

A Lighting Reproduction Approach to Live-Action Compositing

Paul Debevec Andreas Wenger[†] Chris Tchou Andrew Gardner Jamie Waese Tim Hawkins

University of Southern California Institute for Creative Technologies¹

ABSTRACT

We describe a process for compositing a live performance of an actor into a virtual set wherein the actor is consistently illuminated by the virtual environment. The Light Stage used in this work is a two-meter sphere of inward-pointing RGB light emitting diodes focused on the actor, where each light can be set to an arbitrary color and intensity to replicate a real-world or virtual lighting environment. We implement a digital two-camera infrared matting system to composite the actor into the background plate of the environment without affecting the visible-spectrum illumination on the actor. The color response of the system is calibrated to produce correct color renditions of the actor as illuminated by the environment. We demonstrate moving-camera composites of actors into real-world environments and virtual sets such that the actor is properly illuminated by the environment into which they are composited.

Keywords: Matting and Compositing, Image-Based Lighting, Radiosity, Global Illumination, Reflectance and Shading

1 Introduction

Many applications of computer graphics involve compositing, the process of assembling several image elements into a single image. An important application of compositing is to place an image of an actor over an image of a background environment. The result is that the actor, originally filmed in a studio, will appear to be in either a separately photographed real-world location or a completely virtual environment. Most often, the desired result is for the actor and environment to appear to have been photographed at the same time in the same place with the same camera, that is, for the results to appear to be real.

Achieving a realistic composite requires matching many aspects of the image of the actor and image of the background. The two elements must be viewed from consistent perspectives, and their boundary must match the contour of the actor, blurring appropriately when the actor moves quickly. The two elements need to exhibit the same imaging system properties: brightness response curves, color balance, sharpness, lens flare, and noise. And finally,

¹USC ICT, 13274 Fiji Way, Marina del Rey, CA, 90292

Email: debevec@ict.usc.edu, aw@cs.brown.edu, tchou@ict.usc.edu, gardner@ict.usc.edu, jamie_waese@cbc.ca, timh@ict.usc.edu.

[†]Andreas Wenger worked on this project during a summer internship at USC ICT while a student of computer science at Brown University.

See also <http://www.debevec.org/Research/LS3/>



Figure 1: **Light Stage 3** focuses 156 red-green-blue LED lights toward an actor, who is filmed simultaneously with a color camera and an infrared matting system. The device composites an actor into a background environment as illuminated by a reproduction of the light from that environment, yielding a composite with consistent illumination. In this image the actor is illuminated by light recorded in San Francisco's Grace Cathedral.

the lighting of the two elements needs to match: the actor should exhibit the same shading, highlights, indirect illumination, and shadows they would have had if they had really been standing within the background environment. If an actor is composited into a cathedral, his or her illumination should appear to come from the cathedral's candles, altars, and stained glass windows. If the actor is walking through a science laboratory, they should appear to be lit by fluorescent lamps, blinking readouts, and indirect light from the walls, ceiling, and floor. In wider shots, the actor must also photometrically affect the scene: properly casting shadows and appearing in reflections.

The art and science of compositing has produced many techniques for matching these elements, but the one that remains the most challenging has been to achieve consistent and realistic illumination between the foreground and background. The fundamental difficulty is that when an actor is filmed in a studio, they are illuminated by something often quite different than the environment into which they will be composited – typically a set of studio lights and the green or blue backing used to distinguish them from the background. When the lighting on the actor is different from the lighting they would have received in the desired environment, the composite can appear as a collage of disparate elements rather than an integrated scene from a single camera, breaking the sense of realism and the audience's suspension of disbelief.

Experienced practitioners will make their best effort to arrange the on-set studio lights in a way that approximates the positions, colors, and intensities of the direct illumination in the desired background environment; however, the process is time-consuming and difficult to perform accurately. As a result, many digital composites require significant image manipulation by a compositing artist to

convincingly match the appearance of the actor to their background. Complicating this process is that once a foreground element of an actor is shot, the extent to which a compositor can realistically alter the form of its illumination is relatively restricted. As a result of these complications, creating live-action composites is labor intensive and sometimes fails to meet the criteria of realism desired by filmmakers.

In this paper, we describe a process for achieving realistic composites between a foreground actor and a background environment by directly illuminating the actor with a reproduction of the direct and indirect light of the environment into which they will be composited. The central device used in this process (Figure 1) consists of a sphere of one hundred and fifty-six inward-pointing computer-controlled light sources that illuminate an actor standing in the center. Each light source contains red, green, and blue light emitting diodes (LEDs) that produce a wide gamut of colors and intensities of illumination. We drive the device with measurements or simulations of the background environment’s illumination, and acquire a color image sequence of the actor as illuminated by the desired environment. To create the composite, we implement an infrared matting system to form the final moving composite of the actor over the background. When successful, the person appears to actually be within the environment, exhibiting the appropriate colors, highlights, and shadows for their new environment.

2 Background and Related Work

The process of compositing photographs of people onto photographic backgrounds dates back to the early days of photography, when printmakers would create “combination prints” by multiply-exposing photographic paper with different negatives in the darkroom [4]. This process evolved for motion picture photography, meeting with early success using rear projection in which a movie screen showing the background would be placed behind the actor and the two would be filmed together. A front-projection approach developed in the 1950s (described in [10]) involved placing a retro-reflective screen behind the actor, and projecting the background image from the vantage point of the camera lens to reflect back toward the camera. The most commonly used process to date has been to film the actor in front of a blue screen, and process the film to produce both a color image of the actor and a black and white matte image showing the actor in white and the background in black. Using an optical printer, the background, foreground, and matte elements can be combined to form the composite of the actor in front of the background.

The color difference method developed by Petro Vlahos (described in [21]) presented a technique for reducing blue spill from the background onto the actor and providing cleaner matte edges for higher quality composites. Variants of the blue screen process have used a second piece of film to record the matte rather than deriving it from a color image of the actor. A technique using infrared (or ultraviolet) sensitive film [24] with a similarly illuminated backdrop allowed the mattes to be acquired using non-visible light, which avoided some of the problems associated with the background color spilling onto the actor. A variant of this process used a background of yellow monochromatic sodium light, which could be isolated by the matte film and blocked from the color film, producing both high-quality mattes and foregrounds.

