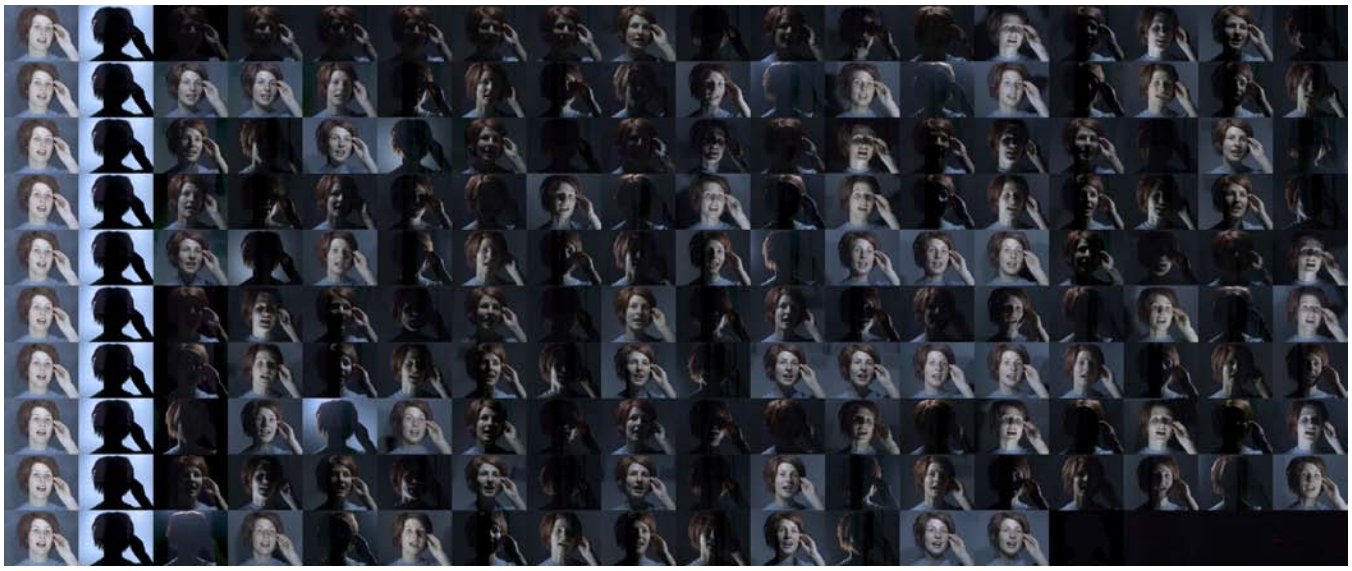


Figure 5: Successive images of an actor lit by the 180-pattern sequence in the span of one twelfth of a second. Tracking and matte frames seen in the two left columns occur ten times each within the lighting basis with an effective rate of 120Hz in the 2160fps sequence.

Image Quality We tested the effective signal to noise ratio of the three bases by capturing 100 successive basis sequences of a greyscale chart placed in the apparatus. For several pixels on the patch corresponding to average skin reflectance, we computed the signal-to-noise ratio (SNR) over the 100-image sequence. The average of these SNRs is tabulated below:

	Single	Triangle	Hadamard
mean μ	613	1327	2015
std dev σ	13.5	16.4	50.2
SNR	45.4	80.9	40.1

The single light images exhibited greater noise than the triangle basis due to the lower light levels. We found that the recovered single-light images from the Hadamard basis exhibited a slightly worse SNR than the single lights. However, in shadowed regions, the Hadamard patterns produced *significantly* greater noise than the single lights as seen in comparing Figure 6(d) to Figure 6(f). This is because these shadowed regions are generally not shadowed in the original Hadamard basis images and thus exhibit the increased photon noise of the larger signal. The triangle basis images produced the best SNR, but it is important to note that this is not a direct comparison since this basis can not resolve as detailed a lighting resolution as either the single lights or the Hadamard patterns. If we were to simulate the triangle basis by adding together triples of single light images or recovered Hadamard directions, their SNR would increase by $\sqrt{3}$ and in each case become much closer to the SNR of the triangle basis, at least for non-shadowed regions.

For our subjects, none of the basis patterns produced significant problems with specularities, except at extreme grazing angles on the skin and in the reflections of the eyes.

