Symmetric Photography: Exploiting Data-sparseness in Reflectance Fields

Gaurav Garg   Eino-Ville Talvala   Marc Levoy   Hendrik P. A. Lensch
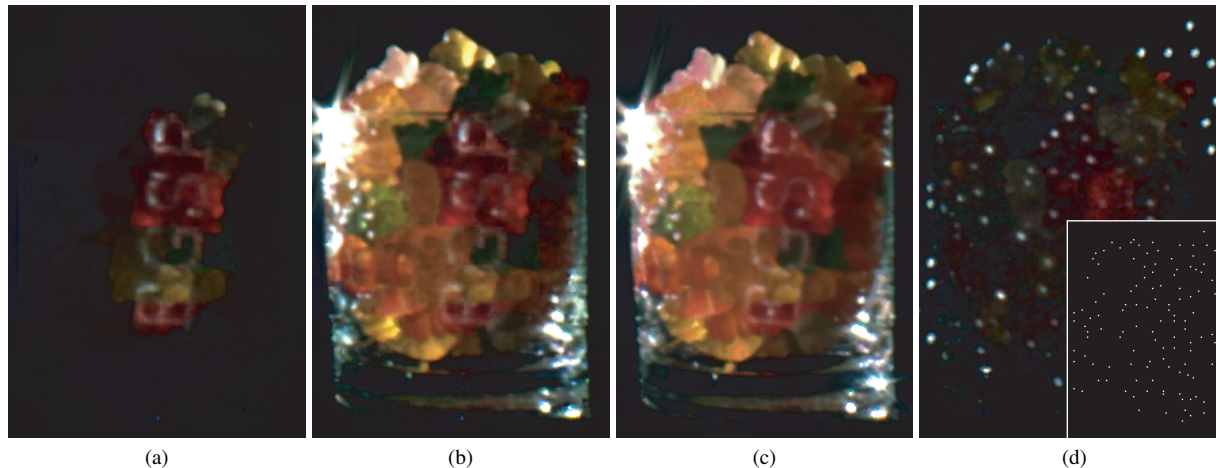
Stanford University

<div align="center">(a)   (b)   (c)   (d)</div>

Figure 1: The reflectance field of a glass full of gummy bears is captured using two coaxial projector/camera pairs placed 120° apart. (a) is the result of relighting the scene from the front projector, which is coaxial with the presented view, where the (synthetic) illumination consists of the letters "EGSR". Note that due to their sub-surface scattering property, even a single beam of light that falls on a gummy bear illuminates it completely, although unevenly. In (b) we simulate homogeneous backlighting from the second projector combined with the illumination used in (a). For validation, a ground-truth image (c) was captured by loading the same projector patterns into the real projectors. Our approach is able to faithfully capture and reconstruct the complex light transport in this scene. (d) shows a typical frame captured during the acquisition process with the corresponding projector pattern in the inset.

Abstract

We present a novel technique called symmetric photography to capture real world reflectance fields. The technique models the 8D reflectance field as a transport matrix between the 4D incident light field and the 4D exitant light field. It is a challenging task to acquire this transport matrix due to its large size. Fortunately, the transport matrix is symmetric and often data-sparse. Symmetry enables us to measure the light transport from two sides simultaneously, from the illumination directions and the view directions. Data-sparseness refers to the fact that sub-blocks of the matrix can be well approximated using low-rank representations. We introduce the use of hierarchical tensors as the underlying data structure to capture this data-sparseness, specifically through local rank-1 factorizations of the transport matrix. Besides providing an efficient representation for storage, it enables fast acquisition of the approximated transport matrix and fast rendering of images from the captured matrix. Our prototype acquisition system consists of an array of mirrors and a pair of coaxial projector and camera. We demonstrate the effectiveness of our system with scenes rendered from reflectance fields that were captured by our system. In these renderings we can change the viewpoint as well as relight using arbitrary incident light fields.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Digitizing and scanning; I.3.6 [Computer Graphics]: Graphics data structures and data types; I.3.7 [Computer Graphics]: Color, shading, shadowing, and texture

1 Introduction

The most complete image-based description of a scene for computer graphics applications is its 8D reflectance field [DHT*00]. The 8D reflectance field is defined as a transport matrix that describes the transfer of energy between a light field [LH96] of incoming rays (the illumination) and a light field of outgoing rays (the view) in a scene, each of which are 4D. The rows of this matrix correspond to the view rays and the columns correspond to the illumination rays. This representation can be used to render images of the scene from any viewpoint under arbitrary lighting. The resulting images capture all global illumination effects such as diffuse inter-reflections, shadows, caustics and sub-surface scattering, without the need for an explicit physical simulation.

Treating each light field as a 2D collection of 2D images, and assuming (for example) 3×3 images with a resolution of 100×100 in each image, requires a transport matrix containing about 10^{10} entries. If constructed by measuring the transport coefficients between every pair of incoming and outgoing light rays, it could take days to capture even at video rate, making this approach intractable.

This paper introduces *symmetric photography* - a technique for acquiring 8D reflectance fields efficiently. It relies on two key observations. First, the reflectance field is data-sparse in spite of its high dimensionality. Data-sparseness refers to the fact that sub-blocks of the transport matrix can be well approximated by low-rank representations. Second, the transport matrix is symmetric, due to Helmholtz reciprocity. This symmetry enables simultaneous measurements from both sides, rows and columns, of the transport matrix. We use these measurements to develop a hierarchical acquisition algorithm that can exploit the data-sparseness.

To facilitate this, we have built a symmetrical capture setup, which consists of a coaxial array of projectors and cameras. In addition, we introduce the use of hierarchical tensors as an underlying data structure to represent the matrix. The hierarchical tensor representation turns out to be a natural data structure for the acquisition algorithm, and it provides a compact factorized representation for storing a data-sparse transport matrix. Further, hierarchical tensors allow fast computation during rendering. Although our current capture system allows us to acquire only a sector of full sphere of incident and reflected directions ($37° \times 29°$), and even then only at modest resolution (3×3 images, each of resolution 130×200 pixels), this subset is truly 8-dimensional, and it suffices to demonstrate the validity and utility of our approach.

2 Related Work

The measurement of reflectance fields is an active area of research in computer graphics. However, most of this research has focused on capturing various lower dimensional slices of the reflectance field. For instance, if the illumination is fixed and the viewer allowed to move, the appearance of the scene as a function of outgoing ray position and direction is a 4D slice of the reflectance field. The light field [LH96] and the lumigraph [GGSC96] effectively describe this *exitant* reflectance field. By extracting appropriate 2D slices of the light field, one can virtually fly around a scene but the illumination cannot be changed. If the viewpoint is fixed and the illumination is provided by a set of point light sources, one obtains another 4D slice of the 8D reflectance field. Various researchers [DHT*00, MGW01, HED05] have acquired such data sets where a weighted sum of the captured images can be combined to obtain re-lit images from a fixed viewpoint only. Since point light sources radiate light in all directions, it is impossible to cast sharp shadows onto the scene with this technique.

If the illumination is provided by an array of video projectors and the scene is captured as illuminated by each pixel of each projector, but still as seen from a single viewpoint, then one obtains a 6D slice of 8D reflectance field. Masselus et al. [MPDW03] capture such data sets using a single moving projector. More recently, Sen et al. [SCG*05] have exploited Helmholtz reciprocity to improve on both the resolution and capture times of these data sets in their work on dual photography. With such a data set it is possible to relight the scene with arbitrary 4D incident light fields, but the viewpoint cannot be changed. Goesele et al. [GLL*04] use a scanning laser, a turntable and a moving camera to capture a reflectance field for the case of translucent objects under a diffuse sub-surface scattering assumption. Although one can view the object from any position and relight it with arbitrary light fields, the captured data set is still essentially 4D because of their assumption. All these papers capture some lower dimensional subset of the 8D reflectance field. An 8D reflectance field has never been acquired before.

Seitz et al. [SMK05], in their recent work on inverse light transport, also use transport matrices to model the light transport. While their work provides a theory for decomposing the transport matrix into individual bounce light transport matrices, our work describes a technique for measuring it.

