

# Fourier Depth of Field

CYRIL SOLER

INRIA, Grenoble University, CNRS

KARTIC SUBR

University of California, Irvine

FRÉDO DURAND

MIT CSAIL

and

NICOLAS HOLZSCHUCH and FRANÇOIS SILLION

INRIA, Grenoble University, CNRS

---

Optical systems used in photography and cinema produce depth-of-field effects, that is, variations of focus with depth. These effects are simulated in image synthesis by integrating incoming radiance at each pixel over the lens aperture. Unfortunately, aperture integration is extremely costly for defocused areas where the incoming radiance has high variance, since many samples are then required for a noise-free Monte Carlo integration. On the other hand, using many aperture samples is wasteful in focused areas where the integrand varies little. Similarly, image sampling in defocused areas should be adapted to the very smooth appearance variations due to blurring. This article introduces an analysis of focusing and depth-of-field in the frequency domain, allowing a practical characterization of a light field's frequency content both for image and aperture sampling. Based on this analysis we propose an adaptive depth-of-field rendering algorithm which optimizes sampling in two important ways. First, image sampling is based on conservative bandwidth prediction and a splatting reconstruction technique ensures correct image reconstruction. Second, at each pixel the variance in the radiance over the aperture is estimated and used to govern sampling. This technique is easily integrated in any sampling-based renderer, and vastly improves performance.

Categories and Subject Descriptors: I.3.7 [Computing Graphics]: Three-Dimensional Graphics and Realism

General Terms: Algorithms

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

**ACM Reference Format:**

Soler, C., Subr, K., Durand, F., Holzschuch, N., and Sillion, F. 2009. Fourier depth of field. *ACM Trans. Graph.* 28, 2, Article 18 (April 2009), 12 pages.  
DOI = 10.1145/1516522.1516529 <http://doi.acm.org/10.1145/1516522.1516529>

---

## 1. INTRODUCTION

The simplistic pinhole camera model used to teach perspective and for most computer graphics rendering results in sharp images because every pixel corresponds to a single ray in the scene. In contrast, real optical systems such as photographic lenses must collect enough light to accommodate the sensitivity of the imaging system, and each pixel combines light rays integrated over a finite-sized aperture. Focusing mechanisms are needed to choose the distance of an in-focus or focal 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 defocus can be dramatic and is used extensively in photography and film.

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 sampled densely to produce a high-quality image. This is particularly frustrating because the defocus produced by the lens is not increasing visual complexity: Quite contrarily, it removes detail. In this article, we exploit the blurriness of out-of-focus regions to reduce computational load. We study defocus from a signal processing perspective and propose a new algorithm that estimates local image

<sup>1</sup>ARTIS project-team, Laboratoire Jean Kuntzmann (UMR 5527)

This work was supported in part by the INRIA Associate Research Team "Flexible Rendering", by an INRIA internship, by the Visitor RTN, by NSF CAREER award 044756, a Microsoft Research New Faculty Fellowship and a Sloan Fellowship.

Authors' addresses: C. Soler, INRIA, Grenoble University, CNRS; ARTIS project-team, Laboratoire Jean Kuntzmann (UMR 5527); email: [cyril.soler@inria.fr](mailto:cyril.soler@inria.fr); K. Subr, University of California at Irvine, Irvine, CA 92697; F. Durand, MIT Computer Science and Artificial Intelligence Laboratory (CSAIL); N. Holzschuch, F. Sillion, INRIA, Grenoble University, CNRS.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© 2009 ACM 0730-0301/2009/04-ART18 \$10.00

DOI 10.1145/1516522.1516529 <http://doi.acm.org/10.1145/1516522.1516529>

bandwidth. This allows us to reduce computation costs 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 of the *lens aperture* 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 defocused 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 can be adapted to the spatial and angular bandwidth of the lightfield.

We emphasize that sparsely sampled images resulting from simulation of depth of field cannot be splatted up to 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.

## 1.1 Related Work

Our work builds on a variety of previous approaches that seek to efficiently simulate depth-of-field effects, such as Potmesil and Chakravarty [1981], Cook et al. [1984, 1987], and 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 postprocess it using various forms of spatially-varying blur, for example, Potmesil and Chakravarty [1981], Kraus and Strengert [2007], Barsky et al. [2003], Zhou et al. [2007], and Kolb et al. [1995]. In this article, we focus on high-quality offline image synthesis that resolves visibility based on a full thin-lens model, not an input pinhole image. We will show a comparison between our technique and such methods in Figure 11 in the results section.

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, for example, Bolin and Meyer [1995], Ferwerda et al. [1997], Bolin and Meyer [1998], Myszkowski [1998], and 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, such as Ward and Heckbert [1992], Suykens and Willem [2001], Shinya et al. [1987], Igehy [1999], and 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*. This also builds on Durand et al.’s analysis of frequency effects in light transport [Durand et al. 2005]. In contrast to the mostly theoretical nature of that 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, such as Isaksen et al. [2000], Chai et al. [2000], and 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 2004; 2001; 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 Section 3.3). The local light field  $\ell$  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

$$\widehat{\ell}(\Omega_x, \Omega_v) = \int_{x=-\infty}^{\infty} \int_{v=-\infty}^{\infty} \ell(x, v) e^{-2i\pi\Omega_x x} e^{-2i\pi\Omega_v v} dx dv \quad (1)$$

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 analyzes the effect of transport phenomena on light fields and shows 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, respectively. 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

