Spectrally Optimal Sampling for Distribution Ray Tracing

Don P. Mitchell

AT&T Bell Laboratories
Murray Hill, NJ 07974

Abstract

Nonuniform sampling of images is a useful technique in computer graphics, because a properly designed pattern of samples can make aliasing take the form of high-frequency random noise. In this paper, the technique of nonuniform sampling is extended from two dimensions to include the extra parameter dimensions of distribution ray tracing. A condition for optimality is suggested, and algorithms for approximating optimal sampling are developed. The technique is demonstrated at low sampling densities, so the characteristics of aliasing noise are clearly visible. At supersampling rates, this technique should move noise into frequencies above the passband of the pixel-reconstruction filter.

CR Categories and Subject Descriptions: I.3.3 [**Computer Graphics**]: Picture/Image Generation; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism

General Terms: Algorithms

Additional Keywords and Phrases: Antialiasing, Distribution Ray Tracing, Nonuniform Sampling, Noise Perception

1. Introduction

In 1979, Whitted demonstrated that ray tracing could be used to simulate a number of realistic shading effects [Whitted80]. Unfortunately, ray tracing has a special difficulty with aliasing, a problem sometimes encountered when sampling signals. To focus on this issue, Whitted's algorithm can be cast into the form of a two-dimensional sampling problem. At each point (x,y) on the image plane, a brightness sample is defined by calculating the radiance of a ray from the viewpoint through that point. Assuming the image-plane coordinates range between zero and one, the image brightness is defined by the mapping:

$$f\colon [0,1]^2 \rightarrow \textbf{radiance} \qquad (1)$$

Any synthetic image might be described as (1), but the details of ray tracing have special implications: the values of f can only be evaluated at a point, and it is virtually impossible to symbolically intergrate or low-pass filter the function. In other words, the signal can be sampled but generally cannot be prefiltered to avoid aliasing. An interesting approach to this problem is *nonuniform sampling* which an yield aliasing in the form of high-frequency random noise [Dippé85, Cook86, Mitchell87].

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

An elegant extension of Whitted's algorithm is *distribution ray tracing* (previously "distributed ray tracing"), introduced by Cook, Porter and Carpenter [Cook84]. Their algorithm simulates motion blur, shadow penumbras from finite-area light sources, depth-of-field effects, and glossy reflections from partially polished surfaces. This is achieved by sampling in an additional set of parameter dimensions. For example, an object in motion will have a position in the scene parameterized by the time t, and motion blurred pixels can be calculated by averaging over many different samples of t. Depth-of-field effects are associated with a finite aperture on the camera and are simulated by deflecting rays through different points on the lens, parameterized by two more variables a,b. Glossy reflection results from varying the direction of a surface normal, as if the surface were made up of randomly distributed microscopic facets parameterized by an orientation θ,ϕ.

With these extra parameters, distribution ray tracing defines a multidimensional brightness function $f'(x, y; t, u, \cdots)$. A sample of this function is evaluated by performing a Whitted-style ray tracing operation. However, first moving objects would be transformed to their location at time t, a point light source is defined by (u,v) representing a sample of the area light, the primary ray from the camera is deflected through a focal point from a position (a,b) on the lens, etc. Once the scene is prepared for a given set of parameter values, a ray-tracing calculation can be done. Assuming x, y, and $D-2$ parameters range from zero to one, we have the brightness mapping:

$$f'\colon [0,1]^D \rightarrow \textbf{radiance} \qquad (2)$$

The two-dimensional image is an integration over the parameters

$$f(x,y) = \int_0^1 \int_0^1 \cdots \int_0^1 f'(x, y; t, u, \cdots)\, dt\, du \cdots \qquad (3)$$

Additional integration or convolution (with a filter) may be done in x and y to define a bandlimited image function $i(x,y)$ suitable for alias-free digitization. The integration in (3) cannot be evaluated analytically, but the process of distribution ray tracing estimates $f(x,y)$ by averaging many samples per pixel. This process can be viewed as a *Monte Carlo integration*, or as a classical *statistical sampling problem* of estimating the mean value of f in a region of the image plane, or it can be viewed as a sampling problem in the *signal processing* sense. These viewpoints are not independent, and all of them can be found with varying degrees of emphasis in discussions of distribution ray tracing [Cook84, Lee85, Cook86, Kajiya86, Shirley90].

The question investigated in this paper is how to extend the techniques of nonuniform sampling, used in Whitted-style ray tracing, to the multiple dimensions of distribution ray tracing. This is not simply the problem of generating a D-dimensional image from samples of f', which might be a obvious extension of the two-dimensional methods. We are still interested in the characteristics of noise in a two-dimensional image, and we expect

©1991
ACM-0-89791-436-8/91/007/0157 $00.75

the parameter dimensions $t, u, v, \cdots$ to play a qualitatively different role than the image coordinates x, y.

