Fourier Depth of Field

Cyril Soler
and
Kartic Subr
and
Fredo Durand
and
Nicolas Holzschuch
and
Francois Sillion

Additional Key Words and Phrases: Depth of field, Fourier analysis of light transport, Sampling

1. INTRODUCTION

The simplistic pinhole camera model used to teach perspective (and computer graphics) produces sharp images because every image element corresponds to a single ray in the scene. Real-life optical systems such as photographic lenses, however, must collect enought light to accomodate the sensitivity of the imaging system, and therefore combine light rays coming through a finite-sized aperture. Focusing mechanisms are needed to choose the distance of an "in-focus" plane, which will be sharply reproduced on the sensor, while objects appear increasingly blurry as their distance to this plane increases. The visual effect of focusing can be dramatic and is used extensively in photography and film, for instance to separate a subject from the background.

Although the simulation of depth of field in Computer Graphics has been possible for more than two decades, this effect is still rarely used in practice because of its high cost: The lens aperture must be densely sampled to produce a high-quality image. This is particularly frustrating because the defocus produced by the lens is not increasing the visual complexity, but rather removing detail! In this paper, we propose to exploit the blurriness of out-of-focus regions to reduce the computation load. We study defocus from a signal processing perspective and propose a new algorithm that estimates local image bandwidth. This allows us to reduce computation costs in two ways, by adapting the sampling rate over both the image and lens aperture domain.

In image space, we exploit the blurriness of out-of-focus regions by downsampling them: we compute the final image color for only a subset of the pixels and interpolate. Our motivation for adaptive sampling over the lens comes from the observation that in-focus regions do not require a large number of lens samples because they do not get blurred, in contrast to out of focus regions where the large variations of radiance through the lens requires many samples. More formally, we derive a formula for the variance over the lens and use it to adapt sampling for a Monte-Carlo integrator. Both image and lens sampling are derived from a Fourier analysis of depth of field that extends recent work on light transport [Durand et al. 2005]. In particular, we show how image and lens sampling correspond to the spatial and angular bandwidth of the lightfield.

1.1 Related work

Our work builds on a variety of previous approaches that seek to efficiently simulate depth of field effects, *e.g.* [Potmesil and Chakravarty 1981; Cook et al. 1984; Cook et al. 1987; Haeberli and Akeley 1990]. A number of approaches, in particular in real-time rendering, start from a pinhole image together with a depth map and post-process it using various forms of spatially-varying blur, e.g. [Potmesil and Chakravarty 1981; Kraus and Strengert 2007; Barsky

et al. 2003; Zhou et al. 2007; Kolb et al. 1995]. In this paper, we focus on high-quality offline image synthesis that resolves visibility based on a full thin-lens model, not an input pinhole image.

The method of multidimensional light cuts [Walter et al. 2006] reduces the cost of estimating a composition of multiple integrals, one of which is over the aperture. However, their work efficiently estimates the integral over the aperture *only in conjunction with* complex illumination. For scenes with simple direct lighting, their method performs no better than the naïve technique of independent stratified sampling of the aperture and image. Our work is complementary to theirs because we seek to reduce the number of image-space samples and lens samples, while they reduce light gathering.

Our approach is related to techniques that adaptively refine computation based on the smoothness of the current estimate and by assessing how well smooth interpolation can predict new simulated data, e.g. [Bolin and Meyer 1995; Ferwerda et al. 1997; Bolin and Meyer 1998; Myszkowski 1998; Stokes et al. 2004]. In contrast, we seek to *predict* the local bandwidth or smoothness of the image.

A variety of approaches compute derivatives of illumination to predict smoothness and improve interpolation, e.g. [Ward and Heckbert 1992; Suykens and Willems 2001; Shinya et al. 1987; Igehy 1999; Chen and Arvo 2000]. In particular, Ramamoorthi et al. [2007] compute 4D gradients of radiance and adaptively subdivide a Whitted ray tracing solution. While they are usually easier to estimate than frequency content, derivatives do not directly provide information about sampling rate and their locality can be both an asset and a drawback. We seek to predict frequency content in small neighborhoods that are not infinitesimally small so as to avoid missing small features and be able to derive bandwidth with a reasonable amount of precomputation.

Our work is complementary to the optimization of sampling patterns, [Mitchell 1991; Ostromoukhov et al. 2004; Agarwal et al. 2003] since we seek to optimize sampling *density*.

It also builds on Durand et al.'s analysis of frequency effects in light transport [2005]. In contrast to the mostly-theoretical nature of this work, we seek to apply bandwidth prediction to accelerate high-quality rendering.

Finally, we build on studies of defocus effects using Fourier analysis over 4D light fields, e.g. [Isaksen et al. 2000; Chai et al. 2000; Ng 2005]. Our derivation of the frequency effect of depth of field is similar to theirs but we use it in a ray-tracing context rather than for image-based rendering and photography.

1.2 Background on the Frequency analysis of Light Transport

Our technique builds on signal processing theory of light transport [Durand et al. 2005], local reflection [Ramamoorthi and Hanrahan 2001; 2004; Basri and Jacobs 2003], and light field sampling [Chai et al. 2000; Isaksen et al. 2000]. We briefly review these theoretical results, following the analysis by Durand et al. [2005] since it addresses both spatial and angular effects in global illumination.

We are interested in the content of a *local light field* characterized by a 4D slice of radiance in the neighborhood of a central ray. Following Durand et al. [2005], we use the flatland counterpart of the 4D radiance function to simplify exposition; for application in 3D scenes, we project the 4D function down to 2D (see Sec.3.3). The local light field ℓ is parameterized by a spatial component x in the plane orthogonal to the central ray and an angular component v, usually the tangent of the angle to the plane normal. We study the Fourier transform of such light fields

$$\widehat{\ell}(\Omega_x, \Omega_v) = \int_{x=-\infty}^{\infty} \int_{v=-\infty}^{\infty} \ell(x, v) \mathrm{e}^{-2\mathrm{i}\pi\Omega_x x} \mathrm{e}^{-2\mathrm{i}\pi\Omega_v v} \, \mathrm{d}x \, \mathrm{d}v \tag{1}$$

and how it is modified by transport phenomena. In what follows, we describe effects in the Fourier domain since this domain enables bandwidth and sampling rate prediction.

Transport in free space is a shear of the Fourier transform of the local light field. Reflection is described by two scale transforms due to the incident and outgoing angles and two shears due to the curvature of the receiver. Shading corresponds to a convolution with a small kernel corresponding to the spectrum of the clamped cosine term followed by a clamping by the BRDF angular bandwidth. Texture mapping is a multiplication of radiance, which is a convolution in the Fourier domain. Occlusion corresponds to a convolution by the spectrum of the blockers.

To summarize, existing literature analyze the effect of transport phenomena on light fields and show that transport through free space, reflection and occlusion can be modeled by simple transformations of the light field spectra [Durand et al. 2005]– shear, convolution and multiplication repectively. We use this theory to analyze the effect due to depth of field and derive an efficient algorithm for image synthesis, taking into account effects due to a finite sized aperture.

