Lighting Sensitive Display

SHREE K. NAYAR and PETER N. BELHUMEUR
Columbia University
and
TERRY E. BOULT
University of Colorado

Although display devices have been used for decades, they have functioned without taking into account the illumination of their environment. We present the concept of a lighting sensitive display (LSD)—a display that measures the incident illumination and modifies its content accordingly. An ideal LSD would be able to measure the 4D illumination field incident upon it and generate a 4D light field in response to the illumination. However, current sensing and display technologies do not allow for such an ideal implementation. Our initial LSD prototype uses a 2D measurement of the illumination field and produces a 2D image in response to it. In particular, it renders a 3D scene such that it always appears to be lit by the real environment that the display resides in. The current system is designed to perform best when the light sources in the environment are distant from the display, and a single user in a known location views the display.

The displayed scene is represented by compressing a very large set of images (acquired or rendered) of the scene that correspond to different lighting conditions. The compression algorithm is a lossy one that exploits not only image correlations over the illumination dimensions but also coherences over the spatial dimensions of the image. This results in a highly compressed representation of the original image set. This representation enables us to achieve high quality relighting of the scene in real time. Our prototype LSD can render 640×480 images of scenes under complex and varying illuminations at 15 frames per second using a 2 GHz processor. We conclude with a discussion on the limitations of the current implementation and potential areas for future research.

Categories and Subject Descriptors: I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism; I.4.2 [**Image Processing and Computer Vision**]: Compression (Coding); I.4.10 [**Image Processing and Computer Vision**]: Image Representation

General Terms: Algorithms, Design, Measurement

Additional Key Words and Phrases: Color, display algorithms; illumination field; image-based rendering; light field; lighting sensitive display; principal component analysis; reactive display; real-time rendering, relighting; shading, shadowing, singular value decomposition; and texture; three-dimensional displays; virtual reality

1. INTRODUCTION

Displays have become a vital part of our everyday lives. They are used to convey information in a wide range of electronic devices including televisions, computers, PDAs and cellular phones. Recently,

This work was done in the Computer Vision Laboratory in the Computer Science Department at Columbia University and was supported in part by an NSF ITR Grant (IIS-00-85864). These results were initially presented as a Technical Sketch and demonstrated at SIGGRAPH 2001 [Nayar et al. 2001].
Authors' addresses: S. K. Nayar and P.N. Belhumeur, Department of Computer Science, Columbia University, New York, NY 10027; email: {nayar,belhumeur}@cs.columbia.edu; T. E. Boult, Department of Computer Science, University of Colorado at Colorado Springs, Colorado Springs, CO 80933-7150; email: tboult@cs.uccs.edu.

high quality digital displays have also emerged as possible replacements for physical media such as photographs, paintings and sculptures.

Research on display technology has made great strides in improving resolution, brightness, and color characteristics. However, current displays can be viewed as being passive devices—they convey visual information without regard to the environment they reside in. We introduce the notion of a lighting sensitive display (LSD), which can sense the illumination of its environment and accordingly render the appearance of its content. This capability can be used to enhance today's displays in several ways. For instance, the power consumed by the display can be minimized by measuring the lighting of the environment and adjusting the content such that it meets the minimum quality requirements of the user/application. The clarity of content displayed on a device can be maximized by using the measured illumination—the colors, brightnesses and fonts of displayed text may be adjusted such that legibility is always maintained. In the context of digital art, the displayed content—be it 2D (a flat surface), 2D+ (a surface with 3D texture) or an arbitrary 3D scene—can be rendered to appear as if it is illuminated by the display's environment.

The LSD concept has broader implications for the future. Today's displays are typically planar and seek to produce images that appear equally bright from all viewing directions. Advances in material science may make it possible to develop displays that are deformable; they may be used to cover objects with arbitrary shapes. In addition, in the future one may be able to control the 2D light field produced at each point on the display. Finally, it may be possible to embed light sensors within the display that can measure the 2D light field incident at each point on the display. If these challenges are met, the LSD would be able to measure the 4D illumination field incident upon it and generate a 4D light field in response to it. Then, any part of the display can be used to emulate a real surface with a particular bidirectional reflectance distribution function (BRDF) or bidirectional texture function (BTF). Such a device can be used to create physical objects whose material properties are programmable.

In this article, we use currently available sensing and display technologies to develop an LSD that demonstrates the proposed concept, albeit in a limited sense. This system uses a 2D measurement of the illumination field and produces a 2D image in response to it. Our goal is to produce a virtual image of a scene that is consistent with the real environmental illumination. We would also like the scene's appearance to faithfully adapt to lighting changes in the environment. Ideally, the appearance of the displayed scene should also vary with the viewpoint of the observer.[1] However, in our work we focus on the lighting aspect of the problem and assume the viewpoint to be fixed.

We first describe several sensing methods that permit dense 2D sampling of the illumination field. In our prototype LSD, we use a specially designed compact video camera with a hemispherical field of view to obtain local but dense directional measurements of the illumination. Such an approach is adequate in the case of distant light sources as long as each source is visible from all points on the display; the display is not partially in shadow with respect to any of the sources.

Next, we describe a relighting algorithm that can produce realistic appearances of the scene, and yet is efficient enough to quickly respond to lighting changes. Note that today's ray-tracing algorithms are far from real-time (standard video-rate) performance. Therefore, we adopt an image-based approach that uses a large set of images (either rendered or captured) of the scene corresponding to different lighting directions. We present a novel lossy compression algorithm that computes a compact representation of the original image set. In comparison with previous compression algorithms, ours simultaneously exploits correlations over the lighting domain as well as coherences over the spatial domain of the image. This compression is time consuming, but it is done off-line and results in a very efficient representation of

[1]Making a displayed scene appear to be a real one has been suggested by Miller [1995] to be one of the "holy grail" problems of computer graphics.

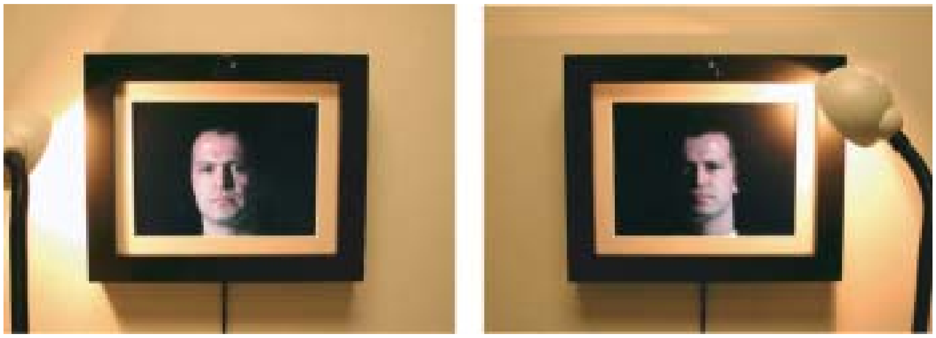

Fig. 1. A lighting sensitive display (LSD) measures its illumination field and modifies the displayed content accordingly. The prototype LSD system shown here uses a hemispherical camera embedded in the frame of the display to approximate the illumination field. It uses an image-based rendering algorithm that can relight the scene under arbitrary illumination at 15 frames a second using a 2 GHz processor. In this example, one can see that the appearance of the face (including shadows and specularities) is consistent with the location of the table lamp.

the original image set. This representation enables real-time relighting of the scene under the display's illumination.

