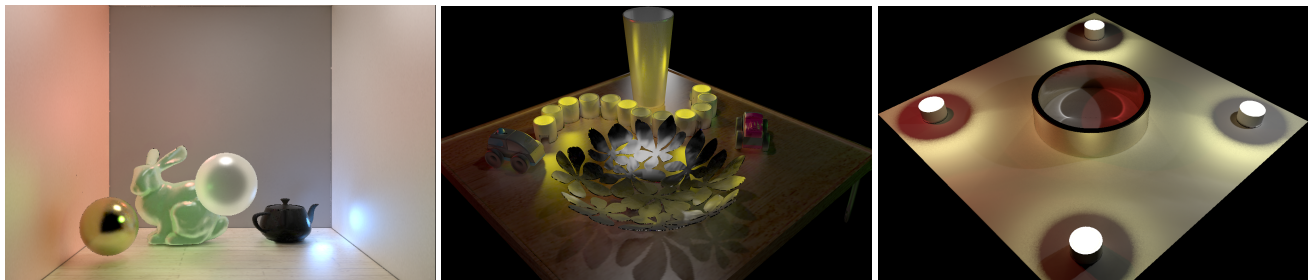


Learning Based Compression of Surface Light Fields for Real-time Rendering of Global Illumination Scenes

Ehsan Miandji *

Joel Kronander
C-Research, Linköping University

Jonas Unger



(a) 34 Frames Per Second

(b) 38 Frames Per Second

(c) 26 Frames Per Second

Figure 1: Rendering results using CEOB for three scenes. (a) Materials: Translucent, measured gold BRDF and glossy. Light sources: one area (disk) and 3 small spherical. (b) Materials: measured silver BRDF and glossy materials with different roughness. Light sources: 10 small spherical. (c) Materials: the ring is measured gold BRDF with a perfect specular component to enable caustics. The rest of the scene is diffuse. Light sources: 6 small spherical. Real-time renderings and comparisons to other methods can be found in the supplementary video.

Abstract

We present an algorithm for compression and real-time rendering of surface light fields (SLF) encoding the visual appearance of objects in static scenes with high frequency variations. We apply a non-local clustering in order to exploit spatial coherence in the SLF data. To efficiently encode the data in each cluster, we introduce a learning based approach, Clustered Exemplar Orthogonal Bases (CEOB), which trains a compact dictionary of orthogonal basis pairs, enabling efficient sparse projection of the SLF data. In addition, we discuss the application of the traditional Clustered Principal Component Analysis (CPCA) on SLF data, and show that in most cases, CEOB outperforms CPCA, K-SVD and spherical harmonics in terms of memory footprint, rendering performance and reconstruction quality. Our method enables efficient reconstruction and real-time rendering of scenes with complex materials and light sources, not possible to render in real-time using previous methods.

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

Keywords: real time rendering, global illumination, compression

1 Introduction

Real-time rendering of complex scenes with full global illumination is one of the main goals of computer graphics. Although many real-time approximations for realistic rendering have been proposed, this is still an unsolved problem. This is especially true for applications like interactive product visualization, where predictive results without any artifacts is an absolute requirement. A popular approach is to use appearance capture methods, [Chen et al. 2002;

Unger et al. 2008; Löw et al. 2009], or to pre-compute detailed surface light fields (SLF) describing the appearance of virtual objects using full global illumination. A key problem, however, is that an SLF data set exhibit a very large memory footprint (often in the order of several GBs per object in the scene) and does not easily lend itself to real-time rendering.

In this paper we present a learning based algorithm for efficient compression of appearance information encoded as SLFs, which enables real-time rendering of virtual scenes exhibiting complex materials and light source configurations with full global illumination. Our algorithm is built upon the following requirements: the complexity of off-line pre-computations should not exceed that of standard methods for pre-computed radiance transfer (PRT) [Sloan et al. 2003; Löw et al. 2009], general high frequency scenes and materials must be handled, the technique must handle arbitrary light source configurations, and finally support real-time rendering.

Surface light fields - An SLF is a 4D function $f(u, v, \phi, \theta)$ that describes the visual appearance of an object as seen from any vantage point. For each point (u, v) on the object surface, it describes the distribution of outgoing, scattered radiance in any direction (ϕ, θ) on the hemisphere around the surface normal, see Figure 2.