2. A FREQUENCY ANALYSIS OF DEPTH OF FIELD

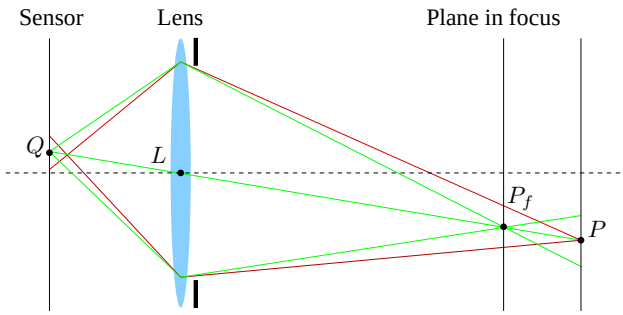


Fig. 1: *Finite aperture (thin lens) camera model: Rays from points that lie in front of (resp. behind) the plane in focus converge behind (resp. in front of) the sensor plane, after passing through the lens, resulting in finite blurry regions on the sensor called "circles of confusion".*

We present a theoretical analysis of the frequency content of the light field at the sensor plane of a camera with a finite sized aperture. For effective exposition, we present a flatland analysis where the lightfield is two dimensional: one spatial and one angular dimension; in 3D space the corresponding quantities and transforms are four dimensional.

Consider a point P in the scene (See Figure 1). We assume that we know the local light field at P, denoted by $\ell_P(x, v)$, and its spectrum, $\widehat{\ell_P}(\Omega_x,\ \Omega_v)$. We describe the transport of ℓ_P to ℓ_Q where Q is in the support plane of the camera sensor and derive the transformations undergone by $\widehat{\ell_P}(\Omega_x,\ \Omega_v)$ corresponding to this transport. The complete process is illustrated on Figure 2.

2.1 Transport from P to the lens:

To begin with, the light from P travels in free space in direction to the lens. From earlier work [Durand et al. 2005], we know that free-space traveling a distance d corresponds to a re-parameterization of the lightfield, *i.e.* a shear in the angular domain of its Fourier spectrum. We define an operator $\mathcal{S}$ to represent this transformation:

$$(\mathcal{S}\ \widehat{\ell_P})\ (\Omega_x,\ \Omega_v)\ \equiv\ \widehat{\ell_P}(\Omega_x,\ \Omega_v + d\Omega_x).$$

If the light from P passes by an occluder en route to L, this occluder also affects the light field. We express this by the operator C. This operator corresponds to a product between the lightfield and the visibility function of the occluder. C is thus a convolution of the spectrum of the local light field with that of the occluder [Durand et al. 2005]. If the occluder were planar, the effect of C is to inject spatial frequencies at the plane of occlusion. For non planar occluders, this is a continuous process through the width of the occluder.

The spectrum of the local light field at the lens after passing by a single occluder is a simple composition of the above operators:

$$\widehat{\ell_L}(\Omega_x,\ \Omega_v)\ =\ (C\,\mathcal{S}\ \widehat{\ell_P})\ (\Omega_x,\ \Omega_v)$$

In the general case, light travelling from P to L will encounter m different occluders, and $m+1$ shears (with different values for the shear parameter d). In this case we can write $\widehat{\ell_L}(\Omega_x,\ \Omega_v)$ as

$$\widehat{\ell_L}(\Omega_x,\ \Omega_v)\ =\ (\mathcal{S}\ (\mathcal{C}\ \mathcal{S})^m\ \ \widehat{\ell_P})\ (\Omega_x,\ \Omega_v) \tag{2}$$

2.2 Lens integration

The result of a finite-sized aperture is that, at each location Q on the sensor, there is an integration of the cone of incident rays from the lens to the scene, defined by the aperture. We choose to model this integration as an operation over the lightfield at the lens (meaning that the lightfield dimensionality is not reduced by this operation). This integration corresponds to a convolution in ray-space at L, and thus the light field just after L is actually

$$\ell_{L_+}(x,v)\ =\ \ell_{L_-}(x,v)\otimes\ a\,(x,v). \tag{3}$$

In this equation L_+ (resp. L_-) represent the lightfield after (resp. before) the lens, and a is the indicator function of the set of rays not blocked by the aperture. The equivalent transform in Fourier space is a product and can be written as

$$\widehat{\ell_{L_+}}(\Omega_x,\ \Omega_v)\ =\ \widehat{\ell_{L_-}}(\Omega_x,\ \Omega_v)\ \widehat{a}(\Omega_x,\ \Omega_v). \tag{4}$$

To understand what $\widehat{a}(\Omega_x,\ \Omega_v)$ looks like, one can notice that the set of rays over which the lightfield is integrated, converge at a point P_f in the plane in focus (See Figure 1). Therefore, at this point, the integration filter is a box in angles and a Dirac in space. It's Fourier transform is thus a sinc in angle and a constant in space. At L, $a\,(x,v)$ is the same function sheared from the distance between P and L. In 3D, the box is circular, and its Fourier transform is consequently a Bessel function in angles.

As a consequence, the light field at L_+ (*i.e.* just after the lens) is *bandlimited* by the spectrum of the aperture response function. Constricting the aperture of a camera spreads the width of $\widehat{a}(\Omega_x,\ \Omega_v)$ resulting in increased angular bandwidth at L_+. The ultimate case of a pinhole camera restricts a to a Dirac in both space and angles at the plane in focus, which means that it's Fourier transform is a constant that retains all frequencies in the lightfield.

Finally, because we have already accounted for the integration at the lens, and because the free-space traveling from the lens to the sensor is usually very small, we will neglect this very last phase of the transport to Q.

2.3 Consequences on lens integration and image-space frequencies

When numerically performing the lightfield integration at the lens, one would rather adapt the integration accuracy to the frequency content of the lightfield at L_- so as to ensure a desirable precision while keeping the computation cost as low as possible. This information is available in $\widehat{\ell_{L_-}}(\Omega_x,\ \Omega_v)$ and will be used in our algorithm to drive the lens sampling.

When computing an image, one would also preferably adapt the image sampling to the frequency content of the image and interpolate between samples, rather than explicitly compute all pixels. At the sensor, the result of the integrated lightfield is the radiance at point Q, corresponding to a pixel into the image. Seen from the lens, image-space frequencies correspond to angular frequencies of the lightfield at L_+ measured at the center of the lens (See Figure 1), times the cosine of the incident angle at the sensor. In Fourier space, this means that we can measure image-space frequencies from the angular frequencies in $\widehat{\ell_{L_+}}(\Omega_x,\ \Omega_v)$ integrated over the spatial domain. This operation of view-extraction is therefore a projection of the spectrum over the angular axis.

3. ADAPTIVE DEPTH OF FIELD RENDERING

We increase the efficiency with which depth of field effects can be simulated by adaptively varying the *image space samples* and the number of samples over the *aperture* at each image sample. The former are obtained according to conservatively predicted bandwidths over the camera sensor and, at each of these samples, the latter are obtained by estimating the variance of the integrand over the aperture. The computation of both, the bandwidth and the estimate

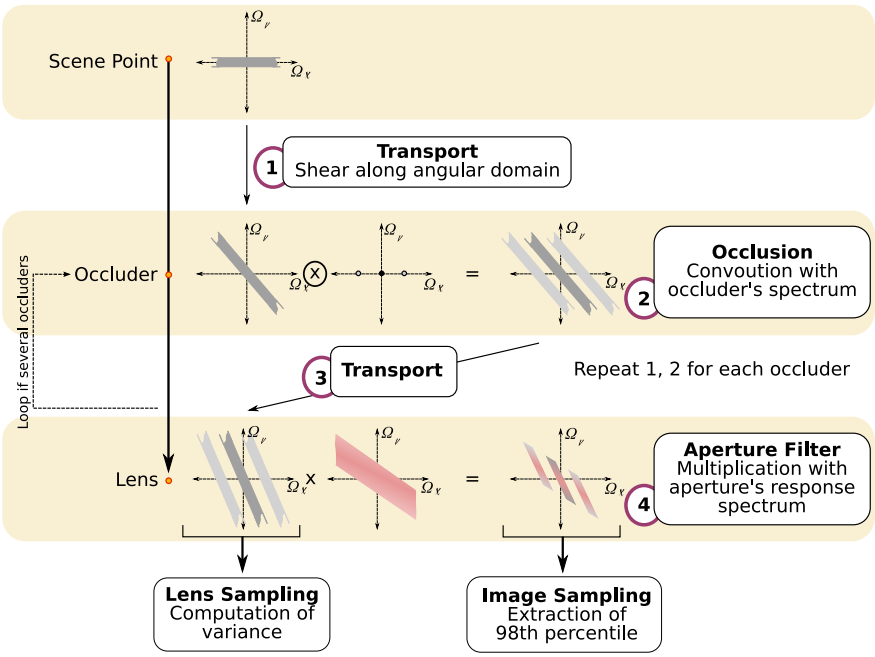


Fig. 2: *Flatland illustration of the transformations at different locations undergone by power spectra of local light fields after last bounce in the scene as they travel to the camera sensor.*

of the variance, are enabled by the propagation of local light field spectra after last bounce off surfaces in the scene towards the camera sensor.