The optical printing process has been refined using digital image processing techniques, with a classic paper on digital compositing presented at SIGGRAPH 84 by Porter and Duff [19]. More recently, Smith and Blinn [21] presented a mathematical analysis of the color difference method and explained its challenges and variants. They also showed that if the foreground element could be photographed with two different known backings, then a correct matte for the object could be derived for any colored object without apply-

ing the heuristic approximations used in other methods. The recent technique of Environment Matting [27] showed that by using a series of patterned backgrounds, the object could be shown properly refracting and reflecting the background behind it. Both of these latter techniques generally require several images of the subject with different backgrounds, which poses a challenge for compositing a live performance using this method. [5] showed an extension of environment matting applied to live video.

Our previous work [7] addressed the problem of replicating environmental illumination on a foreground subject when compositing into a background environment. This work illuminated the person from a large number of incident illumination directions, and then computed linear combinations of the resulting images as in [18] to show the person as they would appear in a captured illumination environment [17, 6]. This synthetically illuminated image could then be composited over an image of the background, allowing the apparent light on the foreground object to be the same light as recorded in the real-world background. Again, this technique (and subsequent work [15, 13]) requires multiple images of the foreground object which precludes its use for live-action subjects.

In this work we extend the lighting reproduction technique in [7] to work for a moving subject by using an entire sphere of computer-controlled light sources to simultaneously illuminate the subject with a reproduction of a captured or computed illumination environment. To avoid the need for a colored backing (which might contribute to the illumination on the subject and place restrictions on the allowable foreground colors), we implement a digital version of Zoli Vidor’s infrared matting system [24] using two digital video cameras to composite the actor into the environment by which they are being illuminated.

A related problem in the field of building science has been to visualize architectural models of proposed buildings in controllable outdoor lighting conditions. One device for this purpose is a *heliodon*, in which an artificial sun is moved relative to an architectural model, and another is a *sky simulator*, in which a hemispherical dome is illuminated by arrays of lights of the same color to simulate varying sky conditions. Examples of heliodons and sky simulators are at the College of Environmental Design at UC Berkeley [25], the Building Technology Laboratory at the University of Michigan, and the School of Architecture at Cardiff University [1].

3 Apparatus

This section describes the construction and design choices made for the structure of Light Stage 3, its light sources, the camera system, and the matte backing.

3.1 The Structure

The purpose of the lighting structure is to position a large number of inward-pointing light sources, distributed about the entire sphere of incident illumination, around the actor. The structure of Light Stage 3 (Figure 1) is based on a once subdivided icosahedron, yielding 42 vertices, 120 edges, and 80 faces. Placing a light source at each vertex and in the middle of each edge allows for 162 lights to be evenly distributed on the sphere. The bottom 5-vertex of the structure and its adjoining edges were left out to leave room for a person, reducing the number of lights in the device to 156. The stage was designed to be two meters in diameter, a size chosen to fit a standard room, to be large enough to film as wide as a medium close-up, and close enough to provide sufficient illumination. We mounted the sphere on a base unit 80cm tall to place the center of the sphere at the height of a tall person’s head. To minimize reflections onto the actor, the structure was painted matte black and placed in a dark-colored room.

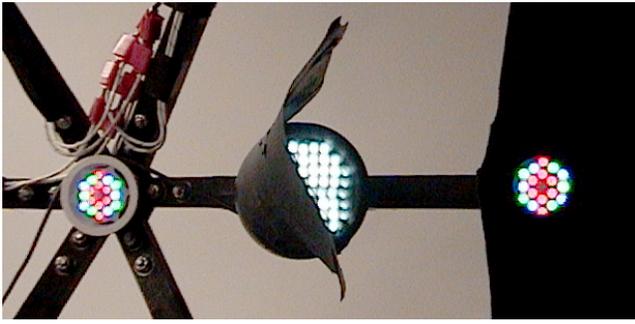


Figure 2: Color and Infrared Lights A detail of Light Stage 3 showing two of the 156 RGB LED lights and one of the six infrared LED light sources. At right is the edge of the infrared-reflective backing for the stage, attached to the fronts of the lights with holes to let the lights shine through. The infrared lights, flagged with black tape to illuminate just the backing, produce light invisible to both the human eye and the color camera used in our system. They do, however, produce a response in the digital camera used to take this picture.

3.2 The Lights

The lights we used were iColor MR lights from Color Kinetics corporation. Each iColor light (see Figure 2) consists of a mixture of eighteen red, green, and blue LEDs. Each light, at full intensity, produces approximately 20 lux on a surface one meter away. The beam spread of each light is slightly narrower than ideal, falling off to 50% of the maximum at twelve degrees from the beam center. We discuss how we compensated for this falloff in Section 4.

The lights receive their power and control signals from common wires, simplifying the wiring. The lights interface to the computer via the Universal Serial Bus allowing us to update all of the light colors at video frame rates. Each color component of each light is driven by an 8-bit values 0 through 255, with the resulting intensities produced through pulse width modulation of the current to the LEDs. The rate of the modulation is sufficiently fast to produce continuous shades to the human eye as well as to the color video camera. The color LEDs produce no infrared light, facilitating the infrared matting process and avoiding heat on the actor.

We found 156 lights to be sufficient to produce the appearance of a continuous field of illumination as reflected in both diffuse and specular skin (Fig. 3), a result consistent with the signal processing analysis of diffuse reflection presented in [20]. For tight closeups, however, the lights are sparse enough that double shadows can be seen when two neighboring lights are used to approximate a single light halfway between them (Fig. 3(d)). The reflection of the lights in extreme closeups of the eyes also reveals the discrete light sources (Fig. 3(f)), but for a medium closeup the lights blur together realistically (Fig. 3(f)). We found the total illumination from the lights to be just adequate for getting enough illumination to the cameras, which suggests that even more lights could be used.

3.3 The Camera System

The camera system (Figure 4) was designed to simultaneously capture color and infrared images of the actor in the light stage. For these we chose a Sony DXC-9000 three-CCD color camera and a Uniq Vision UP-610 monochrome camera. The DXC-9000 was chosen to produce high quality, progressive scan color images at up to 60 frames per second, and the UP-610 was chosen for producing high quality, progressive scan monochrome images with good sensitivity in the near infrared (700-900 nm) spectrum asynchronously up to 110 frames per second. Since the UP-610 is also sensitive to visible light, we used a Hoya R72 infrared filter to transmit only the infrared light to the UP-610. Conveniently, the DXC-9000 color