Hierarchical data structures have been previously used for representing reflectance fields. These representations provide greater efficiency both in terms of storage and capture time. A typical setup for capturing reflectance fields consists of a scene under controlled illumination, as imaged by one or more cameras. Peers and Dutré [PD03] illuminate a scene with wavelet patterns in order to capture environment mattes (another 4D slice of the reflectance field). A feedback loop determines the next pattern to use based on knowledge of previously recorded photographs. The stopping criteria is based on the error of the current approximation. Although their scheme adapts to the scene content, it does not try to parallelize the capture process. Matusik et al. [MLP04] use a kd-tree based subdivision structure to represent environ-

ment mattes. They express environment matte extraction as an optimization problem. Their algorithm progressively refines the approximation of the environment matte with an increasing number of training images taken under various illumination conditions. However, the choice of their patterns is independent of the scene content.

In the dual photography work of Sen et al. [SCG*05], they also use a hierarchical scheme to capture 6D slices of the reflectance field. Their illumination patterns adapt to the scene content, and the acquisition system tries to parallelize the capture process, depending on the sparseness of the transport matrix. However, their technique reduces to scanning if the transport matrix is dense, e.g. in scenes with diffuse bounces. In this paper, we make the observation that the light transport in these cases is data-sparse as well as sparse. By exploiting this data-sparseness, we are able to capture full 8D reflectance fields in reasonable time. In Section 6.1, we compare our technique to dual photography in more detail.

3 Sparseness, Smoothness and Data-sparseness

To efficiently store large matrices, sparseness and smoothness are two ideas that are typically exploited. The notion of data-sparseness is more powerful than these two. A sparse matrix has a small number of non-zero elements in it and hence can be represented compactly. A data-sparse matrix on the other hand may have many non-zero elements, but the actual information content in the matrix is small enough that it can still be expressed compactly. A simple example will help convey this concept. Consider taking the cross product of two vectors, each of length n. Although the resulting matrix (which is rank-1 by construction) could be non-sparse, we only need two vectors ($O(n)$) to represent the contents of the entire ($O(n^2)$) matrix. Such matrices are data-sparse. More generally, any matrix in which a significant number of sub-blocks can have a low-rank representation is data-sparse. Note that a low-rank sub-block of a matrix need not be smooth and may contain high frequencies. A frequency or wavelet-based technique would be ineffective in compressing this block. Therefore, the concept of data-sparseness is more general and powerful than sparseness or smoothness.

Sparseness in light transport has been previously exploited to accelerate acquisition times in the work of Sen at al. [SCG*05]. Ramamoorthi and Hanrahan [RH01] analyze the smoothness in BRDFs and use it for efficient rendering and compression. A complete frequency space analysis of light transport has been presented by Durand et al. [DHS*05]. The idea of exploiting data-sparseness for factorizing high dimensional datasets into global low-rank approximations has also been investigated, in the context of BRDFs [KM99, MAA01, LK02] and also for light fields and reflectance fields [VT04, WWS*05]. By contrast to these global approaches, we use local low-rank factorizations. We tie in the factorization with a hierarchical subdivision scheme (see Section 5). This hierarchical subdivision allows us to exploit the data-sparseness locally.

4 Properties of Light Transport

4.1 Data-Sparseness

The flow of light in a scene can be described by a light field. Light fields, which were introduced in the seminal work of Gershun [Ger36], are used to describe the radiance at each point $\mathbf{x}$ and in each direction ω in a scene. This is a 5D function which we denote by $\widetilde{L}(\mathbf{x}, \omega)$. Under this paradigm, the appearance of a scene can be completely described by an outgoing radiance distribution function $\widetilde{L}_{out}(\mathbf{x}, \omega)$. Similarly, the illumination incident on the scene can be described by an incoming radiance distribution function $\widetilde{L}_{in}(\mathbf{x}, \omega)$. The relationship between $\widetilde{L}_{in}(\mathbf{x}, \omega)$ and $\widetilde{L}_{out}(\mathbf{x}, \omega)$ can be expressed by an integral equation, the well known rendering equation [Kaj86]:

$$\widetilde{L}_{out}(\mathbf{x}, \omega) = \widetilde{L}_{in}(\mathbf{x}, \omega) + \int_V \int_\Omega K(\mathbf{x}, \omega; \mathbf{x}', \omega') \widetilde{L}_{out}(\mathbf{x}', \omega') d\mathbf{x}' d\omega'$$
(1)

The function $K(\mathbf{x}, \omega; \mathbf{x}', \omega')$ defines the proportion of flux from $(\mathbf{x}', \omega')$ that gets transported as radiance to $(\mathbf{x}, \omega)$. It is a function of the BSSRDF, the relative visibility of $(\mathbf{x}', \omega')$ and $(\mathbf{x}, \omega)$ and foreshortening and light attenuation effects. Eq. (1) can be expressed in discrete form as:

$$\widetilde{\mathbf{L}}_{out}[i] = \widetilde{\mathbf{L}}_{in}[i] + \sum_j \mathbf{K}[i, j] \widetilde{\mathbf{L}}_{out}[j]$$
(2)

where $\widetilde{\mathbf{L}}_{out}$ and $\widetilde{\mathbf{L}}_{in}$ are discrete representations of outgoing and incoming light fields respectively. We can rewrite eq. (2) as a matrix equation:

$$\widetilde{\mathbf{L}}_{out} = \widetilde{\mathbf{L}}_{in} + \mathbf{K}\widetilde{\mathbf{L}}_{out}$$
(3)

Eq. (3) can be directly solved [Kaj86] to yield:

$$\widetilde{\mathbf{L}}_{out} = (\mathbf{I} - \mathbf{K})^{-1} \widetilde{\mathbf{L}}_{in}$$
(4)

The matrix $\widetilde{\mathbf{T}} = (\mathbf{I} - \mathbf{K})^{-1}$ describes the complete light transport between the 5D incoming and outgoing light fields as a linear operator. Heckbert [Hec91] uses a similar matrix in the context of radiosity problems and shows that such matrices are not sparse. This is also observed by Börm et al. [BGH03] in the context of linear operators arising from an integral equation such as eq. (1). They show that even though the kernel $\mathbf{K}$ might be sparse, the resulting matrix $(\mathbf{I} - \mathbf{K})^{-1}$ is not. However, it is typically data-sparse. In particular, the kernel is sparse because of occlusions. But due to multiple scattering events one typically observes light transport between any pair of points in the scene, resulting in a dense $\widetilde{\mathbf{T}}$. On the other hand, we observe that a diffuse bounce off a point on the scene contributes the same energy to large regions of the scene. Therefore, large portions of the transport matrix, e.g. those resulting from inter-reflections of diffuse and glossy surfaces, are *data-sparse*. One can exploit this data-sparseness by using local low-rank approximations for sub-blocks of $\widetilde{\mathbf{T}}$. We choose a rank-1 approximation.