3.1 Algorithm

To adaptively distribute effort between sampling the image and aperture, we consider the different transport phenomena between a visible object and the camera sensor. We propagate the spectral information of local light fields after last bounce off visible objects. To do this, we sample the power spectrum of the light field and adjust these samples during the different stages of transport to reflect the power spectrum density locally. Using a depth map to detect occlusion along the transport, we are able to efficiently estimate frequency propagation towards the camera sensor.

Using the frequency information of the light fields at the sensor, we extract a slice to obtain an image space density (see Sec. 3.4) that predicts bandwidth locally over the camera sensor. This operation is performed for a subset of image pixels on a regular grid, namely one every ten to one hundred pixels, and the frequency information is splatted using a max across the image. This makes the whole process very fast. Slices of the spectra at the plane of focus are used to estimate the variance of the integrand over the aperture (see Sec. 3.4). We use the density yielded by this slice to derive the number of lens samples for each pixel.

The next stage of our algorithm samples the image density and estimates the number of lens samples required at each of those sample locations. Given this information, we estimate incident radiance at those locations on the camera sensor using a Monte Carlo path tracer. The final image is reconstructed from the scattered radiance estimates. Figure 3 shows this process on a simple scene with a dramatic depth-of-field effect.

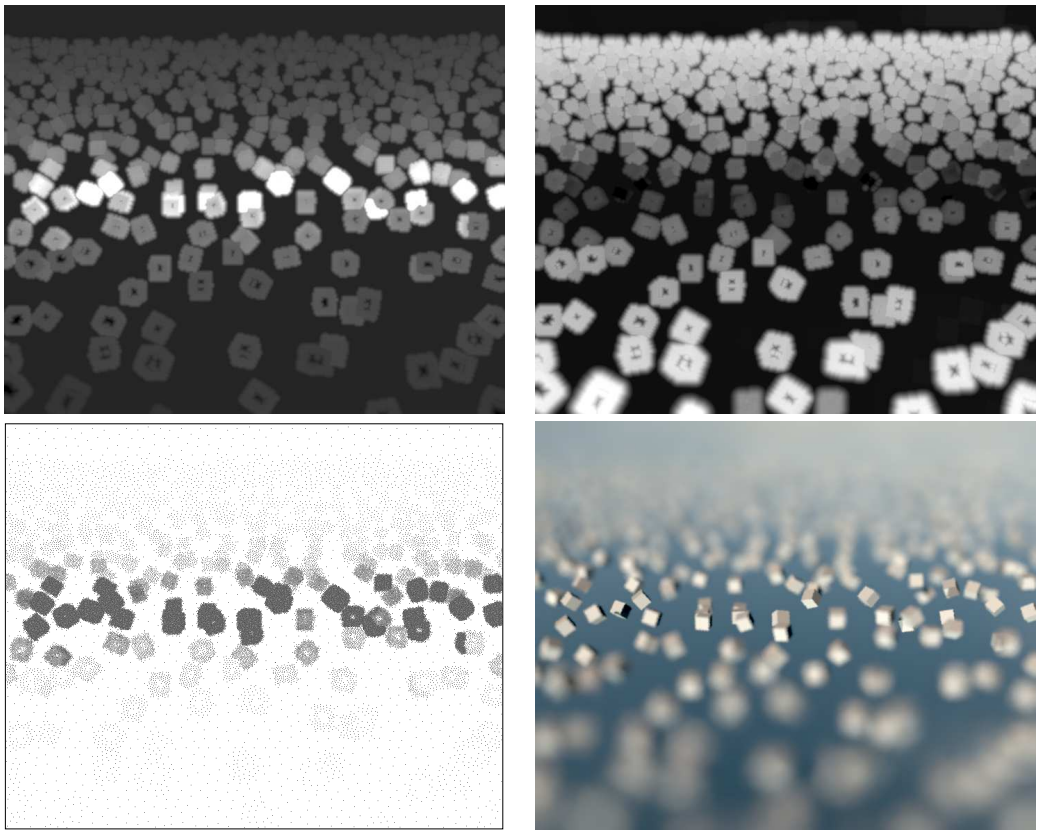


Fig. 3: *Walkthrough into our algorithm:* Top Left*: Image density depicting local bandwidth at each pixel.* Top Right*: Lens density indicating expected variance in the aperture integral.* Bottom Left*: image samples at which incoming radiance is estimated;* Bottom Right*: reconstructed image, using adaptive gaussian splatting. Blurry regions of the image are sampled sparsely, but require profuse sampling of the lens.*

3.2 Sampling local light field spectra

Let Q be a point on the sensor from where a primary ray r is cast (through the center of the lens) and let P be the point of intersection of this primary ray with the scene. We represent the power spectrum of the local light field at P, $\left|\widehat{\ell_p}(\Omega_x,\ \Omega_v)\right|$, by a set of random variables

$$\{({\omega^s}_i,\ {\omega^a}_i)\} \sim \mathcal{P}\left(\left|\widehat{\ell_p}(\Omega_x,\ \Omega_v)\right|\right) \qquad 0 < i < n_s.$$

$|\omega_i^s| < \infty$ and $|\omega_i^a| < \Omega_p$ are independent random variables representing the spatial and angular components of a 2D frequency sample. Ω_p is half the angular bandwidth of the reflectance function at P. $\mathcal{P}$ is a projection of the four dimensional power spectrum down to two dimensions, one in each, namely space and angle. The projection down to two dimensions implies that we assume isotropy independently in space and in angle which makes the computation, representation and propagation of the spectra practicable. In practice this assumption is reasonable since we are only interested in maximum frequencies and not in accurate estimates of the spectra themselves.

Local light fields in the scene can of course be arbitrarily complex, as can their corresponding 4D spectra. The existance of discontinuities in the light field implies that the range of frequencies is infinite. Although, after reflection

they are restricted in the angular domain by the bandwidth of the reflectance function, they could contain arbitrarily high spatial frequencies. This results in a very conservative prediction of bandwidth at Q and thus we generate more samples than the optimal number.

Associated with each primary ray is a set of samples– ray r is initialized with $\{({\omega^s}_i,\ {\omega^a}_i)\}$ from the power spectrum at P as above. The range of useful frequencies in the image plane is always bounded by the maximum number of samples N_s per square pixel in image space, and by the maximum number of lens samples N_l, in angle, which are user defined parameters. Also, in practice, anticipating the shear from the point to the sensor, we can restrict the spatial bounds to be such that the resulting frequencies stay below the maximum angular frequency at the sensor.