2. Incomplete Block Sampling Designs

An important problem in distribution ray tracing is how to choose samples effectively to produce the highest quality image with the fewest rays. We might simply choose samples randomly with a uniform distribution in $[0,1]^D$. In *sequential sampling*, random samples are made until we are statistically confident that the average value has a low variance [Lee85, Kajiya86]. This procedure is usually improved by *stratified sampling*, where the interval $[0,1]$ is divided into N levels, dropping a random sample into each subinterval. This spreads the samples out more evenly and often results in a lower variance of the average.

In the two-dimensional antialiasing problem of Whitted-style ray tracing, each pixel area $[0,1]^2$ can be divided into $N \times N$ subsquares for stratified sampling. This is more or less the same as jittered sampling, a common approach to the antialiasing problem [Dippé85, Cook86, Kajiya86, Painter89]. However, this is not a practical approach to sampling the parameter space of distribution ray tracing because of the high number of dimensions. Stratification of all D dimensions would result in N^D blocks to be sampled. This could easily be tens of thousands of samples per pixel, many more than would be reasonable or necessary. In practice, stratification has been applied to distribution ray tracing, but with *incomplete block sampling designs* that do not fully populate the N^D blocks.

Cook created incomplete block designs by subdividing the pixel area into an $N \times N$ mesh of subsquares. The time dimension was divided into N^2 levels, and pairs of "area-like" parameters like (u,v) and (a,b) were subdividing into $N \times N$ meshes. N^2 samples are then made which projected onto each subsquare of the pixel once, each level of time once, each subsquare of the (u,v) dimension once, etc. Thus only N^2 of the possible N^D blocks are occupied by a sample. Little has been said about how these associations between blocks should be chosen, but it is clear that linear correlation between parameter values should be avoided [Cook86]. Linear correlation would mean a tendency for samples to fall on hyperplanes in $[0,1]^D$ which could cause aliasing. The visual consequences of parameter correlation are objectionable and conspicuous.

Shirley describes another incomplete block design called "N-rook" sampling, where N out of N^D blocks are populated [Shirley90]. Let $\pi_1, \pi_2, ..\pi_{D-1}$ be permutations of the sequence $(0,1,\cdots,N-1)$. Then we choose one sample in each subinterval of each parameter dimension. The n^{th} sample is placed in level n of the x dimension, in level $\pi_1(n)$ of the y dimension, level $\pi_2(n)$ of the t dimension, etc. Once again, little is known about what are good or bad choices for the permutations, except to avoid linear correlation. An example in two dimensions is shown below:

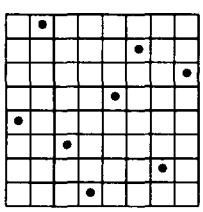

Figure 1. An 8-Rook Sampling Design in Two Dimensions

There seems to remain an important piece of unfinished business. We do not really know very much about what constitutes a good sampling design versus a bad one. Avoiding linear correlation is known to be important. Explicit in the two incomplete block designs described above is the property that the sampling pattern is "good" when projected onto certain lower-dimensional planes or axes. For example, the N-rook patterns are fully populated stratified designs when projected onto any coordinate axis.

Linear correlation could be avoided by randomly choosing sampling designs of either Cook's style or Shirley's. Moreover, sequential sampling (i.e., sampling until statistical confidence is achieved) is probably capable of giving satisfactory image quality [Lee85], whether the sampling design is good or not. The danger is that many more samples might be computed than are necessary.

3. Nonuniform Sampling in Two Dimensions

Before tackling the problem of sampling in D dimensions, it will be useful to review the two-dimensional problem of sampling in the (x,y) dimensions. This is the problem in Whitted-style ray tracing. Typically, the image is "supersampled" at a high rate (by casting rays), and then filtered and resampled to a lower rate to produce the pixels of a digital image. The filter may be an average over the pixel area, or it may be a more sophisticated low-pass filter. The process of sampling is represented mathematically by multiplication of the image signal with delta-function pulses, as diagramed below:

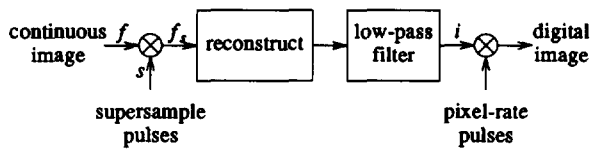

Figure 2. Conversion of Ray-Casting Samples into Pixels

The reconstruction filter interpolates samples to recreate a continuous image. The low-pass filter makes sure that image is bandlimited so aliasing will not result when it is resampled at the pixel rate. These two filters are traditionally combined into one, but when supersampling is nonuniform, it is often the case that reconstruction and low-pass filtering are distinctly separate stages [Mitchell87, Painter89].

If the supersampling pattern is nonuniform, and its spectrum has certain characteristics, the sampling error (or aliasing) will take the form of random noise at high frequencies. This is desirable for two reasons. If noise is concentrated in the high frequencies, more of it will be attenuated by the low-pass filter pictured in Figure 2. Secondly, randomness and high frequency both help to make the noise less perceptible to a human observer.

