

Spatially Varying Image Based Lighting by Light Probe Sequences

Capture, Processing and Rendering

Jonas Unger¹, Stefan Gustavson¹, Anders Ynnerman¹

Visual Information Technology and Applications, VITA
Department of Science and Technology, Linköpings Universitet
601 74 Norrköping Sweden
e-mail: jonas.unger@itn.liu.se

Received: date / Revised version: date

Abstract We present a novel technique for capturing spatially or temporally resolved light probe sequences, and using them for image based lighting. For this purpose we have designed and built a *Real Time Light Probe*, a catadioptric imaging system that can capture the full dynamic range of the lighting incident at each point in space at video frame rates, while being moved through a scene. The Real Time Light Probe uses a digital imaging system which we have programmed to capture high quality, photometrically accurate color images of 512x512 pixels with a dynamic range of 10,000,000:1 at 25 frames per second.

By tracking the position and orientation of the light probe, it is possible to transform each light probe into a common frame of reference in world coordinates, and map each point and direction in space along the path of motion to a particular frame and pixel in the light probe sequence. We demonstrate our technique by rendering synthetic objects illuminated by complex real world lighting, first by using traditional image based lighting methods and temporally varying light probe illumination, and second an extension to handle spatially varying lighting conditions across large objects and object motion along an extended path.

Key words High Dynamic Range Imaging, Image Based Lighting

1 Introduction

One of the ultimate objectives for computer graphics is to generate images of virtual worlds that are indistinguishable from photographs of the real world. This goal poses challenging research questions in all areas of computer graphics. It is widely recognized that one of the

Send offprint requests to: Jonas Unger

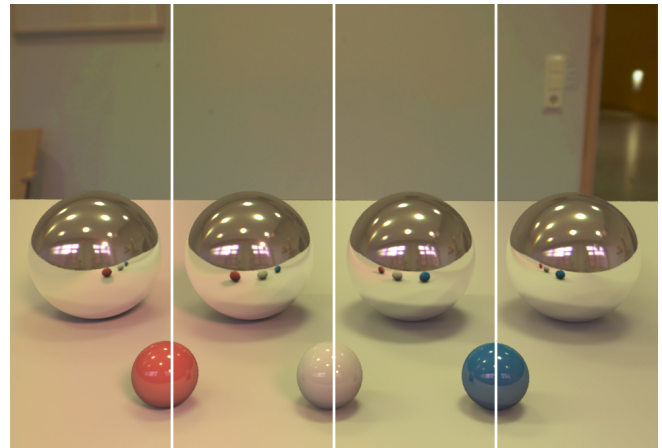


Fig. 1 Excerpts from four frames of an animation of a synthetic scene against a photographic backdrop, rendered with spatially varying image based lighting. The lighting data was a tracked sequence of several hundred light probes from a HDR video sequence, and the rendering was performed using a custom plug-in to Pixar RenderMan. Contrary to traditional image based lighting, this new type of rendering can capture strong local variation in the illumination. Here, the illumination changes continuously with surface position, as can be seen from the different shadows and reflections from the mirror sphere as it moves across the scene.

key factors in generating realistic looking images is accurate modeling of light and its interaction with matter. This has led to the development of increasingly advanced methods for modeling and simulation of light in virtual scenes. Despite the increasing speed of computers and improved algorithms, however, it has proven difficult and time-consuming to realistically model the complex illumination found in most real world environments. Therefore, image based methods have been proposed where *light probes* (omnidirectional *high dynamic range* (HDR) images) are used to capture real world lighting which is then used to illuminate synthetic objects and characters.

During the past decade image based lighting techniques have been the focus of many research efforts, which have made it possible to incorporate basic image based lighting in commercial renderers and production pipelines. Existing methods are still based on images of static lighting at a single point in space, however, while most real world lighting conditions vary over both time and space. This paper addresses this restriction by describing equipment and methods for sampling of spatially and temporally varying lighting situations and how such more detailed measurements of lighting can be used to achieve a higher level of realism in rendered images.

Imaging hardware research and associated development in computer technology are rapidly developing high performance, configurable hardware capable of supporting fast frame rate, and even streaming, HDR capture. We have developed such a digital imaging system to capture high quality color HDR images at video rates. We use this device to capture images with a dynamic range of 10,000,000 : 1 at 25 frames per second, an unprecedented performance both in terms of dynamic range and frame rate. The system performance is highly configurable in terms of frame rate, resolution, exposure times and covered dynamic range, and can be adapted for a wide range of applications. Using this camera hardware, we have designed and built a light sampling device, a Real Time Light Probe, see Figure 2. The HDR video imaging system is mounted on a rig, and records the environment through the reflection in a mirror sphere. The light probe can easily be moved around in a scene, and can rapidly capture the incident illumination at any position in space.

We use the captured data to produce renderings of synthetic objects as illuminated by the lighting found in a real world scene. Using the light probe sequences instead of a static image we can render objects under complex lighting conditions with significant spatial and temporal variation. We foresee many different areas of application for the presented system such as improved lighting for virtual objects and characters in SFX, realistic virtual prototyping in the automotive industry and architectural visualization.

The main contributions of this paper are:

- Description of a photometrically correct, high quality HDR color video capture methodology using a high performing camera programmed for HDR read-out using a rolling shutter technique.
- Design of a light probe rig, based on the HDR camera, enabling high resolution capture of temporally and spatially varying light environments.
- Introduction of novel techniques for rendering of objects in light environments having significant variation on the scale of the size of the object.
- Examples of rendered objects in measured spatially varying lighting, showing the added value of the real-time light probe in image based lighting work flows.

2 Related Work

The work presented in this paper relates to previous work in two different areas and builds on a synthesis and further development of results in these areas. The first area is the design and implementation of the real-time light probe which relates to the field of high speed imaging, camera hardware and sensor technology. The second area is the capture process, data processing, and rendering application which relates to the field of HDR imaging and image based lighting. The key concept that bridges the two fields in this context is the *Plenoptic Function* or, more precisely, the sampling of the plenoptic function.

The plenoptic function $P(\phi, \theta, \lambda, x, y, z, t)$, introduced by Adelson and Bergen [1], describes the radiance of any wavelength λ arriving at any point in space (x, y, z) from any direction (ϕ, θ) at any time t . This function is often simplified by fixing time and using one equation per spectral integral (R, G, B) , and thus three 5D functions, $P(\phi, \theta, x, y, z)$, describe any possible omnidirectional image seen from any fixed point in space.

The idea of environment mapping, as presented by Blinn [2] and subsequently Miller and Hoffman [13] approximates the plenoptic function at one point in space by capturing environment maps as photographs of a mirror sphere placed in a real world environment. Using the environment map they simulated both diffuse and specular materials, and showed how blurring of the map could simulate different reflectance properties. More recently Debevec [3] proposed methods for rendering synthetic objects into real world scenes. He sampled the plenoptic function by capturing panoramic HDR images [25], *radiance maps*, of the incident illumination at one single point in space and used this 2D information, $I(\phi, \theta)$, to render synthetic objects and integrate them into photographs of the real scene. The method was limited to introducing objects into a local scene near the placement of the probe and the effect of the virtual object could only be seen in the vicinity of the object. Sato et al. [18] generalized the approach and, by using an omnidirectional stereo algorithm, a radiance map of the full scene could be reconstructed.