Motion Robustness The recovered single-light directions from the Hadamard patterns occasionally exhibited banding from subject motion during the basis since they are formed as nonconvex linear combinations of images taken of a moving subject. The problem was substantially reduced when the motion compensation techniques described in Sec. 6 were applied, but were still occasionally visible in regions such as eyelid blinks. Neither the single light nor triangle bases exhibited this effect since no basis inversion is required.

For our camera, lighting apparatus, and subjects, we found that the triangle basis provided a reasonable compromise between subject perception, lighting resolution, and image quality. The Hadamard patterns did not yield an improved SNR for our camera and lighting configuration, though they did show the potential to better capture object specularities and to minimize strobing. For systems with different cameras, lights, or lighting resolution requirements, the advantage could be more significant, and further explorations into the space of lighting bases could be fruitful.

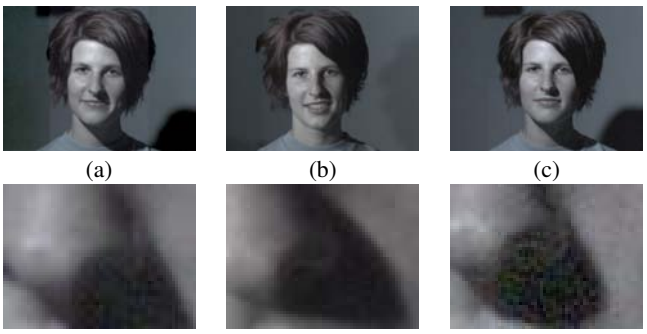


Figure 6: Directional lighting basis frames and detail areas from (a,d) the single light basis, (b,e) the triangle basis, (c,f) the Hadamard basis. For (c,f), the image was recovered as a linear combination of all 155 basis patterns.

6 Motion Compensation

Although we film the actor at a high frame rate, the actor's motion across a full basis sequence can be significant. When these basis images are added together, this can lead to an image smearing effect in the re-illuminated frames as seen in Figure 8(a). For diffuse lighting, the effect is similar to exaggerated motion blur, but when light intensities in the environment correlate with the ordering of the basis, different areas of the face will experience different spatiotemporal displacements, which can appear objectionable.

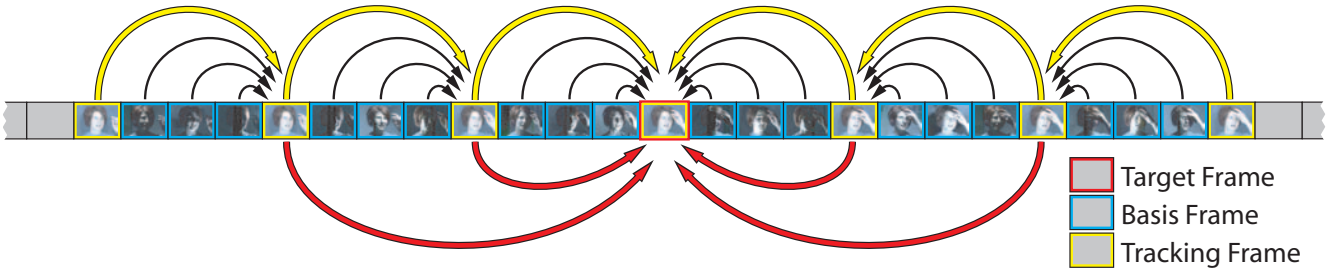


Figure 7: **Motion Compensation Process.** To compensate for the subject motion, basis frames (blue) are warped to match the target output frame (red). Optical flow is calculated between adjacent tracking frames (yellow), and linearly interpolated to warp each basis frame to the nearest tracking frame (black arrows) in the direction of the target output frame. Then, long-range warps (red arrows) are applied to bring each basis frame into alignment with the output frame.