Our prototype LSD system is shown in Figure 1. The current implementation of the relighting algorithm results in a roughly 200:1 compression of the initial image set of the scene. It enables the LSD to output high quality images (including details of shadings, textures, highlights and shadows) at 15 frames a second using just a 2 GHz Pentium processor. We demonstrate the ability of the LSD to display scenes under complex lighting conditions. We conclude with a discussion on the limitations of the existing implementation and areas for future work.

2. MEASURING THE ILLUMINATION FIELD

Today, the only lighting related controls that the user has on a display are global brightness and color adjustments. These can be varied manually, or automatically using one or a few photodetectors (one or a few samples of the 4D incident field). This approach to global brightness/contrast adaptation has been described by several researchers [Heijligers 1962; Thomas 1963; Korda 1965; Biggs 1965; Szermy 1968; Antwerp 1985; Barbier et al. 1991]. Several methods have also been suggested for mapping[2] the photodetector measurements to the display's brightness and contrast [Gibson 1964; Newman 1972; Constable 1978; Fitzgibbon 1982; Prince et al. 1986; Yabuuchi 1990; Ottenstein 1993]. In the context of photorealistic rendering, the above approaches of using one or a few photodetectors do not provide adequate resolution of the environmental illumination.

Using current sensors, it is not possible to measure the complete 4D illumination field. However, depending on the application that the LSD is intended for, a suitable 2D slice through (or mapping of) the 4D field can be measured. For instance, if a more or less flat, matte surface is being displayed, the direction of lighting is less pertinent; the surface can be rendered accurately if the total energy incident at each point of the display is known. In such cases, one may use a 2D array of photodetectors distributed over the surface of the display, as shown in Figure 2(a). Such a covering of detectors can be implemented in many ways; for instance, they could be solid-state detectors that are incorporated into the display device itself. If the 4D field happens to be very smooth, a coarse set of spatial and directional samples can

[2]This mapping problem is a simple form of the "tone mapping" problem in computer graphics, where the mapping from image brightness to displayed brightness can vary over the image to optimize local image features such as contrast (for examples, see Tumblin and Rushmeier [1993]; Ward [1994]; Pattanaik et al. [1998]; Durand and Dorsey [2000]; Reinhard et al. [2002]).

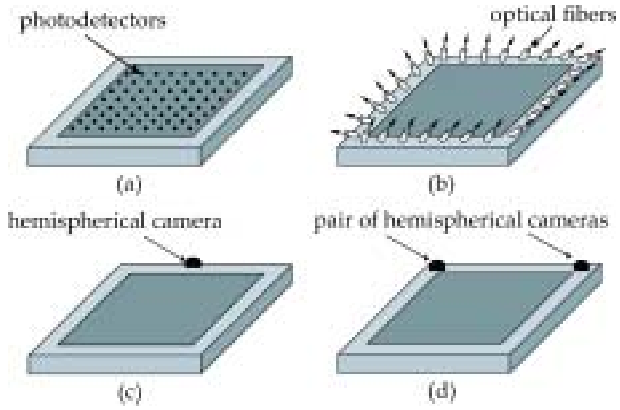

Fig. 2. (a) Dense spatial estimates of the total energy incident upon each point on the display can be obtained using a 2D array of photodetectors distributed over the surface of the display. (b) If the illumination field is very smooth, it can be interpolated using sparse spatial and directional samples obtained using optical fibers distributed over the frame of the display and pointing in different directions. (c) When the sources in the environment are distant compared to the size of the display, the illumination fields at all points on the display have the same directional distribution, which can be measured using a wide-angle camera placed close to the display. (d) If the sources are isotropic over the span of the display, their intensities and locations in 3D can be computed using a stereo pair of wide-angle cameras.

be measured and interpolated to estimate the field. This can be done using fiber optic collectors (each with a narrow cone of sensitivity) distributed around the frame of the display in various orientations, as shown in Figure 2(b).

When the sources of illumination are all distant from the display, and one does not expect partial shadows to be cast upon the display, just directional measurements of the field are sufficient. This is achieved using a wide-angle video camera placed close to the display, as shown in Figure 2(c). Since the sources are distant, the local directional field measured by the camera can be assumed to be the field at all points on the display. If the sources are not distant but are more or less isotropic in their radiant intensities, two wide-angle cameras can be used to estimate the locations of the sources using stereo, as shown in Figure 2(d). This information can be used to determine the contribution of each source to each point on the display. Such a stereo-based method has been used in Sato et al. [1999] to measure the radiance distribution of a scene. In all of the above cases, we obtain a dense 2D (spatial and/or directional) sampling of the illumination field.

The only previous work we have found on sensing display illumination that is somewhat related to ours is Kochanski [1998], where cameras are placed in an environment to measure the locations of dominant light sources as well as the user's head. This information is used to estimate the glare due to specular reflections from the surface of the display in the direction of the user. This glare estimate is then subtracted from the displayed image. The objective of this prior work is different from ours; it is to compensate for reflections that are extraneous to the scene being displayed, while we want the appearance of the displayed scene to be consistent with the lighting of the environment.

