A Survey of Non-Raster Scan Methods with Application to Atomic Force Microscopy

Sean B. Andersson
Dept. of Aerospace and Mechanical Engineering
Boston University
Boston, MA 02215
sanderrss@bu.edu

Daniel Y. Abramovitch
Agilent Laboratories
5301 Stevens Creek Blvd., M/S: 4U-SB
Santa Clara, CA 95051 USA
danny@agilent.com

Abstract—Images in atomic force microscopy (AFM) are built pixel-by-pixel through a raster scan process and can take on the order of minutes to obtain. The problem of imaging a sample can be characterized as using a short-range or point-like sensor to obtain information about a system over a region and is common across a broad range of fields in science and engineering. In many cases, as in most AFM images, the region to be scanned consists primarily of empty or uninteresting space. In this situation raster-scanning, while easy to implement, is extremely inefficient. It can be viewed as an open-loop scheme because no use is made of data being acquired by the sensor. In this paper, we survey results from the literature describing alternative scanning and sampling approaches. These algorithms often use prior information about the system being measured as well as real-time feedback from previously measured points to keep the sensor in the regions of interest.

I. INTRODUCTION

The problem of using a point or short-range sensor to collect data is common to a wide-variety of applications. Such applications include imaging on the nanoscale using scanning probe microscopy (SPM) techniques, display technologies, microfabrication, environmental monitoring, weather, geography, and many more. A generic approach is to raster-scan the sensor throughout the region, collecting data along a regular grid. In the imaging scenario, this leads to a process in which an image is built pixel-by-pixel. Raster scan is an open-loop process and in the absence of any information about the process or sample of interest, it is a rational thing to do.

However, by its very nature, a measurement system is gathering information. Non-raster scan methods make use of this information to alter the measurement process itself. Under this approach, measurements can be made where they will be most effective, namely where the process or sample has spatially or temporally varying features, and can avoid uninteresting areas. The overall measurement time is thereby shortened.

In this paper we discuss non-raster-scan approaches from the literature of several different fields. The goal is to collect these results in the hopes that they will inspire further work on applying similar techniques to scanning probe microscopy (SPM) in general, and to atomic force microscopy (AFM) in particular.

This paper is organized as follows. Section II will discuss non-raster methods used in scanning probe microscopy. Section III will delve into unevenly spaced measurements made in the area of environmental monitoring. Section IV will describe some engineering applications where measurement speeds have been increased through the use of non-raster scan methods. In Sections V and VI, two unifying themes from all of these problems are explored. First, Section V will discuss the issue of extracting data quickly from non-uniformly spaced sample points. Next, Section VI will discuss sample point selection.

II. TECHNIQUES IN SPM

The technologies of SPM, including AFM, scanning tunneling microscopy (STM), and scanning electron microscopy (SEM), generically rely on building an image pixel-by-pixel through a raster-scan of the detector through the sample. This approach has been extremely successful in investigating the structure of systems with nanometer-scale features. However, these same tools are now being called upon to explore dynamic processes. As a result, there is growing interest in developing techniques to improve the temporal resolution of such instruments. While there is considerable work in speeding up the physical measurement process of any given pixel [1]-[3], non-raster scan methods attempt to speed up the overall measurement by measuring fewer pixels and concentrating those measurements in areas where the entity to be measured shows features.

Some of the earliest attempts at non-raster scanning can be found in the context of SEM. In the standard raster pattern, the time difference between two pixels adjacent in the slow-scan direction is much larger than the difference between two pixels in the fast direction, leading to an anisotropy in the image. Alternative patterns were proposed to simplify the temporal relationship between pixels as well as to simplify the control electronics and reduce image distortions by providing for a constant-speed motion of the electron beam across the sample [4]-[6]. In recent work, researchers have attempted to solve the same problem for large images by taking advantage of characteristics of the human visual system to reduce the sampling while maintaining a high-quality image for viewing [7].

Soon after the development of STM, a tracking technique was suggested to lock the motion of the tip to particular surface structures [8]. Under this approach, tracking is achieved by moving the tip in a small circular motion, using the resulting data to estimate the gradient of the property of interest, and then using the gradient to drive the overall
motion of the tip. This idea was later used to study surface diffusion by tracking single atoms [9]-[11] and continues to be developed and utilized in STM [12]-[15]. It has recently found application in AFM as well [16], [17]. Recent work promises to improve the temporal resolution of tracking techniques in AFM by using the transient dynamics to detect the presence or absence of an interaction between the tip and the sample [18].

