DNA Microarray Spot Detection Using Hough Transforms

William H. Horsthemke
Jacob D. Furst
DePaul University
School of Computer Science
Chicago, IL
horsthemke@acm.org
jfurst@cs.depaul.edu

Abstract

Genetic fingerprinting using DNA microarrays has shown significant promise in rapidly detecting dangerous microorganisms. Though primarily a genetic and bio-chemical technology, one stage of the detection process records an image of the genetic signature. This image contains a variety of illuminated spots which indicate the type of microorganism by their location and intensity. Analyzing these very noisy images currently requires a trained biologist to locate the spots for computer measurement. This paper presents initial work on exploiting the roundness of these spots to identify and measure them with a noise resistant technique from the family of generalized Hough Transforms. Initial results illustrate the advantage of this approach and indicate methods and directions for future work.

1 Introduction

Microarray technology shows significant promise in reducing the cost and speeding the application of genetic fingerprinting. By illuminating the biochemical reaction of a self-matching property of genetic material, the microarray generates an illumination pattern representing a genetic signature of a microorganism. This illumination forms the input to an image analysis technique to measure the specific genetic fingerprint. Though the microarray represents a grid of measurement points designed to permit a direct intensity measurement application, the overall process introduces sufficient noise and artifact to confound these simple approaches to fingerprint measurement. The approach of this nascent research endeavor attempts to exploit the radial formation of illumination resulting from the biochemical process of genetic matching. Since this illumination should emanate radially from the source of the match, the areas of interest in the images should form round objects suitable to an edge-based circle detection approach.

Microarrays identify a microorganism by comparing its genetic makeup to an array of shorter genetic sequences, a form of pattern matching. Similarity between organisms is indicated by the correspondence between their sequence matching. Matching an organism to an array of short genetic sequences results from the biochemical process of gene expression where mRNA molecules bind (hybridize) to the (complementary) DNA templates from which they originate. This biochemical matching process of hybridization can be instrumented for measurement by tagging the DNA templates with fluorescent tags. These tags will illuminate where matching

occurs on the microarray and indicate that the genetic sequences stored at that microarray location matches the microorganism DNA. Partial matching between DNA and mRNA can occur. Hybridization does not require a strict sequence match such as a string matching algorithm, but will increase with the degree of matching. Thus, the overall illumination at that microarray location will not be on or off, but vary depending upon the degree of genetic matching.

The microarray consists of a regular grid of spots where the genetic probing takes place and where the illumination will occur. A camera records the hybridization process and records an image of the final result. The degree of illumination at each microarray location indicates the amount and type of genetic expression, fingerprint, of the sample organism.

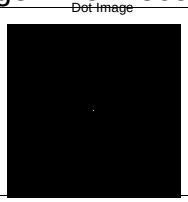
The major challenges result from the irregularity of the grid and the appearance of significant illumination artifacts which closely resemble the expected illumination of the genetic markers where the mRNA probes are placed. As an image analysis problem, the microarray challenge requires finding the location of the microarray probes where the hybridization might occur and the "spots" resulting from the fluorescence of the biochemical reaction. Once the candidate locations have been determined, various spot illumination measurement techniques can be performed to determine their discriminatory power and robustness.

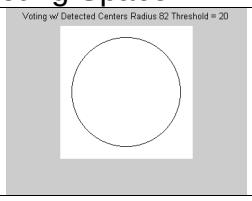
The research described in this paper approaches microarray spot discovery as an image formation problem, where the spots result from biochemical fluorescence which should emanate radially and form a generally round appearance. Assuming a round appearance for spots, the problem can be reduced to an edge followed by a circular detection problem. Though the problem reduces to circle detection, the inherent noise in the image, as well as the expected variability of the biochemical process driving the image formation, requires an approach suitably resilient to the presence of noise and circle irregularity. The approach adopted by this paper follows the general Hough transform for circle detection. .

2 Hough Transform

The Hough transform establishes a voting space constructed from the parameters of a closed form expression describing an object such as a line or circle. Points in the voting space represent objects in the original image. The Hough transform process starts by creating an edge map of an image. For each point on the edge map, votes accumulate in the voting space. For each pixel in the edge map a circle is created in the voting space. The following illustration shows the voting pattern which results from a single pixel in the edge map.