This can be understood by looking at the sampling process in the frequency domain. Let $f(x,y)$ represent the continuous image, $s(x,y)$ represents the sampling pattern (delta functions), and let $r(x,y)$ be the combined reconstruction/low-pass filter. F, S, and R will represent the corresponding spectra. In the spatial domain, the sampling and filtering process is expressed by:

$$i(x,y) = r(x,y) * [f(x,y) \cdot s(x,y)] \qquad (4)$$

And in the frequency domain:

$$I(\omega_x,\omega_y) = R(\omega_x,\omega_y) \cdot [F(\omega_x,\omega_y) * S(\omega_x,\omega_y)] \qquad (5)$$

where $\cdot$ and $*$ represent multiplication and convolution respectively.

The reconstruction filter is described above as a linear low-pass filter, which is ideal for uniform samples but can give a distorted reconstruction of nonuniform samples. Nonuniform reconstruction for images is not perfectly understood; but in practice, nonlinear or space-varying filters (which are not representable by a convolution) give better results [Dippé85, Mitchell87, Marvasti87]. The result is still some type of low-pass filter (i.e., a "smooth" surface interpolating the sample spikes). When the reconstruction and low-pass stages (in Figure 2) are implemented separately, the low-pass stage

could be a linear filter [Painter89]. If the supersampling rate is much higher than the pixel rate, the linear low-pass stage should dominate the behavior of the system. We will model the reconstruction as a linear low-pass filter for the purposes of qualitative analysis.

The spectrum of the sampling pattern S will be a delta-function spike at the origin (the DC component) and some pattern of noise surrounding it. The convolution $F*S$ (shown in Figure 3) of a image spectrum with the nonuniform-sampling spectrum gives a copy of the true image spectrum (the symmetric shape at the center of the figure) and a halo of noise energy (represented by the scattered dots). The low-pass filter R (represented by the dotted box) attenuates energy outside its bounds.

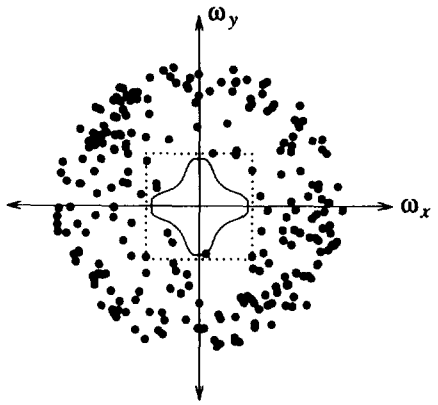

Figure 3. Spectrum of Nonuniformly Sampled Image

If the spectrum of the sampling pattern has energy concentrated in high frequencies, then the halo of noise will be pushed farther out from the origin, and more of it will be outside of the pass band of the filter. The best known patterns having this high-frequency characteristic are the *Poisson-Disk* stochastic point processes. These patterns are random, but include a constraint that no two points can be closer than some minimum distance (as if each point was surrounded by a hard disk) [Ripley77]. The spectral consequences of this sampling pattern were first investigated by Yellott, who found this arrangement in the photoreceptors of monkey retinas [Yellott83].

More commonly used are patterns based on *jitter* processes. These are formed by randomly perturbing the points in a periodic uniform lattice. Jitter sampling contains more low-frequency energy in its spectrum than Poisson-disk patterns, and images produced with it have a more grainy appearance at low sampling rates [Mitchell87]. However, jitter sampling is easy to generate, and straightforward adaptive-sampling schemes exist for generating jitter samples at variable density [Dippé85, Cook86, Kajiya86, Painter89]. Some of these methods could also be described as stratified sampling.

4. Sequential Poisson-Disk Sampling

Poisson-disk samples are typically generated by a "dart-throwing" algorithm which is computationally expensive and which makes it difficult to control the final density of samples (one initially choses the hard-disk diameter, not the desired sample density) [Dippé85, Mitchell87].

With the following new algorithm, it is possible to generate good high-frequency sampling patterns with sequentially increasing density. Begin by choosing the first sample at random in a region. To add the $(n+1)^{th}$ sample, generate mn uniformly distributed random candidate points (where m is a constant parameter). For each of these random points, compute the distance to the closest of the n points already in the pattern. Then chose the candidate point with the largest closest-point distance, and add it to the pattern. By scaling up the number of random candidate points, in proportion to n, we maintain a constant ratio m of candidates to pattern points in the process. Thus we expect the statistics of the pattern (the autocorrelation, etc) to also scale and remain similar as the sample density increases. The high-frequency quality of the pattern increases with m.

