

S. E. Paul, A. A. Goodenough, S. D. Brown, and C. Salvaggio, "Sub-pixel radiometry: A three-part study in generating synthetic imagery that incorporates sub-pixel variation," in *Proceedings of SPIE, SPIE Defense and Security, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI, Modeling and Simulation*, **7695**, SPIE, (Orlando, Florida, United States), April 2010

Copyright 2010 Society of Photo Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic electronic or print reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

<http://dx.doi.org/10.1117/12.850180>

Sub-Pixel Radiometry: A three-part study in generating synthetic imagery that incorporates sub-pixel variation

Sarah E. Paul, Adam A. Goodenough, Scott D. Brown, Carl Salvaggio
Rochester Institute of Technology, Center for Imaging Science, Digital Imaging Remote Sensing Laboratory, Rochester, NY, USA

ABSTRACT

A pixel represents the limit of spatial knowledge that can be represented in an image. It is represented as a single (perhaps spectral) digital count value that represents the energy propagating from a spatial portion of a scene. In any captured image, that single value is the result of many factors including the composition of scene optical properties within the projected pixel, the characteristic point spread function (or, equivalently, modulation transfer function) of the system, and the sensitivity of the detector element itself. This presentation examines the importance of sub-pixel variability in the context of generating synthetic imagery for remote sensing applications. The study was performed using the Digital Imaging and Remote Sensing Image Generation (DIRSIG) tool, an established ray-tracing based synthetic modeling system whose approach to sub-pixel computations was updated during this study.

The paper examines three aspects of sub-pixel variability of interest to the remote sensing community. The first study simply looks at sampling frequency relative to structural frequency in a scene and the effects of aliasing on an image. The second considers the task of modeling a sub-pixel target whose signature would be mixed with background clutter, such as a small, hot target in a thermal image. The final study looks at capturing the inherent spectral variation in a single class of material, such as grass in hyperspectral imagery. Through each study we demonstrate in a quantitative fashion, the improved capabilities of DIRSIG's sub-pixel rendering algorithms.

Keywords: sampling, DIRSIG, PSF, small target radiometry, sub-pixel, synthetic imagery, clutter, hyperspectral, thermal, aliasing, ray-tracing

1. INTRODUCTION

The ability to create synthetic imagery that accurately models the real world in both a spatial and spectral sense is a powerful tool in the remote sensing community. This capability is useful for a range of applications from system tradeoff analysis for sensors that are still in the design phase to algorithm testing. In most cases the user is concerned with accurate modeling on a macroscopic scale, meaning objects that occupy multiple pixels. However, there are situations where accurately capturing the sub-pixel variation in a synthetic image is crucial. This paper presents the improved capability of the Digital Imaging and Remote Sensing Synthetic Image Generation (DIRSIG) tool to model sub-pixel variation.

2. BACKGROUND

2.1 DIRSIG

The Digital Imaging and Remote Sensing Image Generation (DIRSIG) tool was developed by the Digital Imaging and Remote Sensing (DIRS) Laboratory at the Rochester Institute of Technology (RIT). It is a synthetic image generation application that is capable of modeling imaging systems with sensitivity in the visible through thermal infrared regions of the spectrum as well as polarimetric, RADAR, and LIDAR systems.¹ To accurately reproduce the radiometry of a scene, DIRSIG utilizes ray tracing and first principle physics, chemistry, and mathematical theories. First-principle based sub-models, including BRDF prediction, facet temperature prediction, sensor models, and atmospheric models are used to generate the synthetic imagery. All modeled components are combined using a spectral representation and integrated radiance images can be simultaneously produced for an arbitrary number of user defined bandpasses.

2.2 Overview of Sampling Methods

A real sensor has a pixel that has a finite area that is projected onto the ground.² If atmospheric and sensor effects are excluded, the radiance recorded at each pixel is the result of averaging the radiances from the material(s) within the projected pixel. In the case of a mixed pixel, where there is more than one material within a pixel boundary, the sensor automatically takes into account the proportions (and in turn the radiance contributions of each material) of the materials because of this spatial averaging. DIRSIG treats each pixel as a finite sampling element where one sampling ray is cast from the center of each detector element. In other words instead of spatially averaging the radiances as a real sensor would, DIRSIG samples the world at discrete points and calculates the radiometry based on the material that the sampling ray intersects with. This single ray sampling is DIRSIG's standard operating mode. There are two other sampling modes within DIRSIG, a grid sampling mode where a pixel is divided into sub-pixel elements and a sampling ray is cast from the center of each sub-pixel element, and an adaptive sampling mode where the pixel is divided into sub-pixel elements and samples are generated randomly via a Halton sampler.