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

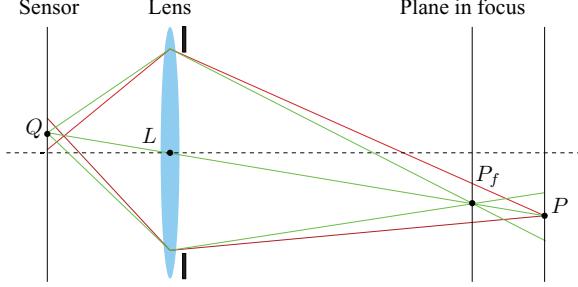


Fig. 1. Finite aperture (thin lens) camera model: Rays from points that lie in front of (respectively, behind) the plane in focus converge behind (respectively, in front of) the sensor plane, after passing through the lens, resulting in finite blurry regions on the sensor called “circles of confusion.”

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  $\ell_P$  to  $\ell_Q$  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 in 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 reparameterization of the lightfield, that is, a shear in the angular domain of its Fourier spectrum. We define an operator  $\mathcal{S}$  to represent this transformation.

$$(\mathcal{S} \widehat{\ell}_P)(\Omega_x, \Omega_v) \equiv \widehat{\ell}_P(\Omega_x, \Omega_v + d\Omega_v)$$

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  $\mathcal{C}$ . This operator corresponds to a product between the lightfield and the visibility function of the occluder.  $\mathcal{C}$  is a convolution of the spectrum of the local light field with that of the occluder [Durand et al. 2005]. If the occluder is planar, the effect of  $\mathcal{C}$  is to inject spatial frequencies at the plane of occlusion. For nonplanar occluders, this is a continuous process through the depth 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 preceding operators.

$$\widehat{\ell}_L(\Omega_x, \Omega_v) = (\mathcal{C} \mathcal{S} \widehat{\ell}_P)(\Omega_x, \Omega_v).$$

In the general case, light traveling 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

$$\widehat{\ell}_L(\Omega_x, \Omega_v) = (\mathcal{S} (\mathcal{C} \mathcal{S})^m \widehat{\ell}_P)(\Omega_x, \Omega_v). \quad (2)$$

## 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

the light field just after  $L$  is

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

In this equation  $L_+$  (respectively,  $L_-$ ) represent the lightfield after (respectively, 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

$$\widehat{\ell}_{L+}(\Omega_x, \Omega_v) = \widehat{\ell}_{L-}(\Omega_x, \Omega_v) \widehat{a}(\Omega_x, \Omega_v). \quad (4)$$

To understand what  $\widehat{a}(\Omega_x, \Omega_v)$  looks like, it can be noticed 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. Its Fourier transform is 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 has a 2D circular support, and its Fourier transform is consequently a 2D Bessel function in angles.

Consequently, the light field at  $L_+$  (i.e., just after the lens) is *bandlimited* by the spectrum of the aperture response function. Narrowing down 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 angle at the plane in focus, which means that its Fourier transform is a constant that retains all frequencies in the lightfield.

## 2.3 Consequences on Lens Integration and Image-Space Frequencies

When numerically performing the lightfield integration at the lens, it is desirable to adapt the integration accuracy to the frequency content of the lightfield at  $L_-$  so as to ensure a given 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, we also seek to 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 of the variance are enabled by the propagation of local light field spectra after the last bounce off surfaces in the scene towards the camera sensor.

### 3.1 Algorithm

To adapt the effort for both image and aperture sampling, we consider the different transport phenomena between a visible object and

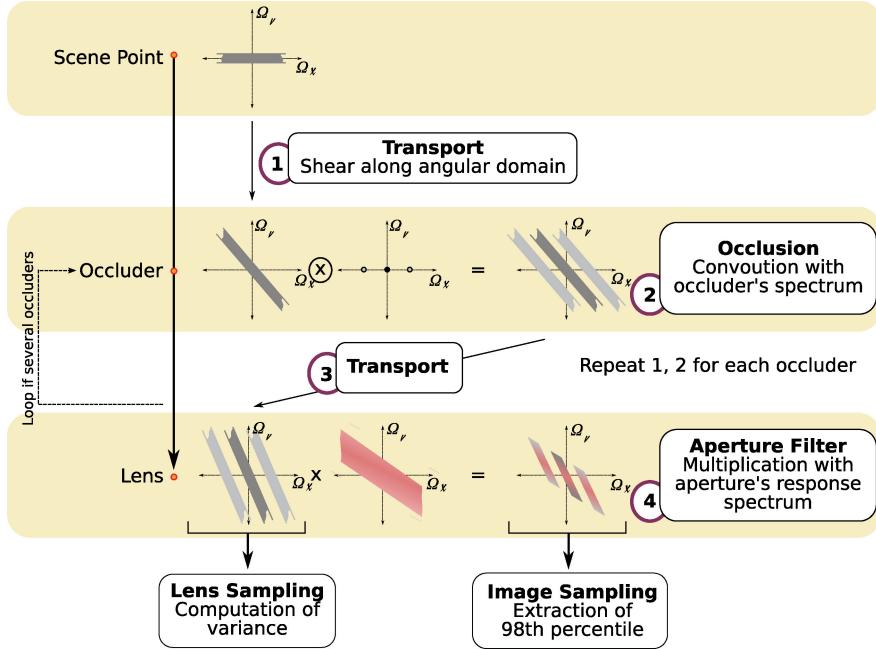


Fig. 2. Flatland illustration of the transformations at different locations undergone by power spectra of local light fields after last bounce in the scene as they travel to the camera sensor.

the camera sensor. We propagate the spectral information of local light fields after the 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 Section 3.4 and Figure 3(a)) that predicts bandwidth locally over the camera sensor. To improve efficiency, this operation is performed for a subset of image pixels on a regular grid, namely for one-tenth to one-hundredth of total pixels, and the frequency information is splatted across the image and combined using a per-pixel max. Slices of the spectra at the plane of focus are used to estimate the variance of the integrand over the aperture (see Section 3.4 and Figure 3(b)). We use the density read from 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 these sample locations. Given this information, we estimate incident radiance at these 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.

