

Sampling Density for Image Analysis

Ian T. Young

Faculty of Applied Physics, Lorentzweg 1

Delft University of Technology

NL-2628 CJ Delft, The Netherlands

e-mail: young@ph.tn.tudelft.nl

WWW: <http://www.ph.tn.tudelft.nl/>

Abstract

We show that the convergence of measures for object size to their "true values" can be analyzed by looking at the pixels (or voxels) on the border of the object. This leads to results for a proper choice for the sampling density of the asymptotic form $CV = \sigma/\mu = k \cdot Q^{-(N+1)/2}$ where Q is the sampling density (pixels per object "diameter") and N is the number of spatial dimensions.

Introduction

For a relatively long time it was thought that the Nyquist sampling theorem was the definitive statement concerning the proper choice of sampling frequency or sampling density for multi-dimensional signal (image) processing [1-5]. With publications in the 1980's, however, it became clear that the issue was somewhat more complex [6-9]. When the goal is image filtering, then the use of the Nyquist criterion is highly appropriate. When the goal is image measurement, another criterion must be used to choose the proper sampling density.

The act of sampling – cutting up an image into rows and columns in 2D and rows, columns, and planes in 3D – is a source of spatial "quantization" noise which is pivotal when the goal is image measurement. This quantization occurs, however, not in the signal amplitude, the dependent variable, but in the spatial coordinates, the independent variables. There are a number of ways to assess the influence of this type of noise on image measurements. In general we seek measurement formulas that, despite the various noise sources, provide reasonable estimates of the analog quantities given the digital representation. In this context a "reasonable" estimate means one that is unbiased (accurate) and whose variance goes to zero (precise) as the sampling density increases.

While the goal of finding an unbiased estimate for a class of objects is laudable, it is not always possible. (What is the "true value" of the volume of an orange?)

As a result the *coefficient-of-variation*, CV , (defined below) is frequently used to characterize the quality of an estimator. In this paper we will provide a simple argument that shows how to choose the sampling density by studying the behavior of the CV as the sampling density Q increases.

The Model

The relationship between the spatial quantization noise and measurement accuracy and precision can be illustrated with the relatively simple problem of measuring the area of a two dimensional object. It has been known for many years [10] that the best measure of the area of an "analog" object given its digital representation is to simply count the pixels associated with the object. The use of the term "best estimate" means that this area estimate is unbiased and the variance of the area estimate goes to zero as the sampling density increases. We assume, here, that the pixels belonging to the object have been labeled thus producing a binary representation of the object. The issue of using the actual gray values of the object pixels to estimate the object area will not be covered here but can be found in [11].

Let us look at a simple example. When a randomly placed (circular) object is digitized, one possible realization is shown in Figure 1. The equation for generating the object is $(x - e_x)^2 + (y - e_y)^2 \leq R^2$ where R is the radius. The terms e_x and e_y are identically-distributed, independent random variables with a uniform distribution over the interval $(-1/2, +1/2)$. They represent the random placement of the object with respect to the *periodic* (unit) sampling grid.

In the realization shown in Figure 1 the area would be estimated at 84 pixels but a slight shift of the circle with respect to the grid could change that, for example, to 81 or 83 or 86. The sampling density of this figure can be expressed as about 10 pixels per diameter. To appreciate what effect the finite sampling density has on the area estimate let us look at the *coefficient-of-*

variation of the estimate, the $CV = \sigma/\mu$ where σ is the standard deviation of the area estimate of the area and μ is the average area estimate over an ensemble of realizations.

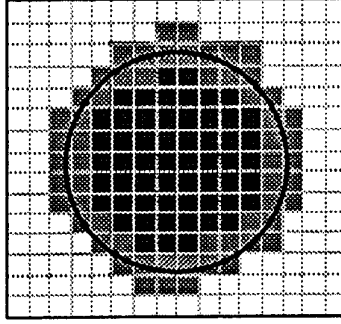


Figure 1: Given small variations in the center position (e_x, e_y) of the circle, pixels that are colored black will always remain part of the object and pixels that are colored white will always remain part of the background. Pixels that are shown in light gray may change from object to background or vice-versa depending on the specific realization of the circle center (e_x, e_y) with respect to the digitizing grid.

If we denote the diameter of the circular object by D and the size of a pixel as $s \times s$, then the sampling density is $Q = D/s$. The area of the circle, A_1 , that is always black (in Figure 1) independent of (e_x, e_y) is given by:

$$A_1 = \frac{\pi}{4}(D - 2s)^2 \quad (1)$$

The number of pixels associated with this is:

$$N_1 = \frac{A_1}{s^2} = \frac{\pi}{4} \left(\frac{D}{s} - 2 \right)^2 = \frac{\pi}{4}(Q - 2)^2 \quad (2)$$