2.2.1 Single Ray

As stated previously, the single ray sampling mode casts one ray from the center of each detector element to sample the scene. This is shown in Figure 1.

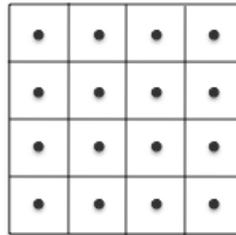
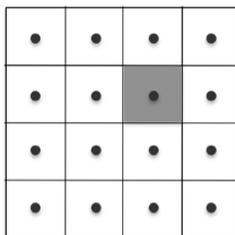
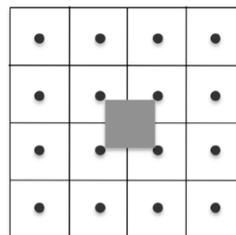


Figure 1. Schematic of DIRSIG single ray sampling. Each box represents one pixel on the focal plane and each dot in the center represents the location of the origin of the sampling ray.

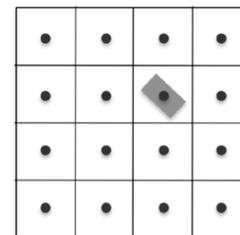
This type of sampling is advantageous in terms of computing time as casting one ray per pixel takes less time than casting multiple rays and calculating an average. However, since only one sampling ray is cast it will intersect with only one material. Thus the radiance value for the entire pixel is calculated based upon the one material even if more than one material occupies the pixel. When a material completely fills a pixel this sampling technique will give the exact radiance value back. This is shown in Figure 2(a) where the gray material completely fills a pixel. The gray area will be referred to as the target material and the white area as the background.



(a) Target completely filling pixel



(b) Mixed pixel where the target area is not sampled



(c) Target completely enclosed in pixel but not filling the entire sensitive area

Figure 2. Single ray sampling issues

A mixed pixel can prove troublesome to the single ray sampling method. In Figure 2(c) the target is contained within the boundary of one pixel but does not completely fill the pixel. The sampling ray will intersect the target material and calculate the radiance using only information about the target even, though there is clearly background material within the projected pixel area. Effectively DIRSIG will overestimate the size of the target. Figure 2(b) presents another problem where the target has been shifted so that it is centered at the intersection of the projected boundaries of four pixels. In this case, the single sampling rays will miss the target material completely and the radiance for those four pixels will be calculated as if there were not target material present.

2.2.2 Grid Sampling

One sampling method DIRSIG uses to account for multiple materials within a pixel is grid sampling. Grid sampling effectively increases the resolution of the image beyond the capabilities of the sensor being modeled by dividing each pixel into sub-elements. A sampling ray is cast from the center of each sub-element thus increasing the number of samples per pixel. Figure 3 shows a single pixel that has been oversampled by a factor of five resulting in 5x5 sub-elements representing one pixel.

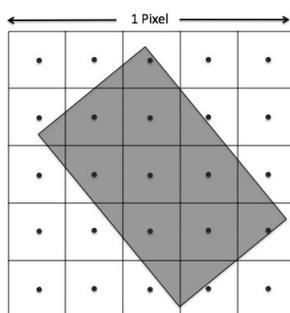


Figure 3. A pixel oversampled by a factor of 5. Each box represents a sub-element of the pixel and each dot the origin of the sampling ray. The gray rectangle is the target material projected on to the virtual detector.

Since more samples are being taken per pixel, this technique better approximates the averaging that occurs in a real world sensor. A drawback of grid sampling is that it is a more computationally time consuming technique because the radiometry is being calculated for more sampling rays. However, with more sampling comes a more accurate radiance value. An issue that arises from grid sampling is aliasing at the edge of two materials in the oversampled space. From a linear systems standpoint, an edge is made of an infinite number of frequencies so to be accurately reconstructed it would have to be sampled an infinite number of times. Due to the averaging that takes place, a real world sensor can sample the edge so that no aliasing occurs. DIRSIG can only cast a finite amount of sampling rays and as a result aliasing will always occur. Figure 4 shows how the oversampled image of Figure 3 would look if filled in with color. The samples have not adequately captured the edge of the target and as a result it appears jagged.