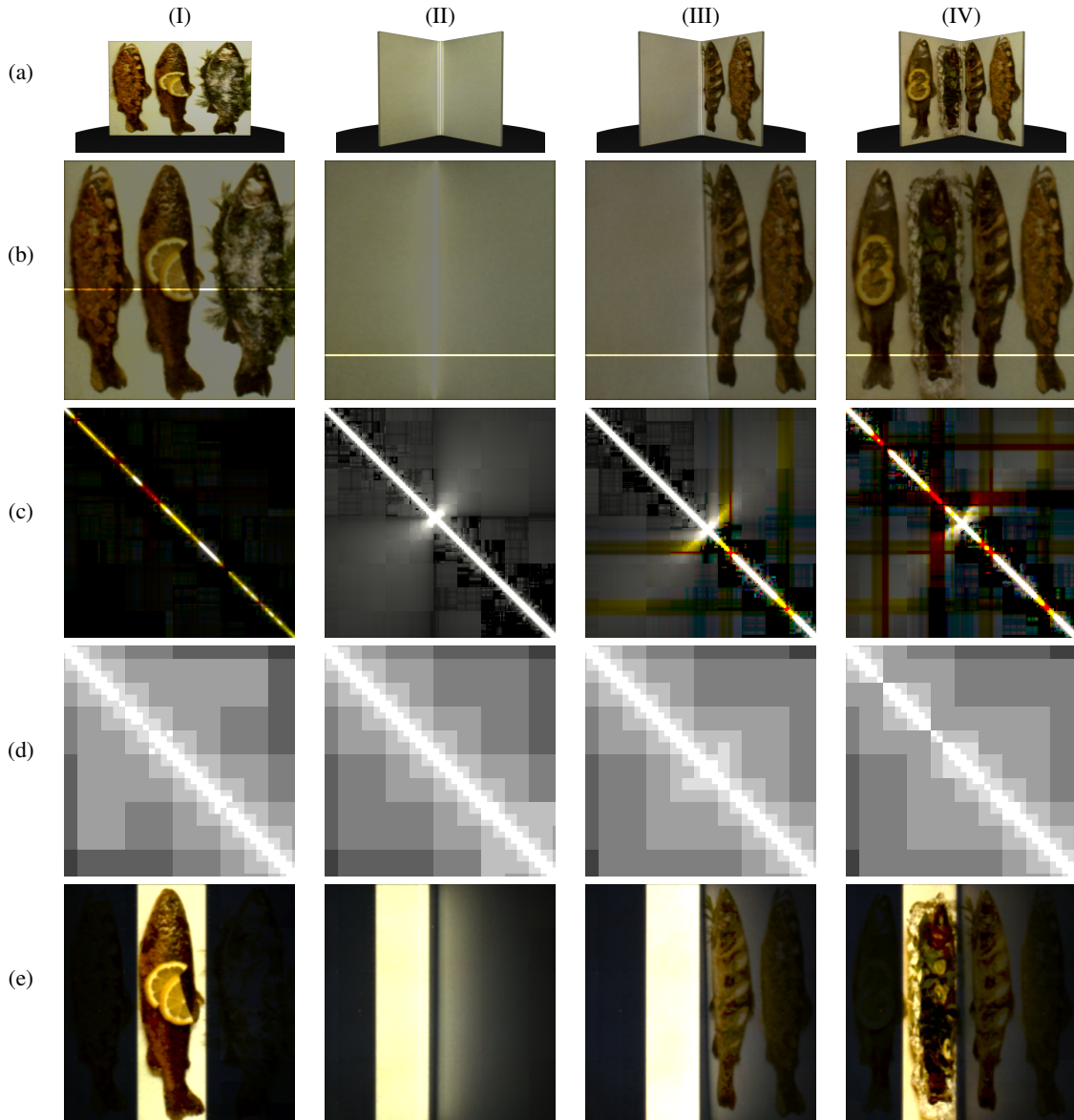Figure 2 illustrates this data-sparseness for a few example

Figure 2: *Understanding the transport matrix. To explain the intrinsic structure of reflectance fields we capture the transport matrix for 4 real scenes shown in row (a) with a coaxial projector/camera pair. The scenes in different columns are: (I) a diffuse textured plane, (II) two diffuse white planes facing each other at an angle, (III) a diffuse white plane facing a diffuse textured plane at an angle, and (IV) two diffuse textured planes facing each other at an angle. Row (b) shows the images rendered from the captured transport matrices under floodlit illumination. A 2D slice of the transport matrix for each configuration is shown in row (c). This slice describes the light transport between every pair of rays that hits the brightened line in row (b). Note that the transport matrix is symmetric in all 4 cases. Since (I) is a flat diffuse plane, there are no secondary bounces and the matrix is diagonal. In (II), (III) and (IV) the diagonal corresponds to the first bounce light and is therefore much brighter than the rest of the matrix. The top-right and bottom-left sub-blocks describe the diffuse-diffuse light transport from pixels on one plane to the other. Note that this is smoothly varying for (II). In case of (III) and (IV), the textured surface introduces high frequencies but these sub-blocks are still data-sparse and can be represented using rank-1 factors. The top-left and bottom-right sub-blocks correspond to the energy from 3rd-order bounces in our scenes. Because this energy is around the noise threshold in our measurements we get noisy readings for these sub-blocks. Row (d) is a visualization of the level in the hierarchy when a block is classified as rank-1. White blocks are leaf nodes, while darker shades of gray progressively represent lower levels in the hierarchy. Finally, row (e) shows the result of relighting the transport matrix with a vertical bar. Note the result of indirect illumination on the right plane in (II), (III) and (IV). Since the left plane is textured in (IV) the indirect illumination is dimmer than in (III). Note that the matrix for a line crossing diagonally through the scene would look similar.*

transport matrices that we have measured, and also demonstrates the local rank-1 approximation. To gain some intuition, let us look at the light transport between two homogeneous untextured planar patches. The light transport between the two is smooth and can be easily factorized. It can be seen in the top-right and bottom-left sub-blocks of the transport matrix for scene (II). Even if the surfaces are textured, it only results in appropriate scaling of either the columns or rows of the transport matrix as shown in (III) and (IV). This will not change the factorization. If a blocker was present between the two patches, it will introduce additional diagonal elements in the matrix sub-blocks. This can only be handled by subdividing the blocks and factorizing at a finer level, as explained in Section 5.

4.2 Symmetry of the Transport Matrix

Capturing full transport matrix is a daunting task. However, $\widetilde{\mathbf{T}}$ is highly redundant, since the radiance along a ray is constant unless the line is blocked. So, if one is willing to stay outside the convex hull of the scene to view it or to illuminate it, then the 5D representation of the light field can be reduced to 4D [LH96, GGSC96, MPDW03]. We will be working with this representation for the rest of the paper. Let us represent the 4D incoming light field by $\mathbf{L}_{in}(\theta)$ and the 4D outgoing light field by $\mathbf{L}_{out}(\theta)$ where θ parameterizes the space of all possible incoming or outgoing directions on a sphere [MPDW03]. The light transport can then be described as:

$$\mathbf{L}_{out} = \mathbf{T}\mathbf{L}_{in} \qquad (5)$$

$\mathbf{T}[i, j]$ represents the amount of light received along outgoing direction θ_i when unit radiance is emitted along incoming direction θ_j. Helmholtz reciprocity [vH56, Ray00] states that the light transport between θ_i and θ_j is equal in both directions, i.e. $\mathbf{T}[i, j] = \mathbf{T}[j, i]$. Therefore, $\mathbf{T}$ is a *symmetric* matrix. We use a coaxial projector/camera setup to ensure symmetry during our acquisitions. Also, note that since we are looking at a subset of rays (4D from 5D), $\mathbf{T}$ is just a sub-block of $\widetilde{\mathbf{T}}$. Therefore, $\mathbf{T}$ is also data-sparse.

5 Data Acquisition

In order to measure $\mathbf{T}$, we use projectors to provide the incoming light field and cameras to measure the outgoing light field. Thus, the full $\mathbf{T}$ matrix can be extremely large, depending on the number of pixels in our acquisition hardware. A naive acquisition scheme would involve scanning through individual projector pixels and concatenating the captured camera images to construct $\mathbf{T}$. This could take days or even months to complete. Therefore, to achieve faster acquisition, we would like to illuminate multiple projector pixels at the same time.