The actuators used in SPM typically suffer from thermal drift and creep [19], [20], leading to difficulties when imaging large samples or imaging samples over extended periods of time. The atom-tracking methods discussed above can compensate for these effects but at the cost of imaging only a single feature. An alternative technique has been proposed which uses surface features as local reference points [21]. Rather than the raster-scan pattern, data is sampled local to the reference points and images are built up from these “patches”. Feature tracking has also been utilized to explore structured samples such as integrated circuits [22].

One of the authors has developed a method for the rapid imaging of string-like structures in AFM [23]. Under this approach, data from the tip is used to estimate the path defined by the structure of the sample. The tip is then steered so that the scan is taken only local to the area of interest. A standard raster-scan image of a carbon nanotube is shown on the left side of Figure 1 while the right side shows the same nanotube imaged using the local raster-scan technique. Notice that most of the raster-scan image is of completely uninteresting substrate. Each such pixel represents wasted time. The local raster-scan image consists primarily of the nanotube. A similar technique was introduced in [24] for imaging boundaries using AFM.

Fluorescence microscopy is an extremely useful technique for understanding molecular processes inside living cells. Single particle tracking techniques use sequences of CCD images together with image processing techniques to determine the motion of isolated molecules [25]. However these approaches are not effective in three dimensions and do not have high enough temporal resolution to explore conformational dynamics of single molecules. Confocal methods, in which the light is passed through a pinhole to block the signal arising from outside of the focal point, and multiphoton microscopy, in which only a small volume of the same is excited to fluoresce, do have the capability to image in 3-D. Moreover, because measurements are performed using single photon counters such as avalanche photo-diodes, the temporal resolution can be orders of magnitude higher than with CCDs [26]. The tradeoff is that the measurement volume is small, on the order of 0.25 femtoliters. Images can be built through raster scanning but this severely reduces the temporal resolution of the device. As a result, in recent years several researchers, including one of the authors, have proposed algorithms to replace the raster scan with tracking algorithms [27]- [32]. This approach allows single molecules to be investigated with the temporal resolution afforded by the confocal approach but across a wide field-of-view.

III. TECHNIQUES IN ENVIRONMENTAL MONITORING

Groups of autonomous robots are increasingly being used to explore and monitor a wide variety of environmental phenomena. In many cases the phenomenon occurs over a vast area and it is not practical to scan the entire region. One general approach is to track a boundary, typically defined as a level set of an appropriate function [33], [34]. Such work is inspired by applications such as monitoring harmful algae blooms [35], [36], oil spills [37], and forest fires [38]. Similar techniques have been utilized to seek the source of a toxic plume [39].

An example effort in this area is the Adaptive Ocean Sampling Network. This project uses small fleets of underwater gliders to collect oceanic data and to modify the sampling strategy in real-time based on model predictions updated by the incoming data together with optimal control laws for steering the gliders [40]. This application requires techniques to investigate relatively small-scale features with high resolution through tracking and sampling [41] as well as approaches for larger-scale sampling patterns [42], [43].
Non-uniform sampling also plays a large role in weather sampling applications. (See Figure 2.) Sampling over land is done by launching weather balloons while sampling over the ocean must be achieved by launching the instruments from ships or aircraft. Because the latter is much more expensive than the former, most of the data is acquired over land. This has inspired such programs as the Aerosonde in which long range miniature robotic aircraft – flying for days at a time – are used to sample weather over the ocean [44]. These aircraft can follow and sample the borders of a weather pattern rather than being restricted to one region.

IV. TECHNIQUES IN ENGINEERING APPLICATIONS

The use of non-uniform sampling is a fairly common technique in instrumentation applications. Dithered sampling shows up in methods to avoid aliasing in both 2D imaging (see, e.g., [45], [46]) and the decimation of time domain signals. For example, in many oscilloscopes, the data is sampled at a far higher rate than can be displayed. To avoid aliasing in the decimated display, the decimation clock can be dithered, so that different phases of the original sampled signal are displayed [47], [48]. However, many of these are applications in which the original data is taken at a high rate and it is the decimated data which must be sampled in a non-uniform method. In such applications the display, rather than the sampling of the data, is the limiting factor.

One non-raster scan example from measurement system design arose from one of the authors’ work in Bit Error Rate Testers (BERT), and specifically in the generation of eye maps [49]. A standard BERT eye map is shown on the left side of Figure 3. The right side of Figure 3 shows a non-raster scan eye map generated by selecting points using a method of fitting ellipses to contours of constant bit error rate [50]. What is important to note in Figure 3 is that in the region of features (in this case a significant change in Bit Error Rate), the non-raster scan method has a higher sampling rate and therefore a higher resolution than the raster scan method. For this reason, the non-raster method captures the essential surface information in 154 rather than 1090 sample points. The non-raster method achieves a second improvement by making far fewer measurements at points with very low BER. These low BER points take far longer to qualify than high BER points.