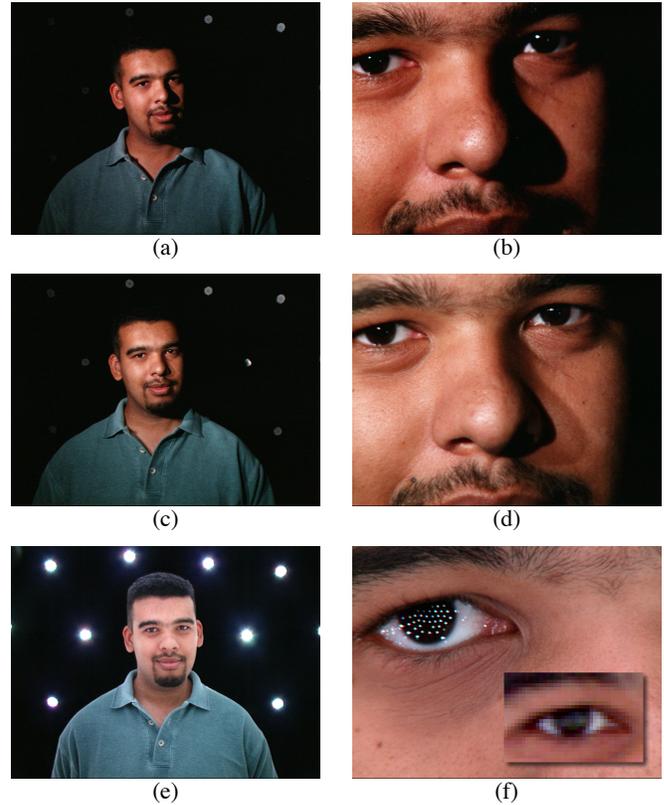


Figure 3: Lighting Resolution (a) An actor illuminated by just one of the light sources. (b) A closeup of the shadows on the actor's face in (a). (c) The actor illuminated by two adjacent light sources, approximating a light halfway between them. (d) A closeup of the resulting double shadow. (e) The actor illuminated by all 156 lights, approximating an even field of illumination. (f) A closeup of the actor's eye in this field, showing the lights as individual point reflections. Inset: a detail of video image (e), showing that the light sources are numerous enough to appear as a continuous reflection for video resolution medium closeups.

camera had no response to infrared light, which is not true of all color video cameras.

The DXC-9000 and UP-610 cameras were mounted at right angles on a board using two translating Bogen tripod heads, allowing the cameras to be easily adjusted forwards and backwards up to four inches. This adjustment allows the camera nodal points to be re-aligned for different lens configurations. A glass beam splitter was placed between the two cameras at a 45 degree angle, reflecting 30% of the light toward the UP-610 and transmitting 70% to the DXC-9000, which worked well for the relative sensitivities of the two cameras. Improved efficiency could be obtained by using a "hot mirror" to reflect just the infrared light toward the infrared camera.

Both cameras yield video-resolution 640×480 pixel images; for future work it would be of interest to design a high-definition version of this system. Each camera is connected to a PC computer with a frame grabber for capturing images directly to memory. With 2GB of memory, the systems can record shots up to fifteen seconds long.

We used the DXC-9000's standard zoom lens at its widest setting (approximately a 40 degree field of view) in order to see as much of the environment behind the person as possible. For the UP-610 we used a 6mm lens with a 42 degree horizontal field of view, allowing it to contain the field of view of the DXC-9000. Ideally, a future

version of the system would employ a single 4-channel camera sensitive to both infrared and RGB color light, allowing changing the focus, aperture, and zoom during a shot.

3.4 The Infrared Lights and Cloth

The UP-610 camera is used to acquire a matte image for compositing the actor over the background. We wished to obtain the matte without restricting the colors in the foreground and without illuminating the actor with colored light from a blue or green backing. We solved both of these problems by implementing an infrared matting system wherein the subject is placed in front of a field of infrared illumination. We chose the infrared system over a polarization-based system [2] due to the relative ease of illuminating the background with infrared rather than polarized light. We used infrared rather than ultraviolet light due to the relative ease of obtaining infrared lights, cameras, and optics. And we chose the infrared system over a front-projection system due to the relative ease of constructing the backing and to facilitate making adjustments to the background image (for example, placing virtual actors in the background) after shooting.

To provide the infrared field we searched for a backing material that would reflect infrared light but not visible light; if the material were to reflect visible light it would reflect stray illumination onto the actor. To find this material, we used a Sony “night shot” camcorder fitted with the Hoya IR-pass filter. With this we found many black fabrics in a typical fabric store that reflected a considerable proportion of incident infrared light (Figure 5). We chose one of the fabrics that was relatively dark, reasonably durable, and somewhat stretchy.

We cut the fabric to cover twenty faces of the lighting structure, filling the field of view of the video cameras with additional latitude to pan and tilt. The cloth was cut to be slightly smaller than its intended size so that it would stretch tightly, and holes were cut for the light sources to shine through. The fabric was attached to the fronts of the lights with Velcro.

To light the backing, we attached six Clover Electronics infrared LED light sources (850nm peak wavelength) to the inside of the stage between the colored lights (Figure 2). The cloth backing, illuminated by the infrared LEDs and seen from the infrared camera, is shown in Figure 6(a). Though darker than the cloth, enough



Figure 4: The Camera System The camera system consists of a 3-CCD Sony DXC-9000 color video camera to record the actor (left) and a Uniq Vision UP-610 monochrome video camera to record the infrared matte. The cameras are placed with coincident nodal points using a 30R / 70T beam splitter.



Figure 5: Infrared Reflection of Cloth (a) Six black cloth samples illuminated by daylight, seen in the visible spectrum. (b) The reflection of the same cloth materials under daylight in the near infrared part of the spectrum. The top sample is the one used for the cloth backing in the light stage, just under it is a sample of Duvateen, and the bottom four are various black T-shirts.

light reflects off the light sources to detect them as part of the background.

4 System Calibration

This section presents the procedures we performed to calibrate the color and intensity of the light sources, the geometric registration of the color and infrared video images, and the brightness consistency of the color and infrared video images.

4.1 Intensity and Color Calibration

Our calibration process began by measuring the intensity response characteristics of both the light sources and the cameras, so that all of our image processing could be performed in relative radiometric units. We first measured the intensity response characteristics of the iColor MR LED lights using a Photo Research PR-650 spectroradiometer, which acquires radiometric readings every 4nm from 380nm to 780nm. We iterated through the 8-bit values for each of the colors of the lights, taking a spectral measurement at each value. We found that the lights maintained the same spectrum at all levels of brightness and that they exhibited a nonlinear intensity response to their 8-bit inputs (Fig. 7(a)). We found that the intensity response for each channel could be modeled accurately by a gamma curve of the form $L = (z/255)^\gamma$ where L is the relative light radiance, z is the eight-bit input to the light, and $\gamma = 1.86$. From this equation we map desired linear light levels to 8-bit light inputs as $z = 255L^{1/\gamma}$. The nonlinear response is a great benefit since it dramatically increases each light’s usable dynamic range: a fully illuminated light is more than ten thousand times as bright as a minimally illuminated one, and this allows us to make use of high dynamic range illumination environments. To ensure that the lights produced no ultraviolet or infrared light, we measured the emission spectra of the red, green, and blue light LEDs as shown in Figure 8.