Propagation of the frequency content along the ray until Q requires that the samples be appropriately updated at each step in the transport from P to Q. These updates are simple and inexpensive to compute (see fig. 4).

3.3 Propagating local light field spectra

Transport through free space shears the power spectrum along the angular direction proportional to the distance transported. Starting from the original samples, obtaining samples that are distributed according to the sheared distribution involves simply shifting each of the samples in the angular dimension. That is, each sample $({\omega^s}_i,\ {\omega^a}_i)$ is updated to be $\left(\omega_i^s,\ \omega_i^a + d\omega_i^s\right)$ as a result of the free space transport by a distance d.

Occlusion involves a convolution of the spectrum with the local light field by the spectrum of the occluder. Random variables representing the spectra of the light field and the occluder when added are representatives of the convolution of the two distributions. Thus if we are able to draw samples $\{({\nu^s}_i,\ {\nu^a}_i)\},\ 0 < i < n_s$ from an occluder's spectrum then we can simply update our samples $({\omega^s}_i,\ {\omega^a}_i)$ to be $\left(\omega_i^s + \nu_i^s,\ \omega_i^a + \nu_i^a\right)$.

For each ray r we use the depth map to build a list of occluders and the points along the ray the occlusions occur. To achieve this, we search the depth map for discontinuities and splat these discontinuities in an occlusion buffer. Each discontinuity is splatted to influence a region as large as its circle of confusion. Given a pixel p and a pixel q in its neighborhood, the test to determine if q corresponds to a discontinuity where occlusion needs to be accounted for is illustrated in Figure 5.

At each occlusion point, the power spectrum of the occluder is assumed to be a Dirac in angle and proportional to $1/\omega_x$ in space. This conservative choice is due to the fact that visibility functions contain zero-order discontinuities and thus produce a spectrum with first-order fall-off. The effect of this is seen in the regions surrounding the foreground cubes in Figure 7 where the predicted effect of occlusion is more conservative than its measured counterpart.

The effect of a finite aperture is to cut off high angular frequencies at the plane in focus. Updating samples to represent the result of applying this operator involves rejecting angular frequencies with a probability defined by the shape of the aperture power spectrum. Although this will increase the variance of the estimate, it is reasonable since we are interested in information about maximum frequencies and not complete spectra.

3.4 Bandwidth, variance and reconstruction

3.4.1 *Sampling the image.* To obtain image space samples, the first step is to conservatively estimate bandwidth over the camera sensor using the incoming local light field spectral information. That is, we project the samples onto the angular axis (view extraction) and compute the highest angular frequency in the local neighborhood of each pixel. In practice, to decrease sensitivity to outliers, we use the 98th percentile of energy ξ_s as a representative of the maximum value at each point $s \in [0, W) \times [0, H)$. Here W and H are the width and height of the image respectively. The distribution of ξ_s over the image serves as an indicator of regions that need to be sampled more densely. Further, since ξ_s represents the maximum local frequency, we can estimate the optimal number of samples required locally (samples per square pixel) at s from the Nyquist limit, as

$$\rho(s) \ = \ \frac{4\,\xi_s^{\,2}\ f_h\ f_v}{W\,H}, \tag{5}$$

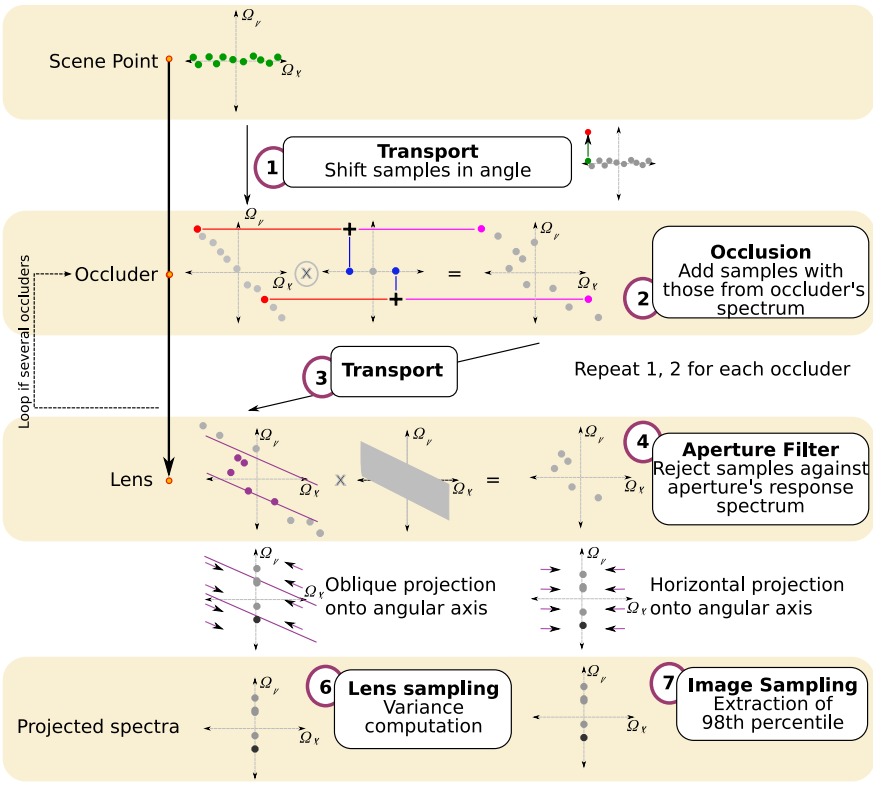


Fig. 4: *Sampled power spectra are propagated from the scene to the camera sensor. Transformations to the spectra are performed by independently modifying each sample.*

where f_h and f_v are the horizontal and vertical fields of view. However, since we predict bandwidth conservatively for increased reconstruction quality, the number of samples over the image may be suboptimal. After computing the density, image samples are generated according to $\rho(s)$ using a technique that produces samples with desireable noise properties [Ostromoukhov et al. 2004]. The total number of samples is dependant on the integral of $\rho(s)$ over the image rather than a user defined parameter.

3.4.2 *Sampling the aperture.* Using Monte Carlo integration over a finite aperture, the variance of the estimates depend on the variance of the integrand. The goal is to sample the aperture more profusely at image locations where the variance of the lens integrand is high. We use the light field spectra at the plane of focus to estimate the angular variance of the light field, since according to Parseval's theorem, the variance of a function is the integral of its power spectrum minus the DC term:

$$\sigma^2 = \int y_p(\Omega_v)^2 - y_p(0)^2$$

In this equation, y_p is the predicted spectrum just before the lens, obliquely projected onto the angular axis. The projection is oblique because of the local parameterization at the lens. Since all rays through the lens intersect at a common point at the plane of focus, the parameterization makes this projection horizontal at this plane. The slope of the projection to apply at the lens is thus given by the shear distance from the lens to the plane in focus.