Debevec et al. [4] then proposed a technique for image based relighting of reflectance fields of human faces, captured using a *light stage*, in which the face is photographed under varying light conditions. Using this technique the subject could be accurately rendered into arbitrary spatially invariant light environments. To capture spatial variations in the lighting environment Unger et al. [23] used light field techniques, as presented by Grotler et. al. [6] and Levoy et al. [9], and captured omnidirectional HDR images of the incident illumination at evenly spaced points on a plane. The 4D captured real world lighting data, $I(\phi, \theta, x, y)$, was then used within a global illumination framework to render synthetic objects illuminated by spatially varying lighting such as

spotlights, dappled lighting and cast shadows. Masselus et al. [12] demonstrated that such light fields were useful for image based relighting of captured reflection data. However, the capture time for such an HDR light field was very long, and the scene to be captured had to be kept stationary during the entire process. This made it impractical to perform a dense sampling of the lighting variation.

The further development and practical use of the techniques described above were hindered by the difficulties of rapid and accurate HDR capture which are largely due to the limited dynamic range of CCD and CMOS sensors. There are a number of commercial sensors and cameras with an extended dynamic range in the order of three to four orders of magnitude, some using sensors with a logarithmic response curve. A nice overview of available cameras and the field of HDR imaging can be found in [16]. However, it should be noted that the extreme dynamic range required for image based lighting can not yet be adequately captured with such systems.

Currently the most common technique for capturing HDR images is to use a series of images of a scene with varying exposure settings such that the full dynamic range of the scene is covered. Most digital cameras have an intrinsic non-linear response function, f , to mimic analogue film and to stretch the dynamic range in the digital, usually 8 bit, output image. This function maps the registered radiance, E , integrated over the exposure time, Δt , to pixel values, Y , where $Y = f(E\Delta t)$. By recovering the camera response function the radiance can be computed as $E = f^{-1}(Y)/\Delta t$, and HDR images can be assembled using multiple exposure techniques. Initial work in this direction was conducted by Madden [10] and Mann and Picard [11]. The nonlinear response function of the camera was recovered through a parametric fitting and the set of low dynamic range images were combined into a high dynamic range image. Subsequently Debevec and Malik [5], Robertson et al. [17] and Mitsunaga and Nayar [14] developed more general, robust and accurate methods for recovering the response function. Nayar and Mitsunaga [15] then presented a technique for extending the dynamic range of a camera by placing a filter mask in front of the sensor, with varying transmittance for adjacent pixels. The values from differently exposed pixels could then be combined into an HDR image.

More recently Kang et al. [7] used a camera from Pt. Grey Research, programming it to alternately capture two different exposures at 15 fps. From each of these pairs of images they could assemble a final image with a slightly extended dynamic range at 7.5 fps. Waese and Debevec [24] demonstrated a real-time HDR light probe where neutral density filters with different attenuation levels were attached to four of five facets of a prismatic lens. By combining the five different images seen by the sensor, HDR frames were assembled and used as lighting information for a rendered sequence. The frame rate was full video rate, but the cost for the high HDR frame

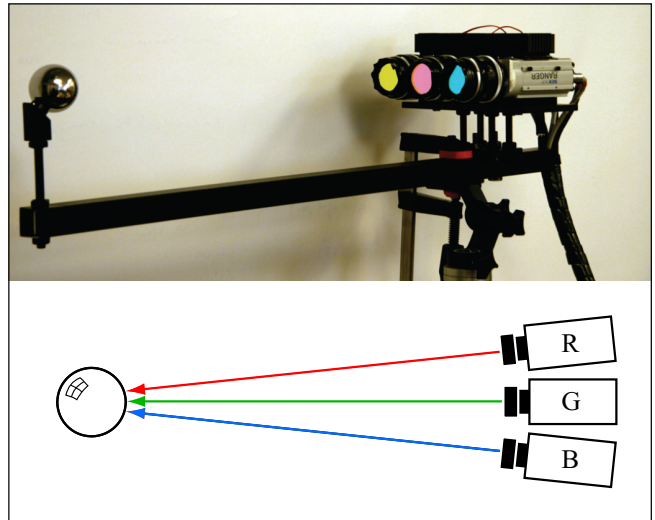


Fig. 2 Top: The RGB capture setup consists of three separate camera units aimed at the same mirror sphere. The camera units are connected to the same host PC that processes the three raw HDR data streams. **Bottom:** Principal sketch of the setup. The angle between the red and green and blue and green camera units is approximately 5° . In the image registration the different view-angles are compensated for by a rotation in the spherical images.

rate was a low spatial resolution of the final light probe image. Another real-time light probe, based on multiple exposures, was presented by Unger et al. [21]. There, a highly programmable imaging system was used to capture HDR images covering 15 *f-stops* at 25 fps. However, that system was monochrome and, because of the time disparity between the different exposures, rapid camera and object motion in the scene could lead to ghosting artifacts in the final HDR image.

3 A Real Time Light Probe

The work presented here overcomes many of the problems with previous methods for rapid HDR imaging, and presents a significant improvement. It is now possible to perform spatial and temporal sampling of a 6D version of the plenoptic function of the form $P(\phi, \theta, x(t), y(t), z(t), t)$, i.e. space and time can be varied in an interdependent fashion. We capture panoramic HDR image sequences of incident lighting, using a catadioptric imaging system consisting of an HDR video camera and a mirror sphere. Our hardware solution for HDR video has been presented in detail in Unger et al. [20], but a summary is given below.

3.1 Imaging Hardware

The HDR video camera, see Figure 2, is based on a commercially available camera platform, the Ranger C50

from the company SICK IVP¹. The camera was originally designed for industrial inspection purposes, but its configurability makes it possible to re-program it to function as a high performance multiple exposure camera for HDR image capture.

The large CMOS sensor, 14.6 by 4.9 mm, has a resolution of 1536 by 512 pixels and an internal and external data bandwidth of 1 Gbit/s. Each column on the sensor has its own A/D converter and a small processing unit. The 1536 column processors working in parallel allow for real-time on-chip image processing. Exposure times can be as short as a single microsecond, and A/D conversion can be performed with 8 bit accuracy. It is also possible to A/D convert the same analogue readout twice with different gain settings for the A/D amplifier. By cooling the sensor, the SNR can be kept low enough to obtain two digital readouts from a single integration time without any significant degradation due to thermal noise.

The camera sensor is monochrome, so color images are acquired through an externally synchronized three-camera system, one for each color channel (R, G, B), see Figure 2. Each camera is connected via a Camera Link interface to a host PC, and the three cameras are mounted in fixed positions aimed at the mirror sphere.

3.2 HDR Capture Methodology

Our HDR capture methodology is similar to the multiple exposure algorithm used for still images, although we have implemented a continuous rolling shutter progression through the image to avoid having the different exposures acquired at widely disparate instants in time, see Figures 3 and 4. This means that a set of rows in a moving window on the sensor are being processed simultaneously. As soon as an exposure is finished for a particular row, the value is A/D converted and the next longer exposure is immediately started for that row, so that at any instant every row on the sensor is either being exposed or processed. All rows are not imaged simultaneously, which yields a slight "curtain" effect for camera and scene motion, but in return all exposures for one particular row of pixels are acquired head to tail within the frame time. Two positive side effects are that almost the entire frame time is used for light integration and that the longest exposure lasts almost the entire frame time.