### 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  $P$ ,  $|\hat{\ell}_p(\Omega_x, \Omega_v)|$ , by a set of random

variables

$$\{(\omega^s_i, \omega^a_i)\} \sim \mathcal{P}(|\hat{\ell}_p(\Omega_x, \Omega_v)|) \quad 0 < i < n_s.$$

$|\omega^s_i| < \infty$  and  $|\omega^a_i| < \Omega_p$  are independent random variables representing the spatial and angular components of a 2D frequency sample.  $\Omega_p$  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 existence 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 conservative prediction of bandwidth at  $Q$ , resulting in 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 before. 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  $\hat{N}_l$ , in angle, which are user-defined parameters. In practice, we anticipate the shear the sensor, and restrict the spatial bounds to be such that the resulting frequencies stay below the maximum angular frequency at the sensor.

There is propagation of the frequency content along the ray until  $Q$  requires that the samples be appropriately updated at each step in

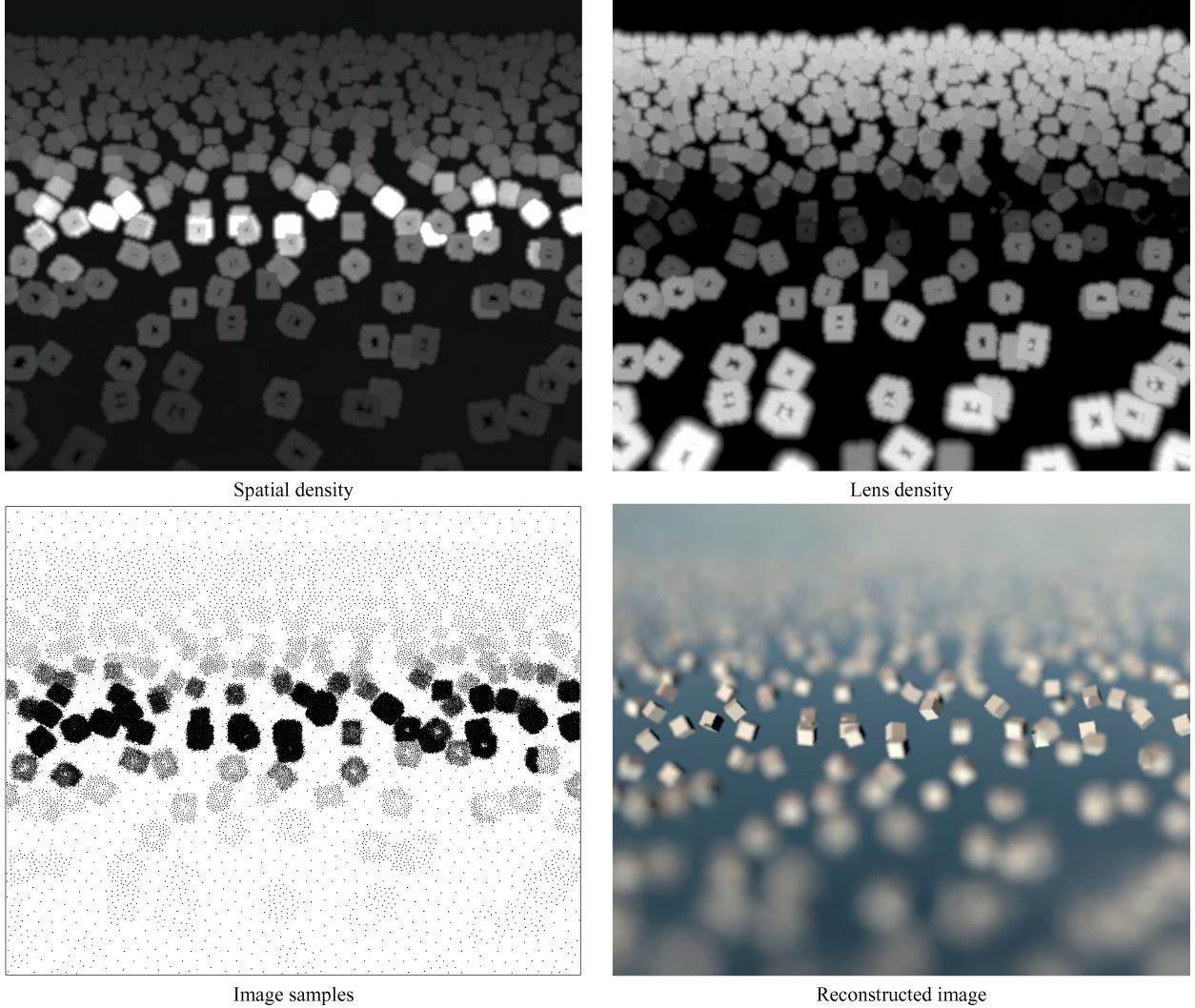


Fig. 3. Overview of our algorithm. Top left: image density depicting local bandwidth at each pixel; top right: lens density indicating expected variance in the aperture integral; bottom left: image samples at which incoming radiance is estimated; bottom right: reconstructed image, using adaptive Gaussian splatting. Blurry regions of the image are sampled sparsely, but require profuse sampling of the lens.

the transport from  $P$  to  $Q$ . These updates are simple and inexpensive to compute (see Figure 4).