3. PROTOTYPE LSD

Our prototype LSD device is shown in Figure 3(a). The display used is a Sony SDM-N50R 15 inch flat panel LCD monitor with a 1024 × 768 native resolution. The controls and the frame of the display are concealed by using a white matting. The display and the matting reside within a black wooden frame. The prototype uses the single camera approach of Figure 2(c) to measure the incident field. Hence it

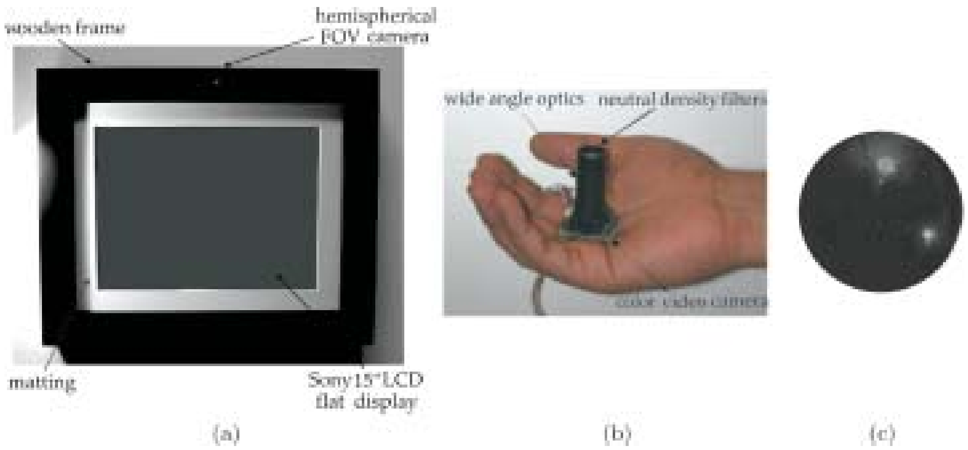

Fig. 3. (a) Prototype lighting sensitive display. (b) Compact camera with a 170 degree field of view that is embedded within the frame of the display in (a). (c) An example image produced by the camera for a scene with two dominant sources.

is only suitable for environments where the sources are not too close to the display and shadows cast on the display do not partially cover the display. The camera is embedded within the wooden frame to conceal it from the observer.

Since we are interested in all sources that appear in front of the display, the camera must have a hemispherical field of view. Current fish-eye lenses are not compact enough for our application. Therefore, we have designed a new imaging lens. It includes a inexpensive peephole lens (used on doors) with a 170 degree field of view and a relay lens to interface the peephole lens with a Computar CM500 1/3 inch color board camera. The complete camera system is shown in Figure 3(b) and an image produced by it for an environment with two dominant sources is shown in Figure 3(c). Since the camera has an 8-bit detector, it cannot measure bright primary sources (ceiling lights, table lamps, monitors, etc.) and secondary sources (walls, objects, etc.) at the same time. This problem can be remedied by using a high dynamic range camera or by using one of the various ways in which high dynamic range images can be computed from low dynamic range ones (see Nayar and Branzoi [2003] for a survey).

4. IMAGE RELIGHTING FOR LSD

To make the LSD visually compelling, we strive to render images that are as close to photorealistic as possible. Furthermore, we impose the requirement that the method works for both real and synthetic scenes. The approach we choose is to prerender for synthetic scenes [Dorsey et al. 1991], or preacquire for real scenes [Hallinan 1994; Epstein et al. 1995; Georghiades et al. 1998; Debevec et al. 2000; Koudelka et al. 2001; Malzbender et al. 2001; Matusik et al. 2002], a collection of images in which the lighting direction is systematically varied. If the sampling of the lighting directions is dense enough, then due to the linearity of scene radiance, images of the scene under complex illumination can be computed simply by superposition of the single light source images, as done in Dorsey et al. [1991].

However, the collection of images needed for relighting is typically too large both to store in memory and to sum in real-time; for example, Debevec et al. [2000] use 2000 images and Koudelka et al. [2001] use more than 4000. (If too few images are used, then the specularities and shadow boundaries will appear to jump during real-time rendering of the scene illuminated by a moving light source.) Thus, with current hardware limitations, it is necessary to compress the data into a form that is small enough to be stored in memory and efficient enough to be summed over all sampled lighting directions. In Debevec et al. [2000], images of each pixel's reflectance function are stored in JPEG format and processed in the compressed domain using the techniques of Smith and Rowe [1996]. However, this only reduces the storage and computation by a factor of 20 or so—not enough to relight the collection of images at near real-time frame rates. Furthermore, if this technique is overused (the individual images are over-compressed) the quality of the renderings will be compromised and JPEG artifacts will be introduced. Similar issues arise in the method described in Lin et al. [2002], where images of a pixel's radiance values are compressed using a 2D DCT.

4.1 Global Dimensionality Reduction

Hallinan [1994] showed that the variation in images of faces due to changes in lighting direction could be approximated by a low-dimensional linear subspace. To do this, Hallinan applied Principal Component Analysis (PCA) Duda and Hart [1973], to find linear bases that best approximated the collection of images of faces (i.e., the bases that minimized the sum of squared differences between the images and the subspace spanned by the bases). Around the same time, Nimeroff et al. [1994] showed that linear bases could also be used for the relighting of scenes under complex, but diffuse illumination. This approach for representing variations in images due to lighting has been analyzed and applied with many variations in the subsequent ten years (see Nayar and Murase [1994]; Epstein et al. [1995]; Belhumeur and Kriegman [1996]; Teo et al. [1997]; Shashua [1997]; Ramamoorthi and Hanrahan [2001]; Basri and Jacobs [2001]). Some of the most recent examples of relighting using linear bases can be found in Malzbender et al. [2001] where polynomial bases are used for 3D textures, Sloan et al. [2002] where spherical harmonics are used for low-frequency lighting of complex scenes, and Matusik et al. [2002] where linear bases are computed for small image blocks and used to relight the scene.

Unlike in Hallinan [1994]; Epstein et al. [1995]; Teo et al. [1997]; Sloan et al. [2002] did not compute the set of bases from the images, instead they used between 9–25 spherical harmonics over the lighting space.[3] The use of spherical harmonics has the advantage that the bases do not need to be precomputed, but has the disadvantage that the choice of bases is sub-optimal—in general these are not the bases that minimize the sum of squared differences between the images and the subspace spanned by the bases.

While the use of linear bases has enabled big strides toward real-time rerendering under complex illumination, it still has been limited to the realm of diffuse, or low-frequency, lighting. In order to capture effects such as sharp specularities and crisp shadow boundaries one requires far more bases than the 9–25 used in Sloan et al. [2002]. This is best illustrated by an example. Figure 6(a) shows one sample image from a collection of 4096 images of Michaelangelo's David, each rendered with a different lighting direction on the frontal hemisphere using the Dali rendering engine provided by Henrik Wann Jensen [Jensen 2001] (on 3D data provided by Marc Levoy [Levoy et al. 2000]). Each of these images is of size 640×480, has 3 color channels and the pixel brightness in each channel is stored using 16 bits. The complete image sequence therefore requires 7.03 Gb for storage. Figure 6(b) shows the closest rendering to the image in Figure 6(a), obtained by using 10 linear bases computed using SVD from the collection of 4096 original images. In this case, the memory required to store the 10 bases is 35.2 Mb and the coefficients corresponding to the 4096 images is 0.469 Mb. However, we see that the

[3] A related approach is taken in Avidan [2002], where linear bases are used to represent an arbitrary ensemble of images. In this case, the brightness profiles of individual pixels along the ensemble are first clustered to find segments (groups of pixels) in the image that have similar profiles. Then, the set of profiles corresponding to each segment is represented using a set of linear bases.

rendered image does not correctly approximate many of the cast shadows; shadows in the neck, near the eye socket, and on the forehead are completely missing.

