Ray Tracing and Irregularities of Distribution

Don P. Mitchell

AT&T Bell Laboratories
Murray Hill, NJ

ABSTRACT

Good sampling patterns for ray tracing have been proposed, based on the theories of statistical sampling or signal processing. It is assumed that such patterns will be more efficient, but in what sense? For some classes of image functions, the theory of irregularities of distribution proves that *N*-point sampling patterns exist which can estimate pixel values asymptotically better than the $O(N^{-1/2})$ of random sampling. Experiments with known sampling patterns from graphics suggests that some already have superior convergence properties.

1. Introduction

Many algorithms for realistic image synthesis employ ray tracing in some form, to sample visibility or to sample the radiance field. The computational cost associated with these operations depends on the cost of intersecting a ray with the model and on the number of rays traced. This paper will discuss the second factor—how many rays must be cast to achieve a given level of image quality. It is understood that some arrangements of rays may yield sampling error with more visually acceptable characteristics [Dippé85, Cook86, Mitchell87]. The study of irregularities of distribution show that some arrangements of rays may also be asymptotically more efficient at estimating pixel values.

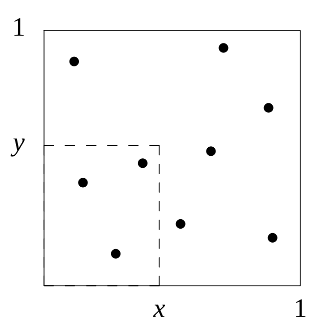
The most sophisticated rendering algorithm that can be easily analyzed is *distribution ray tracing* [Cook84]. This technique can achieve antialiased pixels, motion blur, focusing in depth of field, penumbra shadows from area light sources, and other effects. A pixel value is defined to be a multi-dimensional integral over pixel area, exposure time, lens aperture, the solid angles subtended by area light sources, and possibly other parameters. This integral is typically estimated by averaging over many Whitted-style ray-tracing operations [Whitted80]. Although Whitted's algorithm actually solves a simple transport problem [Kajiya86], and may involve the recursive generation of secondary rays, it is so well understood that we will view it abstractly as a source of point samples in the domain of integration. It is convenient to assume the parameter space is a k-dimensional unit hyper cube $\mathbf{I}^k$, although some pairs of parameters might better be viewed as points on the surface of a sphere.

The problem now is to compute pixel values by sampling the parameter space $\mathbf{I}^k$ and averaging the resulting radiance (or spectral radiance) values. There are several viewpoints for this problem. Many have viewed this as a problem in *statistical sampling theory*, or *Monte Carlo integration* [Lee85, Kajiya86, Purgathofer86, Painter89]. The goal is to estimate the mean value of the radiance from samples. The variance of the sample mean is also estimated, and sampling may be performed until there is sufficient confidence. This viewpoint brings with it a collection of effective variance-reducing techniques, such as importance sampling and block design (stratification). There are two disadvantages of this approach. The implicit measure of image quality is simply average pixel error, and that fails to describe the overall distribution of noise in the image (and thus the visual characteristics of the noise). A second problem is one of analysis. The Central Limit Theorem offers that the image error will decrease with the number of samples as $O(N^{-1/2})$, but variance-reducing techniques may result in errors that can be more tightly bound.

Some researchers have taken a *signal-processing* viewpoint of the image-sampling problem [Dippé85, Cook86, Mitchell87] and the distribution ray-tracing problem [Mitchell91]. The conversion of ray-tracing samples to pixel values is viewed as a resampling problem. If the image spectrum has power which declines with increasing frequency, and a sampling pattern is chosen with power concentrated in the highest possible frequencies, then sampling error will be driven toward higher frequencies. Pushing sampling noise into higher frequencies should cause more of it to be stopped by the frequency response of the pixel resampling system, and it will also minimize the response of the human visual system. A second goal is to randomize the sampling noise and avoid the production of conspicuous Moiré patterns. This viewpoint is not completely unrelated to the statistical one. For example, jitter sampling has desirable (but not optimal) spectral properties and is identical to certain stratification schemes. This view has the advantage of better description of image quality, but it also fails to provide good upper bounds on error as a function of number of samples.

2. Irregularities of distribution

A new viewpoint which can be applied to this problem is the theory of irregularities of distribution [Beck87]. This subject grew out of the study of certain *low-discrepancy* sampling patterns which have been used in quasi-Monte Carlo integration [Halton70, Niederreiter78], and it provides more accurate bounds on sampling error. Assume, for example, we are given a pattern of N samples in the unit square. For some (x,y) in the square, we can estimate the area of the rectangle $[0,x]\times[0,y]$ by counting the number of samples $\nu(x,y)$ within it.