The raster-scan pattern has been used in electron-beam lithography as the fundamental trajectory of the beam. Surface contours are generated by repeatedly scanning regions to increase the depth of the sputter crater. In order to improve the accuracy of the features, the vector scanning technique was introduced [51]. Under this approach, the dwell-time of the beam at a particular location is controlled to create the desired depth, resulting in more accurate feature generation, a minimization of overexposure at crossing points such as the intersections in a grid, and provides the ability to compensate for effects such as increased exposure from back-scattering [52]. As the technology progressed and the accuracy demands increased to the sub-20 nm range, it became necessary to measure the relative position of the electron beam and the substrate. This was generally achieved by patterning a fiducial grid and then sampling that grid with a sub-threshold dose. The data so obtained was used to calculate positioning errors before writing the desired pattern. Because sampling the entire fiducial pattern with a raster-scan would be slow and would likely lead to exceeding the threshold, a sparse-sampling pattern was proposed. Because of the relative spatial frequencies of the sampling and fiducial grids, the measurements yield a two-dimensional Moiré pattern. This pattern can be analyzed to yield the positioning errors [53].

![Fig. 3. Raster scan and non-raster scan eye maps. Note that the non-raster scan eye map on the right produces approximately the same level of accuracy as the raster scan version in far fewer sample points.](image)

![Fig. 4. Map from [54]. Note that in this case, the cell phones report their location, the time of the measurement, and the signal power. The number of measurements from a given area at a given time interval is a measure of the traffic. (Reprinted with permission from [54].)](image)

The engineering application of mapping cell phone strength or traffic patterns [54] is shown in Figure 4. In this application, GPS enabled cell phones transmit their location, the cell phone reception strength, and the time of the measurement to a network of measurement servers. The servers in turn aggregate the data. Not only does this give a measurement of cell phone reception at different locations and times, but the number of measurements at a given area in a given time span gives a measurement of the traffic flow. The necessity of non-raster measurement is clear here: measurements in places with a lot of cell phones – such as major traffic areas – are essentially free. Measurements away from much traffic are much harder to come by. This is not
an issue if one wants to measure traffic speeds over a region, since one only cares about locations with cars. However, if one wants to measure cell phone signal strength, then areas with few measurements must be estimated. With these types of measurements, one of the major costs is the computation to aggregate the data. In Figure 4, major traffic arteries have so many measurements that the display of this data would be improved by reducing the number of points displayed.

V. Extracting Information from Images with Unevenly Spaced Samples

When images are generated from a raster scan, the underlying sample grid has evenly spaced samples. This makes interpretation of the surfaces generated from the scans straightforward. One of the universal characteristics of non-raster scan generated images is the need to interpolate the measured data. In this section, we will discuss two such methods, one coming from computational geometry and one coming from geostatistics. While these two examples are common in the literature, there is by no means a consensus on which methods are best. Furthermore, both of these methods assume noise free sample points. However, they can be used to generate an evenly spaced 2D grid from unevenly spaced samples.

A. Surface Interpolation Via Delaunay Triangulations

Delaunay triangulations are used in the computation and visualization of 3-dimensional surfaces from data taken on 2-dimensional grids. Typically, this data represents a surface height, but in the measurements taken by AFMs, the Z axis can represent phase, friction, or some higher harmonic of a demodulated dynamic mode signal [1].

The premise of interpolation via triangulations is that [55]:

- Any grid of points on a plane that are not all co-linear can be enclosed in a convex hull, where a convex hull is defined as the smallest convex polygon that encloses all the points on the surface.
- This convex hull can be triangulated. That is, the area within the convex hull can be subdivided into a set of non-overlapping triangles that cover the entire convex hull.
- By assigning the measured surface values to the vertices of the triangulation, interpolated surface values can be generated between the measured ones.

There are many theorems used by computer scientists on how many triangles and edges are needed to triangulate a set of planar data and how many operations are needed to compute them. Also, it is the case that for any set of points, there are many possible triangulations. In particular Delaunay triangulations attempt to optimize the interpolation by generating triangles with the largest smallest angle. That is, a Delaunay triangulation will return the most symmetric triangles, and will avoid generating long skinny triangles. The idea is that by choosing a triangulation with the most symmetric triangles, we are interpolating between sample points closest to the interpolation point. While convenient, the Delaunay triangulation assumes each data point is perfect and has no provision for including measurement uncertainty when generating the surface.

Delaunay triangulations are used in Matlab to compute interpolated surfaces of nonuniformly sampled data in the griddata function [56]. There are several options for the interpolation, including linear, cubic, and nearest neighbor interpolation.