The area of the region, A_b , that is light gray (in Figure 1) is given by:

$$A_b = \frac{\pi}{4}(D + 2s)^2 - \frac{\pi}{4}(D - 2s)^2 \quad (3)$$

and the number of pixels, N_b , associated with this region is:

$$N_b = \frac{\pi}{4}(Q + 2)^2 - \frac{\pi}{4}(Q - 2)^2 = 2\pi Q \quad (4)$$

The area of the circle is estimated by counting pixels and the contribution from the black region is clearly N_1 . The total number will be $N_T = N_1 + n$ where n is a random variable. Let us make a simplifying assumption: Let us assume that each of the pixels in the light gray region can be part of the object with probability p and part of the background with probability $(1 - p)$ and that the decision for each pixel is independent of the other neighboring pixels in the light gray region. This, of course, describes a binomial distribution for the pixels in that region. In fact this assumption is not true and the behavior of neighboring pixels, over small distances, is somewhat correlated. But let us see how far we can go with this model. Under this assumption:

$$\begin{aligned} \mu &= E\{N_T\} = E\{N_1 + n\} = N_1 + E\{n\} \\ &= N_1 + N_b p = \frac{\pi}{4}(Q - 2)^2 + 2\pi Q p \end{aligned} \quad (5)$$

and

$$\begin{aligned} \sigma(N_T) &= \sqrt{\sigma^2(N_1) + \sigma^2(n)} = \sigma(n) \\ &= \sqrt{N_b p(1 - p)} = \sqrt{2\pi Q p(1 - p)} \end{aligned} \quad (6)$$

We have made use of the assumption that N_1 is deterministic—the pixels are always black—and that the mean and variance of the binomial distribution for N_b samples with probability p are given by $N_b p$ and $N_b p(1 - p)$, respectively.

This immediately leads to an expression for the CV of our estimate as:

$$CV = \frac{\sigma}{\mu} = \frac{\sqrt{2\pi Q p(1 - p)}}{\frac{\pi}{4}(Q - 2)^2 + 2\pi Q p} \quad (7)$$

We can now study the convergence of the CV as the sampling density increases. As Q increases in this two-dimensional image we have:

$$\lim_{Q \rightarrow \infty} CV(Q) = k_2 \frac{Q^{1/2}}{Q^2} = k_2 Q^{-3/2} \quad (8)$$

This type of argument can easily be extended to the three-dimensional case where the results are:

$$CV = \sigma/\mu = \frac{\sqrt{\left(2\pi Q^2 + \frac{8\pi}{3}\right)p(1-p)}}{\frac{\pi}{6}(Q-2)^3 + \left(2\pi Q^2 + \frac{8\pi}{3}\right)p} \quad (9)$$

and

$$\lim_{Q \rightarrow \infty} CV(Q) = k_3 \frac{Q}{Q^3} = k_3 Q^{-2} \quad (10)$$

Finally, for the N -dimensional case, using the formulas for the volume and surface area of an N -dimensional sphere, we have:

$$\lim_{Q \rightarrow \infty} CV(Q) = k_N Q^{-(N+1)/2} \quad (11)$$

The conclusion is clear. As the sampling density Q increases the precision of our estimates improves as a power of Q . While the independent binomial behavior cannot be strictly true, the arguments presented do show the type of convergence that can be expected and how that varies with Q .

These results have also been found experimentally in a number of publications [9-15]. An example is shown in Figure 2. The measurement is the volume of spheres that have been randomly placed on a sampling grid. The quality of the estimator (voxel counting) is assessed by examining the CV .

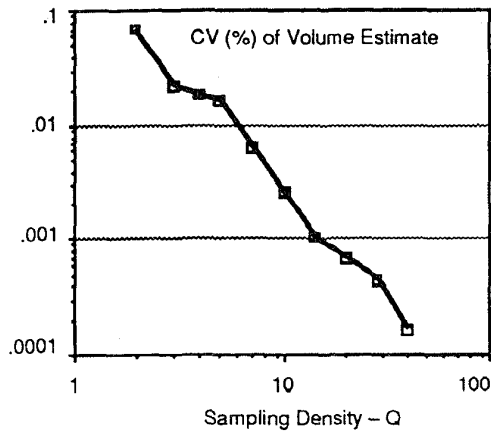


Figure 2: For each sampling density value Q (expressed in voxels per diameter), 16 spheres were generated with randomly placed centers (e_x, e_y, e_z). The volume was measured by counting voxels and the $CV(Q) = \sigma(Q)/\mu(Q)$ calculated accordingly.

It is clear from Figure 2 that, as the sampling density increases by one order of magnitude from $Q=2$ to $Q=20$ samples per diameter, the CV decreases by two orders of magnitude. This illustrates the relation between CV and Q described in equation (10).

Single-Pixel Behavior

We can take this analysis a step further by comparing equation (5) to the known property that pixel counting leads to an unbiased estimate of the area [10]. Because the bias is zero this means that:

$$\left\{ \frac{\pi}{4}(Q-2)^2 + 2\pi Qp \right\} - \left\{ \frac{\pi}{4}Q^2 \right\} = 0 \quad (12)$$

which yields a value for p given by:

$$p = \frac{1}{2} - \frac{1}{2Q} \quad (13)$$

The explanation for this result is as follows. For a circle with a finite radius and using a finite sampling density, the probability, that a given border pixel will be labeled as part of the object, depends on the "coverage" of that pixel. This is illustrated in Figure 3. If the coverage is more than 50% of the pixel area then the pixel will be labeled as object; if less than 50% then the pixel will be labeled as background.

