Consequences Of Stratified Sampling In Graphics

Don P. Mitchell
Microsoft Corporation

ABSTRACT

Antialiased pixel values are often computed as the mean of N point samples. Using uniformly distributed random samples, the central limit theorem predicts a variance of the mean of $O(N^{-1})$. Stratified sampling can further reduce the variance of the mean. This paper investigates how and why stratification effects the convergence to mean value of image pixels, which are observed to converge from N^{-2} to N^{-1}, with a rate of about $N^{-3/2}$ in pixels containing edges. This is consistent with results from the theory of discrepancy. The result is generalized to higher dimensions, as encountered with distributed ray tracing or form-factor computation.

CR Categories and Descriptors: I.3.3 [**Computer Graphics**]: Picture/Image generation.

Additional Key Words: Sampling, Stratification, Discrepancy, Antialiasing, Variance Reduction.

INTRODUCTION

One of the most general solutions to the aliasing problem in image synthesis is to supersample, compute many sample values within a pixel area and average them to estimate the actual integral of the image over an area. Several different theories have been applied to this sampling problem. Shannon's sampling theorem provides the justification for sampling at a higher density when an image is not sufficiently bandlimited for sampling at the pixel rate. The signal-processing viewpoint is not perfectly suited for treating the sampling of discontinuous signals (i.e., an image with sharp edges) or for understanding nonuniform sampling methods, although nonuniform sampling has been analyzed from this standpoint [Dippe85, Cook86, Mitchell87].

Another point of view is the theory of statistical sampling or Monte Carlo integration [Lee85, Kajiya86, Purgathofer86, Painter89]. The pixel value is estimated by the mean of a number of samples taken within the pixel area. If the pixel area is sampled at uniformly distributed random locations, the central limit theorem implies that the variance of the mean is $O(N^{-1})$. This is true even if the pixel area contains edges or if the domain being sampled has an unusual topology (e.g., sampling a function on the surface of a sphere) – cases where signal processing theory is difficult to apply.

A third viewpoint is the theory of discrepancy, which deals with the ability of a sampling pattern to estimate areas of subregions in a pixel. Quasi Monte Carlo methods [Halton70] are based on deterministic sampling patterns with low discrepancy, typically optimized to estimate the area of arbitrary axis-aligned rectangles within a square pixel area. Shirley introduced this sample-pattern quality measure to computer graphics [Shirley91], and Dobkin *et al.* introduced and analyzed a discrepancy measure based on arbitrary edges through a pixel [Dobkin93].

The most commonly used sampling strategy in ray tracing and distributed ray tracing is stratified sampling (often equivalent to the so-called jittered sampling patterns). This type of sampling has been studied from all three theoretical viewpoints mentioned above. This paper presents some theory and observations about the consequences of stratified sampling in computer graphics.

EXPERIMENTAL OBSERVATIONS

In the statistical viewpoint, a pixel value is the mean value of a small square area in an image. This assumes the use of a box filter, which is not ideal. Using a better filter simply involves computing a weighted mean, so for simplicity we will restrict the discussion to pixel-area averaging. A pixel value is estimated by a sample mean, the average of a number of point samples within the pixel area. The variance of the sample mean is a measure of the accuracy of this pixel estimate.

The variance of the mean can be directly measured by repeatedly estimating the same pixel with M independent trials of N samples x_i:

$$\overline{x_j} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma_{\bar{x}}^2 = \frac{1}{M-1}\sum_{j=1}^{M}(\bar{x} - \overline{x_j})^2$$