The system is highly configurable, and there are trade-offs possible between the dynamic range, the number of exposures, the image resolution and the frame rate. The hard limiting factors are the maximum data output rate of 1 Gbit/s, the A/D conversion time of $9.6 \mu s$ per exposure for each row of pixels and the total sum of all exposure times. Because of the rolling shutter methodology, A/D conversion can be performed in parallel with

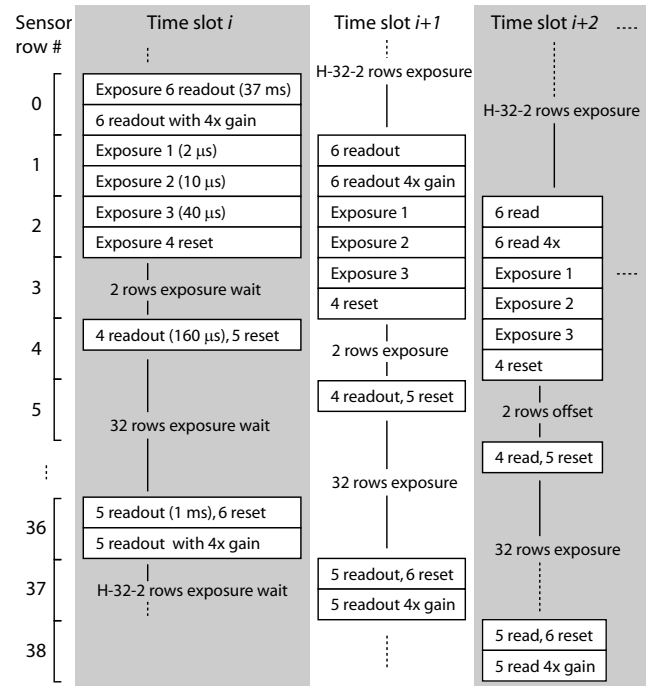


Fig. 3 The progressive image exposure and readout from a rolling shutter algorithm effectively removes any waiting time between subsequent exposures within each HDR frame. For each time slot several exposures and readouts are performed. One full frame is exposed in H time slots, one for each row. In our example $H = 512$.

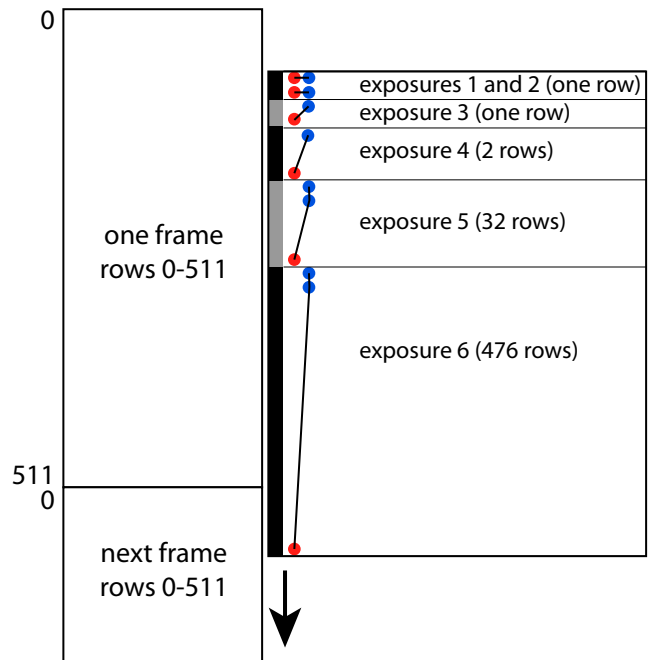


Fig. 4 Another, perhaps more intuitive way of describing the capture algorithm for one time slot: a moving window (right) is positioned over the sensor area (left). Red dots represent resets for each of the six exposures, blue dots represent the eight readouts for the image capture. A full frame is captured as the window is moved 512 rows down.

¹ <http://www.sickivp.se>

exposure. If the number of exposures is N , their exposure times are T_j , and the image resolution is H rows of W pixels each, the resulting minimum frame time is

$$\begin{aligned} T_c &= H \cdot 9.6 \cdot 10^{-6} [s] \\ T_{p,j} &= \max(T_c, T_j) \\ T_d &= (H \cdot W \cdot 8 \cdot 10^{-9}) [s] \\ T_f &= \max\left(\sum_{j=1}^N T_{p,j}, N \cdot T_d\right) \end{aligned} \quad (1)$$

The processing time for each exposure, $T_{p,j}$, is the maximum of the A/D conversion time for one full frame of H rows, T_c , and the exposure time for exposure j , T_j . The frame time, T_f , is the maximum of the processing time and the time required to transfer all N exposures over the 1 Gbit/s data link, T_d .

In our current implementation of the real-time light probe we use eight exposures each 2 to 3 f-stops apart, an image resolution of 512 x 512 pixels and a frame rate of 25 frames per second, which is well within the capabilities of the hardware. The exposure times and gain settings are indicated in Figure 3. This particular choice of parameters results in the frame time being bounded by the A/D conversion time, making a larger image resolution possible.

To estimate the irradiance, E_i , seen by a certain pixel, i , on the sensor, traditional multiple exposure techniques usually use a weighted average of pixel values, $Y_{i,j}$, from the set of captured exposures denoted by j . This averaging is performed in order to reduce artifacts and to produce a robust irradiance estimate.

Because we have a known and carefully calibrated system with direct access to the linear output of the A/D conversion and a high SNR ratio we have no need for averaging. Instead we base our irradiance estimate, E_i , on the most reliable value from the set of exposures, and encode it as a floating-point photometric value. The limiting factors for the accuracy in the irradiance estimate are the increasing relative quantization error and decreasing SNR ratio for low pixel values. This means that we can simply base our estimate on the highest pixel value below the saturation threshold, Y_s , from the set of N exposures. The radiance estimate can then be computed as:

$$E_i = \frac{X_{i,j}}{T_j} \quad (2)$$

where $X_{i,j}$ is the highest non saturated pixel value for pixel i , and T_j is the exposure time of the corresponding j -th and most reliable exposure. This is very hardware friendly since the algorithm is basically a comparison against a saturation threshold, X_s , and a sequential conditional update of a single value over the set of exposures for a particular pixel.

At each readout from the sensor, A/D converted values from two different exposures or two different gain settings are available simultaneously for each pixel. The two shortest exposures are performed on the same row of pixels within the same time slot of 78 μ s, and for the remaining exposures the 1x and 4x dual gain readouts are also performed in rapid succession. Because at most one of these values will be used for the final HDR image, and because the sensor chip itself has considerable processing capability available, we do not transmit every A/D converted value to the PC host. Instead, a simple multiplexing operation is performed on the sensor, so that for each pair of values for one pixel, only the best value is selected for output, and a final 4-bit value is transmitted for each time slot denoting which exposures from each pair that were selected. By this multiplexing operation, we save some bandwidth compared to equation (1) presented above and can transmit a higher resolution image than would have been possible otherwise. We have not taken this optimization to its full potential, only as far as needed to be able to handle the data streams from all three cameras using a single PC host. In the current implementation, a total of 48 bits are transmitted for each row of eight 8-bit exposures, thereby saving 25% on bandwidth. On the host, a final multiplexing operation is performed in software to select the best value for each pixel of HDR output. Once again, the selection is simple: we pick the highest unsaturated value.

With three cameras connected to the same PC, RGB color HDR frames of 640x512 pixels can be streamed to disk with a sustained frame rate of 25 frames per second. The data streams from the three cameras to the PC amount to around 1Gbit/s in total, and the data written to disk is around 300 Mbits/s. Both these figures are well within the bandwidth limits of a standard high-end PC.