### 3.3 Propagating Local Light Field Spectra

**3.3.1 Free-Space Transport.** 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. In other words, each sample  $(\omega_i^s, \omega_i^a)$  is updated to be  $(\omega_i^s, \omega_i^a + d\omega_i^s)$  as a result of the free space transport by a distance  $d$ .

**3.3.2 Occlusion.** Occlusion involves a convolution of the spectrum with the local light field by the spectrum of the occluder. The probability density function of the sum of two independent random

variables is the convolution of their respective density functions. Random variables representing the spectra of the light field and the occluder when added are therefore representatives of the convolution of the two distributions. Thus, if we are able to draw samples  $\{(v_i^s, v_i^a)\}$ ,  $0 < i < n_s$  from an occluder's spectrum then we can simply update our samples  $(\omega_i^s, \omega_i^a)$  to be  $(\omega_i^s + v_i^s, \omega_i^a + v_i^a)$ .

For each ray  $r$  we use the depth map to build a list of occluders and the points along the ray where 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

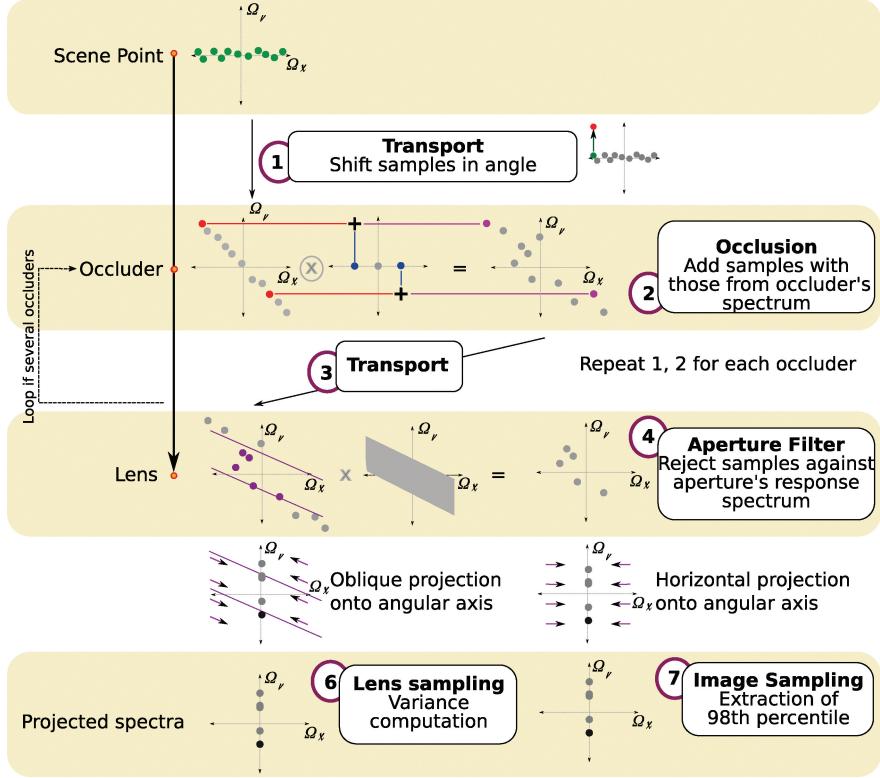


Fig. 4. Sampled power spectra are propagated from the scene to the camera sensor. Transformations to the spectra are performed by independently modifying each sample.

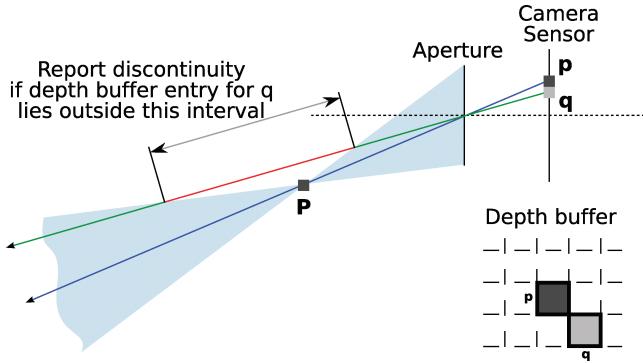


Fig. 5. A depth map of the scene is used to build the lists of occluders, along with their distances, for each primary ray.  $P$  is the point of intersection of the primary ray through pixel  $p$  and the scene. This defines the double cone where a ray from the lens can hit the point  $P$ . The above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

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.