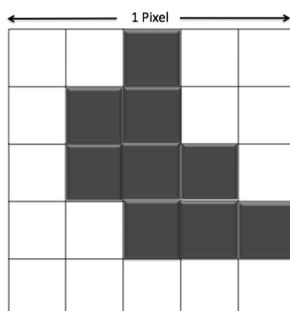


Figure 4. A pixel oversampled by a factor of 5. Inadequate sampling of the edge results in a jagged edge.

As a result of not capturing all of the edge information, the target size is underestimated, and instead of contributing to the radiance of sixteen sub-elements it contributes to nine. As the number of sub-elements increases so does the accuracy of the target size and more importantly the radiance value that is recorded for the pixel.

2.2.3 Adaptive Subsampling

The primary goal of adaptive sampling is to only expend sub-pixel samples (and therefore computation time) when they are needed. When it does sample, it also attempts to distribute those samples as effectively as possible to capture underlying variability of the scene. It therefore needs to work on two levels. First, it attempts to find variability in a projected pixel and to send more samples (rays) to those regions. Second, it uses uniform sampling techniques to generate a distribution of samples that have high uniformity and anti-aliasing characteristics.

For the first part, we sub-divide the problem (i.e. the solution for an individual detector element) into a user-driven number of sub-problems, or sub-elements. This sub-division simply breaks up the effective sensitive area of a pixel along a grid (the sensitive area of a pixel can be larger than the physical detector due to sensor optical effects). Each sub-problem is solved by casting rays into the scene and finding radiance contributions and incorporating the effects of the sensor itself (optics transmission, spectral/spatial responsivity, etc...) and then the components are integrated across the entire pixel to find the final solution.

We assume no a priori knowledge of the content of the scene and rather use information gathered from the scene to look for variability (both within the current sub-element and by comparing to solutions from neighboring sub-elements). Thus, we do not look for an “optimal” sampling of a pixel, but rather try to “adapt” to pixel content as it is gathered. In some cases this is not efficient. For instance, if there is no sub-pixel scene variability, there is no way to know that until enough samples have been gathered to determine that lack of variability – at which point it is too late to use fewer samples. We can only adapt to that information and stop sending samples for the sub-problem. Despite its limitations, this general approach is very flexible for dealing with different types of collection scenarios including non-imaging platforms (e.g. a LIDAR collect measuring time-gated photon counts). The definition of variability is driven by the user, who can set a simple allowed-difference threshold between new samples and a running average. The user is also able to control the number of samples used for each problem, thereby limiting the computational complexity of each pixel solution.

Figure 5 gives a schematic of the adaptive sampling concept.

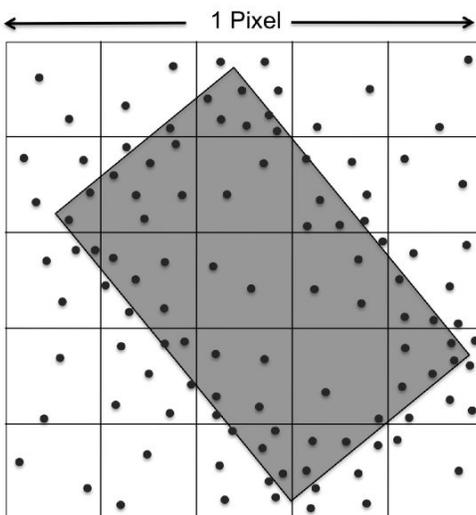


Figure 5. A pixel sampled using adaptive sampling. One sampling ray is cast into a random location in each sub-element. More samples are cast where there is more variability (i.e. the boundary of two materials)

For the second part, we attempt to capture as much information about the scene as possible with the samples we use. If we consider our sampler as sampling function $s(x, y)$ (effectively a collection of delta functions over the projected pixel space) and the area of the scene beneath that pixel to be another 2-D function $i(x, y)$, then the act of sampling can be viewed in frequency space as

$$\mathcal{F}\{s(x, y) \cdot i(x, y)\} = S(\xi, \nu) * I(\xi, \nu). \quad (1)$$

For perfect reconstruction of an arbitrary scene, we would ideally want $S(\xi, \nu) = \delta(0, 0)$ such that

$$\mathcal{F}^{-1}\{\delta(0, 0) * I(\xi, \nu)\} = \mathcal{F}^{-1}\{I(\xi, \nu)\} = i(x, y). \quad (2)$$

However, this ideal sampling would require an infinite number of rays from each pixel in a continuous domain (or many more samples than we can afford in a discrete one). Therefore, we use a standard Halton sequence³ sampler that generates uniform samples to approximate an ideal sampling.