B. Surface Interpolation Via Kriging

At the other end of the spectrum from the computational geometry methods is Kriging, a method that arose from mining and petroleum exploration [57], [58]. In these fields, the act of making a measurement involves drilling a well or digging a mine, so it is important to interpolate as many data points as possible from the measurements already made.

Kriging is controversial in some circles, because of the practice of assuming certain spatial relationships between the data without going through any rigorous confidence tests on those relationships [57], [59]. Still, it remains a widely used technique in a fields for which there is a need to minimize the number of sample points.

For a set of \( N \) sample points on a 2 dimensional grid, Kriging generates a linear estimator for the value at a point, \( p \),

\[
z(x_p, y_p) = \sum_{i=1}^{N} W_i z_i,
\]

where the \( z_i \) are the \( N \) points of the sample being used to estimate the value of new point, \( z(x_p, y_p) \) and the weights \( W_i \) are normalized to sum to 1.

Kriging assigns a spatial variance between sample points called a semi-varigram which is related to some measure of the distance between them. The three most common forms are the spherical, exponential and Gaussian. The idea is that the influence of any sample point on an interpolation point falls off in some proportion to the distance between the sample point and the interpolation point. As with any other interpolation method, the ability of Kriging to predict a new point depends upon how close the previous sample points are to the new point and how much the semi-varigram is affected by distance. The main issue in this method is how the user establishes the spatial correlation between sample points.

Kriging is common in problems where measurements are expensive and so computation is relatively cheap. It does not appear to be a particularly efficient computational method since for any new point, semi-variograms from all the previous measurement points need to be computed. This stands in sharp contrast with the triangulation method discussed in Section V-A which interpolates surfaces between the 3 “nearest” measurement points.

VI. Sample Point Selection

One of the common threads in all of the techniques which have been discussed is that of sample point selection. Unlike raster scan methods, the selection of the next set of sample
points depends upon the previously measured data. The issue of how to use this data to generate the next set of sample points is one of the major topics for such measurements.

The maps generated from the methods in Section V (and similar ones) allow for the extraction of features, typically by noting regions of high gradient in the surface fit. Sample point selection is often done then by finding a region of high gradient and then moving either in the direction of that gradient or orthogonally to it. In the non-raster measurement of a string like object with an AFM [23] described in Section II, a sample is considered either on or off the string (making the estimated derivative infinite at the border). The scan for new measurements has a slow direction (in the direction of the string) and a fast direction (perpendicular to the string) so that the estimated path of the string generates new points. For the environmental or BERT examples measurements are often made in regions of constant height, causing the aircraft, submersible, or BERT to encircle the region of interest with successive scans. These new sample point methods vary between using mostly local data and data extracted from the entire map.

Looking at these applications, there are two basic approaches. In one, demonstrated in [23], the next sample point is a function of the most recent sample points in the region where the system finds itself. In another, demonstrated in [49], the next sample point (or sets of sample points, come from extracting information about the entire image and then generating new sample points. The traffic measurement example stands in sharp contrast to these as sample point selection is passive, but guaranteed to be high in high traffic areas.

Some of the main differences between these two applications are quite illustrative. In the case of the BERT eye diagram, discussed in Section IV, certain measurement positions require considerably longer measurement times than others. Conversely, there is little time lost in selecting consecutive measurement points that are far apart. Note that in this case, there was an implicit assumption that the eye map was of something static. For sample point selection in AFMs, it is worth noting that there is essentially no difference in measurement time between one point and another. There is a time cost in moving in XY range arbitrarily, so there is an incentive to choose measurement points near the previous measurement points.

Similar differences exist in the weather measurement and underwater mapping applications. In the former, communication with extensive computing resources is relatively easy and so decision making about the next sample points can take into account more global information. In the latter, the underwater nature of the measurements severely limits long distance communication, so the decisions about the next sample point are largely made locally. These two applications have the extra factor that some of the most interesting measurement points may involve physical danger to the vehicle. For the weather monitoring aircraft, the edge of a weather pattern is a great place to make measurements. However, if that weather pattern turns out to be a hurricane or a tornado, the result of selecting the most interesting measurement point may be the destruction of the aircraft. This adds new meaning to the phrase “cost of measurement.”

VII. Conclusion

While non-raster scan measurements are quite common in a variety of fields, there appears to be no unified method of characterizing these. Much of the discrepancy seems to come from the tradeoff between the cost and number of measurements versus the cost of computation. An AFM lies in the middle of this space. While the main cost of a sample point is the time of the measurement, there is usually abundant computation available as well. Unlike consumer applications, AFM designers can afford to add extra processing capacity to the problem. The application of non-raster scan methods to AFMs is in its infancy. However, the promise of being able to concentrate measurements at areas of interest makes it a worthwhile area of exploration.

References