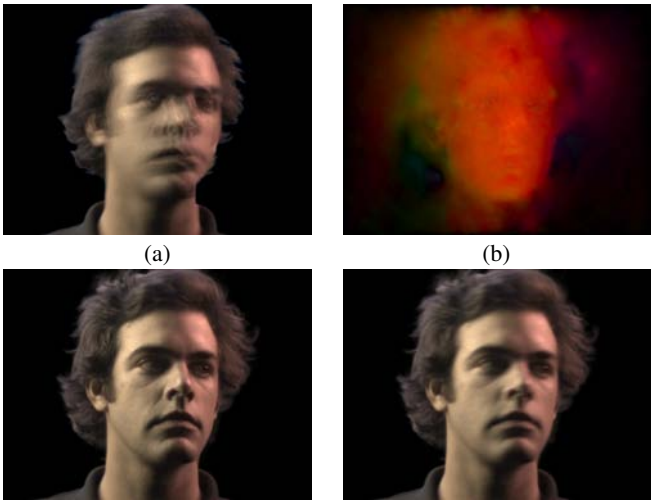


Figure 8: (a) Relit frame without motion compensation, showing smearing. (b) Color-coded optical flow field computed between neighboring tracking frames (c) Stabilized frame where motion compensation has been applied to the basis. (d) Relit frame with synthesized 180-degree shutter motion blur based on the flow field.

We correct for subject motion during the basis by estimating the optical flow between frames and using this flow to warp the image data as if it were taken simultaneously with a *target output frame*. At this level, our technique is similar to that used in [Kang et al. 2003] to correct for motion in a high dynamic range video sequence. By aligning the frames to the target output frame, the relighting results are greatly improved as in Figure 8(c).

Because the lighting conditions change with every frame, our sequences do not obey the intensity constancy assumption made by most optical flow algorithms. To make the frames easier to track, we place evenly lit tracking frames at a regular intervals within the sequence; in this work, every 18th frame in the repeating sequence is a tracking frame, yielding tracking frames at 120Hz (see Figure 5). Though the tracking frames comprise only 6% of the frames in the sequence, they yield sufficient temporal resolution to track most human motion.

We use the robust gradient-based optical flow algorithm described in [Black and Anandan 1993] to calculate a bi-directional optical flow field (Figure 8(b)) between each pair of successive tracking frames (Figure 7, yellow arrows). We linearly interpolate these "local" optical flow fields across the interspersed basis frames, and then use them to warp each basis frame into alignment with the nearest tracking frame in the direction of the target output frame (Figure 7, black arrows).

Subsequently, we calculate "long-range" flow between each tracking frame and the target frame (Figure 7, red arrows). We form an initial estimate of each long-range flow field by concatenating the local flows, and then we refine this flow using the optical flow algorithm. These flows can directly be used to finish warping the basis frame to the target output frame, as no linear interpolation is required.

The long-range warps are calculated through optical flow, using concatenations of the local warps as an initial estimate. In our bases (e.g., Figure 5), there are sixteen basis frames and one matte frame between each pair of successive yellow tracking frames, so 170 images are warped in this manner to align with the target output frame.

Since our tracking frames occur at 120Hz, we can synthesize an output sequence at 24fps, 30fps, or 60fps by choosing every 5th, 4th, or 2nd tracking frame as a target output frame. In this work all sequences have been rendered at 24fps.

6.1 Lazy Bases

The optical flow technique is sufficiently robust to track and warp images of a facial performance for up to $1/24^{th}$ of a second earlier or later, even for relatively rapid motion such as a quickly turning head. Thus, a target frame can safely gather its set of basis frames up to $1/24^{th}$ second earlier and $1/24^{th}$ second later. If we gather frames from both directions , the basis can span a full $1/12^{th}$ of a second. For 24fps output, this allows us to run the basis at *half* the target frame rate, allowing us to record a performance twice as long in the available camera memory and to expose for twice as long for each basis frame, achieving higher light levels. In our experiments, this allowed a frame rate of $12 \times 180 = 2160$fps rather than $24 \times 180 = 4320$fps. We call such a basis whose length exceeds the output frame rate a *lazy basis*. A lazy basis takes advantage of the fact that appearance changes due to reflectance tend to happen more slowly than appearance changes from motion.