Throughout the paper, we will refer to the angular radiance distribution at a single point p_i as a *hemispherical radiance distribution function* (HRDF). An HRDF is defined within the tangent space at the corresponding spatial sample point on the surface. We will denote the discretized SLF function as a set of HRDF matrices H_i , ($i = 1 \dots n$) of size $m_1 \times m_2$, each sampled at a surface location p_i , where n is the number of spatial sample points. The size of an HRDF matrix defines the angular resolution. The placement of spatial sample points p_i is arbitrary. In our implementation, we either place them regularly on a grid in the parametric space (u, v) , or irregularly by placing an HRDF at each vertex of polygonal objects. Each HRDF is represented using concentric mapping [Shirley and Chiu 1997], preserving area in the mapping from a unit square to a unit hemisphere, and vice versa. We evaluate and store the outgoing radiance at each point p_i and along $m_1 \times m_2$ directions of the HRDF. Our pre-computation stage is currently limited to static scenes, but enables us to generate data of arbitrary complexity. The HRDFs are stored in the YCbCr color space. This representation is widely used in digital and analogue video transmission [Poynton 1996], where the chromaticity values are often sub-sampled by a

*e-mail:ehsan.miandji@liu.se

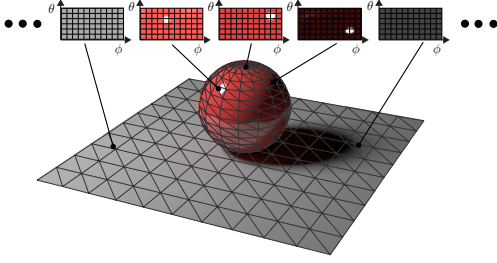


Figure 2: We represent the 4D SLF function $f(u, v, \phi, \theta)$ as a discrete set of hemispherical radiance distribution functions, HRDFs.

factor of two. By compressing each channel separately, we apply a similar approach.

Algorithm overview - Our compression algorithm exploits correlations in both the spatial and angular domains to adapt to the input data in two stages, see Figure 3. First, we analyze the spatial correlation in the data by clustering points with similar HRDFs, see Section 2. In a second step, we then learn a per-cluster compact dictionary exploiting the angular coherence. We explore two different methods for learning the dictionary, the well known Principal Component Analysis (PCA) algorithm, see Section 3, and an adapted version of the recently proposed Exemplar Orthonormal Basis (EOB) method [Gurumoorthy et al. 2010], see Section 4. Both Clustered-EOB (CEOB) and Clustered-PCA (CPCA) allow for real-time reconstruction and rendering on commodity graphics hardware. We show that CEOB in most cases outperforms CPCA [Miandji et al. 2011], K-SVD [Bryt and Elad 2008] and spherical harmonics [Sloan et al. 2003; Löw et al. 2009] in terms of memory footprint, reconstruction quality and rendering performance.

2 Clustering

To group the input HRDFs, H_i , into clusters with similar radiance distributions, we use a non-local clustering based on the $L1$ norm $\|H_i - H_j\|_1$ as distance metric. This metric measures the distance between the HRDFs based on their radiometric properties, and does not consider the spatial distance between the points p_i where the HRDFs were sampled. In this way, we exploit global coherencies in the data. The choice of $L1$ norm is due to its robustness against outliers. We denote the HRDFs in a cluster c as $\{H_i^c\}$, $c = 1..n_c$, $i = 1..n_c$, where n_c is the number of clusters and n_c is the number of HRDFs in cluster c .

For clustering the set $\{H_i\}$, we use the K-Means algorithm, [Lloyd 1982], extended with a robust initialization method, K-Means++ [Arthur and Vassilvitskii 2007]. To initialize centroids for K-Means, we first randomly pick an HRDF as the first centroid. We then compute a discrete probability density function (PDF) proportional to the distance of each HRDF to its closest centroids. We sample the next centroid using this PDF. This process is iterated until we have τ centroids. The result of the clustering is a vector of indices, $\chi \in R^n$, relating each HRDF to its corresponding cluster. This vector is used during reconstruction to find the corresponding dictionary of each HRDF.