The true area is given by the product xy and we will call the error the *local discrepancy* at the point (x,y):

$$\Delta(x,y) = \frac{\nu(x,y)}{N} - xy$$

The irregularity of the distribution of samples can be measured by averaging the local discrepancy over all possible values of x and y in the square (with the obvious extension to $\mathbf{I}^k$). The L^∞-discrepancy is defined to be the maximum absolute value of $\Delta(x,y)$:

$$D_N = \sup_{x,y} |\Delta(x,y)|$$

and the L^2-discrepancy (sometimes called the terpitude) is given by:

$$T_N = \left[\int_0^1 \int_0^1 \Delta^2(x,y)\, dxdy \right]^{1/2}$$

Shirley has already suggested that discrepancy may be a good measure to apply to sampling patterns in distribution ray tracing [Shirley91]. This is motivated by the importance of low-discrepancy sampling in numerical integration. For example, in one dimension, there is the significant result:

Theorem (Koksma 1942). *If f is a function of bounded variation $V(f)$ on the unit interval* **I** *and $x_1, \cdots x_N$ are points in* **I** *with L^∞-discrepancy D_N, then*

$$\left| \frac{1}{N} \sum_{i=1}^{N} f(x_i) - \int_0^1 f(t)\, dt \right| \le V(f)\, D_N$$

This is actually quite easily proven with integration by parts, noting that $\partial \nu(x)/\partial x$ will be a sequence of Dirac delta functions, $\sum \delta(x - x_i)$. Koksma's theorem was extended to higher dimensions [Niederreiter78], but the definition of bounded variation is problematic. Nevertheless, for a given function obeying the bounded-variation conditions, the error of numerical integration is $O(D_N)$. Roth has proven that, in k dimensions, the best sampling patterns have discrepancy tightly bounded by $T_N = \Theta(N^{-1}(\log N)^{(k-1)/2})$ [Beck87].

Sampling patterns have been constructed with discrepancies of $D_N = O(N^{-1}(\log N)^{k-1})$, such as the *Hammersley points.* Let $\phi_r(n)$ be the radical-inverse function of n base r. Its value is a real number from 0 to 1 constructed by taking the integer n, represented in base r, and reflecting its digits about the decimal point to form a fraction, base r. Given the sequence of prime numbers $2, 3, 5, \cdots$, one of N Hammersley points is given by:

$$\mathbf{x}_i = (i/N, \phi_2(i), \phi_3(i), \phi_5(i), \cdots)$$

An improvement suggested by Zaremba and generalized to k dimensions by Warnock is based on the folded radical inverse $\psi_r(n)$ [Warnock72]. Here, the i_{th} most significant digit a_i is replaced by $(a_i + i)$ mod r before the reflection about the decimal point. In the same paper, Warnock presents an algorithm for computing T_N and experimental results for several proposed low-discrepancy patterns.

3. Discrepancies of important sample distributions in graphics

Shirley computed the L^∞- and L^2-discrepancies of a number of commonly used nonuniform sampling patterns. It is worth doing a similar set of experiments again, but with progressively higher densities to observe the asymptotic behavior.

We will consider five sampling patterns (shown in the figures at the end of this paper). The first is Zaremba's low-discrepancy pattern generated with the folded radical inverse function. The second is the jittered sampling pattern obtained by randomly perturbing a regular periodic pattern. The third is the random Poisson-distributed pattern which is approximated by generating N uniformly-distributed points on the unit hypercube (in a section of a true Poisson process, N itself would have a Poisson distribution about the mean sample density). The fourth pattern is a "Poisson-disk" pattern generated by a dart-throwing algorithm on the unit torus (promoted as the "best pattern known" by advocates of the signal-processing viewpoint). The final pattern is a random "N-Rooks" pattern, suggested by Shirley—a random perturbation of points whose projection onto the x and y axis is uniformly spaced.

Process	16 points	256 points	1600 points
Zaremba	0.0358	0.00255	0.000438
jittered	0.0489	0.00633	0.00160
Dart-Throwing	0.0490	0.00799	0.00254
N-Rooks	0.0461	0.0101	0.00391
Poisson	0.0932	0.0233	0.00932

Table 1. 2-Dimensional L^2-discrepancies

These numbers are averages from 100 trials (except for the deterministic Zaremba pattern). The values for $N = 16$ agree fairly well with Shirley's results, and the values for the Poisson process agree with the theoretical value of $T_N^2 = N^{-1}(2^{-k} - 3^{-k})$ for k-dimensional patterns.