We calibrated the intensity response characteristics of the color channels of the DXC-9000 color video camera (Fig. 7(b)) and the monochrome UP-610 camera using the calibration technique in [8]. This allowed us to represent imaged pixel values using linear values in the range $[0, 1]$ for each channel. For the rest of this discussion all RGB triples for lights and pixels refer to such linear values.

The light stage software control system (Fig. 9) takes as input a high dynamic range omnidirectional image of incident illumination [8, 6] in a longitude-latitude format; we have generally used images sampled to 128×64 pixels in the PFM [22] floating-point image format. We determine the 156 individual light colors from such an image as follows. First, the triangular structure of the light stage is projected onto the image to divide the image into triangular cells with one light at each vertex. The color for each light is determined as an average of the pixel values in the five or six triangles adjacent

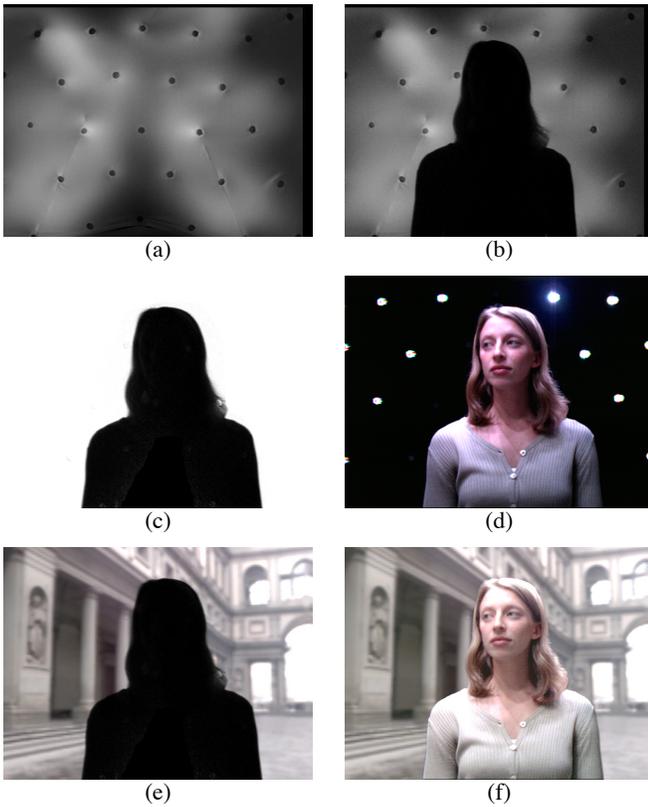


Figure 6: **The Matting Process** (a) The cloth backing, illuminated by the six infrared light sources. The dark circular disks are the fronts of the RGB lights appearing through holes in the cloth. (b) An infrared image of the actor in the light stage. (c) The matte obtained by dividing image (b) by the clean plate in (a); the actor remains black and the background divides out to white. (d) The color image of the actor in a lighting environment. (e) A corresponding environment background image, multiplied by the matte in (c) leaving room for the actor. (f) A composite image formed by compositing (d) onto (e) using the matte (c). The actor’s image has been adjusted for color balance and brightness falloff as described in Section 4.

to its vertex. In each triangle, the pixels are weighted according to their Barycentric weights with respect to the light’s vertex so that a pixel will contribute more light to closer vertices. To treat all light fairly, the contribution of each pixel is weighted by the solid angle that the pixel represents.

As an example, an image of incident illumination with all pixels set to an RGB value of $(0.25, 0.5, 0.75)$ sets each light to the color $(0.25, 0.5, 0.75)$. An image with just one non-zero pixel illuminates either one, two, or three of the lights depending on whether the pixel’s location maps to a vertex, edge, or face of the structure. If such a single-pixel lighting environment is rotated, different combinations of one, two, and three neighboring lights will fade up and down in intensity so that the pixel’s illumination appears to travel continuously with constant intensity.

We calibrated the absolute intensity and color balance of our system by placing a white reflectance standard 11 into the middle of the light stage facing the camera. The reflectance standard is close to Lambertian and approximately 99% reflective across the visible spectrum, and we assume it to have a diffuse albedo $\rho_d = 1$ for our purposes. Suppose that the light stage could produce a completely even sphere of incident illumination with radiance L coming from all directions. Then, by the rendering equation [12], the radiance R of the standard would be:

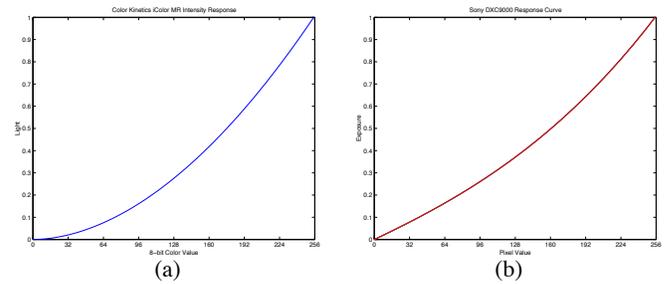


Figure 7: **Intensity Response Curves** (a) The measured intensity response curve of the iColor MR lights. (b) The derived intensity response curve of the DXC-9000 color video camera.

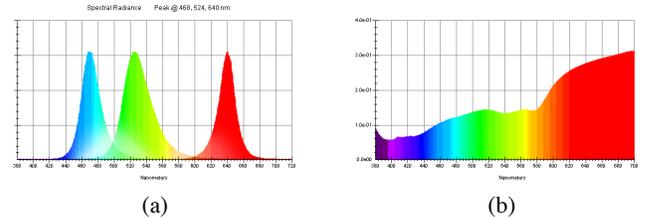


Figure 8: **Light Source and Skin Spectra** (a) The emission spectra of the red, green, and blue LED lights, from 380nm to 700nm. (b) The spectral reflectance of a patch of skin on a person’s face.

$$R = \int_{\Omega} L \frac{\rho_d}{\pi} \cos\theta \, d\omega = L \int_{\Omega} \frac{1}{\pi} \cos\theta \, d\omega = L.$$

This equation simply makes the observation that if a white diffuse object is placed within an even field of illumination, it will appear the same intensity and color as its environment. This observation allows us to photometrically calibrate the light stage: for an incident illumination image with RGB pixel value P , we should send an RGB value to the lights L to produce a reflected value R from the reflectance standard as imaged by the color video camera where $R = P$. If we simply choose $L = P$, this will not necessarily be the case, since the spectral sensitivities of the CCDs in the DXC-9000 camera need not have any particular relationship to the spectral curves of the iColor LEDs. A basic result from color science [26], however, tells us that there exists a 3×3 linear transformation matrix \mathbf{M} such that if we choose $L = \mathbf{M}P$ we will have $R = P$ for all P .