6.2 Synthesizing Motion Blur

Since the basis images are captured at short shutter speeds, there is minimal motion blur once the basis images are aligned to a target output frame. The lack of motion blur can result in an artificially stroboscopic appearance at 24fps. In a manner similar to [Brostow and Essa 2001] which added motion blur to stop-motion animation, we synthesize motion blur by applying a line blur based on our optical flow fields (Figure 8(b)). Because our optical flow is computed at the temporal resolution of the 120Hz tracking frames, the motion blur can be applied in

a piecewise linear manner, producing a more accurate blur profile that can exhibit curves and accelerations in the 24fps output. We improved our results by using modified Wu anti-aliased lines [Wu 1991], yielding a somewhat smoother result than the line algorithm in [Brostow and Essa 2001].

7 Matting

We include a matte frame after each tracking frame in the lighting sequence (see Figure 5) in which the board behind the subject is illuminated by its own light sources. This yields an image of the shadowed subject against the brightly lit board. After filming the subject, we acquire a "clean plate" of the board for one cycle of the full pattern sequence. Dividing each pixel in a matte frame by the clean plate's average matte frame produces an alpha matte image where $\alpha = 0$ represents the foreground and $\alpha = 1$ represents the background. Since there is a matte frame for each tracking frame, we reduce noise in the matte by using a weighted average of the ten closest matte frames to the target output frame. Like the basis images, the matte frames are also motion compensated to align with each other.

Since the matte board receives some stray light in many of the basis frames (see Figure 5.), we use the appearance of the stray light in the clean plate to remove the stray light from the basis images of the actor. Specifically, for a basis image F and corresponding clean plate image C, we compute $F' = F - \alpha C$ to matte the basis image onto black, clamping any negative values to zero. Then, after these basis images are recombined to produce a relit image of the actor L, we composite L over a background B using the [Porter and Duff 1984] "over" operator $I_{final} = L + \alpha B$. We found that this technique produced good matte edges, but note that a rear-projected matte board with low frontal albedo could reduce the problem of stray light.

8 Reflectance Transformation

Following [Malzbender et al. 2001], we can modify both the diffuse and specular reflectance of a performance by processing the reflectance functions. Our reflectance functions are 156-pixel images corresponding to the observed RGB value of a pixel of the performance lit by each lighting direction.

We estimate the surface normal using a variant of photometric stereo [Woodham 1980]. With 156 lighting directions, the system is very overdetermined, which allows us to ignore reflectance function pixels that are likely to be image noise or specular reflections. Specifically, we sort the pixel values, discarding the lowest 50% and top 10% of the values, and use a hat function to weight the lighting directions at the middle of this range the highest when performing the photometric stereo. A surface normal map computed in this way is shown in Figure 9(b). We could alternately have computed the surface normals using a non-Lambertian reflectance model as in [Georghiades 2003].

The photometric stereo technique yields a diffuse albedo value for each pixel as seen in Figure 9(c). We furthermore derive an estimate of the surface point's geometric self-shadowing in the form of an *ambient occlusion* map [Landis 2002]. We first assume that the surface is unoccluded in the direction of the surface normal and that pixels darker than predicted by the lobe are due to shadowing from nearby surfaces. To obtain the map we fit a Lambertian cosine lobe in the direction of the normal to the albedo measurement and we sum the total shortfall of the reflectance function data to the modeled lobe over the hemisphere. An ambient occlusion map for a frame of a dynamic performance is shown in Figure 9(d).

In Figures 9(b-c), we note that the reflectance measurement techniques perform less well for the actor's hair and eyes where the surface is significantly non-Lambertian. For the hair, the behavior is generally stable and the artifacts are relatively unobjectionable in renderings. The reflectance measurements made of the eyes is poor due to the bright specularities from the light sources, and better estimating the reflectance properties of the eyes is an important area for future work.

Unlike [Debevec et al. 2000], we have not estimated a spatially-varying specular channel for the reflectance functions using color space separation. The quality of the results we obtained with the process were significantly reduced by our coarser lighting basis and noisier imaging system.

From the normals, albedo, and occlusion map, we can use reflection mapping [Miller and Hoffman 1984; Greene 1986] to render the performance reflecting different lighting environments. While this technique does not reproduce the complexities of facial reflectance as recorded by the original basis images, it allows the reflectance parameters to be modified in a straightforward manner. For example, using the normals to index into a diffuse convolution of a lighting environment and multiplying the irradiance values by both the albedo and the occlusion map produces a diffuse version of the performance (Figure 9(f)) when compared to the original reflectance (Figure 9(e)). Setting the diffuse lobe to zero and performing specular reflection mapping of the environment yields a synthetic specular channel for the performance (Figure 9(g)). Adding this specular channel to the original performance gives the actor a glossy appearance (Figure 9(h)).