3. METHODOLOGY

3.1 Study 1: Aliasing

For this study an image was constructed using a scene with the classic 1951 USAF tri-bar resolution target showing a high contrast series of bars with different spatial frequencies and scalings. It was rotated with respect to the focal plane to ensure that there would be no fortuitous alignment between lines on the chart and pixels on the detector and to maximize the occurrence of sub-pixel diagonal edges.

3.2 Study 2: Absolute sub-pixel thermal radiometry

This study compared the number of samples used by grid sampling to calculate the correct radiance value for a pixel to the number of samples required by adaptive sampling to do the same. For this study, a 3x3 pixel image was created that contained a hemispheric dome at a temperature of 40°C sitting on top of a flat background that had a temperature of 20°C. The GSD of the image was 1 meter and the dome was scaled such that it filled 50% of the area of the central pixel. The geometry is shown in Figure 6. The radiance was calculated at a wavelength of 8.5μm for the simulations. A point spread function was not added for this demonstration.

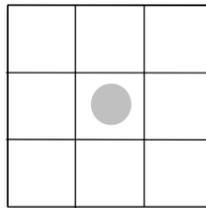


Figure 6. Geometry of Study 2. The gray circle represents a dome with a temperature of 40°C sitting on a background that is 20°C

To generate the grid sampled images, the geometry depicted in Figure 6 was rendered using 2x2 to 300x300 sub-elements in increments of 2 (i.e. 2x2, 4x4, 6x6, etc...). The imagery generated using adaptive sampling was generated in the same way except the number of sub-elements used was scaled by half compared to grid sampling. In other words for an NxN sub-element grid sampled image the corresponding adaptive case would have N/2 x N/2 sub-elements. However the maximum number of samples per pixel allowed for each adaptive run was N². The advantage of adaptive sampling is that more than one sampling ray can be cast per sub-element. By reducing the number of sub-elements in the adaptive sampling runs, more than one ray can be cast per sub-element. Figure 7 shows this graphically.

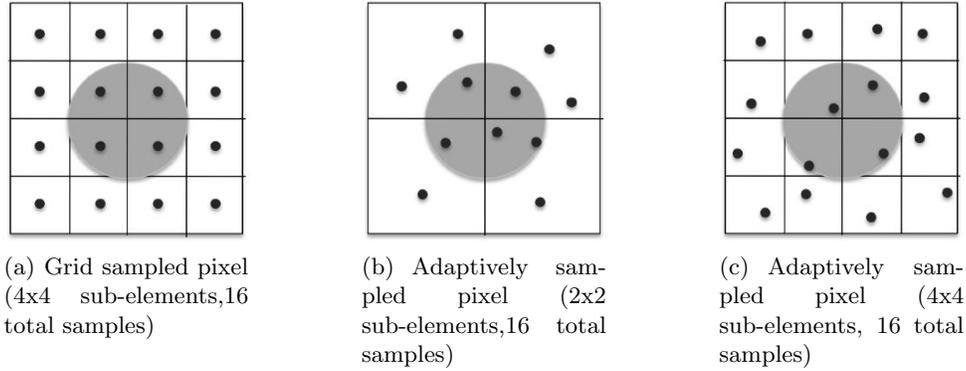


Figure 7. The grid sampling case with 4x4 sub-elements and 16 samples per pixel is shown in (a), the equivalent adaptive sampling case is 2x2 sub-elements with a maximum of 16 elements per pixel and is shown in (b). If the number of sub-elements had been left at 4x4 for the adaptively sampled pixel, the result might look something like (c) which is basically the same as (a) except for the samples being in random locations within each sub-element.

We chose to halve the number of sub-elements in each dimension for the adaptive sampling runs but any fraction of N sub-elements could have been used for the comparison. A summary of the sampling values is shown in Table 1

Table 1. Summary of samples used in study 2.