In order to understand how we can illuminate multiple projector pixels at the same time, let us assume that the trans-

port matrix is:

$$\begin{bmatrix} \mathbf{U1} & \mathbf{M} \\ \mathbf{M^T} & \mathbf{U2} \end{bmatrix} = \begin{bmatrix} \mathbf{U1} & \mathbf{0} \\ \mathbf{0} & \mathbf{U2} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{M} \\ \mathbf{M^T} & \mathbf{0} \end{bmatrix} \qquad (6)$$

where $\mathbf{U1}$ and $\mathbf{U2}$ have not been measured yet. Sen et al. [SCG*05] utilized the fact that if $\mathbf{M} = \mathbf{0}$, then the unknown blocks $\mathbf{U1}$ and $\mathbf{U2}$ are radiometrically isolated, i.e. the projector pixels corresponding to $\mathbf{U1}$ do not affect the camera pixels corresponding to $\mathbf{U2}$ and vice versa. Thus, they can illuminate the projector pixels corresponding to $\mathbf{U1}$ and $\mathbf{U2}$ in parallel in such cases. In this work, we observe that if the contents of $\mathbf{M}$ are known but not necessarily $\mathbf{0}$, we can still radiometrically isolate $\mathbf{U1}$ and $\mathbf{U2}$ by subtracting the contribution of known $\mathbf{M}$ from the captured images. The RHS of eq. (6) should make this clear. We use this fact to illuminate the projector pixels corresponding to $\mathbf{U1}$ and $\mathbf{U2}$ in parallel when $\mathbf{M}$ is known.

Now, consider a sub-block $\mathbf{M}$ of the transport matrix that is data-sparse and can be approximated by a rank-1 factorization. We can obtain this rank-1 factorization by just capturing two images. An image captured by the camera is the sum of the columns in the transport matrix corresponding to the pixels illuminated by the projector. Because of the symmetry of the transport matrix, the image is also the sum of corresponding rows in the matrix. Therefore, by just shining two projector patterns, $\mathbf{p_c}$ and $\mathbf{p_r}$, we can capture images such that one provides the sum of the columns, $\mathbf{c}$ and the other provides the sum of the rows, $\mathbf{r}$ of $\mathbf{M}$ ($\mathbf{c} = \mathbf{M}\mathbf{p_c}$ and $\mathbf{r} = \mathbf{M^T}\mathbf{p_r}$). A tensor product of $\mathbf{c}$ and $\mathbf{r}$ directly provides a rank-1 factorization for $\mathbf{M}$. Thus the whole sub-block can be constructed using just two illumination patterns. This is the key idea behind our algorithm. The algorithm tries to find sub-blocks in $\mathbf{T}$ that can be represented as a rank-1 approximation by a hierarchical subdivision strategy. Once measured, these sub-blocks can be used to parallelize the acquisition as described above. For an effective hierarchical acquisition we need an efficient data structure to represent $\mathbf{T}$. We will describe our data structure now.

5.1 Hierarchical Tensors for Representing T

We introduce a new data structure called *hierarchical tensors* to represent data-sparse light transport. Hierarchical tensors are a generalization of hierarchical matrices (or $\mathcal{H}$-matrices), which have been introduced by Hackbush [Hac99] in the applied mathematics community to represent arbitrary data-sparse matrices. The key idea behind $\mathcal{H}$-matrices is that a data-sparse matrix can be represented by an adaptive subdivision structure and a low-rank approximation for each node. At each level of the hierarchy, sub-blocks in the matrix are uniformly subdivided into 4 children (as in a quadtree). If a sub-block at any level in the tree can be represented by a low-rank approximation, then it is not subdivided any further. Thus, a leaf node in the tree contains a low-rank approximation for the corresponding sub-block, which reduces to just a scalar value at the finest level in the hierarchy.
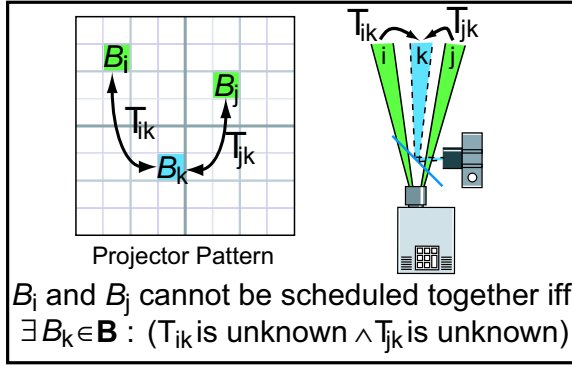
Figure 3: *Determining block scheduling.* *Two blocks B_i and B_j cannot be scheduled in the same frame if and only if, $\exists B_k \in \mathbf{B}$, such that the light transports T_{ik} and T_{jk} are both unknown. This is because upon illuminating B_i and B_j simultaneously, the block B_k in the camera will measure the combined contribution of both T_{ik} and T_{jk}. Since both of these are unknown at this point there is no way to separate them out.*

Consider the 4D reflectance field that describes the light transport for a single projector/camera pair. We have a 2D image representing the illumination pattern and a resulting 2D image captured by the camera. The connecting light transport can therefore be represented by a 4th-order tensor. One can alternatively flatten out the 2D image into a vector, but that would destroy the spatial coherency present in a 2D image [WWS*05]. To preserve coherency we represent the light transport by a 4th-order hierarchical tensor. A node in the 4th-order hierarchical tensor is divided into 16 children at each level of the hierarchy. Thus, we call the hierarchical representation for a 4th-order tensor, a *sedectree* (derived from *sedecim*, Latin equivalent of 16). Additionally, we use a rank-1 approximation for representing data-sparseness in the leaf nodes of the hierarchical tensor. This means that a leaf node is represented by a tensor product of two 2D images, one from the camera side and the other from the projector side.

5.2 Hierarchical Acquisition Scheme

Our acquisition algorithm follows the structure of the hierarchical tensor described in the previous section. At each level of the hierarchy we illuminate the scene with a few projector patterns. We use the captured images to decide which nodes of the tensor in the previous level of hierarchy are rank-1. Once a node has been determined to be rank-1, we do not subdivide it any further as its entries are known. The nodes which fail the rank-1 test are subdivided and scheduled for investigation during the next iteration. The whole process is repeated until we reach the pixel level. We initiate the acquisition by illuminating with a floodlit projector pattern. The captured floodlit image provides a possible rank-1 factoriza-
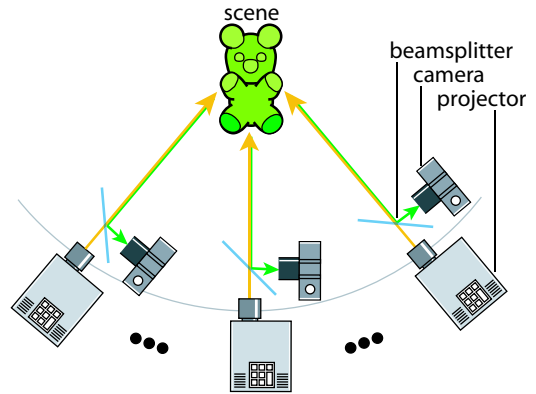

Figure 4: *Schematic of symmetric photography setup.* *A coaxial array of projectors and cameras provides an ideal setup for symmetric photography. The projector array illuminates the scene with an incoming light field. Since the setup is coaxial, the camera array measures the corresponding outgoing light field.*

tion of the root node of the hierarchical tensor. The root node is scheduled for investigation in the first iteration.

For each level, the first step is to decide what illumination patterns to use. In order to speed-up our acquisition, we need to minimize the number of these patterns. To achieve this, our algorithm must determine the set of projector blocks which can be illuminated in the same pattern. To determine this, we divide each scheduled node into 16 children and the 4 blocks in the projector corresponding to this subdivision are accumulated in a list $\mathbf{B} = \{B_1, B_2, ..., B_n\}$. Figure 3 describes the condition when two blocks B_i and B_j cannot be scheduled in parallel. It can be written as the following lemma:

Lemma: Two blocks B_i and B_j cannot be scheduled together if and only if, $\exists B_k \in \mathbf{B}$, such that both T_{ik} and T_{jk} are not known.