This same process can be used to produce more stylized renderings of performances as seen in Figure 9(i-l). A dramatic effect is to set the albedo of the performance to a constant and multiply by the occlusion map which produces the appearance of a moving sculpture (Figure 9(j)); metallic reflection-mapped performances (Figure 9(k)) are possible as well. Several examples are seen in motion in the video.

9 Results

Figure 1 shows a performance relit in several lighting environments. It was captured using the lazy triangle basis (Figure 5 at 640×480 resolution and 2160fps. The top rows show the performance illuminated by two real-world lighting environments. The bottom row shows the performance illuminated by a user-designed three-light environment. The frames are taken from a 4.3 second performance. These and additional relit performances are included in the video.

Figures 9(a-d) show results from the reflectance estimation process, including the recovered surface normals, albedo, and ambient occlusion for a frame of a dynamic performance. Figures 9(e-h) show results from subtly decreasing and increasing the specular component of the performance using the albedo and surface normals. Figures 9(i-l) show several stylized reflectance transformations to various diffuse and specular materials. Several reflectance transformation sequences are included in the video.

10 Discussion and Future Work

Currently, our technique is highly data intensive, which limits our capture time to a few seconds and precludes interactive processing. As seen in [Debevec et al. 2000; Sloan et al. 2002; Ng et al. 2003; Masselus et al. 2004], reflectance functions can be compressed using DCT, spherical harmonic, or wavelet techniques and relit directly from the compressed data. With such techniques, it may be possible to interactively modify the lighting on a preprocessed performance. Ideally, reflectance function compression techniques