The most important feature to notice is the asymptotic behavior as N increases. For Poisson patterns, the $O(N^{-1/2})$ behavior is evidenced by the fact that increasing N by a factor of 100 only decreased the discrepancy by a factor of 10. At the other extreme, the discrepancy of Zaremba's pattern decreases at an impressive rate. The jitter and dart-throwing patterns are intermediate, but it is very intersting to note that their discrepancy seems to be better than $O(N^{-1/2})$. The N-Rooks pattern has lower T_N than Poisson, probably because of its uniform-axis-projection property, but its asymptotic behavior is not much better.

Next, consider several patterns that might be used in three dimensions, to sample time as well as the image plane. Four patterns are studied. The first is random 3D Poisson distributed points. The second is more realistic—jittered samples in (x,y) and random Poisson values of t. The third is the "spectrally optimal" pattern generated by the scanning algorithm proposed by the author [Mitchell91]. The final pattern is the 3D Zaremba-Warnock pattern.

Process	16 points	256 points	1600 points
Zaremba	0.0393	0.00376	0.000761
Scan-Time	0.0473	0.00721	0.00235
Poisson-Time	0.0580	0.0126	0.00512
Poisson-XYT	0.0741	0.0185	0.00741

Table 2. 3-Dimensional L^2-discrepancies

Once again, the asymptotic behavior of the Zaremba pattern is impressive, and the spectrally optimal pattern shows evidence of better convergence than purely random sampling.

Experiments with actual image errors (using a distribution ray tracer) are in progress. In doing such experiments, one can begin with an initial stratification of the image into pixel areas and apply various patterns within each pixel. Alternatively, one can sample the entire image without regard to pixel locations. Shirley performed a number of experiments using the first method [Shirley91], which simulates the approach implicit in most distribution ray tracers. However, given the uniformity created by stratification into pixel areas, we may not see dramatic differences in error at low sampling densities.

4. Isotropic irregularities of distribution

Theorems such as Koksma's and the remarkable $O(N^{-1}(\log N)^{k-1})$ low-discrepancy patterns give an initial impression that vast improvements can be easily made in the efficiency of distribution ray tracing. This section may, to some extent, dash those hopes. First, it should be noted that it is easy to create an image which does not have bounded variation everywhere (a checkerboard viewed in perspective, for example). However, such pathological behavior may often be isolated to small regions of an image.

A more fundamental problem occurs if we notice that synthetic images contain edges and curved boundaries at arbitrary orientations. The existence of $\Theta(N^{-1}(\log N)^{(k-1)/2})$ low-discrepancy patterns is limited to the problem of estimating the area of *axis-aligned* rectangles (and integrating functions having the specialized multi-dimensional bounded variation required by analogs of Koksma's theorem).

Consider the problem of estimating the area of disks or of boxes at arbitrary orientation. This is more like the situation encounted when estimating a pixel value on the edge of an object or containing a small object. Surprisingly, the discrepancy Δ is much larger. Work by Schmidt and others have shown that one can do no better than the lower bound of $\Omega(N^{-1/2k-1/2})$ in k dimensions [Beck87]. Upper bounds on the best discrepancy are known, and are typically larger than the lower bounds by a polylogarithmic factor (for various generalized discrepancy problems). In two dimensions, this still suggests the existence of patterns that are much better than random. As the dimension increases, one can probably not do better than random sampling.

To get some flavor of isotropic discrepancy, some patterns were tested on a set of random disks on the unit torus. Once again, 100 trials were averaged.

Process	16 points	256 points	1600 points
Dart-Throwing	0.0840	0.0120	0.00368
jittered	0.0994	0.0165	0.00394
Zaremba	0.0855	0.0160	0.00511
Poisson	0.104	0.0239	0.00993
N-Rooks	0.0908	0.0224	0.0104

Table 3. Discrepancy of random disks

It is promising to see that the dart-throwing and jitter patterns seem to be asymptotically better than random sampling, and in fact they perform better than Zaremba's low-rectangular-discrepancy pattern.