This is an $O(n^2)$ algorithm, but it is an improvement over the poorly defined termination of the dart-throwing algorithm (which runs until it cannot add new samples). The speed was improved dramatically by using grid methods for the nearest-neighbor calculation. This point process is not strictly hard-disk, because it is possible (although unlikely) for samples to lie very close together. However, the resulting patterns are excellent if m is not too small. The following figure shows some snapshots from this process, using $m = 10$:

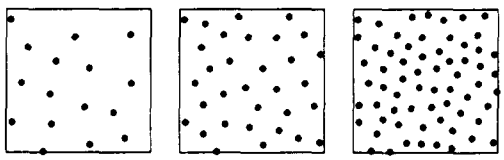

Figure 4. Sequential Generation of High-Frequency Samples

It is sometimes useful to perform this algorithm with wrap-around boundary conditions, so the pattern can be replicated periodically over the plane (with much longer period than the pixel rate, of course). The algorithm can be extended to higher dimensions, and it could also be used to generate isotropic high-frequency sampling patterns on the surface of a sphere. That may be useful because area light sources and glossy reflections require *sampling solid angles*. The alternative of stratified sampling of latitude and longitude is not isotropic because strata near the pole are very different in shape than equatorial strata. The concept of choosing the best samples from random candidates will be used again in the algorithms applied to distribution ray tracing.

By "hard disk" we usually mean a circular region of avoidance around each sample. By using an ellipse or other shape, the spectrum of the pattern can be made anisotropic in some fashion. The human visual sensitivity extends higher into vertical and horizontal frequencies than it does into diagonal frequencies, so a Poisson-diamond pattern might be better than Poisson-disk. Bouatouch *et al* support this idea in their experiments with uniform quincunx sampling [Bouatouch91]. This is an issue that could be studied further.

5. Motion Blur and Spatiotemporal Sampling

If two-dimensional sampling can push noise into high frequencies, can the same effect be obtained while sampling the extra parameters of distribution ray tracing? Let us begin by considering motion blur effects, where a single extra parameter t is added. This is not an obvious three-dimensional generalization of the problem of the previous section. $f'(x,y;t)$ is sampled in *three dimensions*, but we are still concerned with the resulting sampling noise in the *two-dimensional* image $i(x,y)$.

To derive the spectrum of $i(x,y)$, let $f_s(x,y,t) = f'(x,y;t)\cdot s(x,y,t)$ be the sampled multiparameter image function, where $s(x,y,t)$ is a distribution of delta functions in space/time. The sampled image function is low-pass filtered spatially with $r(x,y)$, and integrated over an exposure-time interval for motion blur:

$$i(x,y) = r(x,y) * \int_0^1 f_s(x,y,t)\, dt \qquad (6)$$

The spectrum is a little easier to derive if we replace the integration over a time interval with the equivalent operations of convolution with a box function in t followed by sampling one slice through t. Then using the Convolution Theorem, we find the spectrum of $i(x,y)$ to be:†

$$I(\omega_x,\omega_y) = R(\omega_x,\omega_y) \int_{-\infty}^{\infty} Sinc(\omega_t/2\pi)\cdot F_s(\omega_x,\omega_y,\omega_t)\,d\omega_t \quad (7)$$

The important difference between the expression for the static-image spectrum (5) and the spectrum of the motion-blurred image (7) is the integration over ω_t. This means the three-dimensional spectrum (at least, the portion passed by the R and *Sinc* filters) will be projected onto the spatial (ω_x,ω_y) plane. Ideally, we would like the noisy part of this spectrum to be pushed out of a cylindrical region around the ω_t axis, so its projection will contain only the highest possible spatial noise frequencies.

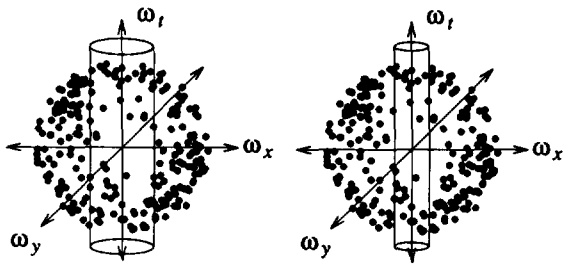

Figure 5. Cylinders of Medium and Low Spatial Frequencies

This suggests that the best general sampling pattern will be one with little power in the low-spatial-frequency region around the ω_t axis. Figure 5 depicts the spectrum of the sampling pattern with cylindrical regions around the ω_t axis enclosing spatial frequencies below some bandlimit. The wide cylinder on the left contains frequencies up to some medium value, and the cylinder on the right represents a lower bandlimit. We would like these cylinders to be as vacant of power as possible. In fact, the practical requirement is to have the power within each cylinder be concentrated at the highest possible frequencies. It is also important to give the highest priority to removing the lowest spatial frequencies, so we require the power in the right-hand cylinder to be concentrated at higher temporal frequencies than in the left-hand cylinder.