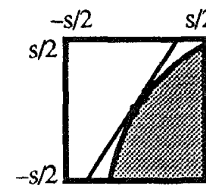


Figure 3: The dark shaded area represents that portion of the circle with radius R that covers a given pixel when the boundary goes through the center of the pixel, $(0,0)$. The straight line is the tangent line to the curve at the pixel center.

As the sampling density Q increases, then for a circle—or any curve with bounded curvature—we can approximate the curve within a given pixel by a straight line. This explains the term $(1/2)$ in equation (13); a straight line going through the center of a pixel has half the area of the pixel on each side of the line. The term $(1/2Q)$ is the correction term associated with the non-

zero curvature of the actual border within the pixel. By examining the error in the straight-line approximation, the area *between* the circle and the straight line in Figure 3, we can compute the (shaded) area that is actually covered by the inside of the curve. As the sampling density increases, as $Q \rightarrow \infty$ and thus $s \rightarrow 0$, the error goes to zero. A plot of $error(Q = 2R/s)$ is shown in Figure 4.

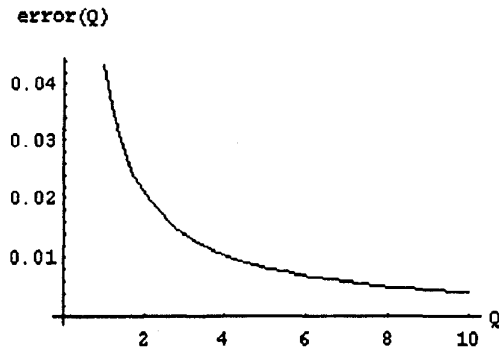


Figure 4: As the sampling density increases the area between the straight line and the shaded region in Figure 3 decreases.

A Taylor series expansion of $error(Q)$ gives the following:

$$error(Q) \propto \frac{1}{Q} + \frac{3}{80 \cdot Q^3} + \frac{3}{896 \cdot Q^5} + O(Q^{-1})^7 \quad (14)$$

This shows that the behavior can, indeed, be approximated by a term of the form $(1/Q)$.

Summary

The question then becomes should we choose the sampling density on the basis of the Nyquist sampling theorem or on the basis of the required measurement precision. The answer lies in the goal of the work. If we are interested in image filtering then the Nyquist theorem should be used. If, however, we are interested in measurements derived from images then the sampling frequencies derived from measurement specifications—as exemplified in Figure 2 and equations (8), (10), and (11)—should be used.

Acknowledgments

This work was partially supported by the Netherlands Organization for Scientific Research (NWO) Grant 900-538-040, the Foundation for Technical Sciences (STW) Project 2987, and the Rolling Grants program of the

Foundation for Fundamental Research in Matter (FOM).

References

- [1] J. M. S. Prewitt, "The selection of sampling rate for digital scanning," *IEEE Transactions on Biomedical Engineering*, vol. BME-12, 1965.
- [2] K. R. Castleman, *Digital Image Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, 1979.
- [3] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, Massachusetts: Addison-Wesley, 1977.
- [4] A. Papoulis, *Systems and Transforms with Applications in Optics*. New York: McGraw-Hill, 1968.
- [5] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, 1984.
- [6] D. Proffitt and D. Rosen, "Metrication errors and coding efficiency of chain-encoding schemes for the representation of lines of finite length," *Computer Graphics and Image Processing*, vol. 10, pp. 318-322, 1979.
- [7] A. M. Vossepoel and A. W. M. Smeulders, "Vector code probabilities and metrication error in the representation of straight lines of finite length," *Computer Graphics and Image Processing*, vol. 20, pp. 347-364, 1982.
- [8] L. Dorst and A. W. M. Smeulders, "Length estimators for digitized contours," *Computer Vision, Graphics, and Image Processing*, vol. 40, pp. 311-333, 1987.
- [9] I. T. Young, "Sampling density and quantitative microscopy," *Analytical and Quantitative Cytology and Histology*, vol. 10, pp. 269-275, 1988.
- [10] B. Matérn, "Precision of area estimation: a numerical study," *Journal of Microscopy*, vol. 153, pp. 269-284, 1989.
- [11] P. W. Verbeek and L. J. Van Vliet, "Estimators of 2D edge length and position, 3D surface area and position in sampled grey-valued images," *Biolmaging*, vol. 1, pp. 47-61, 1993.
- [12] L. J. Van Vliet, "Grey-scale measurements in multi-dimensional digitized images," PhD Thesis: Delft University of Technology, 1993.
- [13] A. W. M. Smeulders and L. Dorst, "Measurement issues in morphometry," *Analytical and Quantitative Cytology and Histology*, vol. 7, pp. 242-249, 1985.
- [14] J. C. Mullikin, "Discrete and continuous methods for three-dimensional image analysis," PhD Thesis: Delft University of Technology, 1993.
- [15] J. C. Mullikin and P. W. Verbeek, "Surface area estimation of digitized planes," *Biolmaging*, vol. 1, pp. 6-16, 1993.