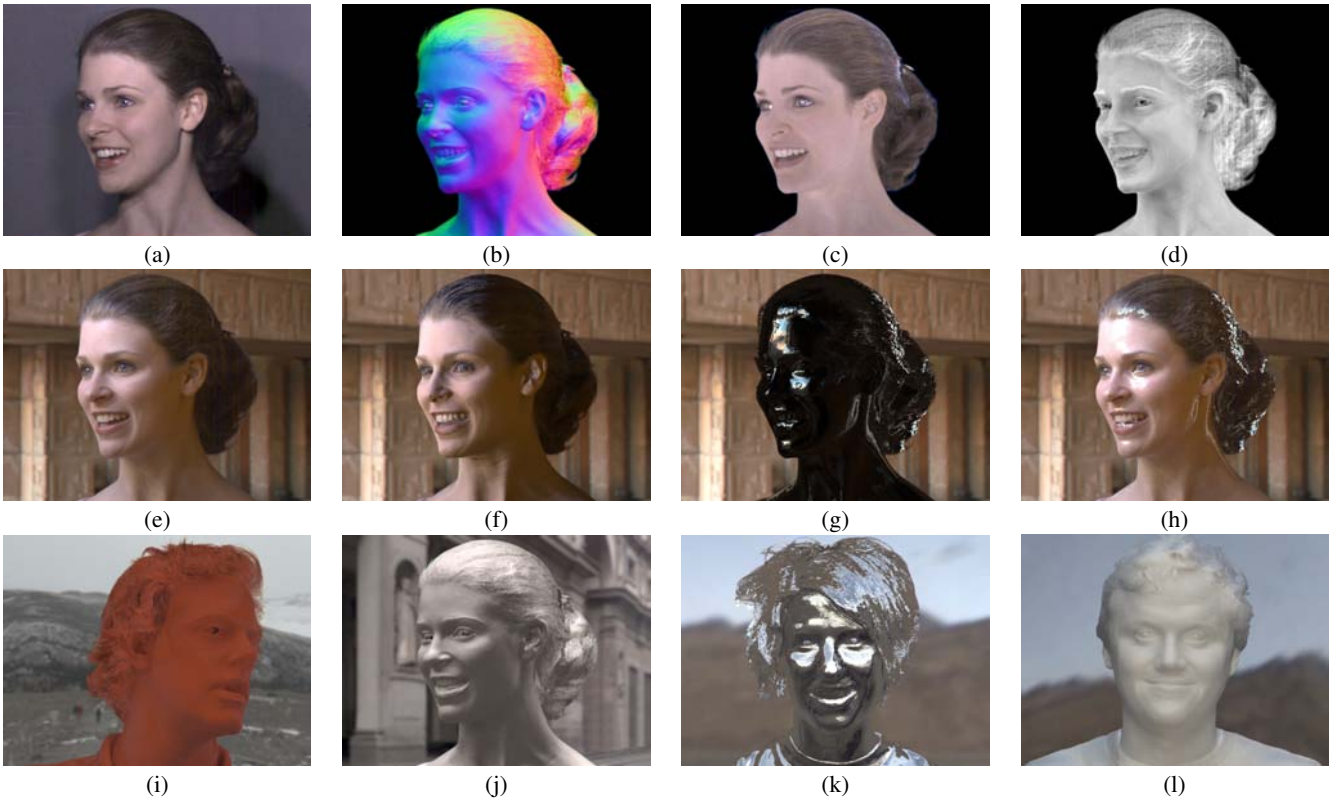


Figure 9: **Reflectance Estimation and Transformation** (a) An original photographed image from the performance with the triangle basis. (b) Estimated surface normals from the full lighting basis (c) Estimated diffuse albedo. (d) Estimated ambient occlusion. (e) Original reflectance re-illuminated by an environment. (f) Diffuse albedo with normals and occlusion illuminated by the environment. (g) Specular reflection in the environment. (h) Specular enhancement. (i) Stylized plastic reflectance with occlusion. (j) Diffuse reflectance with occlusion. (k) Metallic specular reflectance. (l) Diffuse reflectance without occlusion, yielding a translucent appearance.

could be adapted to run within the camera hardware to capture compressed data directly.

The optical flow system performed well for our sequences, but occasionally followed the wrong motion when shadows from the actor's hands passed over their face. We found that this was due to using only a subset of the lights to create the diffuse frames, and would improve if we turned on all the lights for a short period instead. Since several of our lights encircle the camera's viewpoint, it could be straightforward to use the technique of [Raskar et al. 2004] to detect depth edges as a way to provide *a priori* discontinuity information to the optical flow algorithm.

Our technique only reproduces illumination from distant lighting environments and at a resolution no more detailed than our 156 light sources. Adding a geometry capture component to the system would provide some of the information needed to simulate spatially-varying illumination and to synthesize shadows with greater fidelity than can be done directly from our data. Such a system could be based on integrating the normals from photometric stereo as in [Georghiades et al. 1999; Nishino and Nayar 2004] or by multiplexing structured light patterns such as those used in [Zhang et al. 2004] or [Rusinkiewicz et al. 2002] into the lighting basis. Simultaneously captured geometry could also offer the capability of modifying the viewpoint of the performance, though special considerations would need to be applied for hair. At the expense of additional cameras, our system could alternately be combined with a multiple-viewpoint performance capture process such as [Guenter et al. July 1998].

As noted earlier, we believe that additional exploration into the choice of the lighting basis is warranted. Perhaps, an optimal basis could exhibit some of the dynamic range improvements and relative lack of strobing of the Hadamard patterns but also maintain or increase the SNR in the face of photon noise. With a faster microcontroller, we could drive the lights at different illumination intervals allowing grayscale basis patterns such as spherical harmonics to be explored.

Finally, it is possible that the time-multiplexed lighting technique could be applied to traditional studio lighting arrangements rather than our specialized lighting apparatus. LED versions of traditional stage lights could be arranged at the time of filming, but their colors, intensities, and areas of influence could be designed and modified in postproduction. Developing such a hybrid workflow would benefit from on-set visualizations of potential lighting choices.

11 Conclusion

We have presented a technique that uses time-multiplexed illumination and high-speed photography to capture time-varying reflectance properties of a live performance. Our results demonstrate realistic relighting for both real-world lighting environments and manually-designed illumination, and exhibit realistic skin reflectance. While the technique requires specialized equipment and allows for only limited recording time, improvements in memory capacity and on-board camera image processing will likely increase the practicality of the technique. While the relighting capabilities are limited to non-local illumination effects of limited angular res-

olution, the technique holds promise as a novel and useful tool for filmmakers wishing to modify live-action lighting in postproduction.