Regardless of the type of linear bases used—be they spherical harmonics [Sloan et al. 2002] or bases derived from the data—greater accuracy in the renderings requires many more bases. However, as the number of bases increases so do the storage requirements and rendering time.

4.2 Local Dimensionality Reduction

What seems unarguable is that complex illumination effects, such as cast shadows and specularities, cannot be represented by a handful of linear basis functions—at least when the whole image is considered. However, if one looks locally at these illumination effects—at small regions within the image—then the variation within these regions due to changes in illumination is much better behaved. Consider a scene with matte reflectance consisting of a sphere and a cylinder resting on a table top. Consider also a collection of images of this scene in which the lighting direction is systematically varied. Due to shadowing, many more than 10 linear bases would be required to represent the global variation in the images due to changes in lighting. However, if one looks at a small region of the images, say, containing only a portion of the table top, then the variation within this region might be well represented by as few as one basis function. Likewise, a small region of the cylinder might be well approximated by two and the sphere by three (see Shashua [1997] for details). For scenes with more complicated reflectances and/or shadowing effects more bases are needed in each region; however many fewer than if the whole image is considered.

We now describe a method for compressing a collection of images of a scene over varying lighting that exploits the above observation. The method simply divides the image up into square blocks and computes the bases separately for each block. This approach is similar to the one used by Matusik et al. [2002]. It is also related to the patchwise approach used in Nishino et al. [1999] and Wood et al. [2000] for representing variation over viewpoint. This part of our algorithm therefore builds upon previous work. We will describe it in detail as it lays the foundation for the more novel aspect of our algorithm presented in the next section.

Consider the collection of n images of a scene over varying lighting. Each image I_i is an image of the scene illuminated by a single distant point source. Divide the image up into m square blocks each containing p pixels. Let I_i^j denote the jth block in the ith image. For each block in the scene compute a low-dimensional approximation as follows. Create a $p \times n$ matrix I^j as a collection of image blocks in which the ith column of I^j is formed by p pixels from the jth block. Using Singular Value Decomposition (SVD), we find a rank b approximation to I^j as

$$I^j \cong E^j S^j C^{j\,T} \tag{1}$$

where E^j is a $p \times b$ column-orthogonal matrix which we call the block bases, S^j is a $b \times b$ diagonal matrix and C^j is an $n \times b$ column-orthogonal matrix. If we absorb the singular values from S^j into $C^{j\,T}$ we can rewrite Equation 1 as

$$I^j \cong E^j L^j \tag{2}$$

where L^j is a $b \times n$ matrix, which we call the block lighting coefficients.

Let the lighting coefficient matrix L be formed by stacking all of the m L^j block matrices. Likewise, let the collection of image bases for all blocks be denoted by E. The extraction of the image bases E and the lighting coefficient matrix L is shown diagrammatically in Figure 4. The collection of submatrices within E and L contain all the information needed to approximate the collection of images corresponding to the n lighting directions.

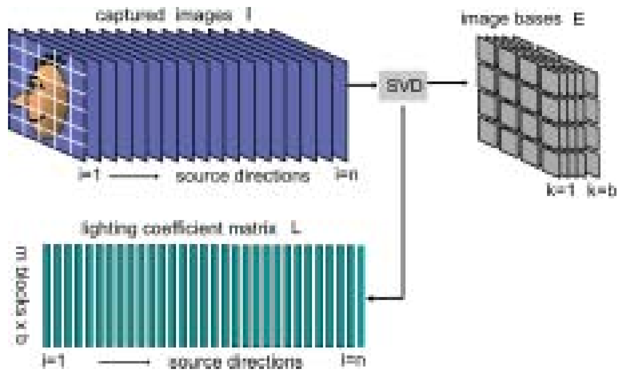

Fig. 4. Images of a scene I_i under different lighting directions are divided into m square blocks. Collections of blocks I^j across all lighting directions are factored and compressed using Singular Value Decomposition (SVD) into block bases E^j and block lighting coefficients L^j. All of block specific E^j and L^j are stored within the matrices E and L, respectively. This computation is done off-line as a pre-processing step.

Note that before the blockwise SVD procedure was applied we had $n \times m \times p$ measurements. After the SVD, we have $b \times m \times p$ elements in E. Thus, for the matrix E the compression ratio is b/n; in general this is a significant saving since, for blockwise SVD, we expect $b \ll n$ (i.e., the number of bases is much less than the number of images) with little compromise in the fidelity to the original data. Figure 6(c) shows the closest approximation to the original image in Figure 6(a) obtained using linear bases computed separately for each block. In this case, the number of images $n = 4096$, the number of blocks $m = 1200$, the number of pixels per block $p = 256$ and the number of bases per block $b = 10$. Note that the rendering in Figure 6(c) correctly approximates the cast shadows and subtle highlights. There is almost no perceptible difference between this image and the original one. However, by using blocks, we have introduced the additional burden of keeping track of L, which contains $n \times m \times b$ elements. In general, we expect that $n > p$, so the cost of storing L dominates. This problem is unique to the blockwise decomposition as the size of L grows linearly with the number of blocks m. In the above example, 300 Mb was needed to store the bases and coefficients.

4.3 Exploiting Spatial Coherence

One can get around the extra burden of keeping track of the lighting coefficient matrix L by exploiting the fact that there is much coherence in the image blocks. We expect that lighting coefficients for each of the blocks are not linearly independent. Consider a scene containing only a Lambertian sphere. Each of m blocks can be well approximated by rank 3 matrices: $b = 3$ for E^j and L^j. In addition, the span of the rows (lighting coefficients) of L^i should be nearly equivalent to the span of the rows of L^j, for all $i, j \le m$. In other words, the $(m \times b) \times n$ matrix L is rank deficient. We can take advantage of this by applying a second stage SVD.