3 CPCA

After spatial clustering, the first step in the CPCA algorithm, [Miandji et al. 2011; Sloan et al. 2003], is to rearrange the collection of n_c HRDFs in each cluster c to form a matrix $F^c \in R^{n_c \times m_1 m_2}$. Then, F^c is normalized by subtracting the mean, μ , computed as the average of rows in F^c . Afterwards, each cluster is projected

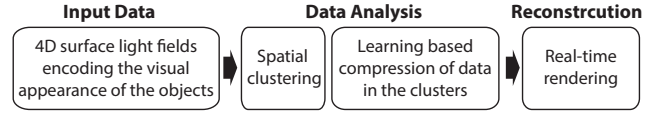


Figure 3: Overview of the proposed algorithm.

onto a low dimensional space spanned by the eigenvectors of the covariance matrix of F^c ; in particular, $F^c = U^c(V^c)^T$, where $U^c \in R^{n_c \times k}$ contains coefficients for each HRDF, $V^c \in R^{m_1 m_2 \times k}$ represents basis functions as rows and k is the number of principal components. Power iteration [Chen et al. 2002] can be used for computing first k singular vectors without the need for performing a full Singular Value Decomposition (SVD). Note that our approach is different from the light field mapping technique [Chen et al. 2002], where PCA is applied on a neighborhood ring of a vertex. Instead we cluster the HRDFs defined over the whole object and then apply PCA on each cluster. Our approach decouples geometrical complexity from radiometric complexity. A robust power iteration algorithm is presented in [Miandji et al. 2011], considering the case when the number of HRDFs in a cluster is smaller than the angular resolution, i.e. $n_c < m_1 m_2$.

4 CEOB

The CEOB method is, after the spatial clustering in Section 2, composed of two stages: training and testing. The training phase learns a set of orthogonal basis pairs which are used during the testing phase to compute a sparse coefficient matrix per HRDF with minimum reconstruction error. We start by considering an HRDF matrix corresponding to a surface point p_i in cluster c , denoted as $H_i^c \in R^{m_1 \times m_2}$, $i = 1..n_c$. Applying SVD on H_i^c , we have $H_i^c = U_c S_i V_c^T$. We can define another pair of orthogonal bases (\bar{U}_c, \bar{V}_c) which satisfies $\bar{H}_i^c = \bar{U}_c \bar{S}_i \bar{V}_c^T$, where \bar{S}_i is not diagonal anymore but is arbitrarily sparse, i.e. $\|\bar{S}_i\|_0 = t$, where t defines the number of non-zero coefficients (sparsity). Given \bar{U}_c and \bar{V}_c , it has been shown, [Gurumoorthy et al. 2010], that the optimal \bar{S}_i can be computed by nullifying the smallest $m_1 m_2 - t$ elements of the estimated projection matrix $\bar{S}_i = \bar{U}_c^T H_i^c \bar{V}_c$.

A single basis pair will not be adequate for approximation of n_c HRDFs of a cluster. As a result, we train a small number of basis pairs. In summary, EOB is based on training a set of $k_c \ll n_c$ full-rank orthogonal basis pairs (exemplars), $\{\bar{U}_a^c, \bar{V}_a^c\}$, $a = 1..k_c$, such that projecting each HRDF onto *one* basis pair would lead to the most sparse coefficient matrix while minimizing the $L2$ -error. This can be formulated as minimizing the following energy function:

$$E(\{\bar{U}_a^c, \bar{V}_a^c, \bar{S}_{ia}^c, M_{ia}^c\}) = \sum_{i=1}^{n_c} \sum_{a=1}^{k_c} M_{ia}^c \|H_i^c - \bar{U}_a^c \bar{S}_{ia}^c (\bar{V}_a^c)^T\|^2 \quad (1)$$

subject to

$$\begin{aligned} (\bar{U}_a^c)^T \bar{U}_a^c &= (\bar{V}_a^c)^T \bar{V}_a^c = I, \quad \forall a, \\ \|\bar{S}_{ia}^c\|_0 &\leq t \text{ and} \\ \sum_a M_{ia}^c &= 1, \quad \forall i, \end{aligned}$$

where $\bar{S}_{ia}^c \in R^{m_1 \times m_2}$ contains coefficients of the i th HRDF when projected onto the a th exemplar; $M \in R^{n_c \times k_c}$ is a binary matrix associating each HRDF to its corresponding exemplar pair $(\bar{U}_a^c, \bar{V}_a^c)$. Since each HRDF is represented using one exemplar pair (the last constraint), each row in matrix M has only one component equal to one. The first two constraints are for orthogonality of the exemplars and sparsity of the coefficients matrix respectively. Equation (1) can be solved efficiently using an iterative approach