We compute this color transformation \mathbf{M} by sending a variety of colors L_i distributed throughout the RGB color cube to the lights and observing the resulting colors R_i of the reflectance standard imaged in the color video camera. Since we desire that $\mathbf{R} = \mathbf{P}$, we substitute R for P to obtain:

$$\mathbf{L} = \mathbf{M}\mathbf{R}.$$

Since M is a 3×3 matrix, we can solve for it in a least squares sense as long as we have at least three (L_i, R_i) color measurements. We accomplished this using the MATLAB numerical analysis software package by computing $\mathbf{M} = \mathbf{L}\mathbf{R}$, the package’s notation for solving for \mathbf{M} using the Singular Value Decomposition method.

With \mathbf{M} determined, we choose light colors $L = \mathbf{M}P$ to ensure that the intensity and color of the light reflected by the standard matches the the color and intensity of any even sphere of incident illumination. For our particular color camera and LED lights, we found that the matrix \mathbf{M} was nearly diagonal, with the off-diagonal elements being less than one percent of the magnitude of the diagonal elements. Specifically, the red channel of the camera responded

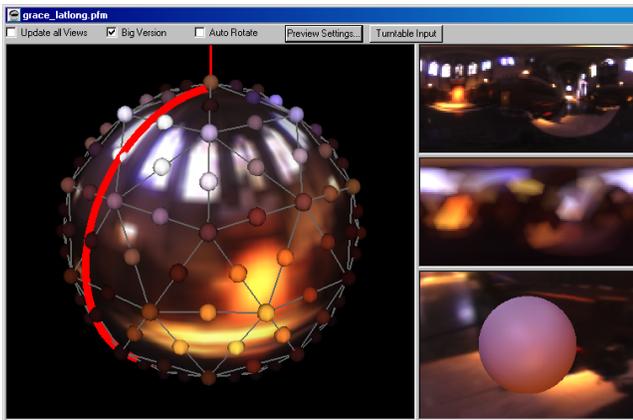


Figure 9: **The Lighting Control Program**, showing the longitude/latitude image of incident illumination (upper right), the triangular resampling of this illumination to the light stage’s lights (middle right), the light colors visualized upon the sphere of illumination (left), and a simulated sphere illuminated by the stage and composited onto the background (lower right).

very slightly to the green of the lights, and the green channel responded very slightly to the blue. Since the off-diagonal elements were so small, we in practice computed the color transformation from P to L as a simple scaling of the color channels of P based on the diagonals of the matrix M .

We note that while this transformation corrects the appearance of the white reflectance standard for all colors of incident illumination, it does not necessarily correct the appearance of objects with more complex reflectance spectra. Fortunately, the spectral reflectance of skin (Figure 8(d)) is relatively smooth, so calibrating the light colors based on the white reflectance standard is a reasonable approach. A spectral analysis of this lighting process is left as future work (see Section 6).

4.2 Matte Registration

Our matte registration process involved first aligning the nodal points of the color and infrared cameras and then computing a 2D warp to compensate for the remaining misalignment. We measured the nodal points of each camera by placing the camera flat on a large horizontal piece of cardboard, and viewing a live video image of the camera output on the computer screen. We then used a long ruler to draw lines on the cardboard radiating out from the camera lens. For each line, we aligned the ruler so that its edge would appear perfectly vertical in live image (Fig. 10a), at times moving the video window partially off the desktop to verify the straightness. We drew five lines for each camera (Fig. 10b) including lines for the left and right edges of the frame, a line through the center, and two intermediate lines, and also marked the position of the front of the camera lens. Removing the camera, we traced the radiating lines to their point of convergence behind the lens and noted the distance of the nodal point with respect to the lens front. Also, we measured the horizontal field of view of each lens using a protractor. For this project, we assumed the center of projection for each camera to be at the center of the image; we realized that determining such calibration precisely would not be necessary for the case of placing organic forms in front of backdrops. The camera focal lengths and centers of projection could also have been determined using image-based calibration techniques (e.g. [23]), but such techniques do not readily produce the position of the nodal point of the camera with respect to the front of the lens.

We determined that for the UP-610 camera that the nodal point was 17mm behind the front of the lens. For the DXC-9000, the

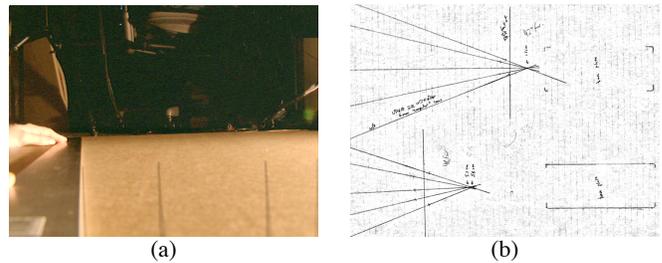


Figure 10: **Measuring Nodal Point Location and Field of View** (a) The nodal point and field of view of each camera were measured by setting the camera on a horizontal piece of cardboard (a) and drawing converging lines on the cardboard corresponding to vertical lines in the camera’s field of view. (b) The field of view was measured as the angle between the leftmost and rightmost lines; the nodal point, with respect to the front of the lens, was measured as the convergence of the lines.

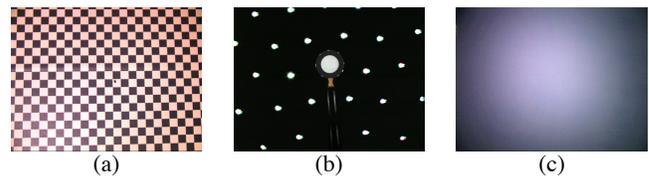


Figure 11: **Geometric and Photometric Calibration** (a) The calibration grid placed in the light stage to align the infrared camera image to the color camera image (b) A white reflectance standard imaged in an even sphere of illumination, used to determine the mapping between light colors and camera colors (c) A white card placed in the light stage illuminated by all of the lights to calibrate the intensity falloff characteristics.

rays converged in a locus between 52mm and 58mm behind the front surface of the lens. Such behavior is not unexpected in a complex zoom lens, as complex optics need not exhibit an ideal nodal point. We chose an intermediate value of 55mm and positioned the two cameras 85mm and 47mm away from the front surface of the beam splitter respectively for the UP-610 and the DXC-9000, placing each camera’s nodal point 102mm behind the glass.