Similar to Equation 2, we apply SVD to the lighting coefficient matrix L and find a rank q approximation to L as

$$L \cong UV \tag{3}$$

where U is an $(m \times b) \times q$ column-orthogonal matrix that we call the lighting coefficient bases, V is a $q \times n$ matrix that we call the compressed coefficient matrix and q denotes the number of linear bases kept to approximate L. Note that we keep the linear bases that best approximate L: the singular vectors corresponding to the largest singular values of L. Also note that the q largest singular values

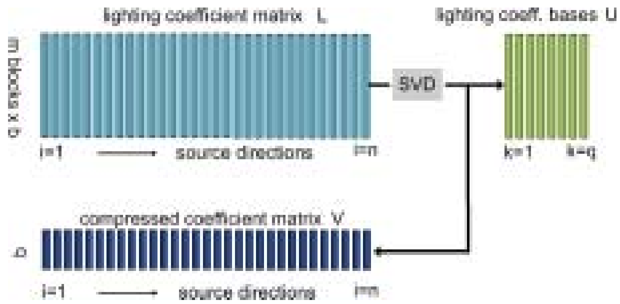

Fig. 5. Due to spatial coherence between blocks, there is much redundancy in the lighting coefficient matrix L. This redundancy can be exploited in a second stage SVD resulting in the compression of L. The $(m \times b) \times n$ matrix L is factored and compressed into a $(m \times b) \times q$ matrix of lighting coefficient bases U and a $q \times n$ matrix of compressed coefficients V. For $n = 4096$, $(m \times b) = 12000$ and $q = 200$, this computation takes 2 hours on a 2 GHz Pentium IV and is done as a pre-processing step.

of L have been folded into V. The factorization and compression of L is shown diagrammatically in Figure 5.

Due to coherence between blocks we can choose q such that $q \ll (m \times b)$ with little compromise in the quality of the approximation. Furthermore, if $q \ll (m \times b)$ we have significantly reduced the cost of keeping track of L. In Figure 6(d), we have shown the closest approximation to the original image in Figure 6(a) using the blockwise bases representation described earlier followed by the compression of the lighting coefficient matrix L described here. For this image, the number of bases to approximate L is $q = 200$. In this case, the total storage required (for the image bases, the lighting coefficient bases, and the lighting coefficient matrix) is 30.25 Mb. Note that there is almost no perceptible difference between the rendered image and the original one even though significant compression of the data is achieved.

Before the two-stage SVD procedure was applied we had $n \times m \times p$ measurements in the collection of images I. After the two-stage SVD, we have $b \times m \times p$ elements in E, $b \times m \times q$ elements in U, and $q \times n$ elements in V. Therefore, the total compression ratio of our representation can be computed as

$$\frac{n\,m\,p}{b\,m\,p + b\,m\,q + q\,n}. \tag{4}$$

If one considers that the elements of I are stored as integers and the elements of E, U, and V should be stored as floating point numbers, then the compression ratio is reduced by a factor of 2–4.

4.4 Real-Time Rendering

All of the data needed to render images of the scene are stored in the matrices E, U and V. Not only do these matrices require significantly less storage than the original set of images I, they also allow for real-time relighting of the scene. Here we detail the steps needed to render a novel image.

The sensor on the LSD measures the environmental illumination. In the case of distant sources, we can represent this illumination as an image, called the illumination field image, in which each pixel corresponds to one of the n lighting directions in the original collection of images I. Let this illumination field image be vectorized and denoted by the $n \times 1$ vector s. To render an image of the scene as if it were illuminated by s, we first compute a compressed coefficient vector as the product Vs. Next, we compute a lighting coefficient vector as the product UVs. Note that the vector UVs has dimension $(m \times b) \times 1$ and that the subvector l^j given by the rows $b \times (j-1)+1$ through $b \times j$ of UVs contain the coefficients necessary for rendering the jth block. Thus, the subvectors can be unstacked and used to render the

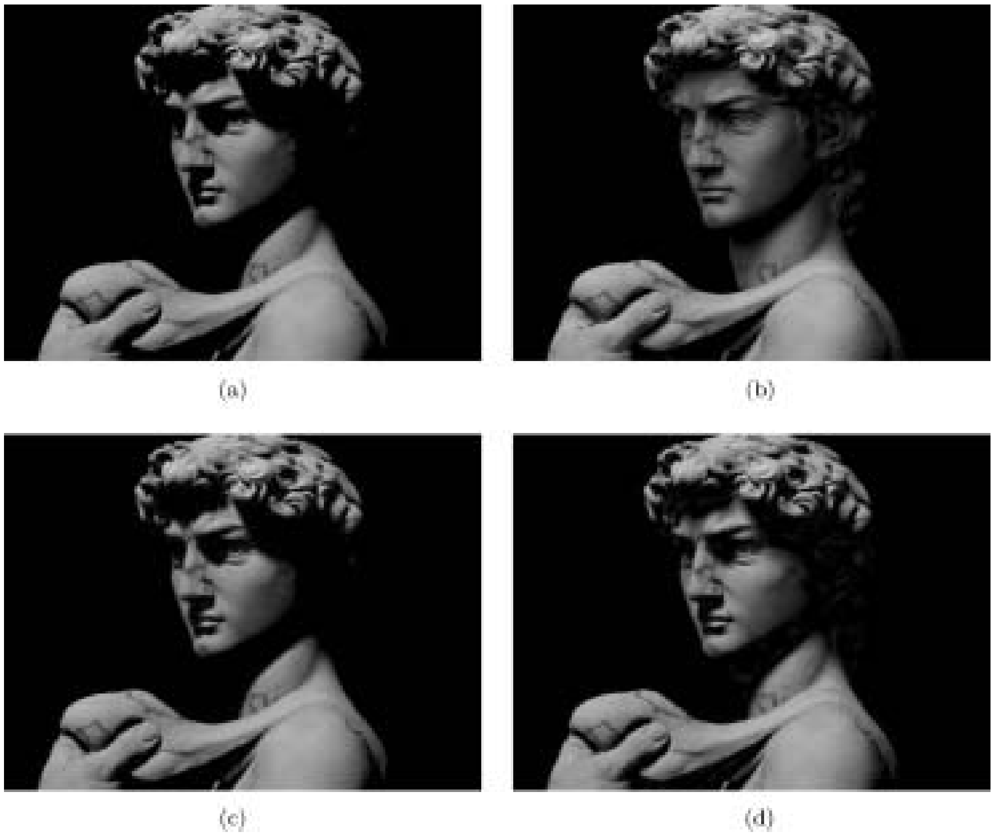

Fig. 6. (a) A sample image from a collection of 4096 rendered (original) images of Michaelangelo's David (courtesy of Marc Levoy, Stanford University). Each of these images was rendered with a different lighting direction on the frontal hemisphere using the Dali rendering engine provided by Henrik Wann Jensen. (b) The closest approximation to the original image shown in Figure 6(a) obtained using 10 linear image bases computed using SVD from the 4096 original images. (c) The closest approximation to the original image obtained using a blockwise representation in which the 10 linear bases are computed separately for each block. (d) The closest approximation to the original image obtained using a blockwise bases representation along with a compression of the lighting coefficient matrix. The memory required to store the representation in each of these cases is (a) 7.03 Gb, (b) 35.669 Mb, (c) 300 Mb, and (d) 30.25 Mb (see text for details).

jth block of the displayed image as the product $E^j l^j$. This process is repeated for all m blocks. This complete rendering process is illustrated in Figure 7.