Since the direct light transport T_{ii} is not known until the bottom level in the hierarchy, any two blocks B_i and B_j for which T_{ij} is not known cannot be scheduled in parallel. For all such possible block pairs for which the light transport has not been measured yet, let us construct a set $\mathbf{C} = \{(B_i, B_j) : B_i, B_j \in \mathbf{B}\}$. Given these two sets, we define an undirected graph $\mathbf{G} = (\mathbf{B}, \mathbf{C})$, where $\mathbf{B}$ is the set of vertices in the graph and $\mathbf{C}$ is the set of edges. Thus, the vertices in the graph have an edge between them if the light transport between the corresponding blocks is not known. In this graph, any two vertices B_i and B_j which do not have an edge between them but have a direct edge with a common block B_k (as shown in Figure 3) also satisfy the lemma. Therefore, we cannot schedule them in parallel. Such blocks correspond to vertices at a distance two from each other in our graph $\mathbf{G}$. In order to capture these blocks as direct edges in a graph, we construct another graph $\mathbf{G}^2$ which is the square of graph $\mathbf{G}$
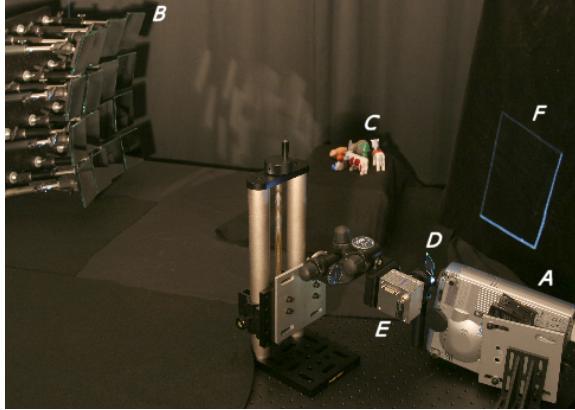
Figure 5: *Coaxial setup for capturing 8D reflectance fields. A pattern loaded into projector at A illuminates a 4×4 array of planar mirrors at B. This provides us with 16 virtual projectors which illuminate our scene at C. The light that returns from the scene is diverted by a beam-splitter at D towards a camera at E. Any stray light reflected from the beam-splitter lands in a light trap at F. The camera used is an Imperx IPX-1M48-L (984×1000 pixels) and the projector is a Mitsubishi XD60U (1024×768 pixels). The setup is computer controlled, and we capture HDR images every 2 seconds.*
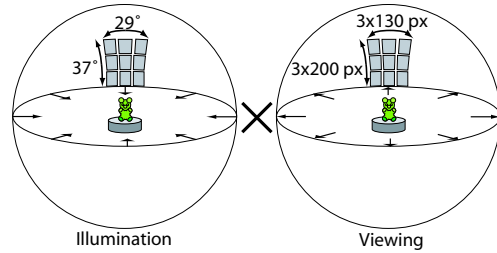

Figure 6: *Region of the sphere sampled by our setup. Our setup spans an angular resolution of $37° \times 29°$ on the sphere both for the illumination and view directions. The spatial resolution in each view is 130×200 pixels. This accounts for about 2% of the total rays in the light field.*

[Har01]. The square of a graph contains an edge between any two vertices which are at most distance two away from each other in the original graph. Thus, in the graph $\mathbf{G}^2$, any two vertices which are not connected can be scheduled together. We use a graph coloring algorithm on $\mathbf{G}^2$ to obtain subsets of $\mathbf{B}$ which can be illuminated in parallel. Once the images have been acquired, the known intra-block light transport is subtracted out for the blocks that were scheduled in the same frame.

In the next step, we use these measurements to test if the tensor nodes in the previous level of the hierarchy can be factorized using rank-1 approximation. We have a current rank-1 approximation for each node from the previous level in the hierarchy. The 8 measured images, corresponding to 4 blocks from the projector side and 4 blocks from the camera side of a node, are used as test cases to validate the current approximation. This is done by rendering estimate images for these blocks using the current rank-1 approximation. The estimated images are compared against the corresponding measured images and an RMS error is calculated for the node. A low RMS error indicates our estimates are as good as our measurements and we declare the node as rank-1 and stop any further subdivision on this node. If on the other hand the RMS error is high, the 16 children we have measured become the new nodes. The 4 images from the projector side and the 4 images from the camera side are used to construct the 16 (4×4) rank-1 estimates for them.

These nodes are scheduled for investigation in the next iteration.

A tensor node containing just a scalar value is trivially rank-1. Therefore, the whole process terminates when the size of the projector block reduces to a single pixel. Upon finishing, the scheme directly returns the hierarchical tensor for the reflectance field of the scene.

5.3 Setup and Pre-processing

In order to experimentally validate our ideas we need an acquisition system that is capable of simultaneously emitting and capturing along each ray in the light field. This suggests having a coaxial array of cameras and projectors. Figure 4 shows the schematic of such a setup. Our actual physical implementation is built using a single projector, a single camera, a beam-splitter and an array of planar mirrors. The projector and the camera are mounted coaxially using the beam splitter on an optical bench as shown in Figure 5, and the mirror array divides the projector/camera pixels into 9 coaxial pairs. Once the optical system has been mounted it needs to be calibrated. First, the center of projection of the camera and projector is aligned. The next task is to find the per pixel mapping between the projector and camera pixels. We use a calibration scheme similar to that used by Han and Perlin [HP03] and Levoy et al. [LCV*04] in their setup to find this mapping. Figure 6 illustrates the angular and spatial resolution of reflectance fields captured using out setup.

The dynamic range of the scenes that we capture can be very high. This is because the light transport contains not only the high energy direct bounce effects but also very low energy secondary bounce effects. In order to capture this range completely, we take multiple images of the scene and combine them into a single high dynamic range image [DM97,RBS99]. Additionally, before combining the images for HDR, we subtract the black level of the projector from our images. This accounts for the stray light coming from the projector even when it is shining a completely black image.
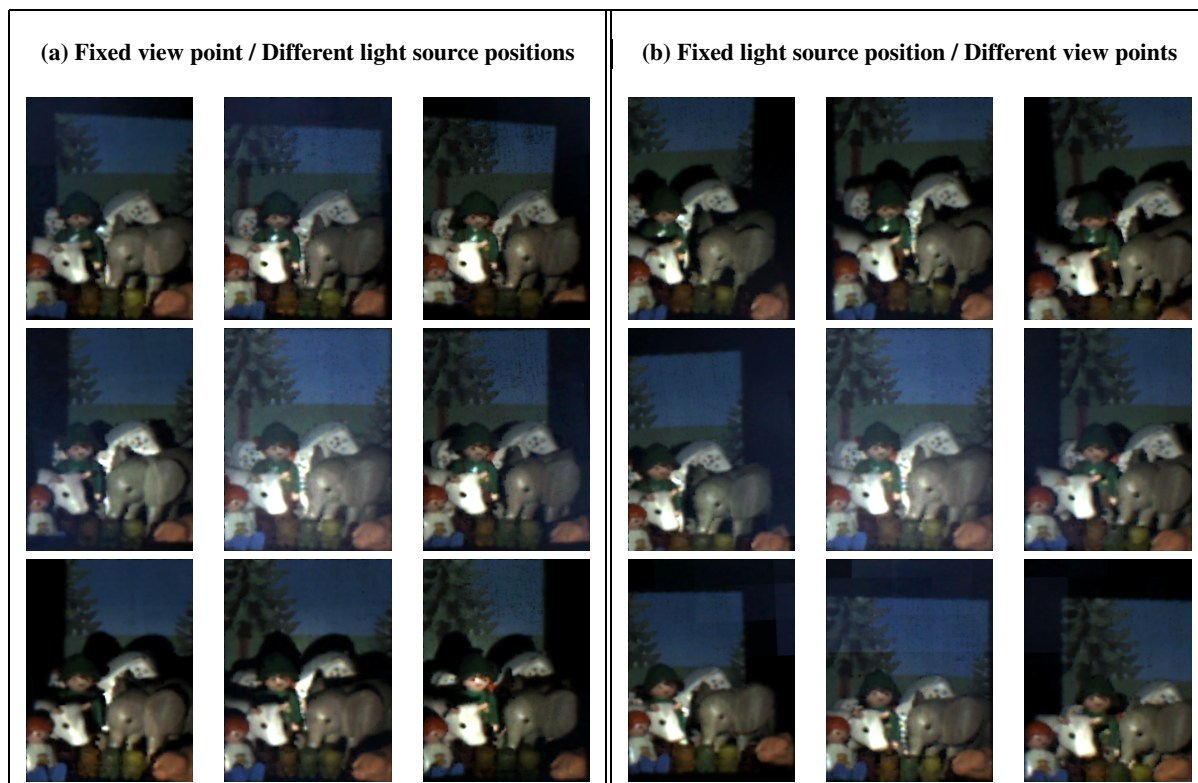
(a) Fixed view point / Different light source positions	(b) Fixed light source position / Different view points