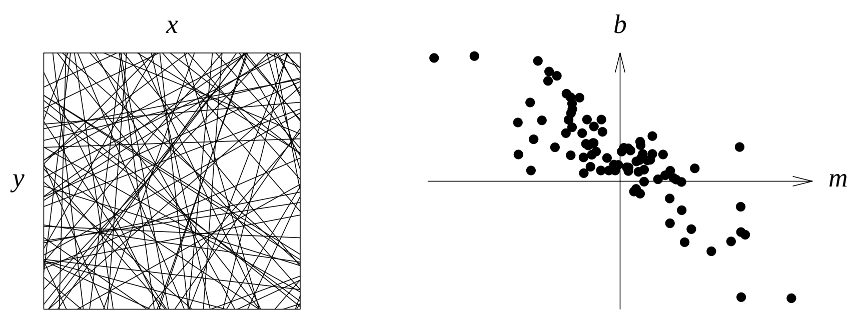
5. Discrepancy of arbitrary edges

A common problem in computer graphics is a pixel in the neighborhood of an edge. Consider a unit square with an arbitrary line passing through it. As before, we have a pattern of samples inside the square. We can define a discrepancy between the area of the square above the line and the fraction of sample points above the line. This may be a good model of sample-pattern quality in the case of box filtered antialiasing (a commonly used, but not very good filter). This arbitrary-edge discrepancy can be estimated by testing random lines against various sample patterns.

Generating random lines is an interesting problem. The chances of a square being hit by a random line from an isotropic distribution is proportional to the square's cross section, $\sin\theta + \cos\theta$. A random point in $[0,1]\times[0,1]$ has this distribution of angle about the origin. Given a slope, m, with this angular distribution, then the y-intercept, b,of the line is uniformly distributed over the allowed values. For lines intersecting the square, the allowed values of m and b form a "bow-tie" shaped region in line space, where $0 < b < 1 - m$ when m is negative, and $-m < b < 1$ when m is positive. Properly distributed random values of m and b can be generated from three uniform variates r_1, r_2, r_3 as follows:

$$m = \frac{r_1 - 0.5}{r_2 - 0.5}\ , \quad b = \begin{cases} (1-m)r_3 & \text{if } m < 0 \\ (-1-m)r_3 + 1 & \text{otherwise} \end{cases}$$

The figure below shows 100 random lines in the primal coordinate space and the corresponding points in the dual line space:



A set of 10,000 random lines was used to estimate the L^2-discrepancy of various sampling patterns. For the stochastic point processes, 100 trials were made as before.

Process	16 points	256 points	1600 points
Zaremba	0.0504	0.00478	0.00111
jittered	0.0538	0.00595	0.00146
Dart-Throwing	0.0613	0.00767	0.00241
N-Rooks	0.0637	0.0123	0.00488
Poisson	0.0924	0.0224	0.00866

Table 4. Discrepancy of random edges

It is interesting to see how well Zaremba's pattern works, although it is not converging as dramatically as it could with axis-aligned rectangles. Once again, random sampling is consistent with $O(N^{-1/2})$ accuracy, but some other patterns seem to do better.

The data above is from a Monte Carlo estimation. Dobkin has developed an algorithm for computing the exact L^{∞}-discrepancy, based on quadratic programming in line space [Dobkin92]. We are currently implementing this algorithm.

6. Conclusions

A factor in the complexity of ray tracing is the number of rays required to achieve a given level of image quality. The ray-tracing problem can be transformed into a multidimensional sampling problem, and the theory of irregularities of distribution suggests that sampling patterns exist which will converge to pixel values asymptotically faster than the $O(N^{-1/2})$ of random sampling.

On pixels containing small disks or arbitrary straight edges, experiments suggest that the best patterns derived from statistical and signal-processing considerations already have fast-convergence properties. In particular, jittered samples and Poisson-disk (blue noise) samples perform well.

Faster asymptotic convergence is an interesting and perhaps comforting result, but I do not believe that the results of these experiments should be used to determine the relative quality of the fast-converging patterns. In particular, the results are not consistent with image quality. The Zaremba pattern leads to Moiré patterns, and jittered sampling leads to grainier images than Poisson-disk (due to a larger component of low-frequency noise).

7. References

[Beck87] Beck, J., Chen, W. W. L. *Irregularities of distribution,* Cambridge University Press (1987).

[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.

[Dobkin92] Dobkin, D., *personal communication.*

[Halton70] Halton, J. H. A retrospective and prospective survey of the Monte Carlo method. *SIAM Review,* 12, 1 (January 1970), 1-63.

[Kajiya86] Kajiya, J. T. The rendering equation. *Computer Graphics* 20, 4, (August 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.

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

[Mitchell91] Mitchell, D. P. Spectrally optimal sampling for distribution ray tracing. *Computer Graphics,* 25, 4 (July 1991), 157-164.

[Niederreiter78] Neiderreiter, H. Quasi-Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society,* 84, 6, (November 1978), 957-1041.