Even with the lenses nodally aligned, the matte and color images will not line up pixel for pixel (after horizontally flipping the reflected UP-610 image) due to differences in the two cameras’ radial distortion, sensor placement, and the fact that infrared light may focus differently than visible light. We corrected for any such misalignments by photographing a checkerboard grid covering both fields of view placed at the same distance from the camera as the actor (Fig. 11(a)). One of the central white tiles was marked with a dark spot, and the corners of this spot were indicated to the computer in each image to seed the automatic registration process. The algorithm convolves each image with a 4×4 image of an ideal checker corner, and takes the absolute value of the output, to produce images where the checker corners appear as white dots on a black background. The subpixel location of the dot’s center is determined by finding the brightest pixel in a small region around the dot and then fitting a quadratic function through it and its 4-neighbors. Using the seed square as a guide, the program searches spirally outward to calculate the remaining tile correspondences between the two images.

With correspondences made between the checker corners in the two images, 2D homographies [9] are computed to map the pixels in each square in the color image to pixels in each square in the IR image. From these homographies, a u - v displacement map is created indicating the subpixel mapping between the color and IR image coordinates. With this map, each matte is warped to align

with the visible camera image for each frame of the captured sequence.

In this work we ran the color camera at its standard rate of 59.94 frames per second, saving every other frame to RAM to produce 29.97fps video. We ran the matte camera at its full rate of 110 frames per second, and took weighted averages of the matte images to produce 29.97fps mattes temporally aligned to the video. (The weight of each matte image was proportional to its temporal overlap with the corresponding color image, and the sequences were synchronized with sub-frame accuracy by observing the position of a stick waved in front of the cameras at the beginning of the sequence.) Though this only approximated having simultaneously photographed mattes, the approach worked well even for relatively fast subject motion. Future work would be to drive the matte camera synchronized to the color camera, or ideally to use a single camera capable of detecting both RGB color and infrared light.

4.3 Image Brightness Calibration

The iColor MR lights are somewhat directional and do not light the subject entirely evenly. We perform a first-order correction to this brightness falloff by placing a white forward-facing card in the center of the light stage covering the color camera's field of view. We then raise the illumination on all of the lights until just before the pixel values begin to saturate, and take an image of the card (Fig. 11(c)). The color of the card image is scaled so that its brightest value is (1,1,1); images from the color camera are then corrected for brightness falloff by dividing these images by the normalized image of the card. While this calibration process significantly improves the brightness consistency of the color images, it is not perfect since in actuality different lights will decrease in brightness differently depending on their orientation toward the card.

4.4 Matte Extraction and Compositing

We use the following technique to derive a clean matte image for creating the composite. The matte background is not generally illuminated evenly (Fig. 6(a)), so we also acquire a "clean plate" of the matte (Fig. 6(b)) without the actor for each shot. We then divide the pixel values of each matte image by the pixel values of the clean plate image (Fig. 6(c)). After this division, the silhouetted actor remains dark and the background, which does not change between the two images, becomes white, producing a clean matte. Pixels only partially covered by the foreground also divide out to represent proper ratios of foreground to background. We allow the user to specify a "garbage matte" [4] of pixels that should always be set to zero or to one in the event that stray infrared light falls on the actor or that the cloth backing does not fully cover the field of view.

Using the clean matte, the composite image (Fig. 6(f)) is formed by multiplying the matte by the background plate (Fig. 6(e)) and then adding in the color image of the actor; the color image of the actor forms a "pre-multiplied alpha" image as the actor is photographed on a black background (Fig. 6(d)). Since some of the light stage's lights may be in the field of view of the color camera, we do not add in pixels of the actor's image where the matte indicates there is no foreground element. Additionally, we disable the few lights which are nearly obscured by the actor to avoid having the actor's silhouette edge contain saturated pixels from crossing the lights. Improving the matting algorithm to correct for such a problem is an avenue for future work.

5 Results

We performed two kinds of experiments with our light stage. For both, we worked to produce the effect of both the camera and the subject moving within the environment in order to best show the interaction of the subject with the illumination.

5.1 Rotating Camera Composites

For our first set of shots, we visualized our actors in differing omnidirectional illumination environments available from the Light Probe Image Gallery at <http://www.debevec.org/Probes/>. For these shots we wished to look around in the environment as well as see the actor, so we planned the shots to show the camera rotating around the actor looking slightly upward. Since in our system the camera stays fixed, we built a manually operated rotation stage for the actor to stand on, and outfitted it with an encoder to communicate its rotation angles to the lighting control program. Using these angles, the computer rotates the lighting environment to match the rotation of the actor, and records the encoder data so that a sequence of synchronized perspective background plates of the environment can be generated. Since the actor and background rotate together, the rendered effect is of the camera moving around the standing actor.

Figure 12 shows frames from two ten-second rotation sequences of subjects illuminated by and composited into two different environments. As hoped, the lighting appears to be consistent between the actors and their backgrounds. In Fig. 12(b-d), the actor is illuminated by light captured in UC Berkeley's Eucalyptus Grove in the late afternoon. The moving hair in (d) tests the temporal accuracy of the matte. In Fig. 12(f-h), the actor is illuminated by an overcast strip of sky and indirect light from the colonnade of the Uffizi gallery in Florence. In Fig. 12(j-l), the actor is illuminated by light captured in San Francisco's Grace Cathedral. In (j) and (k) the yellow altar, nearly behind the actor, provides yellow rim lighting on the actor's left and right sides. In (j-l), the bluish light from the overhead stained glass windows reflects in the actor's forehead and hair. In all three cases, the mattes follow the actors well, and the illumination appears consistent between the background and foreground elements.

There are also some artifacts in these examples. The (b-d) and (j-l) backgrounds are perspective reprojections of lighting environments originally imaged as reflections in a mirrored ball [17, 6], and as a result are lower in resolution than the image of the actor. Also, since the environments are modeled as distant spheres of illumination, there is no three-dimensional parallax visible in the backgrounds as the camera rotates. Furthermore, the metallic gold stripes of the actor's shirt in (b-d) and (j-l) reflect enough light from the infrared backing at grazing angles to appear as background pixels in some frames of the composite. This shows that while infrared spill does not alter the actor's visible appearance, it can still cause errors in the matte.