The central limit theorem predicts that the Monte Carlo estimates of each of these integrals using uniform sampling

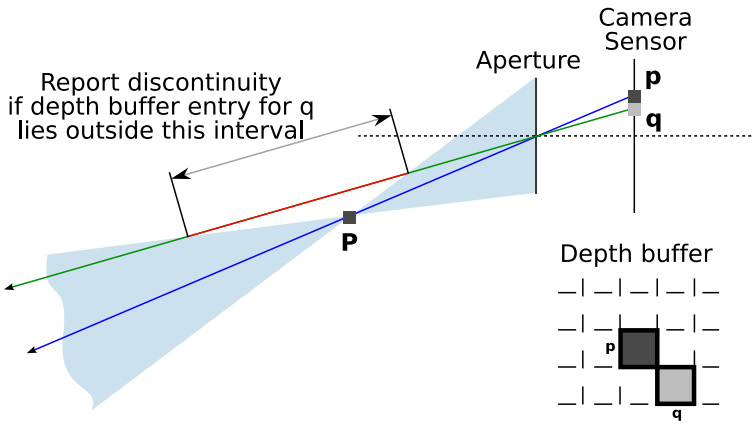


Fig. 5: *A depth map of the scene is used to build the lists of occluders, along with their distances, for each primary ray. P is the point of intersection of the primary ray through pixel p and the scene. This defines the double cone where a ray from the lens can hit the point P. The above figure illustrates the interval of depth values for a neighboring pixel q within which a discontinuity is reported.*

over the aperture has itself a variance of $O(n_s^{-1})$. While, in theory, stratification can improve the variance up to $O(n_s^{-2})$, Mitchell showed [Mitchell 1996] that in practice it is about $O(n_s^{-1.5})$ for pixels with edge boundaries. Using this conservative estimate for stratified sampling of the aperture, we determine the number of samples as

$$n_s = k\left(\sigma^2\right)^{\frac{1}{1.5}} \tag{6}$$

The constant of proportionality, k can be used to control the expected error consistently over the entire image.

3.4.3 *Image reconstruction.* We reconstruct the image using the radiance estimates at each of the image sample locations. The color at each pixel is computed as a weighted average of a constant number of neighboring samples. Since the samples are distributed according to a density, choosing a constant number of neighboring samples involves adaptively varying the radius of contribution of each pixel so that a constant number of samples (independent of the local density) contribute to the color at each pixel. In practice, we use a gaussian weighting term with a variance that is proportional to the square root of the local density.

We emphasize that sparsely sampled images resulting from simulation of depth of field cannot be splatted upto material or depth discontinuities (as is done for pinhole camera simulation), due to the integral over the aperture. Blurred discontinuities in the image need to be sampled adequately, which requires a systematic treatment of occlusion and aperture effects (see Sec.3.3).

4. VALIDATION AND RESULTS

We compare our conservative predictions of the local image bandwidth and lens variance against experimental measurements. To verify our predictions of the image density, at each pixel s_i (in the reference image) we compute a windowed Fast Fourier Transform (FFT) with the window centered at s_i and record the 98th percentile. Figure 6 shows a comparison of such a measured 98th percentile image against our image space sampling density. The measurement is not entirely local due to a fundamental property of the windowed FFT. Depending on the choice of window size the measured frequencies are either heavily blurred (large window) or restricted heavily in the range of measured frequencies (small window). To avoid border effects, the measurements are limited to the interior part of the reference image. From the figure, it is evident that our prediction both appears to qualitatively match the distribution of measured frequency and is of the same order of magnitude. In fact, we obtain a much more local prediction than observed with the windowed FFT.

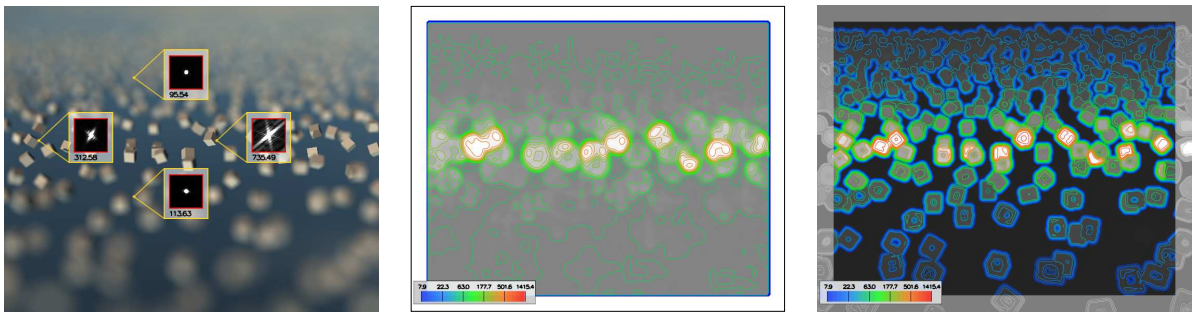


Fig. 6: *Comparison between measured and predicted image-space frequencies.* left: *image space frequencies are measured in the reference image by extracting the maximum* 98 *percentile (radially) in a 2D spatial spectrum computed using a* 64×64 *windowed Fourier transform around the point. Inlays show the spectra and image-space frequencies in pixel*$^{-}$1 *at four points.* Center: *measured values across the image that should be compared to our simulated values (*right*). Our method not only gives qualitatively the same profile of frequencies but also produces a conservative estimate of the actual values. Note that in the domain of low frequencies, the measured frequencies become higher than our estimate since the measurement method can not produce very low frequencies because of the* 64×64 *window resolution. In addition, the windowed fourier transform has an averaging effect whereas we estimate a purely local frequency, hence the difference in blurriness of the two approaches..*

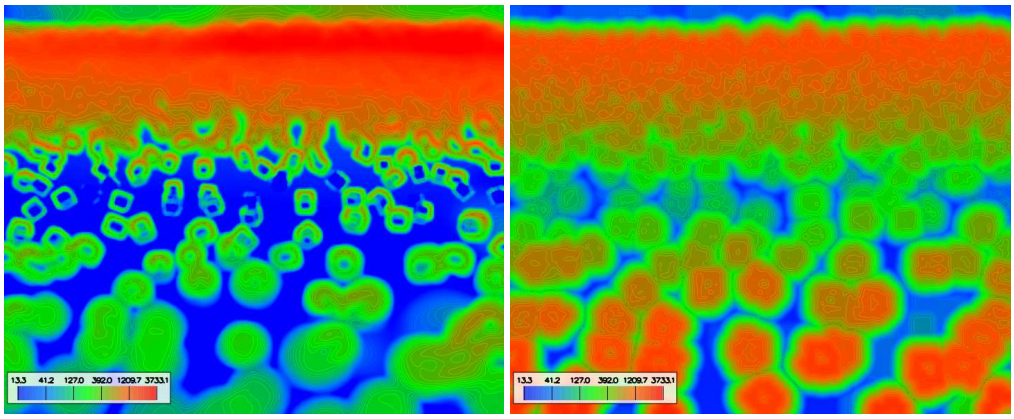


Fig. 7: *Comparison of variance measured over the rays converging to each pixel of the* cubes *scene (*left*), with the variance predicted by our method (*right*). Both images are displayed using the same scale. Our prediction is comparable to the actual measured values both in its distribution over the image, but also qualitatively, except in the foreground where it is a more conservative estimate. This makes it usable for adaptive lens sampling.*

To verify our estimates of the variation of the integrand over the aperture, we use stratified samples to estimate and record the variance in the lens integrals at each pixel. In Figure 7 we compare the predicted variance at each pixel using Eq.6 to the actual variance measured during Monte Carlo integration over the aperture for the reference image. From the comparison we observe that, although our predicted distribution resembles the measured variance, we predict higher frequencies around the blurry cubes in the foreground since our prediction is conservative.

4.1 Computation times

The table in Figure 8 summaries computation cost for the various scenes and focus settings with our algorithm. Kitchen 1 and 2 correspond to the kitchen scene with the plane in focus set on the foreground and background respectively. Clearly, the accumulated cost of propagating, computing and splatting frequency information, along with image reconstruction (using splatting) is quite negligible compared to the cost of naïve stratified Monte Carlo integration over the aperture at all pixels (see table in Fig. 9). This suggests that our adaptive algorithm significantly increases the efficiency of synthesizing images with depth of field effects (at least by an order of magnitude). The shallower the depth of field, the blurrier the image; this is when the adaptive algorithm provides maximum gain.