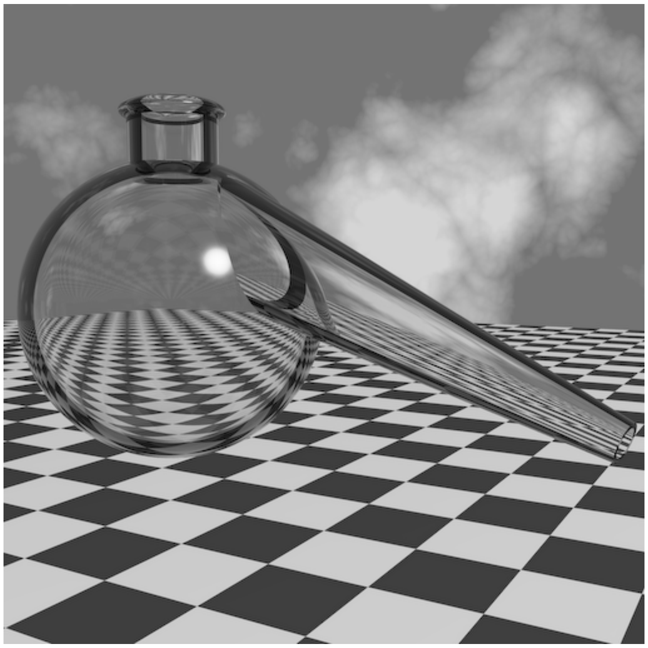

Figure 1. *One of several ray-traced images analyzed.*

A simple experiment demonstrates the $O(N^{-1})$ behavior predicted by the central limit theorem. Choose a set of pixels to study in a ray traced image such as the one shown in Figure 1. For progressively increasing values of N, measure the variance of the mean by performing M trials of uniformly distributed random samples. Plotting the log of the variance versus the log of N shows points fitting closely to a line of slope 1, for any pixel area in the image. Figure 2 shows a histogram of the measured slopes (derived from least-square fits) for test pixels.

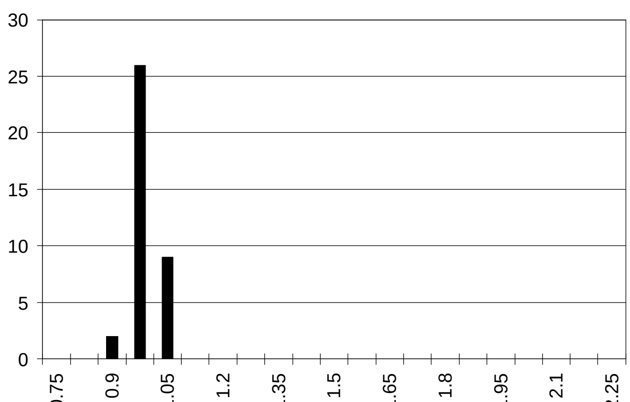

Figure 2. *Convergence of the Mean for Uniformly Distributed Random Sampling of Pixels.*

A standard variance-reducing technique is stratified sampling. Instead of distributing N random samples uniformly within the pixel area, the area can be divided into a grid of $\sqrt{N} \times \sqrt{N}$ cells, with one sample placed randomly within each cell. The literature of Monte Carlo methods contains varying comments about the effectiveness of stratification. Hammersley reports a "general rough working rule" that stratification gives a variance of the mean of $O(N^{-3})$ [Hammersley64]. Hammersley was commenting on the numerical integration of one-dimensional functions. Looking at multi-dimensional radiation transport, Spanier reports that stratification doesn't give much benefit [Spanier69].

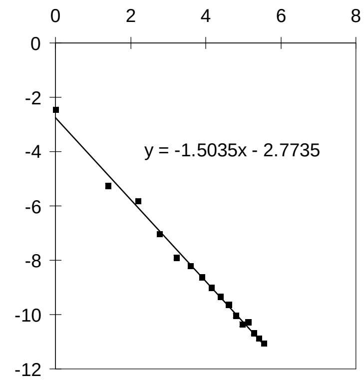

Figure 3. *Measured variance of mean for a pixel versus*

It seems appropriate then to actually measure the variance of the mean in the pixels of an image. Using several ray-traced images, $M = 50$ trials were performed with stratified sampling for N taking on values of 1, 4, 9, 16, … up to 256, and performed a least-square fit in log space to measure the closest fit to a convergence of the mean of the form $O(N^{-p})$. Figure 3 shows the data from one pixel and the least-square fit. Figure 4 shows the histogram of resulting values of p, using pixels from the image in Figure 1. The result is typical: a range of values from 1.0 to 2.0.