[Gurumoorthy et al. 2010]. First the matrices $\{\bar{U}_a^c, \bar{V}_a^c\}$, $\forall a, \forall c$ are initialized with random orthogonal matrices and the matrices M^c are initialized with $1/k_c$. Then the following update rules are applied sequentially:

$$\begin{aligned} \bar{S}_{ia} &= (\bar{U}_a^c)^T H_i^c \bar{V}_a^c, \\ \bar{U}_a^c &= Y_a^c ((Y_a^c)^T Y_a^c)^{-1/2}; Y_a^c = \sum_{i=1}^{n_c} M_{ia}^c H_i^c \bar{V}_a^c \bar{S}_{ia}^T, \\ \bar{V}_a^c &= Z_a^c ((Z_a^c)^T Z_a^c)^{-1/2}; Z_a^c = \sum_{i=1}^{n_c} M_{ia}^c (H_i^c)^T \bar{U}_a^c \bar{S}_{ia}, \\ M_{ia}^c &= \frac{1}{\sum_{b=1}^{k_c} e^{\beta(E_a^c - E_b^c)}}; E_a^c = \|H_i^c - \bar{U}_a^c \bar{S}_{ia} (\bar{V}_a^c)^T\|_2, \end{aligned} \quad (2)$$

where we nullify $m_1 m_2 - t$ elements of \bar{S}_{ia} with smallest absolute value after each update to \bar{S}_{ia} . β is a temperature parameter initialized with a small value. The value of the sparsity parameter, t , is fixed for all HRDFs during training. More details on the derivations of update rules can be found in [Gurumoorthy et al. 2010]. The updates are done sequentially until the changes to matrices $\bar{S}, \bar{U}, \bar{V}, M$ are minimal. Then the temperature parameter is increased and sequential updates are repeated. The algorithm converges when M^c is binary or near binary, e.g. satisfying the following criteria: $\|M^c - [M^c]\|_2 < \epsilon$, where ϵ is a small value.

The training set for each cluster is chosen as a subset of the cluster (typically 20 to 80 percent of the HRDFs in a cluster). Intuitively, this subset should represent all data points of a cluster. A random or pseudo-random selection may lead to missing of important details. Therefore, incorporating the information inside HRDFs is essential. We use the K-Means++ initialization algorithm (described in Section 2) inside each cluster. In this way, it is guaranteed that the selected HRDFs have the maximal (radiometric) distance from each other inside a cluster. Note that for very high frequency materials, the intra-cluster coherence cannot be guaranteed unless we have a very large number of clusters. Due to the fact that we have a dictionary per cluster, we prefer to keep the number of clusters low enough for storage efficiency. Hence, careful selection of training set for each cluster is important.

Having the exemplar pairs for each cluster (the training phase), the testing phase computes the most sparse coefficient matrix resulting in the smallest reconstruction error for each HRDF. Depending on the projection error, each HRDF can have a different sparsity value, i.e. we have variable number of coefficients per HRDF. For this purpose we proceed as follows: Each HRDF is projected onto all exemplars of the corresponding cluster, leading to a set of coefficient matrices that are greedily nullified while the error is below a user defined threshold. The exemplar pair that produces sparsest coefficients with least error is picked and the index is stored in the vector $M \in R^n$. The results of the testing phase is the sparse coefficient matrices $\{\bar{S}_i\}$ and the exemplar membership matrix M .

5 Reconstruction and rendering

In this section, we describe how the appearance (outgoing radiance) of a spatial location p_i on an object, along a direction (ϕ, θ) towards the camera can be reconstructed from the compressed SLF data during rendering. Here we use the notation (ξ_1, ξ_2) to address an element of an HRDF matrix. Note that we do not need to reconstruct the HRDF as a matrix, but only a single element inside it corresponding to the current view direction.

First, the cluster index at the surface point p_i is fetched, $c = \chi_i$. The reconstruction for the proposed compression algorithms is different for CPCA and CEOB, and can be described as follows:

Scene	Raw		CEOB		CPCA		
	Size	Size	FPS	PSNR	Size	FPS	PSNR
Fig. 1a	3180	380	34	65	343	33	61.85
Fig. 1b	13600	1492	38	59.95	1420	34	58.5
Fig. 1c	6300	226	26	67.55	212	23	23.83

Table 1: Statistics for test scenarios shown in Figure 1. The size is in megabytes and the PSNR is calculated as the average of the intensity channel’s PSNR of all objects in the scene

CPCA - To reconstruct an element of a PCA compressed HRDF, H_i^c , in cluster c , we perform a dot product between the coefficient vector of the HRDF in U , with a basis vector in V corresponding to (ξ_1, ξ_2) :

$$H_i^c(\alpha) = \mu(\alpha) + \sum_{j=1}^k U(i, j) V^c(\alpha, j), \quad (3)$$

where $\alpha = \xi_2 + \xi_1 m_2$ is used for addressing a lexicographically ordered matrix. Note that using the power iteration algorithm the singular values, S_i , are pre-multiplied by V^c , hence omitted from the equation above.