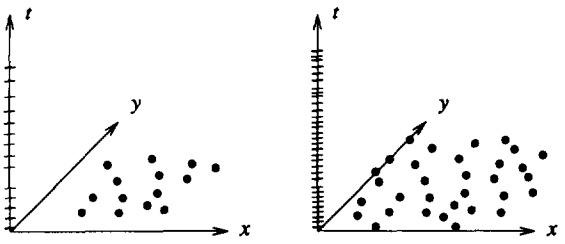

Figure 6. Marginal Distributions of Samples in Space and Time

These conditions in the frequency domain imply some conditions on the arrangement of samples in space and time. The two graphs in Figure 6 illustrate situations corresponding to the spectra in Figure 5. They show the projection of samples onto the spatial plane (dots) and onto the time axis (tick marks). We are not yet certain where these samples should be in space/time, but we will be able to give conditions on their projections into space and time.

If there is no movement in a region of the image, then only the spatial projection of the pattern is important, so we could begin by constraining it to form an optimal distribution, like Poisson-disk.

We are interested in the power contained within cylinders of spatial frequencies, in Figure 5, and in the temporal-frequency distribution of that power. Imagine that we have convolved the spectrum with a cylinder and sampled the result on the ω_t axis (this is equivalent to averaging over spatial frequencies inside the cylinders). That operation corresponds approximately to selecting the samples within a cylindrical region of space, and considering their time distribution. The narrower cylinder of frequencies in Figure 5 corresponds to a wider region of space in Figure 6. The temporal distribution of samples, shown in Figure 6, represent one-dimensional patterns of the highest possible frequency (such as Poisson-rod distributions).

Therefore, the desired property of space/time sampling patterns is that in any cylindrical region of space, the distribution of samples in time will be a high-frequency pattern. An interesting consequence of this is that samples which are adjacent in space should differ greatly in time or other parameter coordinates. This is quite different from the most obvious three-dimensional analog of Poisson-disk sampling. A Poisson-sphere point distribution would not necessarily have this property of high-frequency time distribution.

6. A Scanning Sample-Generation Algorithm

A simple scanning algorithm is one possible way to generate sampling patterns which approximate the conditions described in the previous section. Begin by stratifying the x and y dimensions into a mesh of subsquares, assuming that one jittered sample is contained in each. The goal is to assign each sample a value of the parameter t. This is done in scanning order, from left to right, and top to bottom.

S	S	S	S	S
S	P	P	P	S
S	P	•		

Figure 7. Neighborhood of the Next Unprocessed Subsquare

Figure 7 illustrates the situation at some point in the scanning process. We wish to choose a parameter value for the subsquare containing the dot. In the 5×5 region surrounding the dot, some subsquares above and behind (indicated by "P" or "S") have already been assigned t values. Call these the P-cells and S-cells (meaning primary and secondary).

We would like the new t value to fit into a high-frequency Poisson-rod distribution, as shown in Figure 6. In a manner reminiscent of the sequential Poisson-disk algorithm of section 4, we generate a set of primary candidate t values with uniform random distribution in $[0,1]$. For example, let us say we generate 100 primary candidates. The candidates are sorted by their maximum closest distance to the t values of the P-cells. Distance is defined with wrap-around boundary conditions, so the pattern can used periodically from frame to frame.. From the sorted list, we might pick (for example) the 10 with largest max-min distance. Any one of these 10 values should be a good choice to complete a coarse Poisson-rod distribution as suggested on the left of Figure 6.

The set of 10 values selected above are now considered as secondary candidates. For each secondary candidate, compute the maximum

† This actually corresponds to the integration of time from $-\frac{1}{2}$ to $\frac{1}{2}$. A phase-shift factor could be added to reflect integration from 0 to 1.

$sinc(x) = sin(\pi x)/\pi x$.

closest distance to the t values of the S-cells, and pick the one with the largest max-min distance. This should be a good choice to complete the denser distribution as suggested on the right of Figure 6. We are trying to meet two constraints, picking 100 primary candidates and evaluating how well they match the situation on the left of Figure 6, then selecting 10 secondary candidates to match the conditions on the right.

7. Experiments with Scan-Generated Sampling Patterns

Figures 8 and 9 demonstrate the use of this sampling pattern on a ray-traced scene containing spinning wheels. Figure 8 was made by choosing t values with a uniform random distribution. Figure 9 uses the spectrally optimized t values generated by the scanning algorithm. In fact, a 32×32 pattern of samples was generated and replicated periodically on the plane. Both image were generated with just *one sample per pixel* so the sampling noise can be seen clearly. Essentially, we are looking at the raw supersamples f_s, which would be passed into the filter stages of Figure 2 in order to make an antialiased digital image.

Figures 10a and 10b show the corresponding noise spectra (with lighter shades indicating higher power). These were obtained by subtracting Figures 8 and 9 from a reference image (generated with 100 rays per pixel), to create an error image. The discrete Fourier transform of the error images show a typical "white noise" spectrum in Figure 10a, corresponding to the random sampling. However, Figure 10b shows a considerable concentration of power in the higher frequencies.

Even though the mean square error of Figures 8 and 9 are about the same, the frequency distribution of power has a large impact on subjective appearance. A series of randomly-sampled images like Figure 8 were generated using from 1 to 9 samples per pixel, and several expert observers were asked to select the best comparison with Figure 9. Figure 9 was obviously better looking that 1 sample per pixel and obviously worse than 9. The consensus was that three or four random time samples per pixel were required to match the subjective quality of Figure 9.