Grid sub-elements	# of samples per pixel	Adaptive sub-elements	# of samples per pixel
2x2	4	1x1	1
4x4	16	2x2	16
6x6	36	3x3	36
⋮	⋮	⋮	⋮
298x298	88804	149x149	88804
300x300	90000	150x150	90000

To generate the plot in the results section the number of samples over all nine pixels were summed and plotted against the radiance value for the central pixel. All nine pixels were included in the number of samples calculation because adaptive sampling looks for variation between sub-elements and there could be a case where variation between sub-elements happened to fall across a pixel boundary.

The truth radiance value was calculated analytically via a weighted average of the radiances of the materials. The blackbody radiances of the gray material and the white material calculated at $8.5 \mu\text{m}$ were $12.11 \text{ W/m}^2\text{sr}\mu\text{m}$ and $8.37 \text{ W/m}^2\text{sr}\mu\text{m}$, respectively. The overall radiance for this pixel is computed as

$$\begin{aligned}
 L_{\text{pixel}} &= a_1 L_{\text{gray}} + a_2 L_{\text{white}} \\
 L_{\text{pixel}} &= 0.5(12.11) + 0.5(8.37) \\
 L_{\text{pixel}} &= 10.24 \text{ W/m}^2\text{sr}\mu\text{m}
 \end{aligned} \tag{3}$$

where a_1 and a_2 are the fractional areas of the materials within a pixel.

3.3 Study 3: Capturing sub-pixel scene content in synthetic hyperspectral imagery

We use DIRSIG to simulate hyperspectral sensors where the simulated data is sometimes used to test target detection and classification algorithms. In these situations it is as important to model the background pixels

correctly as it is to accurately model a target pixel. DIRSIG scenes are constructed using a combination of geometry, material maps, and spatial-spectral variation models (texture) and as a result a scene does not feature a single scale that is spatially constant. For example the GSD of the image might be 1 meter but the texture map of a tree within an image might be on the order of 20 cm in terms of resolution. We want to ensure that a given DIRSIG image captures all of the complexity of the scene. To do this, pixels need to be finely sampled to capture the underlying variability as described by the scene database.

In this study a 200x200 pixel image was rendered of an urban residential area. The scene contained grass fields, trees, asphalt, and buildings. A noise free 210 channel hyperspectral sensor with a wavelength range of 400-2500 nm and 10 nm Gaussian channels was modeled. A GSD of 1 meter was used. The rendered image is shown in Figure 8.



Figure 8. Synthetic image used in Study 3.

The image was rendered with 1x1, 5x5, and 10x10 grid sampling as well as adaptive sampling. For the adaptive sampling run, 25 sub-elements were used and the number of samples per pixel was capped at 100 samples.

4. RESULTS

4.1 Study 1: Aliasing

One of the primary reasons to subsample a pixel is to capture variation due to structure in the scene that is undersampled by the detector array. While the examples used here are constructed for testing purposes, real world examples include man-made objects such as building edges, roads, and rails; natural objects such as leaves, rocks, and coastlines; and artifacts of the modeling process such as the pixelated structure of texture and material maps (essentially images mapped onto surfaces) and facetized geometry. In all cases, if variation exists within a pixel, there is a distinct, hard boundary between one type of material and another. The tri-bar charts in Figure 9 have many "hard boundaries"

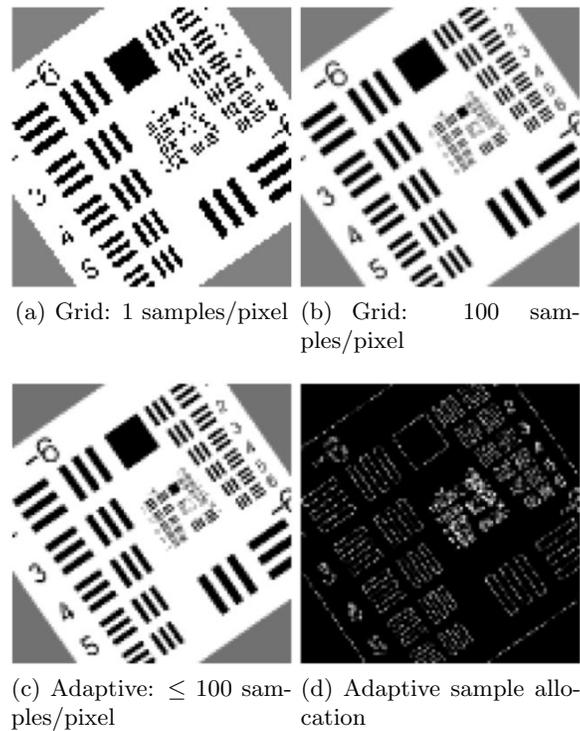


Figure 9. 1951 USAF tri-bar resolution target sampled with both grid and adaptive sampling

Since the chart itself is a binary image, a single ray from a pixel only hits a single value in the target (either white or black). This has the qualitative effect of making the diagonal edges appear jagged in the resulting image (far left), artifacts that are sensitive to the exact alignment of the pixels (and their centers) and the in-scene edges.