3.3 System Evaluation

With 8 bit A/D conversion using a variable gain of 1x to 4x, and exposure times ranging from 2 μ s to 37 ms, the dynamic range of the composite HDR image is comparable to a linear A/D conversion of $8 + \log_2(4 \cdot 37,000/2)$ bits, or more than 24 bits.

Compared to currently available logarithmic sensors, this system has significantly better image quality and accuracy. It should be noted that a hypothetical ideal logarithmic sensor with a similar dynamic range and 10 or 12 bits A/D conversion would exhibit about the same relative quantization error. Currently available logarithmic sensors, however, have problems that are still limiting their practically attainable accuracy for absolute radiometric measurements [8].

This system also compares favorably with traditional multiple exposure techniques. Our final HDR image has a dynamic range comparable with a multiple exposure acquisition using exposures covering as many as 16 f-stops, in effect a dynamic range of 10,000,000 : 1, and

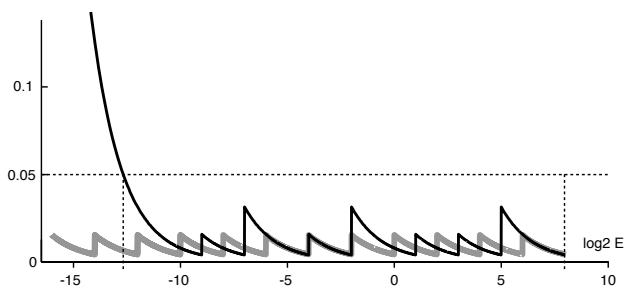


Fig. 5 Black curve: The relative quantization error with our particular choice of exposure times is within a few percent over a wide dynamic range. **Grey curve:** The relative quantization error for a wider range of exposures including very long exposure times, taken 2 f-stops apart using 8 bit linear A/D conversion. This is comparable to what could be achieved using a high quality still image camera, although standard still image cameras rarely allow for exposures in the microsecond range.

a relative quantization error within a few percent for all but the lowest exposure values. Capture of one such HDR frame can be performed in 40 milliseconds. Thus the system is capable of capturing color HDR images, with extreme dynamic range and a spatial resolution similar to standard digital video, at video frame rates. If a frame rate of 24 or 30 fps is desired the system can be reconfigured accordingly.

The sensor has good pixel uniformity and the thermal noise is low for the relatively short exposures used, so our main source of error is the 8 bit quantization. The relative quantization error as a function of irradiance, E , is displayed in Figure 5. For comparison we also display the quantization error for a capture using several more exposures taken only 2 f-stops apart with 8 bit linear A/D conversion. This is similar to what would be practical using a standard digital SLR camera. Towards low radiance values our quantization error peaks because we cannot use exposure times longer than the frame time. However, the sensor has a high light sensitivity and our longest exposure time of 37 ms gives good images even in fairly dim indoor lighting. Thus, longer exposure times are not really needed.

The rolling shutter methodology greatly reduces the potentially serious problem with camera or scene motion during capture. The worst case scenario is a camera or scene motion in the vertical direction in the image. Very small objects that cover only a few sensor pixels must not move through vertical distances comparable to their full height within the frame time, or they will have the wrong size, the wrong intensity or be entirely missed in the shot. In our experiments, this has not been a severe constraint. Rapid vertical scene motion or vertical camera panning is not commonly seen in video footage, and objects so small that they are imaged as a single pixel are also not very common. If this problem arises, it can be alleviated for our application by bringing the

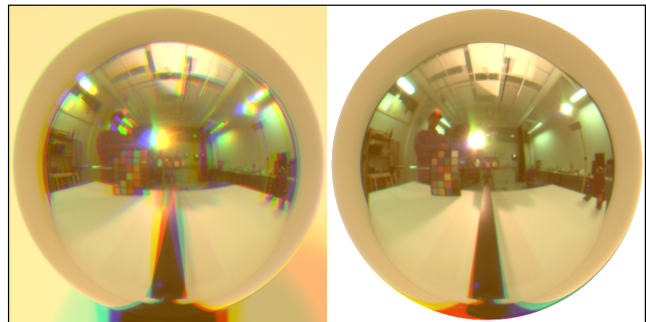


Fig. 6 An HDR image captured with the real time light probe before (**left**) and after (**right**) the alignment rotation of the color channels. The alignment works very well for most parts of the image. Only objects which are close to the mirror sphere and reflected near its edge will be slightly misaligned due to parallax.

camera slightly out of focus, thereby making the problematic object cover more pixels in the captured image. Moreover, in this particular application, the image is a panoramic view through a mirror sphere in a fixed position relative to the camera, so scene motion is typically slow. For direct imaging applications it might still be a problem and should be investigated further.

Because the short exposure times are so extremely short, there will be a problem with flickering light sources. Direct views of fluorescent tubes will not, by default, be measured correctly at shorter exposures. If it is important to capture fluorescent tube lighting correctly, the frame time should be synchronized with the AC supply frequency and the aperture setting for the optics should be adjusted so that one of the longest exposure times gives a valid A/D reading for direct views of the fluorescent lights. This is perfectly possible but requires some extra care. To avoid such problems in our experiments, we used only daylight and incandescent lighting with no significant amount of flicker.

4 Data Processing

In order to produce lighting data that is useful for rendering, we need to process the raw output stream from each camera and assemble the final HDR image sequence. We also need to know the physical light probe position and orientation in the scene for each frame so that they can all be transformed into world coordinates.

4.1 HDR Image Assembly and Image Registration

The output from each camera unit is a data stream with a data rate of 0.315 Gbit/s. Using three cameras, (R, G, B), connected to the same host PC the high data rate, of around 1 Gbit/s, makes it impractical to perform any extensive data processing on the fly. Instead,

the raw data is streamed to disk, and the HDR images are assembled in a post-processing step.

For each pixel, i , the raw pixel data is stored as a mantissa, m_i , and an exponent, e_i , for each color channel. The exponent denotes which of the exposures the mantissa belongs to, that is the exposure with the highest non saturated pixel value.

First, shading correction is performed on the raw mantissa images to remove fixed pattern noise and compensate for the camera’s black and saturation levels.

The shading corrected images are then converted into 16 or 32 bit floating point images using the exponent image. Since the mantissa, m_i , is linear to the observed radiance, no non-linear camera response function needs to be taken into account. By careful calibration of the system, the radiance estimate can then be computed as:

$$\begin{aligned} e_i &= [0, 1, \dots, N - 1] \\ E_i &= m_i \cdot k_j(e_i) \end{aligned} \quad (3)$$

where k_j is inversely proportional to the exposure time T_j for exposure e_i and N is the number of exposures. In this manner the radiance estimate is based on the most reliable sample value only.

Since we are not using a beam splitter and the three color channels are not captured from the same vantage point, see Figure 2, they have to be aligned. In this setup the three cameras are aimed at the same point, the center of the mirror sphere, and positioned at the same distance from this point. This means that the color channels can be aligned by a rotation of the projected directional coordinates for the sphere image. Figure 6 displays an HDR image before and after these rotations of the color channels have been performed. Overall this alignment procedure works very well, but towards the edges of the image objects close to the mirror sphere will not match up perfectly. The reason for this is that the cameras have different viewpoints and see slightly different things. An additional effect from the three-camera setup is that for some angles, there will be a slight tracking disparity between the color channels, as the red, green and blue rays for a realigned image actually emanate from slightly different positions on the reflective sphere. After alignment, the processed images are stored on disk.