**3.3.3 Aperture Effect.** 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 of the spectrum, it is justifiable 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. Specifically, 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 98<sup>th</sup> percentile of energy  $\xi_s$  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  $\xi_s$  over the image serves as an indicator of regions that need to be sampled more densely. Further, since  $\xi_s$  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

$$\rho(s) = \frac{4\xi_s^2 f_h f_v}{WH}, \quad (5)$$

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 hierarchical importance sampling

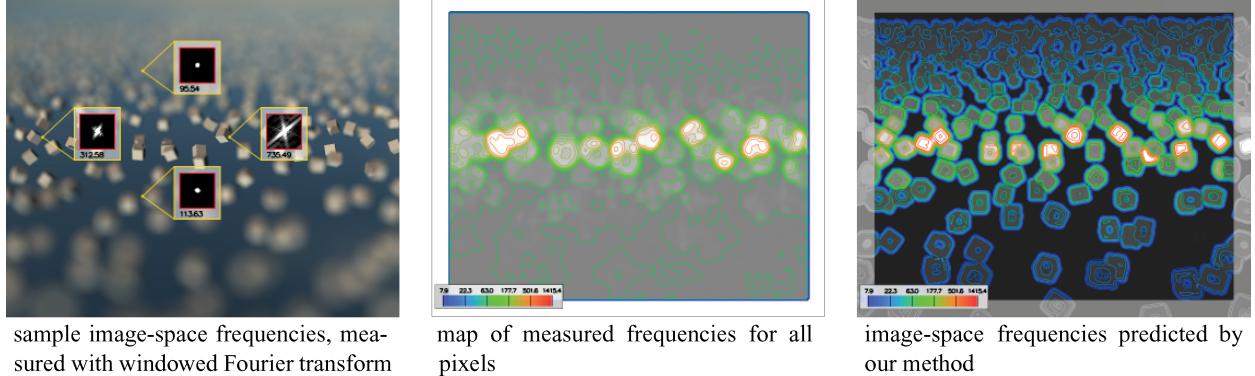


Fig. 6. Comparison between measured and predicted image-space frequencies: Left: image space frequencies are measured in the reference image by extracting the maximum 98th percentile (radially) in a 2D spatial spectrum computed using a  $64 \times 64$  windowed Fourier transform around the point. Insets show the spectra and image-space frequencies in  $\text{pixel}^{-1}$  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  $64 \times 64$  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.

[Ostromoukhov et al. 2004], which produces samples with desirable noise properties. The total number of samples is dependent 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 variance of a Monte Carlo estimator using uniform sampling over the aperture converges as  $O(n_s^{-1})$ . While, in theory, stratification can improve the variance up to  $O(n_s^{-2})$ , Mitchell [1996] showed 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

$$n_s = k (\sigma^2)^{\frac{1}{1.5}}. \quad (6)$$

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.

## 4. VALIDATION AND RESULTS

### 4.1 Validation

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 98<sup>th</sup> percentile. Figure 6 shows a comparison of such a measured 98<sup>th</sup> 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 of the reference image. From the figure, it appears that our prediction qualitatively matches 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.

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.

### 4.2 Computation Times and Memory Requirements

The table in Figure 8 sums up computation cost for the various scenes and focus settings with our algorithm. Kitchen 1 and 2 correspond

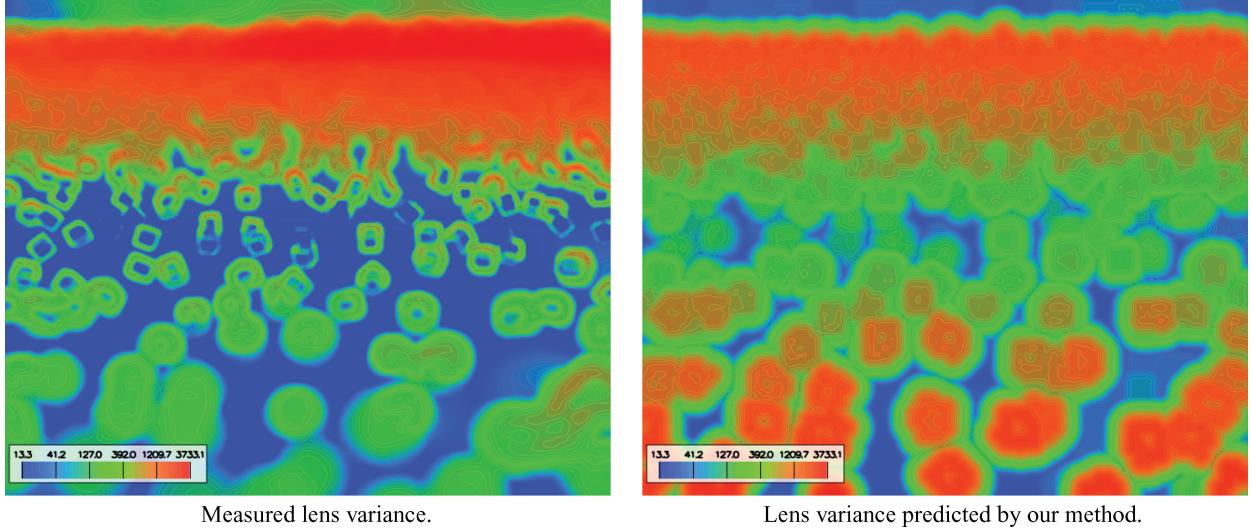


Fig. 7. Comparison of variance measured over the rays converging to each pixel of the cubes scene (left), with the variance predicted by our method (right). Both images are displayed using the same scale. Our prediction is comparable to the actual measured values both in its distribution over the image,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

Scene	Size	Frequency computation (seconds)	Path tracing (seconds)	Reconstruction (seconds)	Image space samples	Primary rays
Cubes	721 × 589	45	3,150	3	76,000	13 M
Snooker	904 × 806	90	4,500	10	119,335	25 M
Kitchen 1	897 × 679	60	7,401	8	867,000	44 M
Kitchen 2	897 × 679	60	6,849	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.

to the kitchen scene with the plane in focus set on the foreground and background, respectively. The accumulated cost of propagating, computing and splatting frequency information, along with image reconstruction (using splatting) is negligible compared to the cost of naïve stratified Monte Carlo integration over the aperture at all pixels (see table in Figure 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.