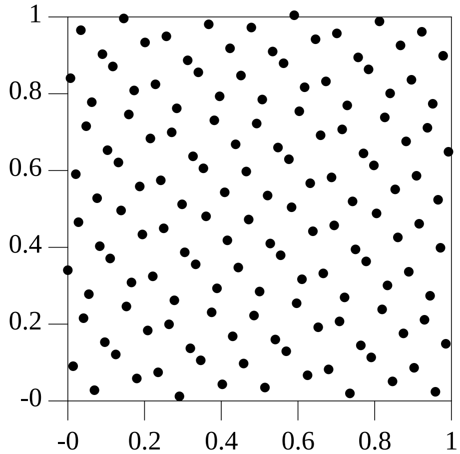
[Painter89] Painter, J., and Sloan, K. Antialiased ray tracing by adaptive progressive refinement. *Computer Graphics,* 23, 3 (July 1989), 281-288.

[Purgathofer86] Purgathofer, W. A statistical model for adaptive stochastic sampling. *Eurographics '86* (1986), 145-152.

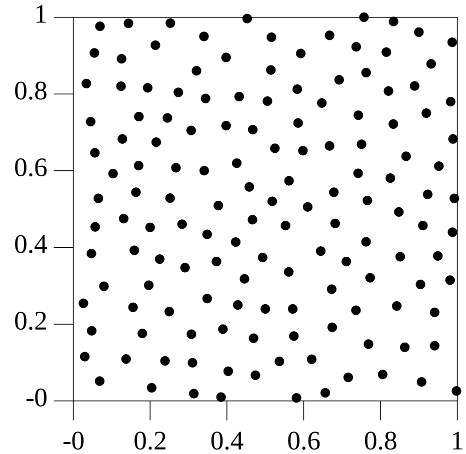
[Shirley91] Shirley, P. Discrepancy as a quality measure for sample distributions. *Proceedings of Eurographics '91,* (1991) 183-193.

[Warnock72] Warnock, T. T. Computational investigations of low-discrepancy point sets. *Applications of Number Theory to Numerical Analysis,* Academic Press (1972).

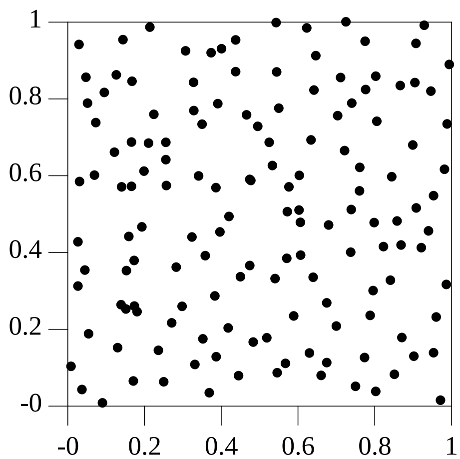
[Whitted80] Whitted, T. An improved illumination model for shaded display. *Comm. ACM*, 23, 6 (June 1980), 343-349.



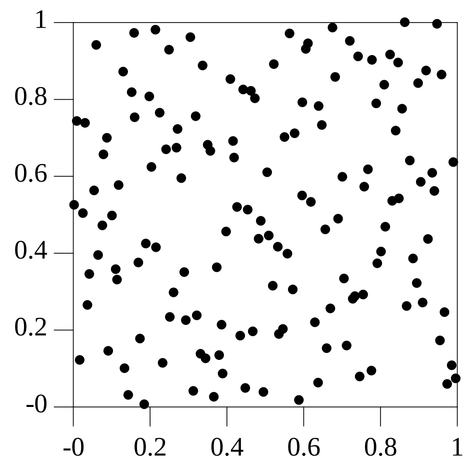
Zaremba-Hammersley Points



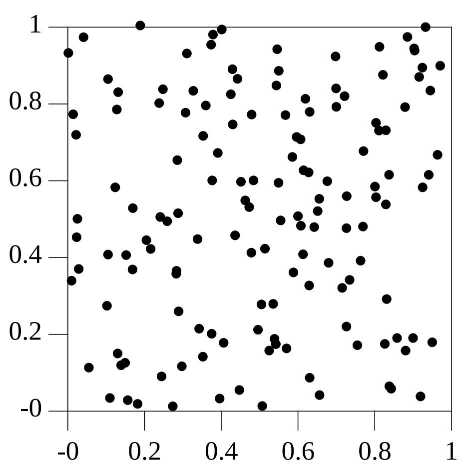
Poisson-Disk Sampling



Jittered Sample Points



N-Rooks Sampling



Poisson-Distributed Points