CEOB - For CEOB, we use the relation $H_i^c = \bar{U}_a^c \bar{S}_i \bar{V}_a^c$, where $a = M_i$. The sparsity of \bar{S}_i , along with the fact that we need only a single element of the HRDF matrix for each view direction allows for a compact reconstruction formula for CEOB:

$$H_i^c(\xi_1, \xi_2) = \sum_{t=1}^{T_i} \bar{U}_a^c(S_i(t, 1), \xi_1) S_i(t, 3) \bar{V}_a^c(S_i(t, 2), \xi_2), \quad (4)$$

where we represent \bar{S}_i as a matrix of size $t \times 3$; the first two columns describe the index of a non-zero element and the third stores its value. Note that the reconstruction cost for a single element in the compressed HRDF is directly proportional to the sparsity factor and that we for each non-zero coefficient only have two scalar multiplications followed by an addition.

6 Results and Evaluation

To measure reconstruction quality, we use a modified version of the Peak Signal to Noise Ratio (PSNR):

$$20 \log_{10} \left(\frac{l}{MSE} \right), \quad MSE = \sqrt{\frac{1}{n} \sum_c \sum_j^{n_i} \|H_j^c - \hat{H}_j^c\|_2^2} \quad (5)$$

where l is the intensity of the brightest light in the scene and \hat{H}_j^c is the reconstructed HRDF after being compressed using CPCA or CEOB. Figures 4a-4c illustrate a simple scene illuminated by three small spherical light sources. The ground plane and the sphere share the same glossy material with three roughness values (0.5, 0.05 and 0.005) in order to evaluate the efficiency of the proposed algorithms. The SLF for the glossy sphere consists of 32×32 spatial and 32×16 angular samples. Figures 4d-4f compare CEOB, CPCA, K-SVD [Bryt and Elad 2008] and spherical harmonics [Sloan et al. 2003] in terms of the number of coefficients and PSNR (an analysis of CEOB parameters and the effect of clustering is included in the supplementary material).

For CEOB we used 16 exemplar pairs, i.e. $k = 16$ in Equation 1; for K-SVD we trained a dictionary of 64 atoms. Note that in our test scenario, the dictionary for K-SVD is twice as large as the one for CEOB. In addition, the dictionary for CPCA grows proportionally to the number of coefficients (see Section 3). We observe that CEOB outperforms other methods for high frequency data while CPCA and K-SVD can achieve a slightly better reconstruction quality for a near-diffuse material. Even in the latter case, CEOB achieves a higher rendering performance and a smaller