Acknowledgements

The authors wish to thank Marissa Sellers (Figure 1), Charles-Felix Chabert (Figure 8), Elizabeth Franklin (Figure 9), Johan Toerne (Figure 9(l)), John Biondo, Mark Ollila, Maya Martinez, John Lai, Andrew Jones, Laurie Swanson, Peter James, Mark Bolas, Tom Pereira, Lora Chen, Krishna Mamidibathula, Larry Vladic, Mark Gangi, Tony Lucatorto, Paul Laureano, Vision Research, Inc., Alexander Singer, Randal Kleiser, Bill Swartout, David Wertheimer, and Neil Sullivan for their support and assistance with this work. This work was sponsored by the University of Southern California Office of the Provost and the U.S. Army Research, Development, and Engineering Command (RDECOM). The content of the information does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

References

BLACK, M. J., AND ANANDAN, P. 1993. A framework for the robust estimation of optical flow. In *Fourth International Conf. on Computer Vision*, 231–236.

BROSTOW, G. J., AND ESSA, I. 2001. Image-based motion blur for stop motion animation. In *Proceedings of ACM SIGGRAPH 2001*, Computer Graphics Proceedings, Annual Conference Series, 561–566.

CALAHAN, S. 2000. *Advanced Renderman: Creating CGI for Motion Pictures*. Morgan Kaufman Publishers, San Francisco, ch. Storytelling through lighting, a computer perspective, 337382.

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. *Proceedings of SIGGRAPH 2000* (July), 145–156.

DEBEVEC, P., WENGER, A., TCHOU, C., GARDNER, A., WAESE, J., AND HAWKINS, T. 2002. A lighting reproduction approach to live-action compositing. *ACM Transactions on Graphics 21*, 3 (July), 547–556.

FIELDING, R. 1985. *The Technique of Special Effects Cinematography*, 4th ed. Hastings House, New York.

GEORGHIADES, A., BELHUMEUR, P., AND KRIEGMAN, D. 1999. Illumination-based image synthesis: Creating novel images of human faces under differing pose and lighting. In *IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes*, 47–54.

GEORGHIADES, A. S. 2003. Recovering 3-d shape and reflectance from a small number of photographs. In *Eurographics Symposium on Rendering: 14th Eurographics Workshop on Rendering*, 230–240.

GERSHBEIN, R., AND HANRAHAN, P. M. 2000. A fast relighting engine for interactive cinematic lighting design. In *Proceedings of ACM SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, 353–358.

GREENE, N. 1986. Environment mapping and other application of world projections. *IEEE Computer Graphics and Applications 6*, 11 (November), 21–29.

GUENTER, B., GRIMM, C., WOOD, D., MALVAR, H., AND PIGHIN, F. July 1998. Making faces. *Proceedings of SIGGRAPH 98*, 55–66.

HAEBERLI, P. 1992. Synthetic lighting for photography. Available at http://www.sgi.com/grafica/synth/index.html, January.

HAGER, G. D., AND BELHUMEUR, P. N. 1996. Real-time tracking of image regions with changes in geometry and illumination. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 403–410.

HALL, M. 1998. *Combinatorial Theory*, 2nd ed. Wiley, New York.

HARWIT, M., AND SLOANE, N. J. A. 1979. *Hadamard transform optics*. Academic Press, New York.

HAWKINS, T., WENGER, A., TCHOU, C., AND DEBEVEC, A. G. F. G. P. 2004. Animatable facial reflectance fields. In *Eurographics Symposium on Rendering: 15th Eurographics Workshop on Rendering*.

HEWLETT-PACKARD COMPONENTS GROUP. 1998. Noise sources in cmos image sensors. Tech. rep., Hewlett-Packard.

KANG, S. B., UYTTENDAELE, M., WINDER, S., AND SZELISKI, R. 2003. High dynamic range video. *ACM Transactions on Graphics 22*, 3 (July), 319–325.

LANDIS, H., 2002. Production-ready global illumination. Course Notes for SIGGRAPH 2002 Course 16, RenderMan in Production.