4.2 Light Probe Tracking and Transformation

To spatially relate the light probe images to each other and to the scene, we track the probe position and orientation through the sequence. Tracking could be performed by either physical or image-based tracking methods, even directly from the spherical light probe image data, but in this experiment two external video cameras were used to track feature points placed on the light probe rig and register the light probe motion in the scene, and tracking was performed with standard

commercial video tracking software. By tracking the motion of the light probe the temporal variation of the sequence is related to the spatial variation of the plenoptic function, i.e. we sample the function on the form $P(\phi, \theta, x(t), y(t), z(t), t)$.

The tracking data is stored together with the light probe sequence image data and used during rendering to determine the spatial relationship between the scene and the light field data set.

5 Rendering

Densely sampled sequences of real world illumination open the door to a whole new area of rendering techniques using image based lighting with high frequency variations, either spatially or temporally. Here we demonstrate two rendering techniques making use of light probe sequences captured along a tracked 1D path in space.

First we show renderings of a small synthetic scene moving along the captured path. In this case the scene is rendered in the traditional way using only one light probe for the entire scene, but a different light probe is used for each frame. This is in itself an improvement over using the same light probe for all positions.

Second, and more importantly, we show renderings where the scene and the objects in it span a region with significant local variation in lighting. In this case we use data from several hundred light probes to render each object point, depending on both the position of the point in world space and the direction of the incident rays on the surface. This is a significant improvement over existing methods.

5.1 Traditional Rendering

The traditional image based lighting method is to use one light probe captured at one single point in space as an approximation of the plenoptic function over a fairly large volume in space, thus removing any spatial variation and reducing $I(\phi, \theta, x, y, z)$ to $I(\phi, \theta)$. Using a light probe sequence, we can render objects as if illuminated by the incident light at each point along the captured path, by simply illuminating the virtual scene using the corresponding light probe. Figure 7 shows frames from an animated sequence where a synthetic scene is illuminated by one single light probe image at different positions along the captured path in a moderately complex indoor lighting situation.

This method corresponds exactly to traditional image based lighting. The extra consideration is that the determination of which light probe to use is dependent on the spatial position and so accurate tracking is required during the capture of the light probe sequence to ensure that accurate position information is available. Also, some speedups which are commonly employed, like

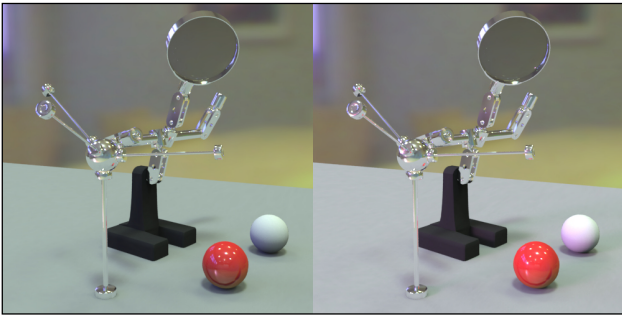


Fig. 7 Two adjacent frames from a rendering using light probe data from an HDR video sequence. The lighting for the real scene was not uniform - there was a relatively slight but noticeable variation between light probes captured only millimeters apart. Even though each frame by itself exhibits a high realism, the rendering fails to capture any spatial variation in lighting across the scene. As a synthetic object moves through the spatially varying light field, the entire object is affected instead of having shadows and streaks of light move across it, and the animation flickers badly.

a caching of the importance sampling of the environment map, will no longer be relevant when each rendered frame makes use of a different light probe image. Apart from that, each rendered frame uses standard image based lighting techniques, available in most modern commercial renderers.

This type of rendering is sufficient for small objects under low frequency spatial variations in the lighting. Using the real time light probe, lighting information can be sampled along complex paths of rapid motion under temporally and spatially varying lighting, which in itself presents new opportunities for special effects purposes. Under real lighting conditions, however, the illumination can vary so rapidly that even across the extent of the object the change will be noticeable. In almost any scene there will be significant variation in lighting across the extent of the rendered view. Traditional image based lighting cannot capture such effects. When used for animation, the result from using a light probe sequence captured under strong local variations in lighting is a disturbing global flickering instead of an impression of continuous motion through a lit scene. Therefore, image based lighting using a single light probe per frame is unsatisfactory for anything but small objects or very smooth and slow variations in lighting.

5.2 Rendering Large Objects

To demonstrate our ability to capture high frequency spatial variations in the illumination over a larger distance, we have used a full sequence of light probes for rendering a single scene. The light probes were captured with a mirror sphere of 5 cm diameter, and the path was approximately 70 cm long. Along that path we captured

700 light probes, one per millimeter. Our real time light probe system made this capture in under 30 seconds.

For the purpose of demonstration, the light field was deliberately chosen to have its main variation along the direction of the path. Any variation orthogonal to that direction would not be properly sampled by a 1D sequence and was therefore avoided. In a real world situation, variations in directions other than along the capture path might or might not be significant, and this should be taken under consideration in determining whether a 1D sequence is sufficient to capture the relevant properties of the light field at hand.

The 1D capture is immediately useful for some situations, and it is also a proof of concept that demonstrates the fundamental principle of capturing a spatial variation in the light field. The real time light probe can be used to perform 2D and even 3D captures in reasonable time, and we will investigate that further in future work.

Using the data set from a 1D light probe sequence, the detailed spatial variation along one dimension can be accurately captured and reproduced in rendering. The rendering method is very similar to a regular environment lookup using a single HDR light probe, with the difference that the influence from the environment depends not only on the incident direction, but also on the point of incidence.

5.3 Single-viewpoint reprojection

As noted by Swaminathan et al. [19], a mirror sphere does not perform a single viewpoint projection. In traditional image based lighting that fact is always ignored because it is inherently assumed that the light field is constant over all spatial dimensions. Even if it is not exactly constant, the variation of the light field is assumed to be negligible over at least the size of the mirror sphere, otherwise classic image based lighting is not applicable.

In our case, the spatial variation is captured at a high resolution, with a spatial sampling considerably more dense than the diameter of the mirror sphere. Therefore, the actual viewpoint needs to be considered so that each ray of incident light may be associated with its correct projection reference point. The 1D motion along the path for acquisition was performed along the optical axis of the camera system. This simplifies the reprojection, because all rays reflected from the mirror sphere surface have an intersection with the optical axis and, if the path of motion is coincident with that axis, the single viewpoint resampling breaks down to a simple angle dependent offset along the path. The incident ray intersects the optical axis with an offset from the center of the sphere according to:

$$\begin{aligned} \frac{\theta}{2} &= \arcsin \frac{r}{R} \\ z &= z_0 - r \cdot \cos \frac{\theta}{2} - r \cdot \tan \left(\theta - \frac{\pi}{2} \right) \end{aligned} \quad (4)$$

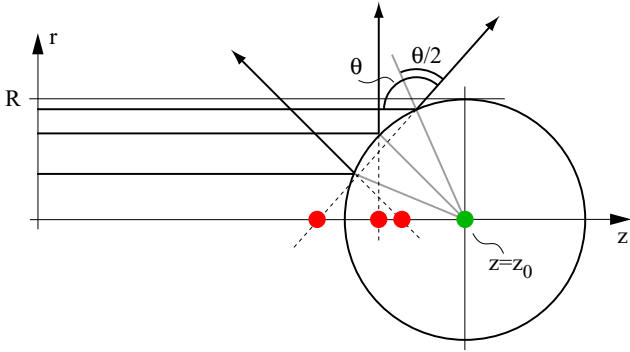


Fig. 8 Principle for the single-viewpoint remapping. The center of the mirror sphere (green) is tracked for each frame, and each ray of incidence (arrows) is reprojected from its position on the mirror sphere surface to its intersection with the optical axis (red) to put all rays in a common frame of reference.

where z is the actual viewpoint, z_0 is the center of the mirror sphere, r is the radial distance from the center of the image and R is the radius of the mirror sphere, see Figure 8.