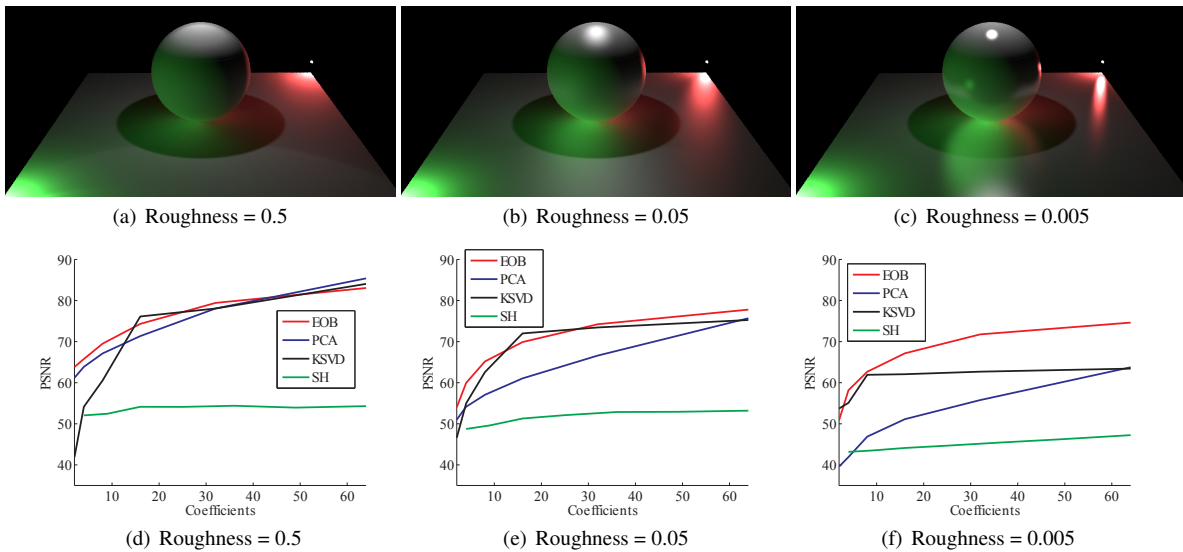


Figure 4: The scene used in the evaluation of the CPCA, CEOB, K-SVD and Spherical Harmonics (SH) compression schemes. The roughness parameter of the material on the sphere and ground plane is systematically varied to introduce a variation in the frequency content of HRDFs.

memory footprint. Spherical harmonics has the least PSNR due to its inefficiency for handling high frequency signals.

Figure 1 presents rendering results using CEOB for three scenes including complex materials and light sources. Visual quality comparison between CEOB, CPCA and a reference rendering is included in supplementary materials. The spatial resolution for objects vary from 128×128 to 512×512 and the angular resolution is 32×16 . The number of clusters is 32 for all objects except for the ring in Figure 1c where we used only one cluster. The material of the ring includes a perfect specular reflection component to cast caustics on the near-diffuse ground plain. Figure 1 presents statistics for these three scenarios compressed using CEOB and CPCA. For all these cases, the number of exemplars is set to 16 or 32 and we used maximum of 32 coefficients for intensity channel and 16 coefficients for chromaticity channels. Parameters for CPCA are then set to match the storage cost of CEOB. Performance was measured using a PC with a NVIDIA GeForce GTX 560.

7 Conclusion

We presented a SLF based method for real-time photo-realistic rendering of static scenes with arbitrary materials illuminated under general lighting conditions. We proposed a learning based compression technique, CEOB, and discussed the application of CPCA on SLF data. Our results show an overall advantage of CEOB in terms of memory footprint, rendering performance and reconstruction error. In addition, we analysed different parameters for our compression techniques. In the future, we would like to explore the possibility of including temporal domain in our data. Hence, learning a compact dictionary that exploits coherence on spatial, angular and temporal domains.

Acknowledgements

This project was funded by the Swedish Foundation for Strategic Research through grant IIS11-0081, and Linköping University Center for Industrial Information Technology.

References

ARTHUR, D., AND VASSILVITSKII, S. 2007. k-means++: the

advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, SIAM, SODA 07, 1027–1035.

BRYT, O., AND ELAD, M. 2008. Compression of facial images using the k-svd algorithm. *J. Vis. Commun. Image Represent.* 19, 270–282.

CHEN, W.-C., BOUGUET, J.-Y., CHU, M. H., AND GRZESZCZUK, R. 2002. Light field mapping: efficient representation and hardware rendering of surface light fields. *ACM Trans. Graph.* 21, 447–456.

GURUMOORTHY, K., RAJWADE, A., BANERJEE, A., AND RANGARAJAN, A. 2010. A method for compact image representation using sparse matrix and tensor projections onto exemplar orthonormal bases. *IEEE Transactions on Image Processing* 19, 2, 322–334.

LLOYD, S. 1982. Least squares quantization in pcm. *IEEE Transactions on Information Theory* 28, 2, 129–137.

LÖW, J., YNNERMAN, A., LARSSON, P., AND UNGER, J. 2009. Hdr light probe sequence resampling for realtime incident light field rendering. In *Proceedings of the 25th Spring Conference on Computer Graphics*, 43–50.

MIANDJI, E., KRONANDER, J., AND UNGER, J. 2011. Geometry independent surface light fields for real time rendering of precomputed global illumination. In *SIGRAD 2011*, Linköping University Electronic Press.

POYNTON, C. 1996. *A Technical Introduction to Digital Video*. Wiley.

SHIRLEY, P., AND CHIU, K. 1997. A low distortion map between disk and square. *J. Graph. Tools* 2, 3, 45–52.

SLOAN, P.-P., HALL, J., HART, J., AND SNYDER, J. 2003. Clustered principal components for precomputed radiance transfer. *ACM Trans. Graph.* 22, 382–391.

UNGER, J., GUSTAVSON, S., LARSSON, P., AND YNNERMAN, A. 2008. Free form incident light fields. In *Proceedings of the Nineteenth Eurographics conference on Rendering*, EGSR’08, 1293–1301.