4.5 Rendering Color Images

In the case of color images, we treat each color channel as a separate image and the parameter n describing the number of images increases to $3n$ for both stages of the SVD. However, as there is great redundancy in the color channels from image to image, we do not increase the size of the first and second stage bases as specified by b and q, respectively. Thus, the matrices E and U remain the same size, while the matrix V triples in size to $q \times 3n$. This matrix is then broken into three submatrices V_r, V_g and V_g representing the compressed coefficient vectors for each of the color channels separately. For color images, the illumination sensor of the LSD measures one illumination field vector for each color

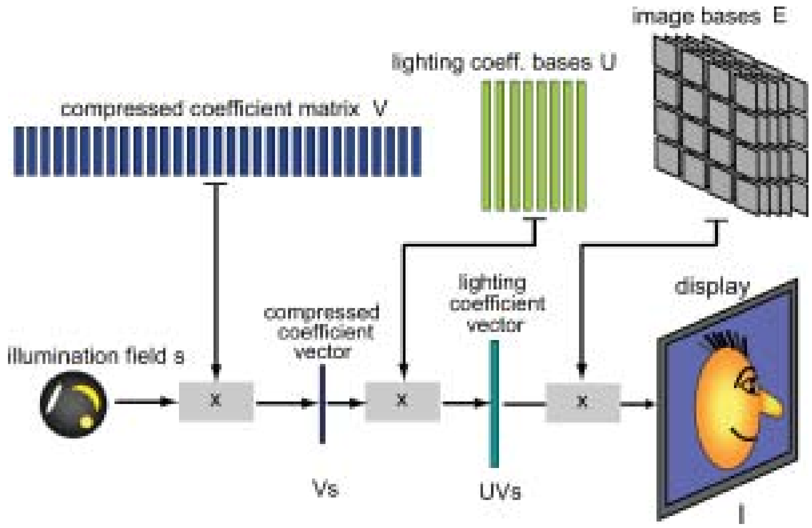

Fig. 7. To render an image of the scene as if it were illuminated by a field specified by a vector s we appeal to the previously calculated matrices E, U and V. First, we compute a compressed coefficient vector as the product Vs. Next, we compute a lighting coefficient vector as the product UVs. Note that the vector UVs has dimension $(m \times b) \times 1$ and that the subvector l^j given by the rows $b \times (j-1)+1$ through $b \times j$ of UVs contain the coefficients necessary for rendering the jth block. Thus, the subvectors can be unstacked and used to render the jth block of the displayed image as the product $E^j l^j$. This process is repeated for all m blocks and an image is displayed on completion.

channel: s_r, s_g and s_b. To relight the image, we proceed as above, computing the image for each color channel separately.

The compression ratio improves in the case of color images (due to the redundancy between channels) and can be written as

$$\frac{3nmp}{bmp + bmq + 3qn}. \tag{5}$$

In the case of David and the still life scene shown in the next section, the values of the parameters are as follows: $n = 4096$, $m = 1,200$, $p = 256$, $b = 10$ and $q = 200$. This yields a compression ratio of 476:1. In practice, the ratio is lower as the bases need to be represented with higher precision than the input images. If the input images have 16 bits per color channel (high dynamic range) and the bases are stored as floating point numbers (4 bytes each) then the compression ratio drops by a factor of 2 to 238:1. If the input images have 8 bits per color channel, then one can maintain the same compression ratio by storing the bases as 2-byte fixed point numbers. If one were concerned about storage of the computed representation (the matrices E, U and V), further gains can be achieved by applying one of the various available compression schemes to the matrices themselves.

5. IMPLEMENTATION

The LSD relighting algorithm was implemented on a Dell Precision 340 PC with a 2 GHz processor and 512 Mb of RAM. The illumination images from the camera embedded in the LSD (see Figure 3) are acquired using a Matrox Meteor capture card.

In the previous section, we have shown why the representation we use is compact to store as well as efficient to render new images with. To ensure that the LSD can respond to illumination changes at

15 frames per second (fps), however, we have used some additional optimizations. Note that typically only a fraction of the captured illumination field image will have active sources. Therefore, only a fraction of the columns of the compressed coefficient matrix V (see Figure 7) are needed to compute the compressed coefficient vector. Furthermore, when the illumination varies, an even smaller fraction of the illumination image changes. Therefore, the computation of a new compressed coefficient vector can be done using the previously computed vector and an even smaller fraction of the columns of the matrix V. We found that these savings are sufficient to almost always relight the scene at 15 fps. When there is a sudden and substantial change in the illumination, these savings drop. In such cases, we maintain the relighting rate of 15 fps by reducing the number of image bases in E that are used in the final stage of the rendering (see Figure 7).

Figure 8 shows results for the model of David in Figure 6. As mentioned earlier, the 4096 original (rendered) images were compressed with a 238:1 ratio. The output of the LSD is always consistent with the lighting condition/direction. In particular, the details of the shadows and shading are noteworthy.

In the case of complex scenes, pre-rendering a large set of high quality images can prove impractical. In such cases, it is more convenient to use real images of the scene captured under different lighting directions. Rather than using a complex setup with a large number of fixed sources, we used an interactive method (called “light sketch”), which is illustrated in Figure 9. A camera observes the scene from the desired viewpoint.[4] An additional “source” camera is placed close to the scene (see Figure 9(a)). When a user waves a light source around the scene, the source directions are computed in real-time from the known projection model of the source camera and are displayed on a monitor (see Figure 9(b)). The circle represents the field of view of the source camera and the dots are the discrete source directions for which images are desired. The solid curve is the detected path of the source and the red dots are the directions for which images have been captured. This simple interface allows one to quickly scan the desired set of lighting directions.

A total of 4096 color images (640×480 in size) of the still life scene in Figure 10 were captured using the light sketch method. This image set was also compressed with a 238:1 ratio. In Figure 10, we show LSD outputs for three lighting conditions. Although a single lamp is used, the lamp acts like an area source as it is close to the LSD. The detailed views (right column) show the high quality of the rendered images. The shadows cast by the pepper dispenser, the fish and the knife are sharp and consistent with the lighting. As seen in Figure 9(a), the original scene was placed within a box. This causes the large shadow covering almost half the red velvet background in the first example (source on the left). Note how the highlights on the pepper dispenser, the knife, the orange and the bottle of vinegar vary with the lighting. The texture due to the scales of the fish and the translucent appearance of the grapes are visible. In the second image one can also see the red light cast on the grape basket due to light passing through the bottle of vinegar.