The obvious way to deal with this is to send more rays per pixel (sub-sampling) to attempt to find the right proportions of each material in the pixel. This works quite well, as can be seen in the second generated image using 10x10 grid sampling (100 samples/pixel). However, it is clear that in constant areas (open space away from the tri-bars) we do not need a hundred samples to capture what is occurring in a constant, uniform region and many samples (and the associated computation time) are wasted. In contrast, adaptive sampling (third image), attempts to only generate samples where they are needed, generating the proportions needed around boundaries, yet saving time when they are not (in this case running in approximately a sixth of the time as the equivalent grid sampling). The reason for the speed up is obvious when looking at the final image in Figure 9(d), which shows the number of samples that end up being allocated per pixel, ranging from 9 (the minimum for this case) to 93 (seven less than the maximum number allowed and the count used for every pixel in the grid case).

4.2 Study 2: Absolute sub-pixel thermal radiometry

The radiance as a function of the number of samples taken in a 3x3 grid using both grid and adaptive sampling is shown in Figure 10.

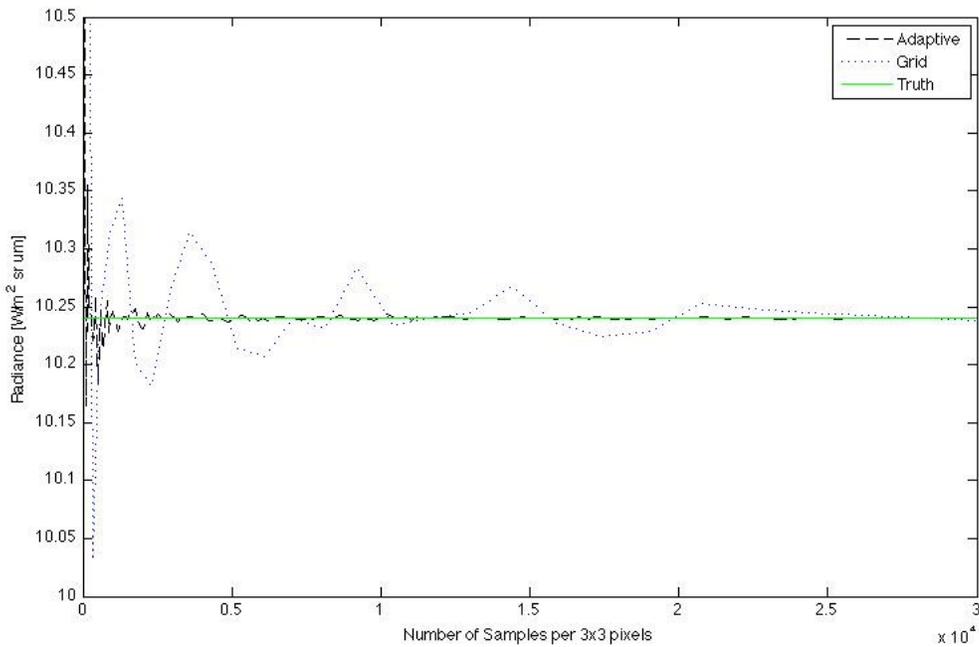


Figure 10. Plot of radiance at $8.5 \mu\text{m}$ vs. number of samples per 3×3 pixels.

The radiance value for the target pixel from the grid sampled images oscillates around the truth radiance value more than the radiance value obtained from the adaptive sampled images. The radiance value from the adaptive sampled images converges and stays close to the truth with fewer samples than the grid sampled images. The small oscillations around the truth line in both cases represents $\pm 0.01 \text{ W/m}^2 \text{ sr } \mu\text{m}$ which translates to a change in temperature of 0.053°C at $8.5 \mu\text{m}$. There is no magic number in terms of the number of samples that need to be cast to achieve a certain amount of radiometric accuracy. The plot in Figure 10 is only valid for the geometry described in Section 3.2. However, the trend of adaptive sampling approaching the correct radiance value with less samples than grid sampling is true for all geometries. Figure 11 summarizes the number of samples and the error in temperature from the truth for the entire pixel.