Scene	**Size**	**Frequency computation** (seconds)	**Path tracing** (seconds)	**Reconstruction** (seconds)	**Image space samples**	**Primary rays**
Cubes	721×589	45	3150	3	76 000	13 M
Snooker	904×806	90	4 500	10	119 335	25M
Kitchen 1	897×679	60	7401	8	867 000	44 M
Kitchen 2	897×679	60	6849	3	2 000 000	77 M

Fig. 8: *Execution times for the different steps in our algorithm and number of primary rays cast are shown for different scenes.*

The number of image samples is indicative of the number of pixels where radiance needs to be estimated. For images with larger regions in focus (large depth of field), this number would be very close to the number of pixels in the image. In those regions, the gain from using our algorithm is due to the extremely sparse lens sampling, again implying that fewer radiance estimates are required. Note that focused images are reconstructed faster, which is consistent since samples require smaller splatting radii.

We use the total number of primary rays cast to compare our technique with the non-adaptive stratified sampling technique. By distributing the total number of primary rays cast in our method amongst all pixels for the stratified sampling method, we generate images of similar computational cost. The table shown (see Fig. 9) shows the number of rays cast for similar image quality as those images used for measurements in Fig. 8. We also tabulate the theoretical speedup by dividing the number of primary rays in the reference technique by the number of primary rays cast by our algorithm.

Scene	**Number of lens rays/pixel**	**Number of . primary rays**	**Speedup due to our method**
Cubes	450	191M	14.7
Snooker	600	437M	17.3
Kitchen 1	1100	2 719M	61.0
Kitchen 2	1100	2 719M	35.3

Fig. 9: *Number of rays cast using stratified sampling Monte Carlo integration for similar appearance quality as for the images tabulated in Fig.8. The last column shows the speedup gained by using our method, obtained by dividing the middle column by the last column in Fig.8.*

Finally, we compare our approach to adaptive lens sampling based on *a priori* variance estimation: For each pixel, we trace a fixed (and small) number of rays and use their radiance value to estimate their variance σ accross the lens. Using Equation 6 we compute for each pixel the required number of rays to reduce the variance of the integrated light through the lens under a given threshold. We setup this threshold so that the total number of primary rays is the same than the number of rays used by our method. In figure 4.1 we compare the two methods on the kitchen (foreground focus setup) at different locations.

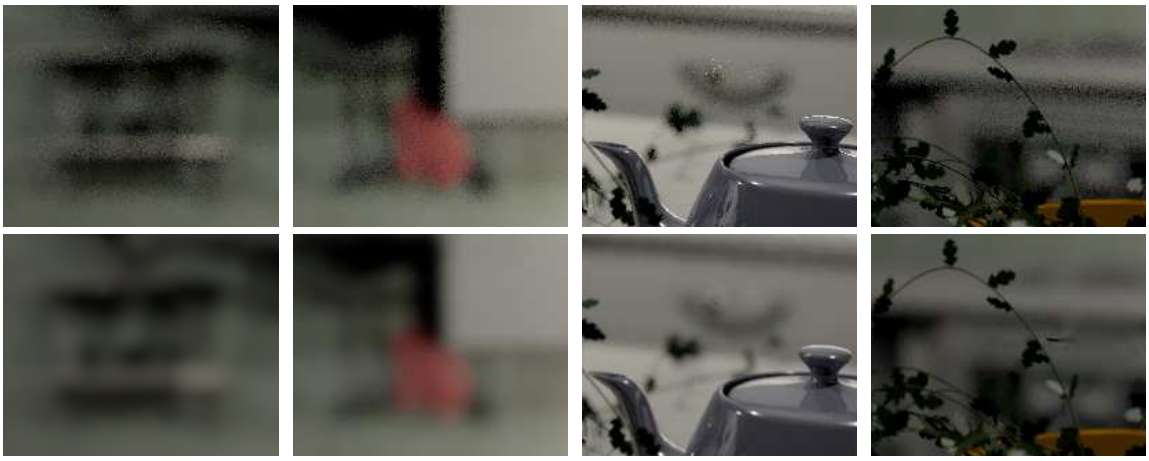


Fig. 10: *Comparison between our method (*botton row*) and adaptive lens sampling based on variance estimation from a small number of samples (*top row*). Both bethods use the same number of primary rays for the entire image. While the adaptive lens sampling manages to uniformize the variance accross the image, it needs to send rays for each and every pixel while our method only samples a few pixels in the most blurry regions. This makes the adaptive lens sampling unable to compete with our method at equal number of primary rays.*

4.2 Examples

We present, in Figure 11, example renderings with direct illumination of a scene lit by area and point light sources. The frequency maps conservatively capture the various effects which can produce high image-space and lens frequencies such as focused regions for the former, and highly curved specular regions for the later. The image samples as well as the lens samples are automatically adapted so as to produce an image of constant quality. The image resolution is 897×679, and we used maximum values of $N_s = 4$ image samples per square pixel and $N_l = 2500$ lens samples per pixel. The total number of primary rays is 44, 000, 000 and 77, 000, 000 in the two settings respectively.

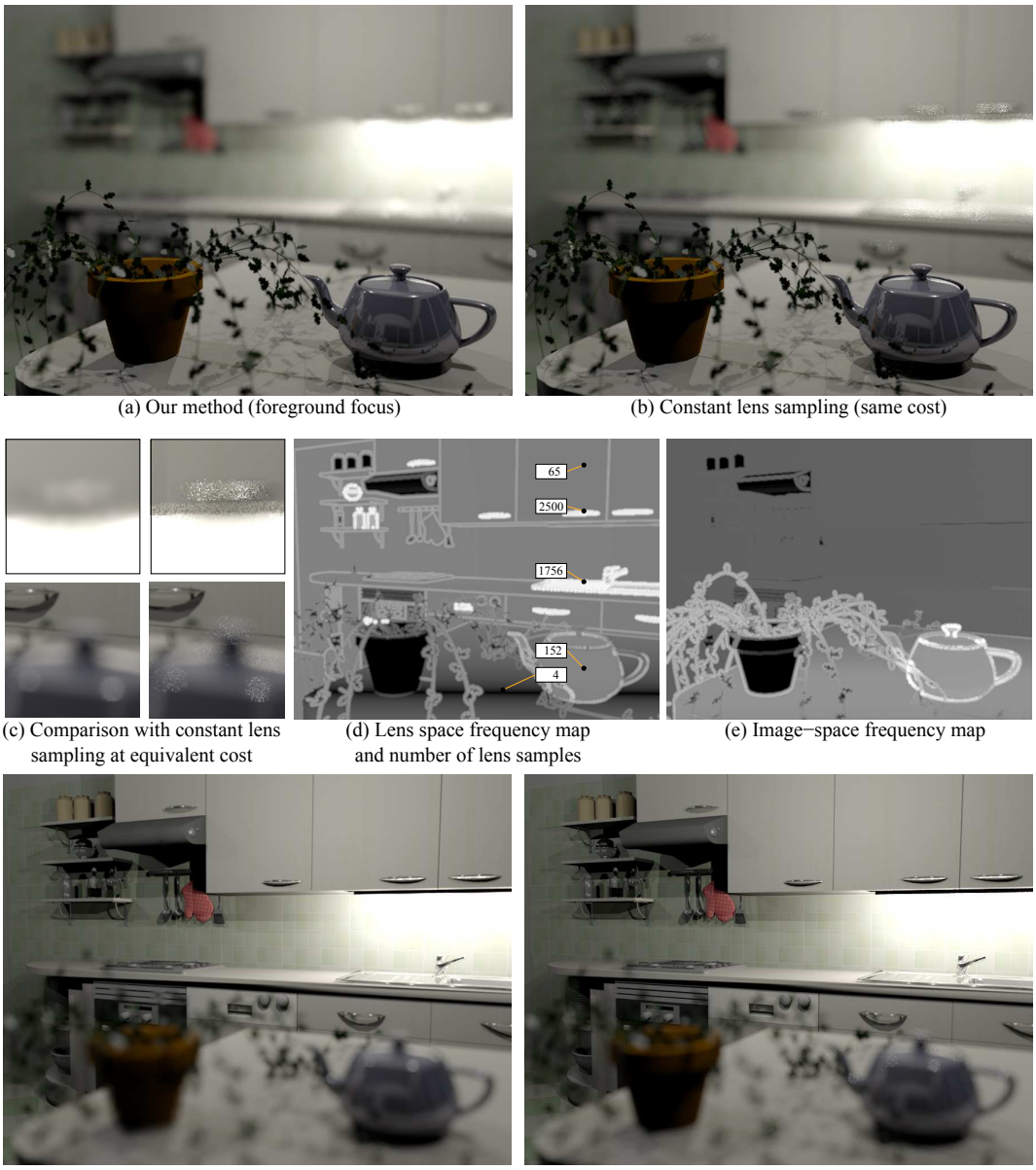
We compare our results to what we can obtain using a stratified lens sampling (with image space stratification for antialiasing) for the same computation cost. We do this by setting the number of lens samples so that the total number of primary rays is the same as with our method samples (70 and 129 for the foreground and background focus settings respectively). In both cases our algorithm results in images that are less noisy. Our algorithm performs particularly well in regions of high angular variance such as the handles of the cabinet. Despite the total cost being the same, the reason that the naïve method does not produce regions with less noise, is that many regions of the image are wastefully oversampled because of its non-adaptive nature.