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 because samples require smaller splatting radii.

We use the total number of primary rays cast to compare our technique with the nonadaptive 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 (see Figure 9) shows the number of rays cast for similar image quality as those images used for measurements in Figure 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	1,100	2,719M	61.0
Kitchen 2	1,100	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 Figure 8. The last column shows the speedup gained by using our method, obtained by dividing the middle column by the last column in Figure 8.

Finally, the memory overhead of our algorithm is small, as we only need to store the density images for the lens and spatial sampling, which most of the time is negligible as compared to the scene, textures, and BRDF information.

### 4.3 Comparison with Adaptive Lens Sampling

We compare our approach to adaptive lens sampling based on variance estimation: For each pixel, we trace a fixed (and small) number of rays and use their radiance value to estimate their variance  $\sigma$  across the lens. Using Eq. (6) we compute, for each pixel, the required number of rays to reduce the variance of the integrated radiance through the lens under a given threshold. We set up this threshold so that the total number of primary rays is the same as the number of rays used by our method. In Figure 10 we compare the two methods on the kitchen (foreground focus setup) at different locations.

### 4.4 Comparison with Image-Based Methods

We finally compare our approach to blurring a pinhole camera image, based on the depth map. The blur is performed using a kernel of the same size as the circle of confusion for each pixel. It appears that the image-space blur solution fails in some configurations, for example, where small blurred objects are surrounded by focused regions (see the leaves of the plant for instance in Figure 11(a)),



Fig. 10. Comparison between our method (bottom row) and adaptive lens sampling based on variance estimation from a small number of samples (top row). Both methods use the same total number of primary rays for the entire image. While the adaptive lens sampling manages to uniformize the variance across 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. This explains why the images are noisier than the ones produced using our method.



Fig. 11. Comparison between blurring a pinhole camera image according to depth (a) and our technique (b). The lack of visibility information in image-space methods is a source of bias.

whereas our technique (Figure 11(b)) gives a result which is much closer to the actual brute-force algorithm.

While we admit that various improvements over this naive approach may increase the quality of the output [Potmesil and Chakravarty 1981; Kraus and Strengert 2007; Barsky et al. 2003; Zhou et al. 2007; Kolb et al. 1995], image-based algorithms always lack visibility information and are therefore necessarily biased. These methods produce approximations which are acceptable when real-time images are required, while ours produces an unbiased result.

#### 4.5 Examples

We present, in Figure 12, 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, and highly curved specular regions, respectively. 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  $897 \times 679$ , 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 for the same computation cost to stratified lens sampling with image, space stratification for antialiasing. 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. Although the total cost is the same, the naïve method exhibits more noise because it cannot adapt to the local image blurriness and wastes image samples in defocused regions.

In Figure 13 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.