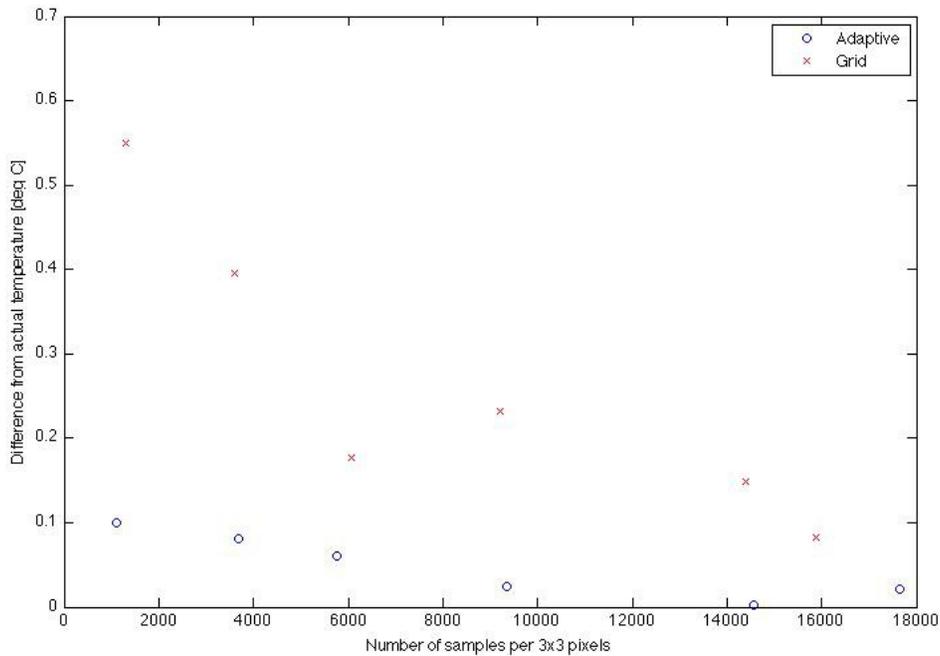


Figure 11. Grid sampling vs. adaptive sampling: temperature error comparison (for 1 pixel)

In this example, if a temperature error of 0.1°C was acceptable then an image that had been adaptively sampled needs only 1,118 samples over a 3×3 pixel area as opposed to 15,876 grid samples. These values are for a 3×3 pixel area and one could imagine a much larger discrepancy in the number of samples used as the area of interest grows.

4.3 Study 3: Capturing sub-pixel scene content in synthetic hyperspectral imagery

A major advantage of adaptive sampling over grid sampling is that the user does not need to be aware of the different scales within the DIRSIG database in order to get a radiometrically accurate image. With grid sampling one has to determine the smallest scale within the database and then determine what that corresponds to in terms of the number of sub-elements needed to capture all of the variability. When adaptive sampling is used, the code determines whether or not there is a change in radiance values within a pixel and therefore if that pixel needs to be sampled with more than one ray per sub-element. The image in Figure 12 shows the adaptive sampling density, with bright areas being highly sampled regions and dark areas sampled less frequently. Geometrically “interesting” targets cause the adaptive sampling routine to spend more time examining these areas.



Figure 12. Samples map from adaptive sampling run.

The darkest areas include the tennis courts where there is little texture and the brightest areas include the trees in the top of the image where there is a lot of spatial variation.

Figure 13 shows the projection of all of the pixels from the image in ENVT's n-Dimensional Visualization tool.

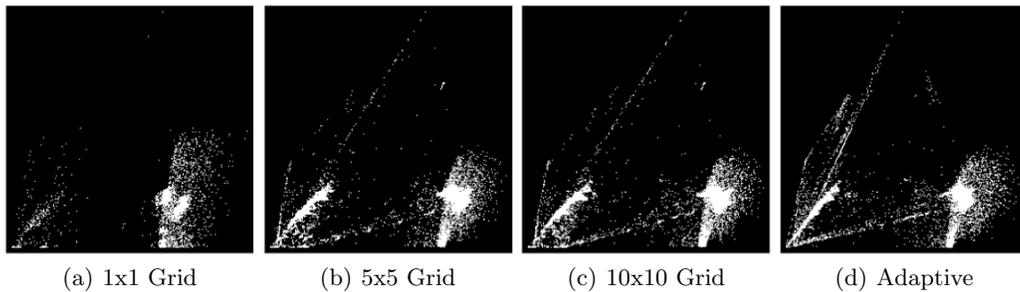


Figure 13. Projection of all 40,000 pixels in the image. Sampling increases from left to right.