In Figure 12 we show another configuration where ray tracing benefits from our method: in particular, the lens sampling densities and the image-space sampling densities adapt to the BRDFs of the shiny balls and the specular trumpet.

4.3 Discussion of the various approximations

Ignoring phase information of the local light field spectra, as we do in our model, implies approximations in the computation of convolutions between spectra. In practice, this means that we neglect the relative positions of multiple obstacles close to the same ray. The convolution is then over-estimated, and tends to produce higher frequencies when multiple obstacles lie between the eye and the scene. This approximation is therefore conservative with respect to image-space frequency and lens variance.

By reducing dimensionality from 4D spectra to 2D spectra, we implicitly make assumptions about the isotropy in the spatial and angular domains independently. In practice, since we only use the spectra to conservatively predict



(a) Our method (foreground focus)

(b) Constant lens sampling (same cost)

(c) Comparison with constant lens sampling at equivalent cost

(d) Lens space frequency map and number of lens samples

(e) Image−space frequency map

(f) Our method (backgound focus)

(g) Constant lens sampling (same cost)

Fig. 11: *Example of renderings using our method, with two settings of the focus plane* (a) *and* (f)*. In both cases, we compare our result to sampling the lens constantly throughout the image and by shooting the same number of total rays than in our method. The images obtained are much more blurry in regions of high variance, such as door handles which are highly curved very specular materials. In (c) we zoom on specific image locations and compare our method (at left) to the uniform constant sampling (at right). In* (d) *and* (e) *we show the lens and image-space frequency maps (logarithmic tone mapping) that we used to sample the lens and image, as well as the number of lens samples used at some locations.*

bandwidth, we do not observe artifacts that could be due to this projection.

Our choice of using conservative spectra such as maximum spatial frequencies when a textured surface is detected and angular frequencies equal to the bandwidth of the BRDF on all surfaces results in suboptimal sampling. Thus we are not able to take special advantage of knowing the local bandwidth of a region with texture. In addition we do not take illumination into account while sampling.

5. CONCLUSION AND FUTURE WORK

We have proposed a practical scheme that adaptively distributes effort between sampling the image and the aperture, in order to simulate depth of field effects in image synthesis. The gain in efficiency due to our method is independent of the technique used to estimate radiance at each pixel. We presented an algorithm for conservative predictions of regions in the image where the effect of a finite sized aperture is dominant, using existing theory that analyzes analysis light fields in the Fourier domain. Our algorithm predicts both pixels where the image is bandlimited and hence can be sampled sparsely, and pixels where there is a significant expected variance in the integral over the aperture.

Our algorithm yields a sparse yet sufficient sampling of the image in conjunction with a number of lens samples at each pixel that reduces variance drastically. We have shown a significant reduction in the number of primary rays required, in comparison with a uniform sampling of the image with stratified sampling of the aperture.

Since our estimates of bandwidth and expected variance over the aperture are both conservative, the number of samples are sub optimal. We believe, however, that initializing our algorithm with more intelligent spectral samples will further improve the efficacy of our method. One possibility would be to approximately predict light field spectra at points in the scene, due to global illumination effects. Although this might be an interesting path to explore, it is unclear if the gain due to this optimization will outweigh the cost of estimating spectral information due to illumination effects.

Another interesting avenue would be to explore the possibility of seeding the Metropolis light transport algorithm [Veach and Guibas 1997] with carefully chosen paths according to frequency predictions.

REFERENCES

AGARWAL, S., RAMAMOORTHI, R., BELONGIE, S., AND JENSEN, H. W. 2003. Structured importance sampling of environment maps. *ACM Transactions on Graphics 22*, 3, 605–612.

BARSKY, B. A., HORN, D. R., KLEIN, S. A., PANG, J. A., AND YU, M. 2003. Camera models and optical systems used in computer graphics: Part II, image based techniques. In *International Conference on Computational Science and its Applications*.

BASRI, R. AND JACOBS, D. 2003. Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell. 25*, 2, 218–233.

BOLIN, M. R. AND MEYER, G. W. 1995. A frequency based ray tracer. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 409–418.

BOLIN, M. R. AND MEYER, G. W. 1998. A perceptually based adaptive sampling algorithm. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 299–309.

CHAI, J.-X., CHAN, S.-C., SHUM, H.-Y., AND TONG, X. 2000. Plenoptic sampling. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 307–318.

CHEN, M. AND ARVO, J. 2000. Theory and application of specular path perturbation. *ACM Transactions on Graphics 19*, 4, 246–278.

COOK, R. L., CARPENTER, L., AND CATMULL, E. 1987. The reyes image rendering architecture. *Computer Graphics (Proc. SIGGRAPH '87) 21*, 4 (Oct), 95–102.

COOK, R. L., PORTER, T., AND CARPENTER, L. 1984. Distributed ray tracing. *Computer Graphics (Proc. SIGGRAPH 84) 18*, 3 (July), 137–145.

DURAND, F., HOLZSCHUCH, N., SOLER, C., CHAN, E., AND SILLION, F. X. 2005. A frequency analysis of light transport. *ACM Transactions on Graphics 24*, 3 (Aug.), 1115–1126.

FERWERDA, J. A., SHIRLEY, P., PATTANAIK, S. N., AND GREENBERG, D. P. 1997. A model of visual masking for computer graphics. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 143–152.

HAEBERLI, P. AND AKELEY, K. 1990. The accumulation buffer: hardware support for high-quality rendering. *Computer Graphics (Proc. SIGGRAPH '90) 24*, 4.

IGEHY, H. 1999. Tracing ray differentials. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 179 – 186.

ISAKSEN, A., MCMILLAN, L., AND GORTLER, S. J. 2000. Dynamically reparameterized light fields. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 297–306.

KOLB, C., HANRAHAN, P. M., AND MITCHELL, D. 1995. A realistic camera model for computer graphics. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 317–324.