For grazing angles on the sphere, i.e. rays reflected from near the outer rim of the sphere, the offset to the true projection point becomes large and any uncertainty in the angle becomes a potential problem. However, grazing angles represent rays of incidence where the spatial variation is very small along the path, so the numerical inaccuracy of the reprojection for those angles gives a negligible error.

It is worth noting that this single-viewpoint resampling in z is not performed explicitly. Instead, the light probe images are kept intact and the angle dependent offset is used for each lookup to find the correct point in the 3D data set $I(\phi, \theta, z)$.

5.4 Ray projection

A standard, single-probe IBL environment lookup is performed based on ray direction only. For a spatially varying light field, the lookup should be performed based on both the direction and the point of incidence. Our 1D sampling captures the variation along one dimension, and we assume that, at least at the scale of the objects we wish to render, there is no significant variation in the light field in the other two dimensions. This seems like a strong constraint, but in many situations it can be valid. For comparatively small objects moving along a path in a large scene, it is a good approximation which captures the most prominent effects well.

A naïve approach is to find the point on the principal axis of our captured data set which is closest to the point of incidence, and use the light probe acquired at that position to perform the environment lookup. This is the approach used in Unger et al. [22]. Even though

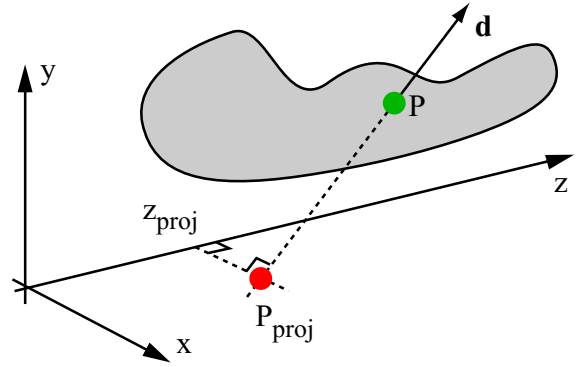


Fig. 9 Principle for the ray projection to find the correct light sample to use for an incident ray with direction \mathbf{d} at point \mathbf{P} . Each ray to be sampled for illumination is projected to where it is closest to a point in the data set which was captured along the z axis.

it reproduces a detailed variation, that variation is not physically correct. Parallax effects for oblique angles of illumination are not reproduced, so the detailed pattern of light and shadow on the object surface is correct only for incident light directions orthogonal to the principal axis of the capture. (Figure 10, middle row) The remedy for this is quite simple, and the result is a much more accurate reproduction of the lighting (Figure 10, bottom row).

For each point which is queried in the environment lookup, we project the incident rays at that point to seek the point along each ray which is closest to the principal axis for our light field capture. The point is found by simply seeking the minimum of the distance between the z axis and a point along the ray:

$$\begin{aligned} \mathbf{q}(u) &= \mathbf{P} + u\mathbf{d} \\ \Delta &= \sqrt{\mathbf{q}_x^2 + \mathbf{q}_y^2} \\ \mathbf{P}_{proj} &= \mathbf{P} + (\arg \min \Delta(u))\mathbf{d} \\ z_{proj} &= \mathbf{P}_{proj,z} \end{aligned} \quad (5)$$

where \mathbf{P} is the point being rendered and \mathbf{d} is the direction of incidence, see Figure 9. At the projected z position, z_{proj} , a direct environment lookup is performed in the viewpoint-adjusted 3D data set to find the incident light intensity from the direction \mathbf{d} at the point \mathbf{P} . To reduce sampling artifacts, trilinear interpolation is performed for the two angular and the single spatial coordinate. If the ray projection extends beyond the spatial extent of the capture, the outermost sample from the data set is used.

By this ray projection scheme, most rendered points will use lighting information from a wide range of positions in the capture. In fact, most points will use samples from the entire spatial extent of the data set even for simple, first-hit diffuse illumination. The drawback

of this is that for every point, the renderer requires information on the full data set for lighting. However, this is a common problem with global illumination in general and is not specific to our method.

Figure 11 shows a rendering of a diffuse cylinder illuminated by a light probe sequence captured under spatially varying lighting along with a reference photograph of the same scene. The synthetic image was rendered using the single-viewpoint adjustment and ray reprojection described above. The rendering shows correct behavior and good correspondence to the original scene. The camera angle and the object position were not matched exactly for these two images, and no attempts were made to match the tone mapping in the rendering with that of the digital camera. If these issues are addressed, this type of rendering will be very close to reality, as demonstrated by Figure 10. To demonstrate that this lighting algorithm can be used for arbitrarily complex scenes, Figures 1 and 12 shows more visually interesting scenes rendered using light probe sequences. The diffuse and specular reflections shows a correct dependence of position.

5.5 Implementation for a commercial renderer

Rendering with a spatially varying light field is not fundamentally different from using a single environment map, and not significantly more computationally complex. For each incident direction, it is still a simple lookup operation to find a representative sample in the data set. The main difference is that more data is required to represent the environment but, for reasonably sized light probe images and a moderate number of spatial positions, it is not a big problem. For our experiments, we have used at most 1,000 angular maps with a resolution of 512x512 pixels or less, resulting in a total of amount of data which can be accommodated in core memory of a standard personal computer, and our rendering times are not significantly higher than would arise from using standard IBL rendering. Significantly less data than we have used in our experiments is sufficient to render images of high quality but with less crisp specular reflections and less sharp transitions between light and shadow. There are also numerous possibilities for strong data compression and adaptive sampling, none of which have been investigated yet.

The renderings shown in this article were all created in Pixar’s PRMan, with a Renderman plug-in to read in the large 3D data set and handle the spatially varying light field illumination. The only thing that was changed compared to traditional IBL was the environment lookup function. At the code level, the implementation consists of an RSL plug-in named `environment1D(point, direction)` which replaces the standard call to `environment(direction)` in the illumination calculations in the regular global illumination rendering pipeline. No other changes are re-