5.2 Moving Camera Composite

For our second type of shot, we composited the actor into a 3D model of the colonnade of the Parthenon with a moving camera. The model was rendered with late afternoon illumination from a photographically recorded sky [6] and a distant circular light source for the sun (Figure 13(a-c)). The columns on the left produced a space of alternating light and shadow, and the wall to the right holds red and blue banners providing colored indirect illumination in the environment. The background plates of the environment were rendered using a Monte Carlo global illumination program (the "Arnold" renderer by Marcos Fajardo) in order to provide physically realistic illumination in the environment.

After the background plates, a second sequence was rendered to record the omnidirectional incident illumination conditions upon the actor moving through the scene. We accomplished this by placing an orthographic camera focused onto a virtual mirrored ball moving along the desired path of the actor within the scene (Figure 13(d-f)). These renderings were saved as high dynamic range images to preserve the full range of illumination intensities incident upon the ball. We then converted this 300-frame sequence of mirrored ball images into the longitude-latitude omnidirectional



Figure 12: **Rotation Sequence Results** (a) A captured lighting environment acquired in the UC Berkeley Eucalyptus Grove in the late afternoon. (b-d) Three frames from a sequence of an actor illuminated by the environment in (a) and composited over a perspective projection of the environment. (e) A captured lighting environment taken between two stone buildings on a cloudy day. (f-h) Three composited frames from a sequence of an actor illuminated by (e). (i) A captured lighting environment inside a cathedral. (j-l) Three frames from a sequence of an actor illuminated by the environment in (i) and composited over a perspective projection of the environment. Rim lighting from the yellow altar and reflections from the bluish stained glass windows can be seen in the actor’s face and hair.

images used by our lighting control program.

The virtual camera in the sequence moved backwards at 4 feet per second, and our actor used a treadmill to become familiar with the motions they would make at this pace. We then recorded our actor walking in place in our light stage at the same time that we played back the incident illumination sequence of the corresponding light in the environment, taking care to match the real camera system’s azimuth, inclination, and focal length to those of the virtual camera.

Figure 13(g-i) shows frames from the recorded sequence of the actor, Figure 13(j-l) shows the derived matte images, and Figure 13(m-o) shows the actor composited over the moving background. As hoped, the composite shows that the lighting on the actor dynamically matches the illumination in the scene; the actor steps into light and shadow as she walks by the columns, and the side of her face facing the wall subtly picks up the indirect illumination from the red and blue banners as she passes by.

The most significant challenge in this example was to reproduce the particularly high dynamic range of the illumination in the environment. When the sun was visible, the three lights closest to the sun’s direction were required to produce ten times as much illumination as the other 153 lights combined. This required scaling down the total intensity of the reproduced environment so that the brightest lights would not be driven to saturation, making the other lights quite dim. While this remained a faithful reproduction of the illumination, it strained the sensitivity of the color video camera, requiring 12 decibels of gain and a fully open aperture to register the

image. As a result, the image is mildly noisy and slightly out of focus. Initial experiments using a single repositionable high-intensity light to act as the sun show promise for solving this problem.

6 Discussion and Future Work

The capabilities and limitations of our lighting apparatus suggest a number of avenues for future versions of the system. To reduce the need to compensate for the brightness falloff, using lights with a wider spread or building a larger light stage to place the lights further from the actor would be desirable. To increase the lighting resolution and total light output, the system could benefit from both more lights and brighter lights. We note that in the limit, the light stage becomes an omnidirectional display device, presenting a panoramic image of the environment to the actor.

In the current stage, the actor can not be illuminated in dappled light or partial shadow; the field of view of each light source covers the entire actor. In our colonnade example, we should actually have seen the actor’s nose become shadowed before the rest of her face, but instead the shadowing happens all at once. A future version of the light stage could replace each light source with a video projector to reproduce spatially varying incident fields of illumination. In this limit, this light stage would become an immersive light field [14] display device, immersing the actor in a holographic reproduction of the environment with full horizontal and vertical parallax. Recording the actor with a dense array of cameras would also allow arbitrary virtual viewpoints of the actor to be generated using the



Figure 13: **Moving Sequence Results** (a-c) Three frames of an animated background plate, rendered using the Arnold global illumination system. The camera moves backwards down the collonade of the Parthenon illuminated by late afternoon sun. (d-f) The corresponding light probe images formed by rendering an orthographic view of a reflective sphere placed where the actor will be in the scene. The sphere goes into and out of shadow and reflects the red and blue banners on the right wall of the gallery. (g-i) The corresponding frames of the actor from the live-action shoot, illuminated by the lighting in (d-f), before brightness falloff correction. (j-l) The processed mattes obtained from the infrared compositing system, with the background in black (m-o) The final composite of the actor into the background plates after brightness falloff correction. The actor begins near the red banner, moves into the shadow of a column, and emerges into sunlight near the blue banner. (The blue banner appears directly to the right in the lighting environment (f) but has not yet entered the field of view of the background plate in (c) and (o).) Comparing (m) to (o), it can be seen that the shadowed half of the actor's face receives subtle indirect illumination from the red and blue banners.

light field rendering technique.

To create a general-purpose production tool, it would be desirable to build a large-scale version of the device capable of illuminating whole bodies and multiple actors. Once we can see an entire person, however, it will become necessary to reproduce the shadows that the actor should cast into the environment. Having an approximate 3D model of the actor's performance, perhaps obtained using a silhouette-based volumetric reconstruction algorithm (e.g. [16]) or one or more range imaging cameras (such as the 3DV Systems Z-Cam), could be useful for this purpose. The matting process could be improved by using a single camera to detect both the RGB and IR light; for this purpose a single-chip camera with an R,G,B, and IR filter mosaic over each group of four pixels or a 4-chip camera using a prisms to send light to R, G, B, and IR image sensors could be constructed.

In this work we calibrated the light stage making the common trichromatic approximation [3] to the interaction of incident illumination with reflective surfaces. For illumination and surfaces with complex spectra, such as fluorescent lights and certain varieties of fabric, the material's reflection of reproduced illumination in the light stage could be noticeably different than its actual appearance under the original lighting. This problem could be addressed through multispectral imaging of the incident illumination [11], and by illuminating the actor with additional colors of LEDs. Adding yellow and turquoise LEDs as a beginning would serve to round out our illumination's color gamut.