Figures 11 and 12 show a similar comparison of the technique applied to depth-of-field effects. Figure 11 was generated with uniformly random values of (a,b), the parameters controlling the deflection of primary rays through the camera aperture. Figure 12 used parameter values generated by the scanning algorithm. The only difference from scanning generation of t values is the use of a two-dimensional Euclidean distance for the max-min distance selections. Once again, both figures were generated with one ray per pixel. Figure 11 shows the clumpy pattern of sampling error characteristic of white noise, and Figure 12 shows the finer structure of high-frequency noise.

A critical observer may notice, from the point-spread, that the simulated camera has a square lens. There is no special problem in simulating a round lens, which should have been done if this were not a simple experimental ray tracer.

Figures 13 and 14 demonstrate another two-parameter experiment, using parameter values to perturb the normal vector of a surface and simulate glossy reflection. Figure 13 uses random perturbations and Figure 14 uses scan-generated parameters.

These images provide evidence that the condition for optimal parameter sampling is correct. The scanning sample-generation algorithm should not be thought of as a definitive way to generate optimal samples, however. It is an *ad hoc* way to generate a pattern with approximates the conditions defined in section 5, but only in a 5×5 region, and probably not with perfect isotropy. There is a great deal of opportunity for experiment and improvement.

8. N-Parameter Sampling

Suppose an image of the spinning-wheels picture (seen in Figures 8 and 9) is generated with motion blur and also an area light source, creating shadows with penumbras. Using an optimized pattern of t parameters ensures that the spinning spokes of the wheels are well sampled as in Figure 9. Using an optimized pattern of (u,v) parameters ensures that the penumbra around the rim of the wheels has good high-frequency sampling noise. However, in regions where both distribution ray-tracing effects are combined—in a moving penumbra—the sampling noise has a coarser white-noise appearance.

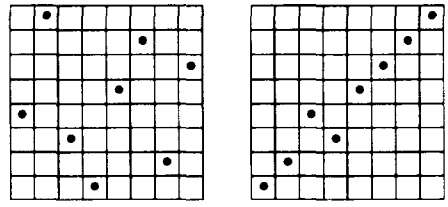

Figure 15. 8-Rook Sampling Patterns

It is not sufficient to optimize the t and (u,v) distributions of samples alone. The joint distribution of (t,u,v) matters. Figure 15 demonstrates this concept. Here, two patterns both have the same projected ("marginal") distributions in u and v (in this case, uniform periodic). However, the overall joint distributions of the two patterns are very different. Suppose a signal was sampled with the pattern on the right. If the signal contains only variations in the u or only in the v dimensions, it may be sampled well enough. But if the image contains variation along the diagonal perpendicular to the row of samples, severe aliasing might occur. In distribution ray tracing, a similar situation can occur, and aliasing caused by poor joint distribution of the parameter samples can be projected onto the image.

It is also not sufficient to just optimize the joint distribution of (t,u,v) without considering the marginal distributions of t and (u,v). The scanning algorithm was used to generate samples in (t,u,v) with Poisson-sphere joint distributions, and this resulted in relatively poor image quality. The region of moving penumbra was much improved, however.

The best image quality in the moving-penumbra test was achieved by generating sampling patterns in which both the joint distribution and the marginal distributions are spectrally optimized. Figure 16 shows the spinning wheels image, using a sampling pattern which combines a joint distribution of (t,u,v) which is Poisson-sphere, and a marginal distribution of t which is Poisson-rod. This suggests that as the parameter space becomes higher in dimension, sampling patterns must be found which meet the conditions of Figure 6 in a combination of marginal and joint distributions. This combination of conditions was met by extending the scanning algorithm to select a series of primary, secondary, and tertiary candidates.

9. Adaptive Sampling

No matter how optimal a supersampling pattern may be, we cannot ignore the computational efficiency of adaptive sampling. It is often the case (except in the most complex scenes) that many portions of an image can be sampled at relatively low density. A simple solution might be to use a few discrete levels of sampling density. The two-level sampling algorithm described by the author in [Mitchell87] was very easily adapted to use a five-dimensional (x,y,t,u,v) periodic pattern of 1024 samples.

In that scheme, an image is sampled at a low base rate. It is a good idea to make this base rate selectable by the user, and typically one or a few samples per pixel area are sufficient. The results of the

base-rate sampling are then used to estimate local bandwidth and identify regions that require sampling at a higher rate.

Variable sampling rates can be achieved by simply scaling the stored pattern. If the pattern is optimal in the sense discussed above, it will only be necessary to scale the pattern in x,y. As the sampling rate per pixel area increases, the rate per parameter dimension should also increase and the distribution should remain spectrally optimal, as indicated in Figure 6.