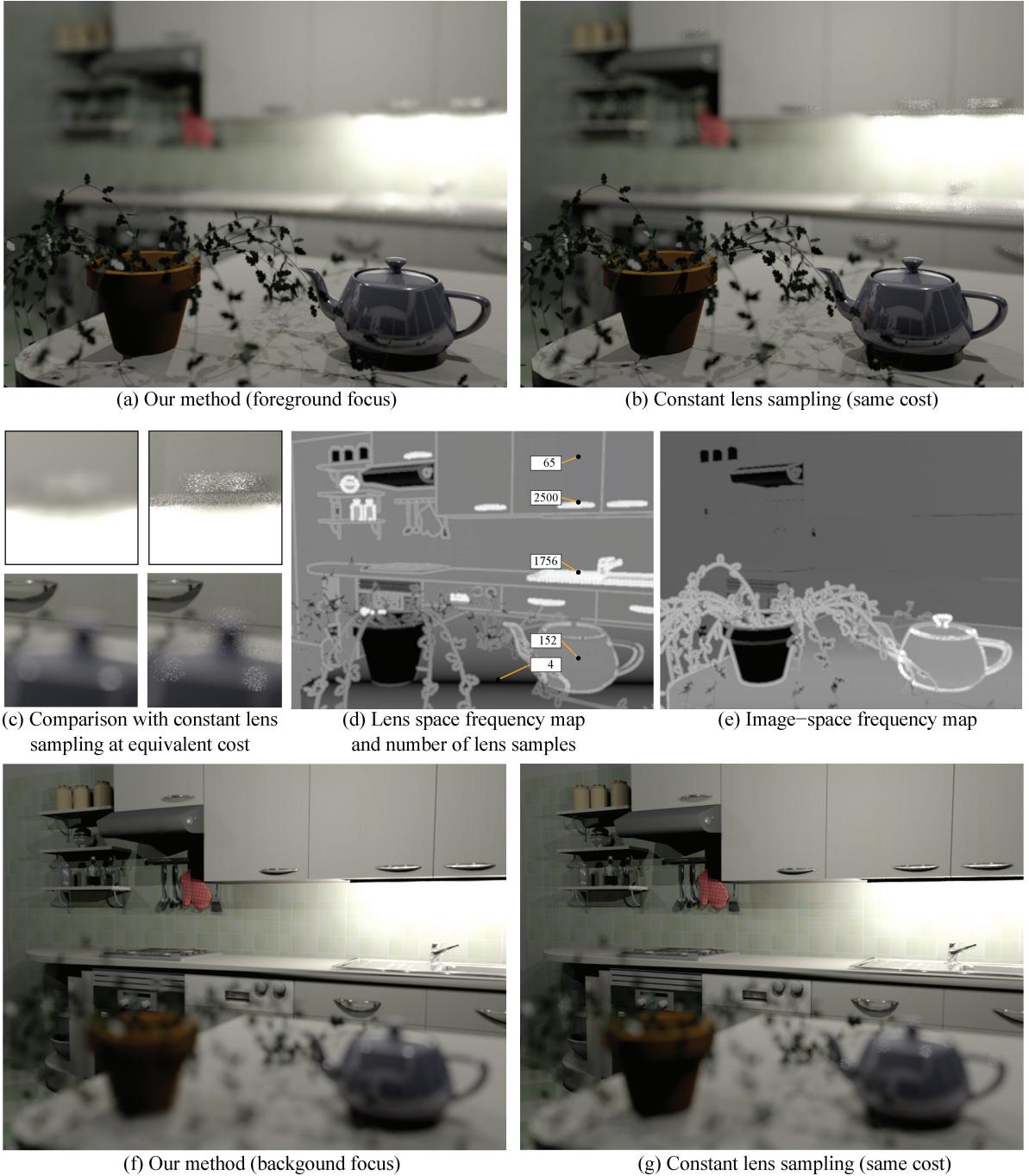


Fig. 12. Example of renderings using our method, with two settings of the focus plane (a) and (f). In both cases, we compare our result to sampling the lens constantly throughout the image and by shooting the same number of total rays as in our method. The images obtained are much more blurry in regions 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.

#### 4.6 Discussion of the Various Approximations

Our model ignores the phase information in local light field spectra, which produces 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, which could in some configurations result in no light passing at all. The convolution is then overestimated, and tends to produce higher frequencies

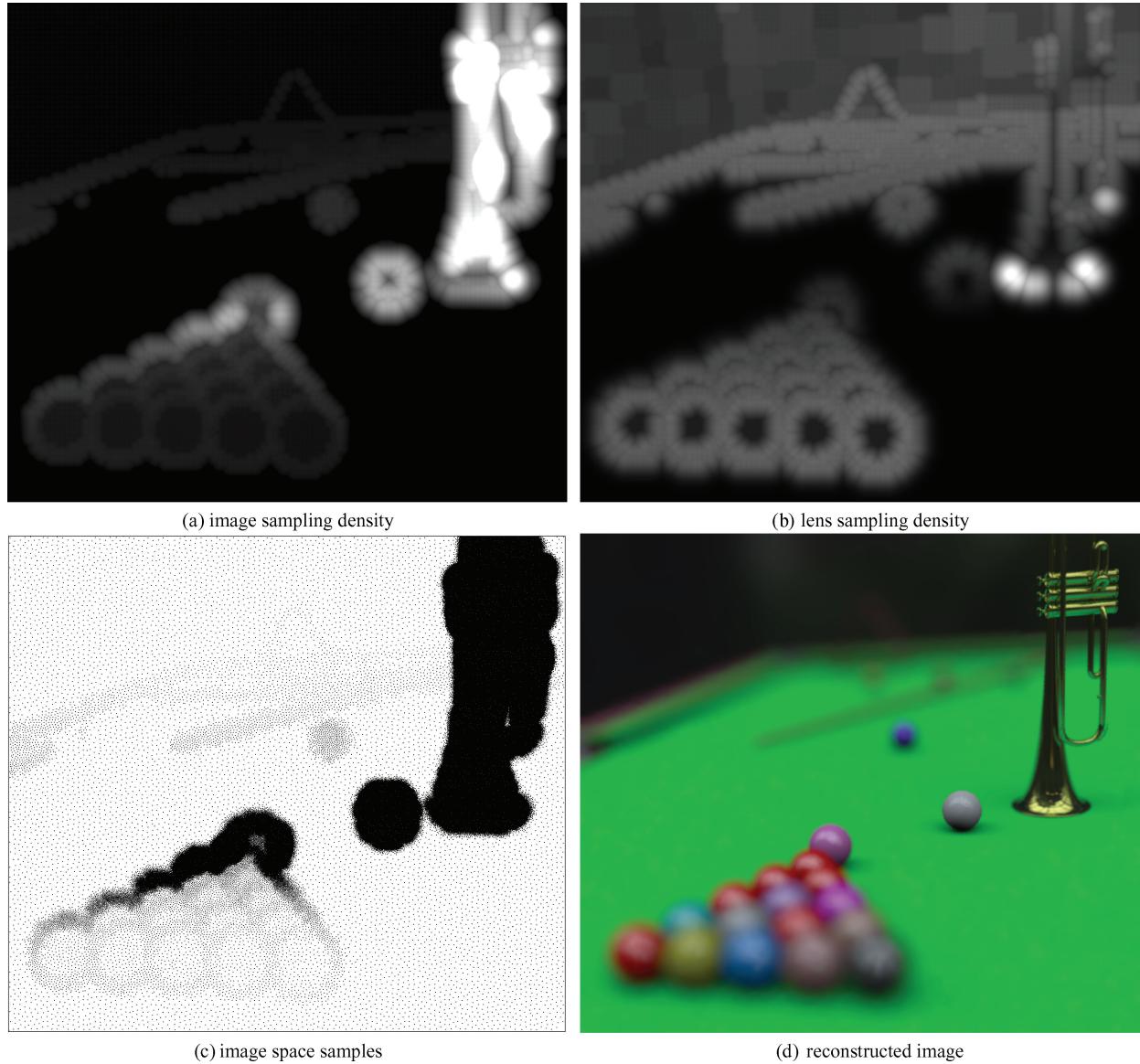


Fig. 13. (a) The image sampling density predicts that the shiny regions of the trumpet, with high curvature and in focus, need to be sampled most profusely in the image; (b) the aperture density predicts that defocused regions need to be sampled densely while the ball in focus requires very few samples over the aperture; (c) the image samples obtained from the image sampling density; (d) the image is reconstructed from scattered radiance estimates.