The projection in Figure 13(a) shows two distinct clusters with the hyperspace between the clusters being mostly empty. With one sample being taken per pixel it is intuitive that the clusters would be distinct regions because sampling with one ray per pixel does not allow for mixture of materials. A clustering algorithm would do well in this situation, however this is not indicative of what is seen in real data. As the sampling increases (Figures 13(b) and 13(c)) the hyperspace fills in and simplexes start to emerge. These simplexes indicate the mixture of the radiance values from the materials within pixels. In Figure 13(d) to the left of the large simplex there is a smaller, thinner simplex which is difficult to see even in the 10x10 grid sampling image. A simple clustering algorithm, for example, would have a more difficult time separating out classes in Figure 13(d) than in Figure 13(b). However, Figure 13(d) better approximates what is seen in real data.

Table 2 summarizes the number of samples per pixel, the average number of materials found per pixel, and the relative run time for the grid sampling cases as well as the adaptive sampling case.

Table 2. Comparison of grid sampling and adaptive sampling using the number of samples cast per pixel, the average number of materials found in a pixel, and the run time relative to the 1x1 grid sampling case.

	Min/Avg/Max #Samples/Pixel	Average Materials/Pixel	Relative Run Time
1x1	1/1/1	1.00	1
5x5	25/25/25	1.51	25x
10x10	100/100/100	1.58	100x
Adaptive	25/52/100	1.65	47x

Adaptive sampling used 52 samples per pixel on average, which is half the number used in the 10x10 grid sampling case. Recall that the maximum number of samples allowed for the adaptive sampling case was 100, which is the same number allowed in the 10x10 grid sampling case. In some areas (i.e. trees) at least 100 samples were needed to capture the variability whereas in areas like the tennis courts (excluding the lines) only 25 samples were needed to characterize the pixel. The average number of materials per pixel was calculated for each of the sampling cases by averaging the number of materials in each pixel of the materials map. As the sampling increases so does the average number of materials per pixel. The average number of materials from adaptive sampling is larger than the 10x10 grid sampling case meaning that the grid sampling cases have not yet captured all of the variability in the scene. It does not necessarily mean that the “true” average is 1.65, but it does mean that adaptive sampling is closer to completely characterizing the sub-pixel variation than the 10x10 grid sampling. If the maximum number of samples had not been capped at 100 adaptive sampling might have, on average, found more materials per pixel. The last column of Table 2 shows how long it took each sampling scenario to run relative to the 1x1 grid sampling case. The adaptive sampling run took approximately half the time of the 10x10 grid sampling run and did slightly better at capturing the sub-pixel variability.

5. SUMMARY

The DIRSIG tool has two sampling methods with which to calculate the radiometry for mixed pixels. One is a grid sampling approach where each pixel in the image is divided into a user-defined number of sub-elements and a sampling ray is cast from the center of each sub-element. The second is an adaptive sampling approach where each pixel is divided into sub-elements and a sampling ray is cast into a random location in each sub-element. Adaptive sampling looks for variation both within and between sub-elements and sends more rays where it finds variation.

Study 1 shows that DIRSIG is accounting for sub-pixel structure efficiently using the adaptive sampling technique. Study 2 shows that sub-pixel structure is being accurately modeled in a radiometric sense. Study 3 shows that adaptive sampling is an effective tool for getting radiometrically accurate predictions without *a priori* knowledge of inherent scales and spectral complexity.

6. ACKNOWLEDGEMENTS

This work was carried out using funding received from the United States Department of Energy under BAA PDP08 Grant Number DE-AR52-07NA28115. Data visualization was accomplished using ITTVIS ENVI software. RIT is designated as a Geospatial Innovation Center for Excellence by ITT Visual Information Solutions.

REFERENCES

1. Digital Imaging and Remote Sensing Laboratory, Rochester, NY, *The DIRSIG User’s Manual*, March 2008.
2. M. Pharr and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*, Elsevier, Inc., 2004.
3. H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, Society for Industrial and Applied Mathematics, 1992.