KRAUS, M. AND STRENGERT, M. 2007. Depth-of-field rendering by pyramidal image processing. *Computer Graphics Forum (Proc. EG 2007) 26*, 3, 645–654.

MITCHELL, D. P. 1991. Spectrally optimal sampling for distributed ray tracing. *Computer Graphics (Proc. of SIGGRAPH '91) 25*, 4 (July), 157–164.

MITCHELL, D. P. 1996. Consequences of stratified sampling in graphics. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 277–280.

MYSZKOWSKI, K. 1998. The visible differences predictor: applications to global illumination problems. In *Rendering Techniques '98 (Proc. EG Workshop on Rendering '98)*. Eurographics, 223–236.

NG, R. 2005. Fourier slice photography. *ACM Transactions on Graphics (Proc. SIGGRAPH 2005) 24*, 3, 735–744.

OSTROMOUKHOV, V., DONOHUE, C., AND JODOIN, P.-M. 2004. Fast hierarchical importance sampling with blue noise properties. *ACM Transactions on Graphics (Proc. SIGGRAPH 2004) 23*, 3 (Aug.), 488–495.

POTMESIL, M. AND CHAKRAVARTY, I. 1981. A lens and aperture camera model for synthetic image generation. *Computer Graphics (Proc. SIGGRAPH '81)*, 297–305.

RAMAMOORTHI, R. AND HANRAHAN, P. 2001. A signal-processing framework for inverse rendering. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 117–128.

RAMAMOORTHI, R. AND HANRAHAN, P. 2004. A signal-processing framework for reflection. *ACM Transactions on Graphics 23*, 4, 1004–1042.

RAMAMOORTHI, R., MAHAJAN, D., AND BELHUMEUR, P. 2007. A first order analysis of lighting, shading, and shadows. *ACM Transactions on Graphics 26*, 1 (Jan.).

SHINYA, M., TAKAHASHI, T., AND NAITO, S. 1987. Principles and applications of pencil tracing. *Computer Graphics (Proc. SIGGRAPH '87) 21*, 4.

STOKES, W. A., FERWERDA, J. A., WALTER, B., AND GREENBERG, D. P. 2004. Perceptual illumination components: a new approach to efficient, high quality global illumination rendering. *ACM Transactions on Graphics 23*, 3 (Aug.), 742–749.

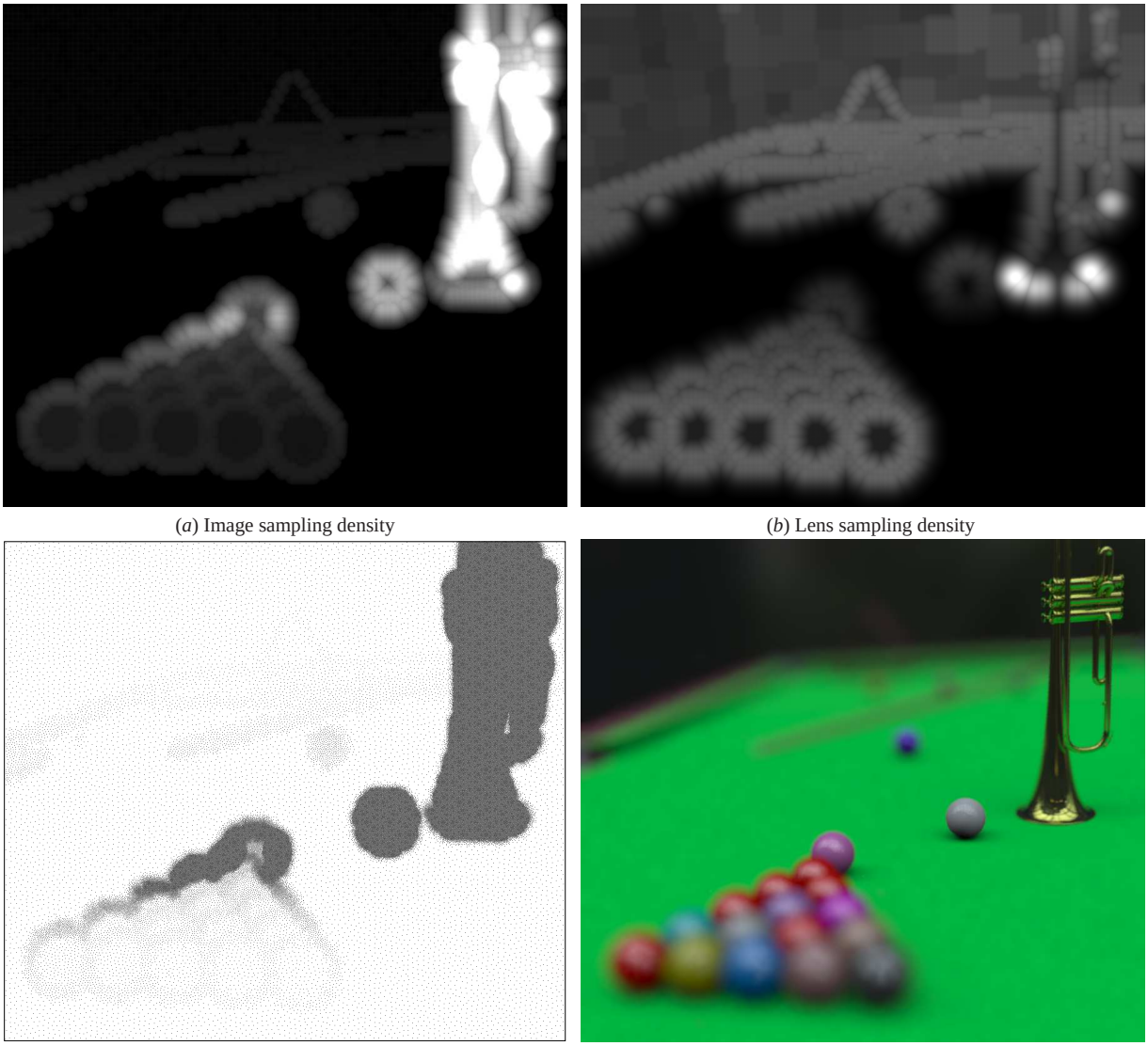
SUYKENS, F. AND WILLEMS, Y. 2001. Path differentials and applications. In *Rendering Techniques '01 (EG Workshop on Rendering)*. Eurographics, 257–268.

VEACH, E. AND GUIBAS, L. J. 1997. Metropolis light transport. In *Computer Graphics* Proceedings. Annual Conference Series. ACM SIGGRAPH, 65–76.

WALTER, B., ARBREE, A., BALA, K., AND GREENBERG, D. P. 2006. Multidimensional lightcuts. *ACM Transactions on Graphics 26*, 3, 1081–1088.

WARD, G. J. AND HECKBERT, P. 1992. Irradiance gradients. In *Proc. of EG Workshop on Rendering '92*. Eurographics, 85–98.

ZHOU, T., CHEN, J., AND PULLEN, M. 2007. Accurate depth of field simulation in real time. *Computer Graphics Forum 26*, 1 (Jan.), 15–23.



(*a*) Image sampling density (*b*) Lens sampling density

(*c*) Image space samples (*d*) Reconstructed image

Fig. 12: (*a*) *The image sampling density predicts that the shiny regions of the trumpet, with high curvature and in focus need to be sampled most profusely in the image.* (*b*) *The aperture density predicts that defocused regions need to be sampled densely while the ball in focus requires very few samples over the aperture.* (*c*) *the image samples obtained from the image sampling density.* (*d*) *The image is reconstructed from scattered radiance estimates.*