Figure 7: *8D reflectance field of an example scene. This reflectance field was captured using the setup described in Figure 5. A 3 × 3 grid of mirrors was used. In (a) we see images rendered from the viewpoint at the center of the grid with illumination coming from 9 different locations on the grid. Note that the shadows move appropriately depending upon the direction of incident light. (b) shows the images rendered from 9 different viewpoints on the grid with the illumination coming from the center. In this case one can notice the change in parallax with the viewpoint. Note that none of these images were directly captured during our acquisition. The center image in each set looks slightly brighter because the viewpoint and lighting are coincident in this case.*

Also, upon illuminating the scene with individual projector pixels, we notice that the captured images appear darker and have a significantly reduced contrast. This is because an individual projector pixel would be illuminating very few pixels on the Bayer mosaiced sensor of the camera, leading to an error upon interpolation during demosaicing. This problem was also noticed by Sen et al. [SCG*05]. To remove these artifacts, we employ a solution similar to theirs, i.e. the final images are renormalized by forcing the captured images to sum up to the floodlit image.

6 Results

We capture reflectance fields of several scenes. For reference, Table 1 provides statistics (size, time and number of patterns required for acquisition) for each of these datasets.

In Figure 2, we present the results of our measurement for four simple scenes consisting of planes. This experiment has been designed to elucidate the structure of the **T** matrix. A coaxial projector/camera pair is directly aimed at the scene in this case. The image resolution is 310×350 pixels. Note the storage, time and number of patterns required for the four

scenes (listed in Table 1). A brute-force scan, in which each projector pixel is illuminated individually, to acquire these **T** matrices would take at least 100 times more images. Also, since the energy in the light after an indirect bounce is low, the camera would have to be exposed for a longer time interval to achieve good SNR during brute-force scanning. On the other hand, in our scheme the indirect bounce light transport is resolved earlier in the hierarchy, see rows (c) and (d) in Figure 2. At higher levels of the hierarchy, we are illuminating with bigger projector blocks (and hence throwing more light into the scene than just from a single pixel), therefore we are able to get good SNR even with small exposure times. Also, note that the high frequency of the textures does not affect the data-sparseness of reflectance fields. The hierarchical subdivision follows almost the same strategy in all four cases as visualized in row (d). In row (e), we show the results of relighting the scene with a vertical bar. The smooth glow from one plane to the other in column (II), (III) and (IV) shows that we have measured the indirect bounce correctly.

Figure 1 demonstrates that our technique works well for

	Symmetric photography			Brute-force
SCENE Fig.	SIZE (MB)	TIME (min)	PATTERNS (#)	PATTERNS (#)
1	337	151	2,417	91,176
2(I)	255	44	809	108,500
2(II)	371	70	1,085	108,500
2(III)	334	65	1,081	108,500
2(IV)	274	46	841	108,500
7	1,470	484	3,368	233,657

Table 1: *Table of relevant data (size, time and number of patterns) for different example scenes captured using our technique. Note that our algorithm requires about 2 orders of magnitude fewer patterns than the brute-force scan.*

acquiring the reflectance fields of highly sub-surface scattering objects. The image (240×340 pixels) reconstructed from relighting with a spatially varying illumination pattern (see Figure 1(b)) is validated against the ground-truth image (see Figure 1(c)). We also demonstrate the result of reconstructing at different levels of the hierarchical tensor for this scene in Figure 11. This figure also explains the difference between our hierarchical tensor representation and a wavelet based representation.

Figure 7 shows the result of an 8D reflectance field acquired using our setup. The captured reflectance field can be used to view the scene from multiple positions (see Figure 7(b)) and also to relight the scene from multiple directions (see Figure 7(a)). The resolution of the reflectance field for this example is about $3 \times 3 \times 130 \times 200 \times 3 \times 3 \times 130 \times 200$. The total size of this dataset would be 610 GB if three 32-bit floats were used for each entry in the transport matrix. Our hierarchical tensor representation compresses it to 1.47 GB. A brute force approach would require 233,657 images to capture it. Our algorithm only needs 3,368 HDR images and takes around 8 hours to complete. In our current implementation, the processing time is comparable to the actual image capture time. We believe that the acquisition times can be reduced even further by implementing a parallelized version of our algorithm. Rendering a relit image from our datasets is efficient and takes less than a second on a typical workstation.

6.1 Symmetric vs. Dual Photography

It is instructive to compare the symmetric photography technique against dual photography [SCG*05]. Dual photography reduces the acquisition time by exploiting only sparseness (the fact that there are regions in a scene that are radiometrically independent of each other). These regions are detected and measured in parallel in dual photography. However, for a scene with many inter-reflections or sub-surface scattering, such regions are few and the technique performs poorly. In order to resolve the transport at full resolution, the technique would reduce to brute-force scanning for such scenes. Illuminating with single pixel for observing multiple scattering events has inherent SNR problems because indi-
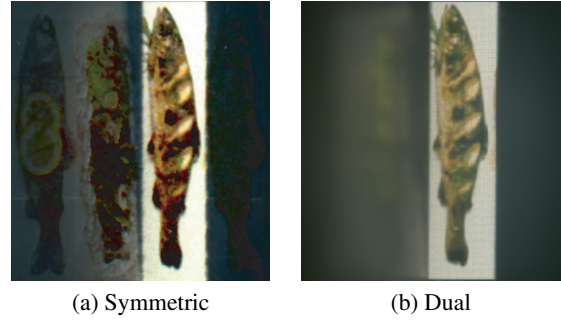

(a) Symmetric (b) Dual

Figure 8: *Symmetric vs. Dual Photography. The figure illustrates the strength of symmetric photography technique (a) when compared against the dual photography technique (b) of Sen et al. [SCG*05]. The setup is similar to the book example of Figure 2 (IV). In both cases, the right half of the book is synthetically relit using the transport matrices captured by the respective techniques. Note that in the case of symmetric photography (a), the high frequencies in the left half of the book are faithfully resolved while in dual photography (b), the frequencies cannot be resolved and just appear as a blur. The light transport for (a) was acquired using 841 images while that for (b) was acquired using 7382 images.*

rect bounce light transport coefficients could be extremely low. The measurement system, which is limited by the black level of the projector and dark noise of the camera cannot pick up such low values. The scheme therefore stops refining at a higher level in the hierarchy and measures only a coarse approximation of the indirect bounce light transport. This essentially results in a low-frequency approximation for indirect bounce light transport. The comparison of the two techniques in Figure 8 confirms this behavior. Since symmetric photography is probing the matrix from both sides, the high frequencies in indirect bounce light transport are still resolved whereas dual photography can only produce a low frequency approximation of the same. Furthermore, while symmetric photography took just 841 HDR images for this scene, dual photography required 7382 HDR images.