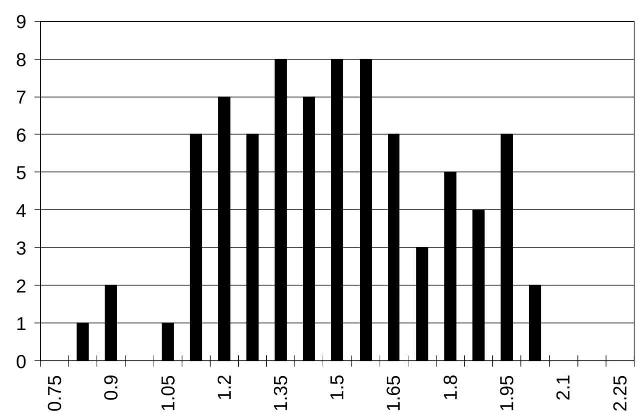

Figure 4. *Convergence of the Mean for Stratified Sampling of Pixels.*

By inspecting various pixels and their associated rates of convergence, three fairly distinct types can be found. Areas containing extremely complex features converge with a variance proportional to N^{-1}. Areas containing smooth regions of the image exhibit a variance of the mean that converges more rapidly, as N^{-2}. Pixels that exhibit $N^{-3/2}$ variance are found to contain smooth areas delimited by a few edges. Figure 5 shows several typical examples:

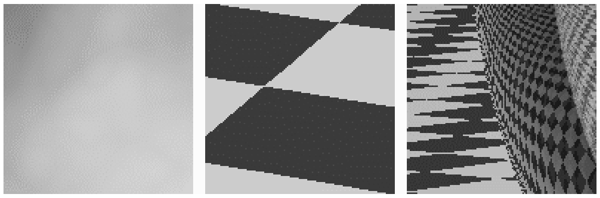

Figure 5. *Pixels with p = 1.89, 1.45, and 1.12.*

THEORETICAL ANALYSIS

The N^{-1} behavior in highly complex regions of the image is not surprising, since this is what would result from sampling a randomly varying function, according to the central limit theorem. There is no benefit from stratification in this case.

The $N^{-3/2}$ behavior in pixels with edges is consistent with the arbitrary-edge discrepancy of stratified/jittered sample patterns, as proven in Beck and Chen [Beck87]. In their derivation of the discrepancy, they note that edge discontinuities are one-dimensional features. As N (the number of samples and the number of strata) increases, an edge intersects only $O(\sqrt{N})$ strata. In computing the variance of the mean, the samples in the edge-crossing strata will dominate. Each sample x_i is given a weight of $1/N$, and so the total variance of the mean is:

$$\sigma_{\bar{x}}^2 = \frac{1}{N^2}\sum_{i=1}^{\sqrt{N}} \sigma_i^2 \approx \frac{\sigma^2}{N^{1.5}}$$

This is equivalent to a convergence of the standard deviation (square root of variance) proportional to $N^{-3/4}$. This is in line with Beck and Chen's bounds for arbitrary edge discrepancy are $\Omega(N^{-3/4})$ and $O(N^{-3/4}\log N^{1/2})$ [Beck87]. Dobkin *et al*. and Cross have used simulated annealing to generate sampling patterns with nearly half the discrepancy of jittered patterns [Dobkin94, Cross95]. These patterns may yield smaller pixel error, but Beck and Chen's lower bound proves that no pattern will be asymptotically better than jittered samples.

This analysis can be generalized to the case of N strata in d dimensions, with a sharp discontinuity of k dimensions. In this case, we expect $O(N^{k/d})$ strata to be cut by the discontinuity and dominate the variance. In that case, the variance of the mean should converge as $O(N^{k/d-2})$.

The $O(N^{-2})$ convergence in smooth regions of the image can be justified if we make the fairly general assumption that the image obeys a Lipshitz smoothness condition within the pixel area. That is, the range of values of the image function $f(x)$ (where x is a point in d dimensions) is no more than a constant factor times the diameter of the region:

$$\frac{|f(x) - f(y)|}{|x - y|} \le C$$

In d dimensions, the diameter of each strata is proportional to $N^{-1/d}$ and so we expect the standard deviation (root of the variance) of the function to be lower by the same proportion. The variance of the average of samples taken from each strata should therefore be:

$$\frac{1}{N^2}\sum_{i=1}^{N} \sigma_i^2 (N^{-1/d})^2 = O(N^{-1-2/d})$$

For $d=2$, this agrees with the observed result of $O(N^{-2})$. It may also explain Hammersley's "working rule" for one-dimensional smooth functions, while Spanier was working with radiation transport problems of high dimension and saw less benefit from stratification.