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. This assumption practically states that angular and spatial frequencies on a 2D slice, containing the sampling direction, do not depend on the orientation of the slice. In practice, since we only use the spectra to conservatively predict bandwidth, we do not observe artifacts due to this projection.

Our use of conservative spectra, such as maximum spatial frequencies for a texture or the angular bandwidth of a BRDF, can result in suboptimal sampling. For example, our implementation

cannot take advantage of the local bandwidth of a texture. In addition we do not take illumination into account while sampling.

## 5. CONCLUSION AND FUTURE WORK

We have proposed a practical scheme that adapts the sampling rate of both the image and the aperture, in order to simulate depth-of-field effects in image synthesis. For this, we have extended prior work on the frequency analysis of light transport to handle depth-of-field effects. We have presented a new algorithm that locally predicts both the image bandwidth as well as the variance of the radiance impinging on the lens aperture. This allows us to discover image regions that can be sampled sparsely because they are out of focus,

as well as pixels where the lens integration can be computed with a smaller number of samples because the incoming radiance has low variance. Our adaptive sampling of the image and the aperture is complementary: In the focal plane, image sampling must be high because the scene is sharp, but the lens sampling is usually low because all rays for a pixel come from the same scene point and have roughly the same radiance. In contrast, out-of-focus regions can be subsampled because they are heavily bandlimited by the depth-of-field effect, but they require more lens samples because the light rays come from different scene points.

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.

Our estimates of bandwidth and expected variance over the aperture are, however, conservative and the number of samples can be suboptimal. An exciting avenue of future work is to initialize our algorithm with more intelligent spectral samples in order to further improve efficacy. In particular, it might be desirable to predict light field spectra at points in the scene taking into account global 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.

## ACKNOWLEDGMENTS

We are grateful to the MIT/ARTIS pre-reviewers and to L. Boissieux for the Kitchen model.

## REFERENCES

AGARWAL, S., RAMAMOORTHI, R., BELONGIE, S., AND JENSEN, H. W. 2003. Structured importance sampling of environment maps. *ACM Trans. Graph.* 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 *Proceedings of the 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*. Annual Conference Series. ACM SIGGRAPH, 409–418.

BOLIN, M. R. AND MEYER, G. W. 1998. A perceptually based adaptive sampling algorithm. In *Computer Graphics*. 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*. Annual Conference Series. ACM SIGGRAPH, 307–318.

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

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

COOK, R. L., PORTER, T., AND CARPENTER, L. 1984. Distributed ray tracing. *Comput. Graphics (SIGGRAPH'84)* 18, 3, 137–145.

DURAND, F., HOLZSCHUCH, N., SOLER, C., CHAN, E., AND SILLION, F. X. 2005. A frequency analysis of light transport. *ACM Trans. Graph.* 24, 3, 1115–1126.

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

HAEBERLI, P. AND AKELEY, K. 1990. The accumulation buffer: Hardware support for high-quality rendering. *Comput. Graph.* 24, 4.

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

ISAKSEN, A., McMILLAN, L., AND GORTLER, S. J. 2000. Dynamically reparameterized light fields. In *Computer Graphics*. 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*. Annual Conference Series. ACM SIGGRAPH, 317–324.

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

MITCHELL, D. P. 1991. Spectrally optimal sampling for distributed ray tracing. *Computer Graphics* 25, 4, 157–164.

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

MYSZKOWSKI, K. 1998. The visible differences predictor: Applications to global illumination problems. In *Proceedings of the Workshop on Rendering*. Eurographics, 223–236.

NG, R. 2005. Fourier slice photography. *ACM Trans. Graph.* 24, 3, 735–744.

OSTROMOUKHOV, V., DONOHUE, C., AND JODOIN, P.-M. 2004. Fast hierarchical importance sampling with blue noise properties. *ACM Trans. Graph.* 23, 3, 488–495.

POTMESIL, M. AND CHAKRAVARTY, I. 1981. A lens and aperture camera model for synthetic image generation. *Comput. Graph.*, 297–305.

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

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

RAMAMOORTHI, R., MAHAJAN, D., AND BELHUMEUR, P. 2007. A first order analysis of lighting, shading, and shadows. *ACM Trans. Graph.* 26, 1.

SHINYA, M., TAKAHASHI, T., AND NAITO, S. 1987. Principles and applications of pencil tracing. *Comput. Graph.* 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 Trans. Graph.* 23, 3, 742–749.

SUYKENS, F. AND WILLEMS, Y. 2001. Path differentials and applications. In *Proceedings of the EG Workshop on Rendering*. Eurographics, 257–268.

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

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

ZHOU, T., CHEN, J., AND PULLEN, M. 2007. Accurate depth of field simulation in real time. *Comput. Graph. Forum* 26, 1, 15–23.

Received May 2008; accepted January 2009