In Figure 11, an example of complex lighting is shown where there are two strong lamps close to the LSD as well as weak ambient lighting. As expected, the pepper dispenser casts two strong shadows and has two distinct highlights on it. We see a slight yellow/orange tinge in the left bottom corner of the image. This is not a rendering artifact but rather due to glare from the display's surface caused by the left lamp.

6. DISCUSSION

We emphasize that the LSD prototype we have presented here is no more than an initial proof-of-concept. Several interesting problems came to the forefront during our implementation. First, current image sensors do not have the dynamic range and spectral resolution needed to measure the wide

[4]Multiple cameras may be used if multiple viewpoints need to be captured.

Fig. 8. LSD outputs for the David model corresponding to different illumination conditions. The complex shadows in the hair, eyes and neck regions are reproduced with the desired sharpness.

range of colors and brightnesses encountered in the real world. As a result, the LSD's camera had to be tuned to respond to only bright sources with broad spectra. Fortunately, camera technology is rapidly improving with respect to dynamic range and spectral resolution (see Nayar and Branzoi [2003] for a recent survey), and we expect this to be less of an issue in the near future. Another important factor is the dynamic range of the display itself. If our ultimately goal is to emulate any surface encountered in practice, the display must be able to generate a very wide range of brightnesses and colors. Fortunately, significant strides are being made in this direction as well (see Sunnybrook Technologies [2003] for example).

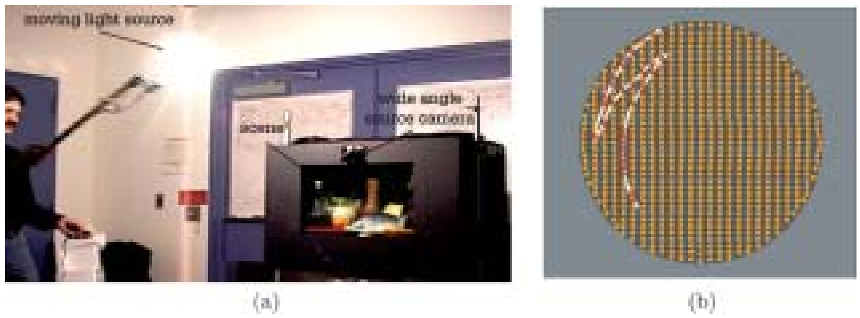

Fig. 9. (a) The "light sketch" method for capturing images of a scene under a large number of known lighting directions. A wide-angle camera placed close to, or within, the scene measures the direction of a hand-held source in real time. (a) The trajectory of the source (white curve), the sampled source directions (red dots) and the remaining desired source directions (yellow dots) are displayed to the user so he/she can quickly cover the space of source directions.

Fig. 10. LSD outputs for a still life scene with different illuminations. Note the shadows cast by the pepper dispenser, the fish and the knife; the specularities on the dispenser, knife, bottle, orange and grapes; and the textures on the fish, the grape basket and the red velvet cloth.

Fig. 11. LSD output for illumination that includes two lamps and some ambient light. As expected, the pepper dispenser casts two strong shadows and has two distinct highlights on it.

Another problem we have not addressed is the glare caused by the surface of the display. Our rendering algorithm only seeks to make the image consistent with the lighting of the display and does not account for the appearance of the display itself. One could use the method in Kochanski [1998] to address this problem. However, compensating for glare within the displayed image makes sense only when there is a single user whose location is known. Perhaps, the only effective way to address the problem of glare is to design the display such that it produces minimal reflections.

We have used a single camera at a fixed location to measure the illumination field. This approach does not work when the incident field varies spatially over the display. For such cases, a somewhat better approximation may be obtained by using a stereo camera system [Sato et al. 1999]. However, a better approach would be to use a large number of cameras distributed around the display. This would provide a dense sampling of the field over the display's periphery, which can be interpolated to obtain better estimates of the field on the display.

Finally, our rendering algorithm may be improved in several ways. Our primary objective here was to take advantage of inherent redundancies in the images of a scene taken under different lighting conditions. However, we have not fully explored how the image blocks are best chosen and whether there are other (possibly non-linear) bases that can capture the lighting variations in a more efficient manner. All this said, the rendering algorithm we proposed does perform well and may be used in its current form for other relighting applications.

7. CONCLUSIONS

We have introduced the concept of a lighting sensitive display that constantly monitors the illumination of its environment and modifies its content accordingly. We presented an initial implementation of this concept. As mentioned earlier, an ideal LSD would be one that can sense the complete 4D illumination field and produce a fully controllable 4D light field in response. Such a device would make it possible to create surfaces whose material properties can be programmed to mimic a wide variety of real-world surfaces. It is hard to predict if such an implementation will indeed be possible in the future. However, if an ideal LSD can be developed, it will facilitate new and powerful ways of merging real and virtual worlds.

ACKNOWLEDGMENTS

The authors thank Melissa Koudelka and Rahul Swaminathan for their help in collecting the image sets used in this work.

REFERENCES

ANTWERP, J. C. V. 1985. Automatic Brightness Control Apparatus. *United States Patent* 4,514,727 (April).

AVIDAN, S. 2002. Eigensegments: A spatio-temporal decomposition of an ensemble of images. In *European Conference on Computer Vision (ECCV)*. Copenhagen, Denmark, III: 747 ff.

BARBIER, B., EDIAR, S., AND BRUN, J. 1991. System for the Display of Luminous Data with Improved Readability. *United States Patent* 5,057,744 (October).

BASRI, R. AND JACOBS, D. 2001. Lambertian reflectance and linear subspaces. In *International Conference on Computer Vision*. 383–390.

BELHUMEUR, P. N. AND KRIEGMAN, D. J. 1996. What is the set of images of an object under all possible lighting conditions? In *IEEE Conference on Computer Vision and Pattern Recognition*. 270–277.

BIGGS, A. J. 1965. Compensator for Color-Television Receivers for Chromacity Variations in Ambient Light. *United States Patent* 3,200,193 (August).

CONSTABLE, D. W. 1978. Ambient Light Contrast and Color Control Circuit. *United States Patent* 4,090,216 (May).

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *SIGGRAPH 00*. 145–156.

DORSEY, J., SILLION, F. X., AND GREENBERG, D. P. 1991. Design and simulation of opera lighting and projection effects. In *SIGGRAPH 91*. 41–50.

DUDA, R. AND HART, P. 1973. *Pattern Classification and Scene Analysis*. Wiley, New York.

DURAND, F. AND DORSEY, J. 2000. Interactive tone mapping. In *Rendering Techniques 2000: 11th Eurographics Workshop on Rendering*. 219–230.