A FOUR-DIMENSIONAL EXPERIMENT

An additional experiment was done to test these results in higher dimensions. The calculation of the form factor between two parallel unit squares (two units apart and aligned) was computed by Monte Carlo integration in four dimensions. Once again, the result was recomputed with $M = 50$ independent trials for values of $N = 1^4, 2^4, \ldots 10^4$, and an estimation of the variance of the mean was found for each value of N. Since this is a smooth function with a four-dimensional domain, we expect a convergence of $O(N^{-3/2})$. The measured least-square fit to the data (in log space) gave $p = 1.430$. Removing the first (least accurate) point from the set gave $p = 1.501$.

The same experiment was then performed, with a smaller occluding square between the two planes. For each point on a given (two-dimensional) plane, the perimeter of the occluding square presents a one-dimensional discontinuity in the differential form factor. Thus, the overall discontinuity in the form factor is three-dimensional. Thus for $d = 4$ and $k = 3$, we expect a convergence of the mean of $O(N^{-5/4})$. The measured result was $p = 1.233$, and with the $N = 1$ point removed we found $p = 1.245$.

CONCLUSIONS

Stratified sampling is commonly used in ray tracing and distributed ray tracing, but its benefit has not been fully analyzed. Pixel accuracy is strictly improved by using stratification. For $N = 1$ samples per pixel, uniformly distributed random sampling and stratified sampling are the same, and as N increases, stratified sampling will often converge to the mean asymptotically faster than uniform random sampling.

The improvement in pixel accuracy depends on the nature of the image within the pixel area. In the worst case, stratification is no better (but no worse) than uniform random sampling, . If a finite number of edges pass through the pixel area, we expect an variance of the mean to be lower by a factor of $N^{1/2}$. If the image is smooth within the pixel area, we expect a variance of the mean to be lower by a factor of N.

The absolute pixel error will actually be proportional to the square root of the variance (ie., the standard deviation). The asymptotic reduction of error due to stratification is a little less impressive when we take the square root. The benefits of stratification are probably a mix of genuine error reduction and the spectral consequences of jittering as described by Dippe, Cook and Mitchell (i.e., the tendency of these sampling patterns to distribute the error in a high frequency pattern over the image).

ACKNOWLEDGEMENTS

I would like to thank Steve Drucker, Steven Gortler, Mike Marr, and my SIGGRAPH reviewers for their helpful comments. Thanks also to Josef Beck, Bernard Chazelle and David Dobkin for discussions about discrepancy.

REFERENCES

[Beck87] J. Beck and W. W. L. Chen. *Irregularities of Distribution*, Cambridge University Press, 1987.

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

[Cross95] R. A. Cross. Sampling Patterns Optimized for Uniform Distributed Edges. *Graphics Gems V*, Academic Press, 1995, 359-363.

[Dippe85] M.A.Z. Dippe and E. H. Wold. Antialiasing through stochastic sampling. *Computer Graphics* 19:3 (1985) 69-78.

[Dobkin93] D. P. Dobkin and D. P. Mitchell. Random-edge discrepancy of supersampling patterns. *Graphic Interface*, York, Ontario (1993).

[Dobkin94] D. P. Dobkin, D. Eppstein and D. P. Mitchell. Computing the Discrepancy with Applications to Supersampling Patterns. *Trans. Graphics (to appear).*

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

[Hammersley64] J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods.* Methuen & Co., London, 1964.

[Kajiya86] J. T. Kajiya. The Rendering Equation. *Computer Graphics* 20 (1986) 143-150.

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

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

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

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

[Shirley91] P. Shirley. Discrepancy as a quality measure for sample distributions. *Proc. Eurographics* (1991) 183-193.

[Spanier69] J. Spanier and E. M. Gelbard. *Monte Carlo Principles and Neutron Transport Problems.* Addison-Wesley, Reading, MA, 1969.