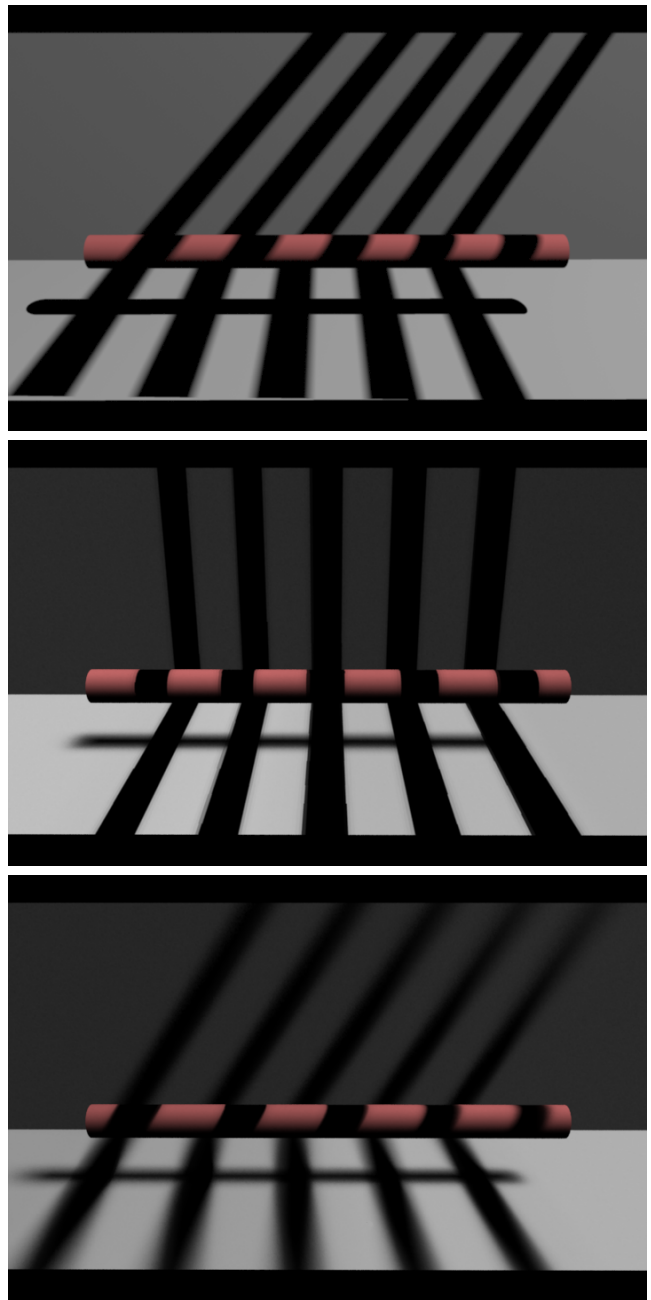


Fig. 10 Renderings from synthetic light probe data to show the benefits of our method. The synthetic scene is illuminated by a single projector light with a stripe pattern. **Top:** direct rendering with a traditional method to show the actual lighting for reference. **Middle:** naive nearest-neighbor rendering from a synthetic light probe sequence from the scene, showing incorrect behavior. **Bottom:** viewpoint-adjusted and reprojected rendering from the same light probe sequence, showing correct behavior and very good spatial and photometric correspondence to the original scene. The light field is sampled and has a limited spatial and angular resolution, hence the slight blurring of the lighting in the bottom image compared to the top image.

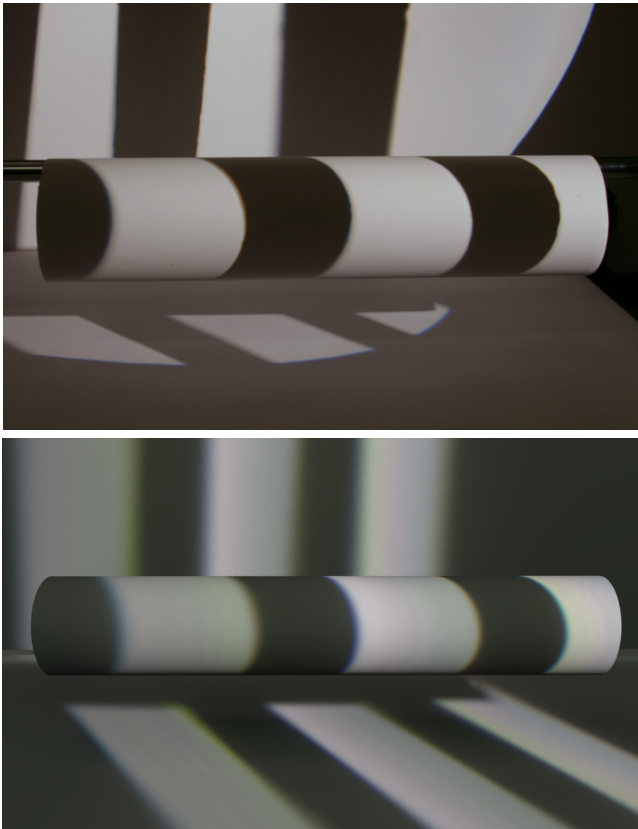


Fig. 11 Renderings from real world data to show the applicability of our method. The real scene is illuminated by a projector light with a stripe pattern. **Top:** photograph of the real scene to show the actual lighting. **Bottom:** viewpoint-adjusted and reprojected rendering from the same light probe sequence, showing correct behavior and good correspondence to the original scene. The three-camera setup causes a slight positional misalignment between the color channels for some angles, which can be seen by a close inspection of the edges between light and shadow. The camera response curve, the view angle and the object position were not matched exactly for these two images.

quired, and all existing features like stochastic and distributed sampling, shadow mapping, photon mapping, ambient occlusion and ray tracing are still available for use if needed, with the same performance as usual. The functionality is not dependent on the exact architecture of RenderMan plug-ins. The same functionality could be implemented in any renderer that supports image based lighting and has an open and flexible plug-in architecture.

6 Conclusion

The presented technique for capturing video sequences with an extreme dynamic range is not limited to the particular imaging hardware used, nor to this particular application to image based lighting. A similar capture methodology could be implemented in other pro-

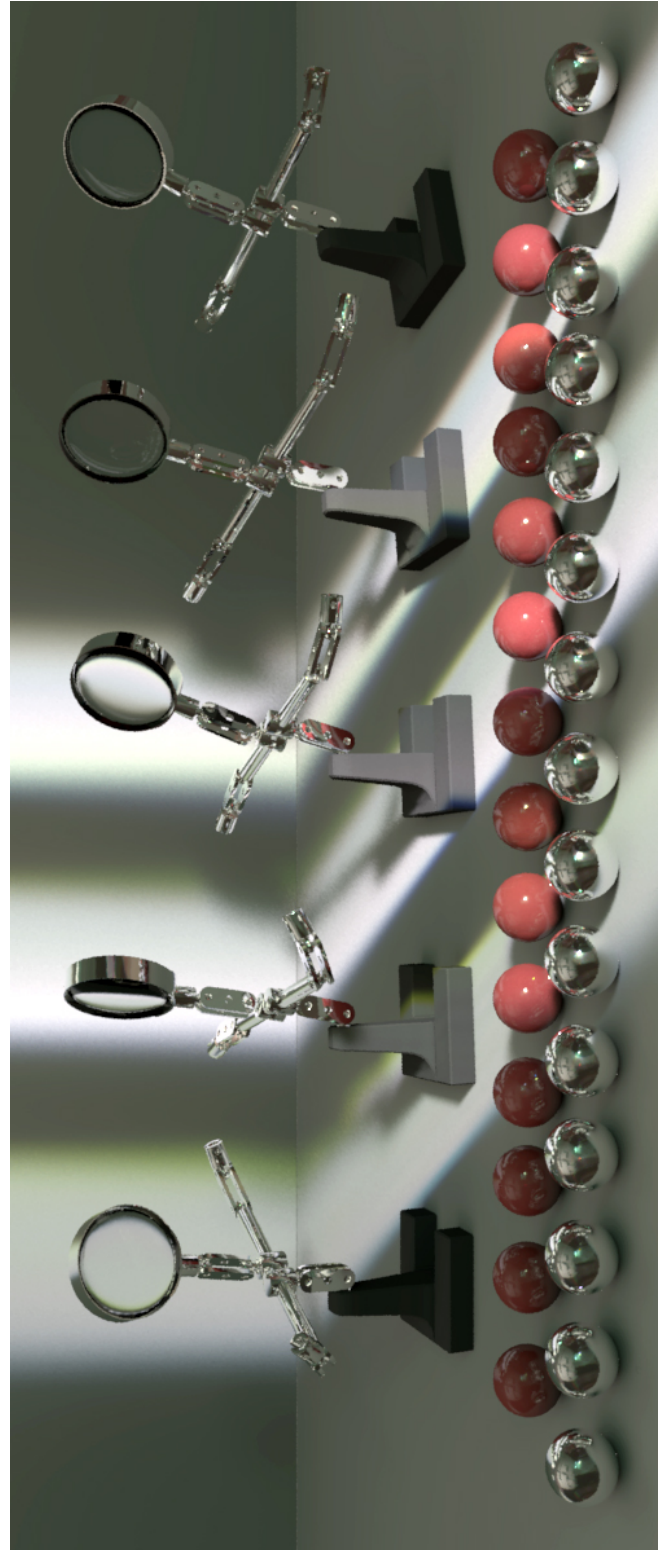


Fig. 12 A more complex synthetic scene with both specular and diffuse objects, rendered with the same spatially varying light field data as the simple scene in Figure 11.