MALZBENDER, T., GELB, D., AND WOLTERS, H. 2001. Polynomial texture maps. *Proceedings of SIGGRAPH 2001* (August), 519–528.

MARKS, J., ANDALMAN, B., BEARDSLEY, P. A., FREEMAN, W., GIBSON, S., HODGINS, J. K., KANG, T., MIRTICH, B., PFISTER, H., RUML, W., RYALL, K., SEIMS, J., AND SHIEBER, S. 1997. Design galleries: A general approach to setting parameters for computer graphics and animation. In *Proceedings of SIGGRAPH 97*, Computer Graphics Proceedings, Annual Conference Series, 389–400.

MARSCHNER, S., GUENTER, B., AND RAGHUPATHY, S. 2000. Modeling and rendering for realistic facial animation. In *Rendering Techniques 2000: 11th Eurographics Workshop on Rendering*, 231–242.

MASSELUS, V., PEERS, P., DUTRE, P., AND WILLEMS, Y. D. 2004. Smooth reconstruction and compact representation of reflectance functions for image-based relighting. In *15th Eurographics Symposium on Rendering*, no. Norrkoping, Sweden.

MILLER, G. S., AND HOFFMAN, C. R. 1984. Illumination and reflection maps: Simulated objects in simulated and real environments. In *SIGGRAPH 84 Course Notes for Advanced Computer Graphics Animation*.

NG, R., RAMAMOORTHI, R., AND HANRAHAN, P. 2003. All-frequency shadows using non-linear wavelet lighting approximation. *ACM Transactions on Graphics 22*, 3 (July), 376–381.

NIMEROFF, J. S., SIMONCELLI, E., AND DORSEY, J. 1994. Efficient re-rendering of naturally illuminated environments. In *Fifth Eurographics Workshop on Rendering*, 359–373.

NISHINO, K., AND NAYAR, S. K. 2004. Eyes for relighting. *ACM Transactions on Graphics 23*, 3 (Aug.), 704–711.

PETROVIĆ, L., FUJITO, B., WILLIAMS, L., AND FINKELSTEIN, A. 2000. Shadows for cel animation. In *Proceedings of ACM SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, 511–516.

PORTER, T., AND DUFF, T. 1984. Compositing digital images. In *Computer Graphics (Proceedings of SIGGRAPH 84)*, vol. 18, 253–259.

RASKAR, R., TAN, K.-H., FERIS, R., YU, J., AND TURK, M. 2004. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *ACM Transactions on Graphics 23*, 3 (Aug.), 679–688.

RUSINKIEWICZ, S., HALL-HOLT, O., AND LEVOY, M. 2002. Real-time 3d model acquisition. *ACM Transactions on Graphics 21*, 3 (July), 438–446.

SCHECHNER, Y. Y., NAYAR, S. K., AND BELHUMEUR, P. 2003. A theory of multiplexed illumination. In *International Conference on Computer Vision*.

SLOAN, P.-P., KAUTZ, J., AND SNYDER, J. 2002. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Transactions on Graphics 21*, 3 (July), 527–536.

SLOAN, N. J. A., 1999. A library of hadamard matrices. http://www.research.att.com/~njas/hadamard/.

SMITH, A. R., AND BLINN, J. F. 1996. Blue screen matting. In *Proceedings of SIGGRAPH 96*, 259–268.

TRUMBULL, D. 2000. Personal communication. January.

WEN, Z., LIU, Z., AND HUANG, T. S. 2003. Face relighting with radiance environment maps. In *2003 Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, 158–165.

WILLIAMS, L. 1991. Shading in two dimensions. In *Graphics Interface '91*, 143–151.

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

WU, X. 1991. An efficient antialiasing technique. In *SIGGRAPH 91: Proceedings of the 18th annual conference on Computer graphics and interactive techniques*, ACM Press, 143–152.

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Spacetime faces: high resolution capture for modeling and animation. *ACM Transactions on Graphics 23*, 3 (Aug.), 548–558.

ZONGKER, D. E., WERNER, D. M., CURLESS, B., AND SALESIN, D. H. 1999. Environment matting and compositing. *Proceedings of SIGGRAPH 99* (August), 205–214.