10. Conclusions

Optimal nonuniform sampling is a familiar approach to the aliasing problem in Whitted's ray-tracing algorithm. This paper takes a first step in extending this technique to the multidimensional algorithm of distribution ray tracing. This is nontrivial for two reasons. First, a simple extension of the stratified/jitter sampling techniques to higher dimensions requires a number of samples exponential in the dimension. Secondly, this is not simply the problem of generating a D-dimensional image, which would be an obvious extension of the two-dimensional theory.

An analysis of the sampling problem in distribution ray tracing suggests a criteria for sampling patterns that can force aliasing noise into higher frequencies. Samples contained in any circular region in space (on the x,y image plane) should have parameter values which form a pattern of the highest possible frequency. It appears that in addition to requiring the overall joint distribution of parameter values to be high-frequency (e.g., a Poisson-hypersphere distribution), it is important to insure that certain marginal distributions are of the highest frequency (e.g., time values being Poisson-rod, (u,v) parameters being Poisson-disk, etc.). This problem could be studied further.

A scanning sample-generation algorithm is proposed, which gives sampling patterns which locally approximate the optimal. This was good enough to demonstrate the correctness of the optimality condition in a number of test images. Further work could be done on better sample-generation algorithms, perhaps using exhaustive Monte Carlo search.

Much more difficult sampling problems arise in the current most advanced rendering algorithms. These problems are made explicit in several recent works [Kajiya86, Heckbert90, Shirley90].

11. Acknowledgements

I would like to thank John Amanatides, Pat Hanrahan, Paul Heckbert, Peter Shirley, and the SIGGRAPH reviewers for their helpful comments on this work and discussions of sampling issues.

12. References

[Bouatouch91] Bouatouch, K., Bouville, C., Tellier, P. Low sampling densities using a psychovisual approach. *Eurographics '91*, to appear.

[Cook84] Cook, R. L., Porter, T., Carpenter, L. Distributed ray tracing. *Computer Graphics*, 18, 3 (July 1984), 137-145.

[Cook86] Cook, R. L. Stochastic sampling in computer graphics. *ACM Trans. Graphics*, 5, 1 (January 1986), 51-72.

[Dippé85] Dippé, M. A. Z. and Wold, E. H. Antialiasing through stochastic sampling. *Computer Graphics*, 19, 3 (July 1985), 69-78.

[Heckbert90] Heckbert, P. S. Adaptive radiosity textures for bidirectional ray tracing. *Computer Graphics*. 24, 4 (August 1990), 145-154.

[Kajiya86] Kajiya, J. T. The rendering equation. *Computer Graphics*, 20, 4 (July 1986), 143-150.

[Lee85] Lee, M., Redner, R. A., Uselton, S. P. Statistically optimized sampling for distributed ray tracing. *Computer Graphics*, 19, 3 (July 1985), 61-67.

[Marvasti87] Marvasti, F. A. *A Unified Approach to Zero-Crossings and Nonuniform Sampling*, Nonuniform Press (1987).

[Mitchell87] Mitchell, D. P. Generating antialiased images at low sampling densities. *Computer Graphics*, 21, 4 (July 1987), 65-72.

[Painter89] Painter, J., and Sloan, K. Antialiased ray tracing by adaptive progressive refinement. *Computer Graphics*, 23, 3 (July 1989), 281-288.

[Ripley77] Ripley, B. D. Modeling spatial patterns. *J. Roy. Statist. Soc. B*, 39, (1977), 172-212.

[Shirley90] Shirley, P. Physically based lighting calculations for computer graphics. PhD Thesis, University of Illinois, (1990).

[Whitted80] Whitted, T. An improved illumination model for shaded display. *Comm. ACM*, 23, 6 (June 1980), 343-349.

[Yellott83] Yellott, J. I. Jr. Spectral consequences of photoreceptor sampling in the rhesus retina. *Science*, 221, (1983), 382-385.