EPSTEIN, R., HALLINAN, P., AND YUILLE, A. 1995. 5 plus or minus 2 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *IEEE Workshop on Physics-Based Modeling in Computer Vision*. 108–116.

FITZGIBBON, J. J. 1982. Dimmer and Dimmer Override Control for a Display Device. *United States Patent* 4,355,334 (October).

GEORGHIADES, A., KRIEGMAN, D., AND BELHUMEUR, P. 1998. Illumination cones for recognition under variable lighting: Faces. In *IEEE Conference on Computer Vision and Pattern Recognition*. 52–59.

GIBSON, V. R. 1964. Automatic Brightness-Contrast Control Using Photoresistive Element to Control Brightness and AGC Voltages in Response to Ambient Light. *United States Patent* 3,147,341 (September).

HALLINAN, P. 1994. A low-dimensional representation of human faces for arbitrary lighting conditions. In *IEEE Conference on Computer Vision and Pattern Recognition*. 995–999.

HEIJLIGERS, H. 1962. Circuit Arrangement for Automatically Adjusting the Brightness and the Contrast in a Television Receiver. *United States Patent* 3,027,421 (March).

JENSEN, H. W. 2001. *Realistic Image Synthesis Using Photon Mapping*. A. K. Peters, Natick, Massachusetts.

KOCHANSKI, G. P. 1998. Retro-Focus Type Lens and Projection-Type Display Apparatus. *United States Patent* 5,854,661 (December).

KORDA, P. 1965. Automatic Contrast and Brilliance Control System for Television Receivers. *United States Patent* 3,165,582 (January).

KOUDELKA, M., BELHUMEUR, P., MAGDA, S., AND KRIEGMAN, D. 2001. Image-based modeling and rendering of surfaces with arbitrary BRDFs. In *IEEE Conference on Computer Vision and Pattern Recognition*. 568–575.

LEVOY, M., PULLI, K., CURLESS, B., RUSINKIEWICZ, S., KOLLER, D., PEREIRA, L., GINZTON, M., ANDERSON, S., DAVIS, J., GINSBERG, J., SHADE, J., AND FULK, D. 2000. The Digital Michelangelo Project: 3D scanning of large statues. In *SIGGRAPH 00*. 131–144.

LIN, Z., WONG, T., AND SHUM, H. 2002. Relighting with the reflected irradiance field: Representation, sampling and reconstruction. *Int. J. Comput. Vision 49*, 2-3 (September), 229–246.

MALZBENDER, T., GELB, D., AND WOLTERS, H. 2001. Polynomial texture maps. In *SIGGRAPH 01*. 519–528.

MATUSIK, W., PFISTER, H., NGAN, A., BEARDSLEY, P., ZIEGLER, R., AND MCMILLAN, L. 2002. Image-based 3d photography using opacity hulls. In *SIGGRAPH 02*. 427–437.

MILLER, G. 1995. Volumetric hyper-reality, a computer graphics holy grail for the 21st century? In *Graphics Interface '95*. Quebec, Canada, 56–64.

NAYAR, S. AND MURASE, H. 1994. The dimensionality of illumination manifolds in eigenspace. Tech. Rep. CUCS021-94, Columbia University, New York. August.

NAYAR, S. K., BELHUMEUR, P. N., AND BOULT, T. E. 2001. Lighting sensitive display. In *Technical Sketch, SIGGRAPH 01*. Los Angeles.

NAYAR, S. K. AND BRANZOI, V. 2003. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *International Conference on Computer Vision*. 1168–1175.

NEWMAN, A. J. 1972. Method and Means for Providing a Lighted Display Having a Constant Contrast Ratio. *United States Patent* 3,649,755 (March).

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

NISHINO, K., SATO, Y., AND IKEUCHI, K. 1999. Eigen-texture method: Appearance compression based on 3D model. In *IEEE Conference on Computer Vision and Pattern Recognition*. 618–624.

OTTENSTEIN, J. 1993. Arrangement for Automatically Controlling Brightness of Cockpit Displays. *United States Patent* 5,270,818 (December).

PATTANAIK, S. N., FERWERDA, J. A., FAIRCHILD, M. D., AND GREENBERG, D. P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of SIGGRAPH 98*. Computer Graphics Proceedings, Annual Conference Series. 287–298.

PRINCE, J. S., HERZ, H. L., AND COLEMAN, E. P. 1986. Brightness Control System for CRT Video Display. *United States Patent* 4,589,022 (May).

RAMAMOORTHI, R. AND HANRAHAN, P. 2001. A signal-processing framework for inverse rendering. In *SIGGRAPH 01*. 117–128.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Trans. Graph. 21,* 3 (July), 267–276.

SATO, I., SATO, Y., AND IKEUCHI, K. 1999. Acquiring a radiance distribution to superimpose virtual objects onto a real scene. *IEEE Trans. Visual. Comp. Graph. 5,* 1, 1–12.

SHASHUA, A. 1997. On photometric issues in 3d visual recognition from a single 2d image. *Int. J. Comput. Vision 21,* 1-2, 99–122.

SLOAN, P., KAUTZ, J., AND SNYDER, J. 2002. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *SIGGRAPH 02*. 527–536.

SMITH, B. AND ROWE, L. 1996. Compressed domain processing of JPEG encoded images. *Real-Time Imaging 2 2,* 2, 3–17.

SUNNYBROOK TECHNOLOGIES. 2003. High dynamic range displays. http://www.sunnybrook.com.

SZERMY, N. 1968. Automatic Contrast and Brightness Control for Television Receiver Utilizing a Light Dependent Resistor. *United States Patent* 3,404,226 (October).

TEO, P., SIMONCELLI, E., AND HEEGER, D. 1997. Efficient linear rerendering for interactive lighting design. Tech. Rep. STAN-CS-TN-97-60, Stanford.

THOMAS, L. P. 1963. Television Receiver Circuits. *United States Patent* 3,096,399 (July).

TUMBLIN, J. AND RUSHMEIER, H. E. 1993. Tone reproduction for realistic images. *IEEE Computer Graphics & Applications 13,* 6 (Nov.), 42–48.

WARD, G. 1994. A contrast-based scalefactor for luminance display. In *Graphics Gems IV*. 415–421.

WOOD, D., AZUMA, D., ALDINGER, K., CURLESS, B., DUCHAMP, T., SALESIN, D., AND STUETZLE, W. 2000. Surface light fields for 3D photography. In *SIGGRAPH 00*. 287–296.

YABUUCHI, S. 1990. Display System with Luminance Calculation. *United States Patent* 4,952,917 (August).

Received April 2003; revised January 2004; accepted May 2004