grammable camera architectures. The focus application in this paper was rapid capture of light probe sequences in high frequency spatially varying illumination. We have displayed high quality renderings from a commercial renderer using the captured real world lighting, and showed that it is now possible to capture such illumination in a rapid and practical way. Rendering high quality images with this approach requires little more computation than is required for traditional image based lighting, only more memory. Even though the lighting was only captured along a 1D path, the data set manages to capture variations which would be impossible to handle with traditional image based lighting, and it is evident that spatially variant light field illumination provides a powerful and useful extension to image based lighting.

7 Future Work

The successful results obtained with the real time light probe system open up new research questions in several areas. Light fields with spatial variation in more than one dimension can also be captured in reasonable time. Rendering methods using such higher-dimensional light field data is an interesting area that we will investigate further. In the experiments presented here the tracking was performed on a video stream captured by an external camera. A more accurate tracking with direct feedback would be very useful, and we are designing a new mechanical tracking system to make it feasible to capture 2D and 3D light field data with more or less freehand camera motion. Given the large number of omni-directional images and the fact that light sources are features that can be detected easily in HDR images, tracking could also possibly be carried out directly on the light probe images.

Although the artifacts introduced by scene and camera motion are not a problem in the light probe setup the issue should be investigated if the camera system is to be used for direct imaging applications. The prototype RGB filters, in conjunction with the non-uniform spectral response of the camera, presents a color synchronization problem compared with commercial cameras, and this of course needs to be solved to make the lighting data useful in a production context.

Acknowledgements We gratefully acknowledge Per Larsson and Nils Högborg for their help with light probe capture, data processing and rendering. We would like to thank Mattias Johannesson at SICK IVP for the discussions and insightful suggestions and Anders Murhed at SICK IVP for the support of this project. We also would like to thank Matthew Cooper for proofreading the article. The first author was supported by the Science Council of Sweden through grant VR2002 621-2001-2623.

References

1. Adelson, E.H., Bergen, J.R.: Computational Models of Visual Processing, chap. 1. MIT Press, Cambridge, Mass. (1991). The Plenoptic Function and the Elements of Early Vision
2. Blinn, J.F.: Texture and reflection in computer generated images. *Communications of the ACM* **19**(10), 542–547 (1976)
3. 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* (1998)
4. Debevec, P., Hawkins, T., Tchou, C., Duiker, H.P., Sarokin, W., Sagar, M.: Acquiring the reflectance field of a human face. *Proceedings of SIGGRAPH 2000* pp. 145–156 (2000)
5. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: *SIGGRAPH 97*, pp. 369–378 (1997)
6. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The Lumigraph. In: *SIGGRAPH 96*, pp. 43–54 (1996)
7. Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R.: High dynamic range video. *ACM Trans. Graph.* **22**(3), 319–325 (2003). DOI <http://doi.acm.org/10.1145/882262.882270>
8. Krawczyk, G., Goesele, M., Seidel, H.P.: Photometric calibration of high dynamic range cameras. *Tech. Rep. Research Report MPI-I-2005-4-005* (2005)
9. Levoy, M., Hanrahan, P.: Light field rendering. In: *SIGGRAPH 96*, pp. 31–42 (1996)
10. Madden, B.C.: Extended intensity range imaging. *Tech. rep.*, GRASP Laboratory, University of Pennsylvania (1993)
11. Mann, S., Picard, R.W.: Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In: *Proceedings of IS&T 46th annual conference*, pp. 422–428 (1995)
12. Masselus, V., Peers, P., Dutre, P., Willems, Y.D.: Relighting with 4d incident light fields. *ACM Trans. Graph.* **22**(3), 613–620 (2003). DOI <http://doi.acm.org/10.1145/882262.882315>
13. Miller, G.S., Hoffman, C.R.: Illumination and reflection maps: Simulated objects in simulated and real environments. In: *SIGGRAPH 84 Course Notes for Advanced Computer Graphics Animation* (1984)
14. Mitsunaga, T., Nayar, S.: Radiometric Self Calibration. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 374–380 (1999)
15. Nayar, S., Mitsunaga, T.: High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 472–479 (2000)
16. Reinhard, E., Ward, G., Pattanaik, S., Debevec, P.: High Dynamic Range Imaging, Acquisition, Display and Image-Based Lighting. Morgan Kaufmann, San Francisco, CA (2006)
17. Robertson, M.A., Borman, S., Stevenson, R.L.: Dynamic range improvement through multiple exposures. In: *IEEE International Conference on Image Processing*, pp. 159–163 (1999). URL [cite-seer.ist.psu.edu/robertson99dynamic.html](http://cseer.ist.psu.edu/robertson99dynamic.html)

18. Sato, I., Sato, Y., Ikeuchi, K.: Acquiring a radiance distribution to superimpose virtual objects onto a real scene. *IEEE Transactions on Visualization and Computer Graphics* **5**(1), 1–12 (1999)
19. Swaminathan, R., Grossberg, M.D., Nayar, S.K.: Non-Single Viewpoint Catadioptric Cameras: Geometry and Analysis. *International Journal of Computer Vision* **66**(3), 211–229 (2006)
20. Unger, J., Gustavson, S.: High-dynamic-range video for photometric measurement of illumination. In: *In Proceedings of Sensors, Cameras and Systems for Scientific/Industrial Applications X, IS&T/SPIE 19th International Symposium on Electronic Imaging*, vol. 6501 (2007)
21. Unger, J., Gustavson, S., Ollila, M., Johannesson, M.: A real time light probe. In: *In Proceedings of the 25th Eurographics Annual Conference*, vol. Short Papers and Interactive Demos, pp. 17–21 (2004)
22. Unger, J., Gustavson, S., Ynnerman, A.: Densely sampled light probe sequences for image based lighting. In: *In Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, vol. Conference Proceedings (2006)
23. Unger, J., Wenger, A., Hawkins, T., Gardner, A., Debevec, P.: Capturing and rendering with incident light fields. In: *EGRW '03: Proceedings of the 14th Eurographics workshop on Rendering*, pp. 141–149. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2003)
24. Waese, J., Debevec, P.: A real-time high dynamic range light probe. In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques: Conference Abstracts and Applications*, p. 247. ACM Press/Addison-Wesley Publishing Co. (2002)
25. Ward, G.: *Real pixels. Graphics Gems II* pp. 80–83 (1991). ISBN 0-12-064481-9. Held in Boston.