Finally, Figure 9 illustrates the relative percentage of rank-1 vs empty leaf nodes at various levels of the hierarchy for the transport matrices that we have captured. The empty leaf nodes correspond to sparse regions of the matrix while the rank-1 leaf nodes correspond to data-sparse regions of the matrix. While dual photography only exploits sparseness and hence culls away only empty leaf nodes at a particular level, symmetric photography exploits both data-sparseness and sparseness and culls away both rank-1 and empty leaf nodes. Note that between levels 4 and 9, there is a significant fraction of rank-1 nodes which are culled away by symmetric photography in addition to empty leaf nodes. This results in large reduction of nodes that still have to be investigated and results in a significantly faster acquisition as compared to dual photography.
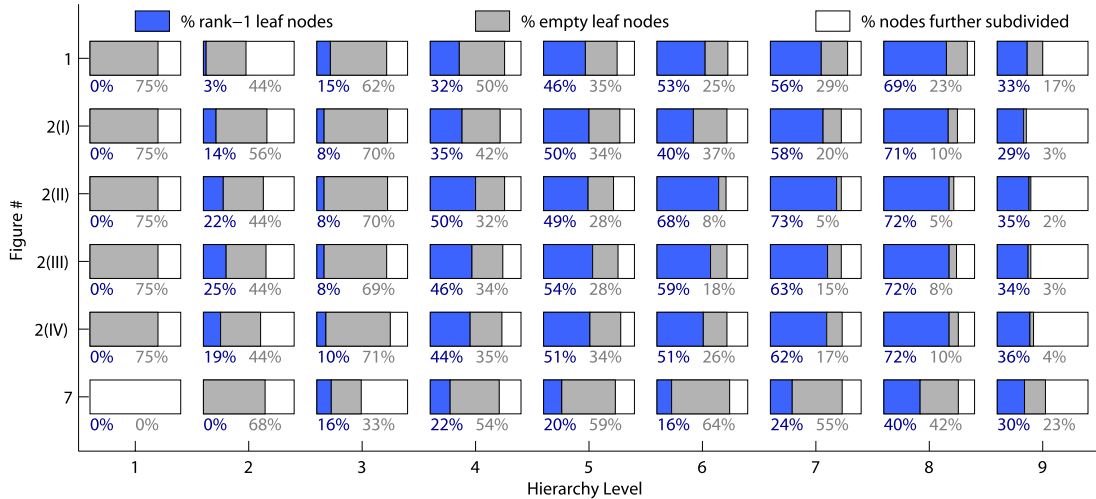
Figure 9: *Comparison of rank-1 vs empty leaf nodes. The figure empirically compares the percentage of rank-1 vs. empty leaf nodes in the hierarchical tensor at different levels of the hierarchy for various scenes captured using our acquisition scheme. The blue area in each bar represents the percentage of rank-1 nodes while the gray area corresponds to the percentage of empty nodes. The white area represents the nodes which are subdivided at next level. Note that at levels 4, 5, 6, 7, 8 and 9 a significant fraction of leaf nodes are rank-1. Also note that for Figures 2(I) and 2(II), at levels 6, 7, and 8, there are far more empty nodes in 2(I) than in 2(II). This is what we expected as the transport matrix for 2(I) is sparser than that for 2(II).*

7 Discussion and Conclusions

In this paper we have presented a framework for acquiring 8D reflectance fields. The method is based on the observation that reflectance fields are data-sparse. We exploit the data-sparseness to represent the transport matrix by local rank-1 approximations. The symmetry of the light transport allows us to measure these local rank-1 factorizations efficiently, as we can obtain measurements corresponding to both rows and columns of the transport matrix simultaneously. We have also introduced a new data structure called a hierarchical tensor that can represent these local low-rank approximations efficiently. Based on these observations we have developed a hierarchical acquisition algorithm, which looks for regions of data-sparseness in the matrix. Once a data-sparse region has been measured we can use it to parallelize our acquisition resulting in tremendous speedup.

There are limitations in our current acquisition setup (Figure 5) that can corrupt our measurements. To get a coaxial setup we use a beam-splitter. Although we use a 1mm thin plate beam-splitter, it produces the slight double images inherent to plate beam-splitters. This, along with the light reflected back off the light trap, reduces the SNR in our measurements. The symmetry of our approach requires projector and camera to be pixel aligned. Any slight misalignment adds to the measurement noise. Cameras and projectors can also have different optical properties. This can introduce non-symmetries such as lens flare, resulting in artifacts in our reconstructed images (see Figure 10).

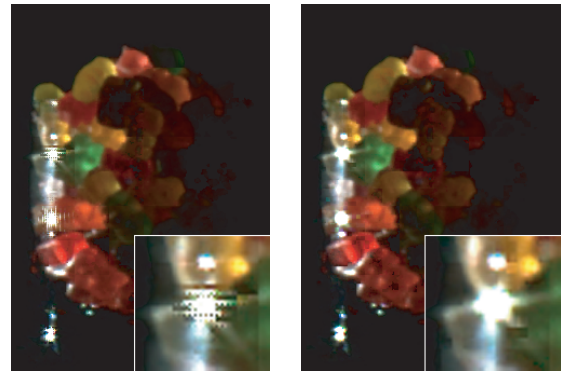By way of improvements, in order to keep our implementation simple, we use a 4th order hierarchical tensor. This

Figure 10: *Artifacts due to non-symmetry in measurement. The lens flare around the highlights (right image) is caused by the aperture in the camera. Since this effect does not occur in the incident illumination from the projector, the measurements are non-symmetric. Applying a strong threshold for the rank-1 test subdivides the region very finely and produces a corrupted result in the area of the highlights (left image). If the inconsistencies in measurement are stored at a higher subdivision level by choosing a looser threshold for rank-1 test, these artifacts are less noticeable (right image) .*

means that we are flattening out 2 of the 4 dimensions of the light field, thereby not exploiting the full coherency in the data. An implementation based on 8th order tensor should be able to exploit it and make the acquisition more efficient.

Since we use a 3×3 array of planar mirrors, the resolution of our incoming and outgoing light fields is low. Therefore, the reflectance fields that we can capture are sparse

and incomplete. Regarding sparseness, techniques have been proposed for interpolating slices of the reflectance fields, both from the view direction [CW93] and from the illumination direction [CL05], but the problem of interpolating reflectance fields is still open. By applying flow-based techniques to the transport matrix, one should be able to create more densely sampled reflectance fields. One can also imagine directly sampling incoming and outgoing light fields more densely by replacing the small number of planar mirrors with an array of lenslets or mirrorlets [UWH*03]. This will increase the number of viewpoints in the light field but at the cost of image resolution.

Regarding completeness, if the setup used for capturing Figure 7 is replicated to cover the whole sphere, then extrapolating from the numbers in Table 1, we would expect the transport matrix to be around 75 GB in size. Acquisition would currently take roughly 2 weeks. Note that although it is not practical to build such a setup now, faster processing and the use of an HDR video camera could reduce this time significantly in the future.

Finally, although we introduce the hierarchical tensor as a data structure for storing reflectance fields, the concept may have implications for other high dimensional data-sparse datasets as well. The hierarchical representation also has some other benefits. It provides constant time access to the data during evaluation or rendering. At the same time it maintains the spatial coherency in the data, making it attractive for parallel computation.

References

[BGH03] BÖRM S., GRAYSEDYCK L., HACKBUSH W.: Introduction to Hierarchical Matrices with Applications. *Engineering Analysis with Boundary Elements 27*, 5 (2003), 405–422.

[CL05] CHEN B., LENSCH H. P. A.: Light Source Interpolation for Sparsely Sampled Reflectance Fields. In *Vision Modeling and Visualization (VMV'05)* (2005), pp. 461–468.

[CW93] CHEN S. E., WILLIAMS L.: View interpolation for Image Synthesis. In *SIGGRAPH '93* (1993), pp. 279–288.

[DHS*05] DURAND F., HOLZSCHUCH N., SOLER C., CHAN E., SILLION F. X.: A Frequency Analysis of Light Transport. In *SIGGRAPH '05* (2005), pp. 1115–1126.

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the Reflectance Field of a Human Face. In *SIGGRAPH '00* (2000), pp. 145–156.

[DM97] DEBEVEC P., MALIK J.: Recovering High Dynamic Range Radiance Maps from Photographs. In *SIGGRAPH '97* (1997), pp. 369–378.

[Ger36] GERSHUN A.: The Light Field, Moscow. *Journal of Mathematics and Physics (1939)* (1936). Vol. XVIII, MIT. Translated by P. Moon and G. Timoshenko.

[GGSC96] GORTLER S. J., GRZESZCZUK R., SZELISKI R., COHEN M. F.: The Lumigraph. In *SIGGRAPH '96* (1996), pp. 43–54.