Finally, we note that cinematographers often – in fact usually – augment and modify environmental illumination with bounce cards, fill lights, absorbers, reflectors, and many other techniques to artistically manipulate the illumination on an actor; skilled use of these techniques can greatly increase the aesthetic and emotional impact of a shot without degrading its realism. Our lighting approach could facilitate such manipulations to the lighting; an avenue for future work is to collaborate with lighting artists to develop useful tools for performing this sort of virtual cinematography.

7 Conclusion

In this paper we have presented a system for realistically compositing an actor's live performance into a background environment by illuminating the actor with a reproduction of the illumination of the light in the environment. Though not fully general, our device has produced results which we believe demonstrate potential use for the technique. We look forward to the continued development of the light stage technique and working with filmmakers to explore how it can evolve into a useful production tool.

Acknowledgements

We thank Brian Emerson and Mark Brownlow for their 3D environment renderings, Maya Martinez for making our infrared cloth backing, Anshul Panday for rendering the light probe backgrounds, Gordie Haakstad, Kelly Richard, and Woody Omens, ASC for cinematography consultations, Randal Kleiser and Alexander Singer for directorial guidance, Yikuong Chen, Rippling Tsou, Diane Suzuki, and Shivani Khanna for their 3D modeling, Diane Suzuki and Mark Brownlow for their illustrations, Elana Livneh and Emily DeGroot for their acting, Marcos Fajardo for the Arnold renderer used for our 3D models, Van Phan for editing our video, the anonymous reviewers for their helpful comments, and Dick Lindheim, Bill Swartout, Kevin Dowling, Andy van Dam, Andy Lesniak, Ken Wiatrak, Bobbie Halliday, Linda Miller, Ayten Durukan, and Color Kinetics, Inc. for their additional help, support, and suggestions for this project. This work has been sponsored by the University of Southern California, U.S. Army contract number DAAD19-99-D-0046, and TOPPAN Printing Co., Ltd. The content of this information does not necessarily reflect the position or policy of the sponsors and no official endorsement should be inferred.

References

- [1] ALEXANDER, D. K., JONES, P. J., AND JENKINS, H. The artificial sky and heliodon facility at Cardiff University. In *Proc. International Building Performance Simulation Association (IBPSA)* (September 1999). (Abstract).
- [2] BEN-EZRA, M. Segmentation with invisible keying signal. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.* (June 2000), pp. 568–575.
- [3] BORGES, C. F. Trichromatic approximation for computer graphics illumination models. In *Computer Graphics (Proceedings of SIGGRAPH 91)* (July 1991), vol. 25, pp. 101–104.
- [4] BRINKMANN, R. *The Art and Science of Digital Compositing*. Morgan Kaufmann, 1999.
- [5] CHUANG, Y.-Y., ZONGKER, D. E., HINDORFF, J., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. Environment matting extensions: Towards higher accuracy and real-time capture. In *Proceedings of SIGGRAPH 2000* (July 2000), pp. 121–130.
- [6] DEBEVEC, P. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH 98* (July 1998).
- [7] DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. Acquiring the reflectance field of a human face. *Proceedings of SIGGRAPH 2000* (July 2000), 145–156.
- [8] DEBEVEC, P. E., AND MALIK, J. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH 97* (August 1997), pp. 369–378.
- [9] FAUGERAS, O. *Three-Dimensional Computer Vision*. MIT Press, 1993.
- [10] FIELDING, R. *The Technique of Special Effects Cinematography*, 4th ed. Hastings House, New York, 1985, ch. 11, pp. 290–321.
- [11] GAT, N. Real-time multi- and hyper-spectral imaging for remote sensing and machine vision: an overview. In *Proc. 1998 ASAE Annual International Mtg.* (Orlando, Florida, July 1998).
- [12] KAJIYA, J. T. The rendering equation. In *Computer Graphics (Proceedings of SIGGRAPH 86)* (Dallas, Texas, August 1986), vol. 20, pp. 143–150.
- [13] KOUDELKA, M., MAGDA, S., BELHUMEUR, P., AND KRIEGMAN, D. Image-based modeling and rendering of surfaces with arbitrary brdfs. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.* (2001), pp. 568–575.
- [14] LEVOY, M., AND HANRAHAN, P. Light field rendering. In *SIGGRAPH 96* (1996), pp. 31–42.
- [15] MALZBENDER, T., GELB, D., AND WOLTERS, H. Polynomial texture maps. *Proceedings of SIGGRAPH 2001* (August 2001), 519–528. ISBN 1-58113-292-1.
- [16] MATUSIK, W., BUEHLER, C., RASKAR, R., GORTLER, S. J., AND MCMILLAN, L. Image-based visual hulls. In *Proc. SIGGRAPH 2000* (July 2000), pp. 369–374.
- [17] MILLER, G. S., AND HOFFMAN, C. R. Illumination and reflection maps: Simulated objects in simulated and real environments. In *SIGGRAPH 84 Course Notes for Advanced Computer Graphics Animation* (July 1984).
- [18] NIMEROFF, J. S., SIMONCELLI, E., AND DORSEY, J. Efficient rendering of naturally illuminated environments. *Fifth Eurographics Workshop on Rendering* (June 1994), 359–373.
- [19] PORTER, T., AND DUFF, T. Compositing digital images. In *Computer Graphics (Proceedings of SIGGRAPH 84)* (Minneapolis, Minnesota, July 1984), vol. 18, pp. 253–259.
- [20] RAMAMOORTHI, R., AND HANRAHAN, P. A signal-processing framework for inverse rendering. In *Proc. SIGGRAPH 2001* (August 2001), pp. 117–128.
- [21] SMITH, A. R., AND BLINN, J. F. Blue screen matting. In *Proceedings of SIGGRAPH 96* (August 1996), pp. 259–268.
- [22] TCHOU, C., AND DEBEVEC, P. HDR Shop. Available at <http://www.debevec.org/HDRShop>, 2001.
- [23] TSAI, R. A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation* 3, 4 (August 1987), 323–344.
- [24] VIDOR, Z. An infrared self-matting process. *Society of Motion Picture and Television Engineers* 69 (June 1960), 425–427.
- [25] WINDOWS AND DAYLIGHTING GROUP. The sky simulator for daylighting studies. Tech. Rep. DA 188 LBL, Lawrence Berkeley Laboratories, January 1985.
- [26] WYSZECKI, G., AND STILES, W. S. *Color Science: Concepts and Methods, Quantitative and Formulae*, 2nd ed. Wiley, New York, 1982.
- [27] ZONGKER, D. E., WERNER, D. M., CURLESS, B., AND SALESIN, D. H. Environment matting and compositing. *Proceedings of SIGGRAPH 99* (August 1999), 205–214.