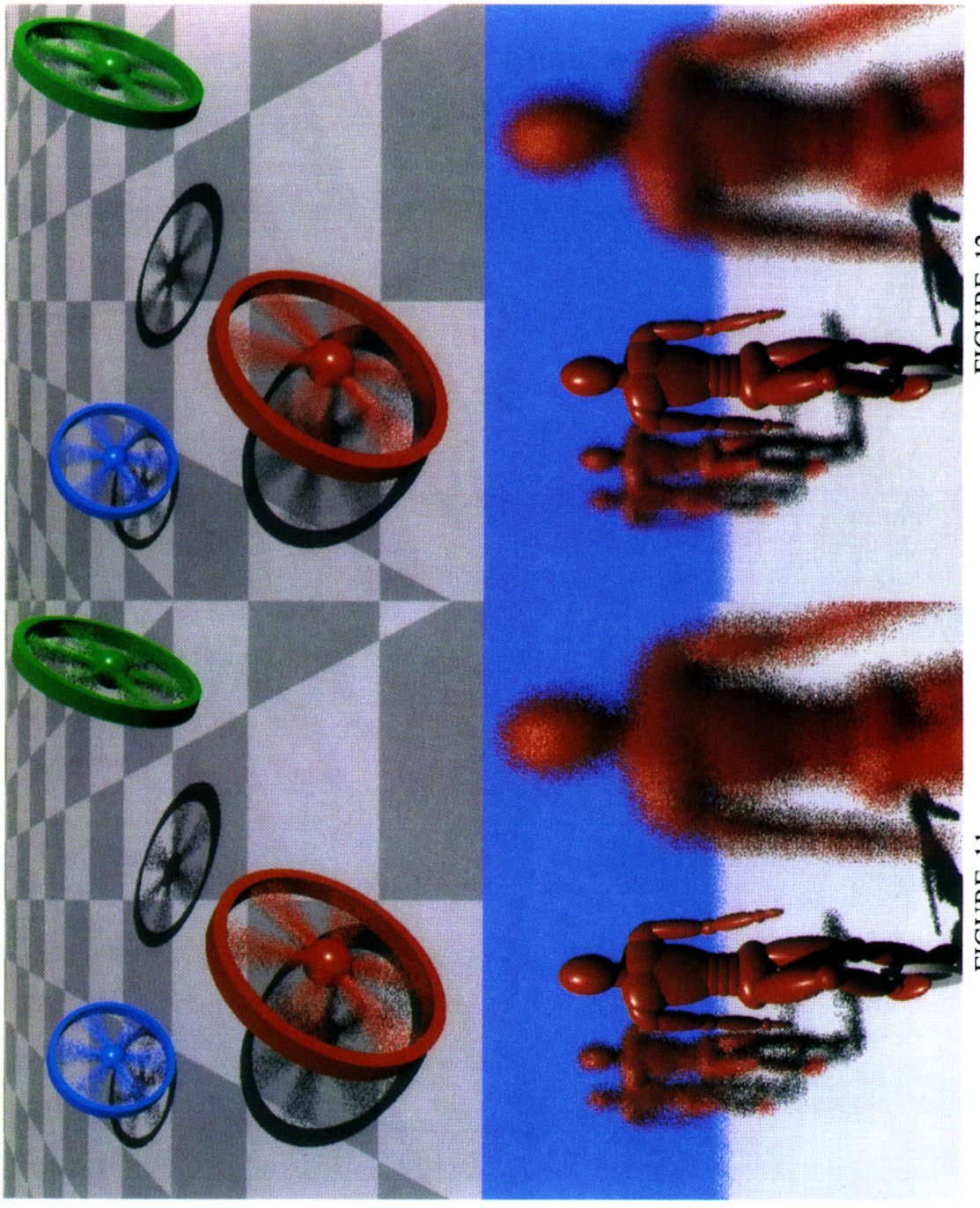

FIGURE 8

FIGURE 9

FIGURE 11

FIGURE 12

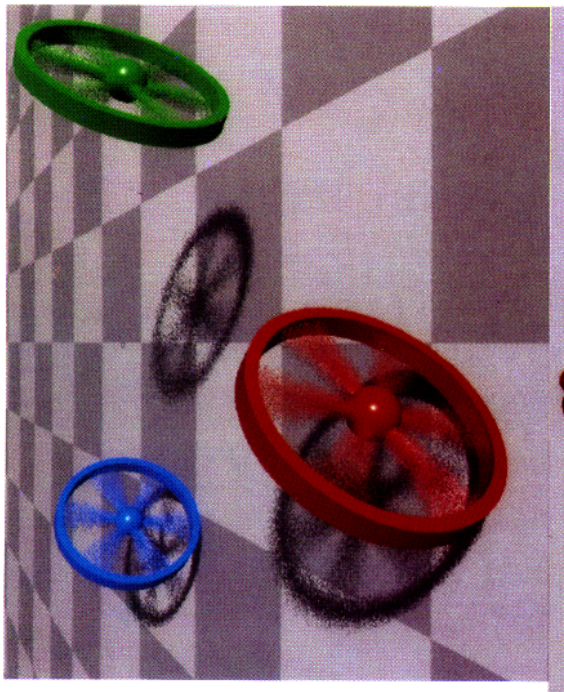

FIGURE 16

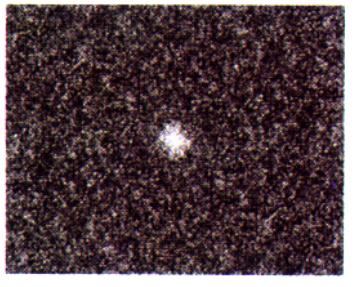

FIGURE 10a

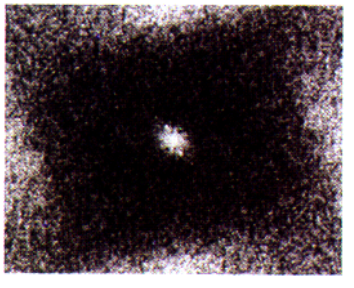

FIGURE 10b

FIGURE 14

FIGURE 13