[GLL*04] GOESELE M., LENSCH H. P. A., LANG J., FUCHS C., SEIDEL H.-P.: DISCO: Acquisition of Translucent Objects. In *SIGGRAPH '04* (2004), pp. 835–844.

[Hac99] HACKBUSCH W.: A Sparse Matrix Arithmetic based on $\mathcal{H}$-Matrices. Part I: Introduction to $\mathcal{H}$-matrices. *Computing 62*, 2 (1999), 89–108.

[Har01] HARARY F.: *Graph Theory*. Narosa Publishing House, 2001.

[Hec91] HECKBERT P. S.: *Simulating Global Illumination Using Adaptive Meshing*. PhD thesis, University of California, Berkeley, 1991.

[HED05] HAWKINS T., EINARSSON P., DEBEVEC P.: A Dual Light Stage. In *Eurographics Symposium on Rendering (EGSR '05)* (2005), pp. 91–98.

[HP03] HAN J. Y., PERLIN K.: Measuring Bidirectional Texture Reflectance with a Kaleidoscope. In *SIGGRAPH '03* (2003), pp. 741–748.

[Kaj86] KAJIYA J. T.: The Rendering Equation. In *SIGGRAPH '86* (1986), pp. 143–150.

[KM99] KAUTZ J., MCCOOL M. D.: Interactive Rendering with Arbitrary BRDFs using Separable Approximations. In *Eurographics Workshop on Rendering (EGWR '99)* (1999), pp. 247–260.

[LCV*04] LEVOY M., CHEN B., VAISH V., HOROWITZ M., MCDOWALL I., BOLAS M.: Synthetic Aperture Confocal Imaging. In *SIGGRAPH '04* (2004), pp. 825–834.

[LH96] LEVOY M., HANRAHAN P.: Light Field Rendering. In *SIGGRAPH '96* (1996), pp. 31–42.

[LK02] LATTA L., KOLB A.: Homomorphic Factorization of BRDF-based Lighting Computation. In *SIGGRAPH '02* (2002), pp. 509–516.

[MAA01] MCCOOL M. D., ANG J., AHMAD A.: Homomorphic Factorization of BRDFs for High-Performance Rendering. In *SIGGRAPH '01* (2001), pp. 171–178.
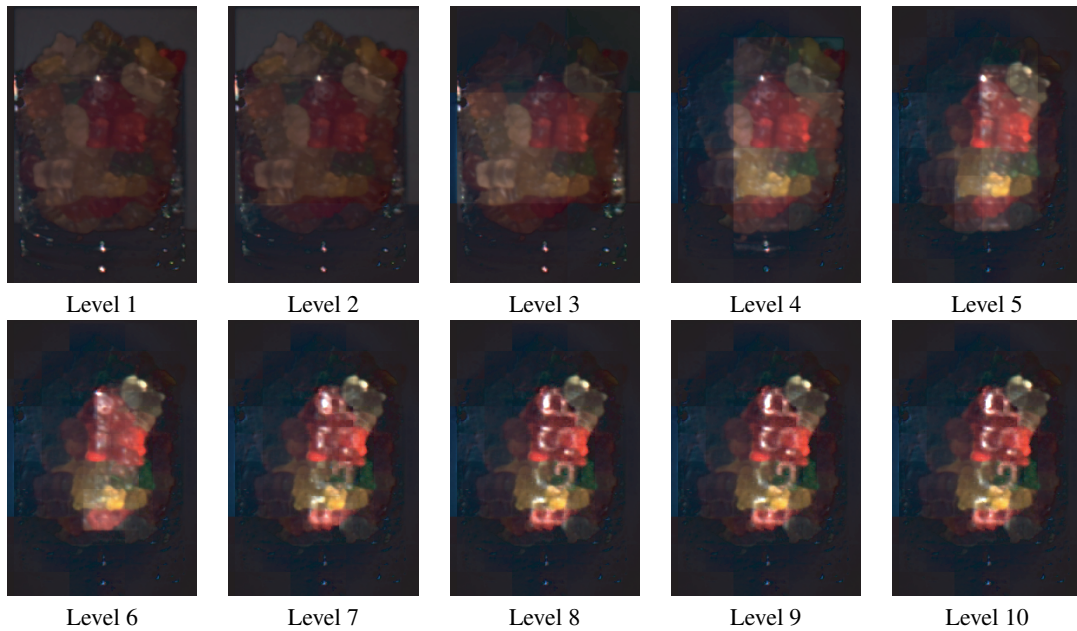
[MGW01] MALZBENDER T., GELB D., WOLTERS H.:

Figure 11: *Reconstruction results for different levels of hierarchy. This example illustrates the relighting of the reflectance field of gummy bears scene (Figure 1) with illumination pattern used in Figure 1(a), if the acquisition was stopped at different levels of the hierarchy. Note that at every level, we still get a full resolution image. This is because we are approximating a node in the hierarchy as a tensor product of two 2-D images. Therefore, we sill have a measurement for each pixel in the image, though scaled incorrectly. This is different from wavelet based approaches where a single scalar value is assigned for a node in the hierarchy implying lower resolution in the image at lower levels. Note that at any level, the energy of the projected pattern is distributed over the whole block that it is illuminating. This is clear from the intensity variation among blocks, especially in the images at levels 3, 4, and 5.*

Polynomial Texture Maps. In *SIGGRAPH '01* (2001), pp. 519–528.

[MLP04] MATUSIK W., LOPER M., PFISTER H.: Progressively - Refined Reflectance Functions for Natural Illumination. In *Eurographics Symposium on Rendering (EGSR '04)* (2004), pp. 299–308.

[MPDW03] MASSELUS V., PEERS P., DUTRÉ P., WILLEMS Y. D.: Relighting with 4D Incident Light Fields. In *SIGGRAPH '03* (2003), pp. 613–620.

[PD03] PEERS P., DUTRÉ P.: Wavelet Environment Matting. In *Eurographics Symposium on Rendering (EGSR '03)* (2003), pp. 157–166.

[Ray00] RAYLEIGH J. W. S. B.: On the Law of Reciprocity in Diffuse Reflexion. *Philosophical Magazine 49* (1900), 324–325.

[RBS99] ROBERTSON M. A., BORMAN S., STEVENSON R. L.: Dynamic Range Improvement through Multiple Exposures. In *IEEE Intl. Conference on Image Processing (ICIP'99)* (1999), pp. 159–163.

[RH01] RAMAMOORTHI R., HANRAHAN P.: A Signal-Processing Framework for Inverse Rendering. In *SIGGRAPH '01* (2001), pp. 117–128.

[SCG*05] SEN P., CHEN B., GARG G., MARSCHNER

S. R., HOROWITZ M., LEVOY M., LENSCH H. P. A.: Dual Photography. In *SIGGRAPH '05* (2005), pp. 745–755.

[SMK05] SEITZ S. M., MATSUSHITA Y., KUTULAKOS K. N.: A Theory of Inverse Light Transport. In *IEEE Intl. Conference on Computer Vision (ICCV '05)* (2005), pp. 1440–1447.

[UWH*03] UNGER J., WENGER A., HAWKINS T., GARDNER A., DEBEVEC P.: Capturing and Rendering with Incident Light Fields. In *Eurographics Symposium on Rendering (EGSR '03)* (2003), pp. 141–149.

[vH56] VON HELMHOLTZ H.: *Treatise on Physiological Optics (1925)*. The Optical Society of America, 1856. Electronic edition (2001): University of Pennsylvania http://www.psych.upenn.edu/backuslab/helmholtz.

[VT04] VASILESCU M. A. O., TERZOPOULOS D.: TensorTextures: Multilinear Image-Based Rendering. In *SIGGRAPH '04* (2004), pp. 336–342.

[WWS*05] WANG H., WU Q., SHI L., YU Y., AHUJA N.: Out-of-Core Tensor Approximation of Multi-Dimensional Matrices of Visual Data. In *SIGGRAPH '05* (2005), pp. 527–535.