Figure 1. Edge Pixel Becomes Circle in Voting Space



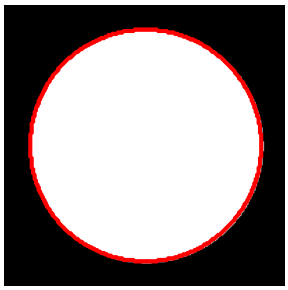


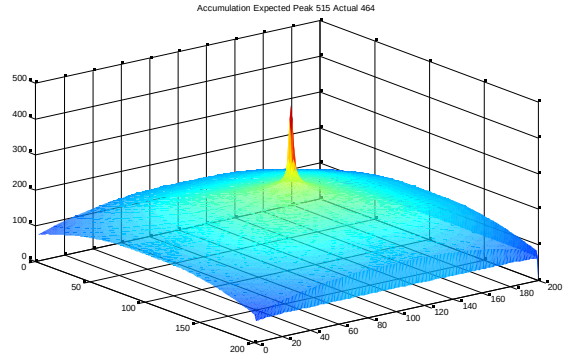
Single Edge Pixel | **Possible Circle Centers**

As shown in Figure 1, a Hough transform to find a circle begins with an edge map and a target circle radius (radius = 82 in the example). When a point is found in the edge map, a set of votes are tallied in the voting space. In fact, for circle detection, a vote in the original edge map of an image becomes a circle in the voting space. This circle represents the possible center points of a candidate circle which could have produced the original edge in the original image. After the final tally, the voting space contains tallies for various potential circles, some of which are circles while others result from other non-circular edges. A fully described circle would have a vote tally of 2*Radius*Pi, while a half-circle would have a vote of (2*Radius*Pi)/2.

The following illustration in Figure 2 shows the voting results from an edge map containing a single circle. The following illustration shows the voting results from a single circle in an edge map. As each pixel point in the edge map votes for a circle in the voting space, a conical shape emerges with high vote tallies in the center and lower votes near the edges. The high values in the center of the voting space indicate likely locations for the center of the edge circle. Several candidate centers will likely exceed the threshold for circle detection, but only one must be selected as the location as the center of a DNA probing spot.

Figure 2. Circle Edge Becomes Conical in Voting Space





Ideal Circle Edge Map | **Voting Results from Hough Transform**

This voting mechanism of the Hough transform offers a form of circle similarity measurements suitable for half-circles, irregular circles, as well as circles with missing data values. This resistance to noise and tolerant of irregular or incomplete objects allows the Hough transform approach to detect objects such as circles in noisy images such as the microarray.

3 Experiment and Initial Results

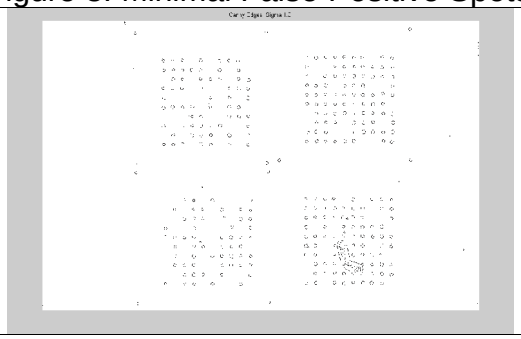
The experiment analyzed the images collected from a DNA microchip containing four hundred (400) probes arranged in four (4) grids of 100x100. The microchip also contains twelve (12) control points along the top, center, and bottom. Each image represented one (1) strain from a set of 28 strains of Anthrax bacillus. A visual survey of the images found noticeable variation in the orientation of the images, missing control markers, and numerous artifacts and a variety of noise.

The methodology created an edge map of each image, searched that edge map for circles with radii in the range of 5 to 20 pixels, removed partial circles containing less than 50% of possible pixels where total possible equals *2 * Radius * Π*, then removed all but one of the final candidate circle centers. For visual analysis a red circle with the specific radius is overlaid upon the raw slide image.

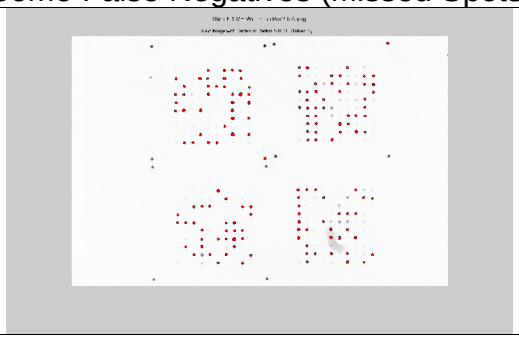
The analysis of this experiment used visual observation to assess its effectiveness in finding circles near the expected probe locations and ignoring other locations. A basic peak detection approach was also applied to consider other reference models.

Initial visual analysis confirms the effectiveness of this approach in detecting probing spots by finding circles in the presence of significant noise. However, many images contain significant circular artifacts not easily separable from the DNA spots. Rejecting these false positive spot locations might be achievable by tuning the voting threshold parameter, though this likely will reject many true spot locations and create false negative spot locations. Failure to find spots within the grid is another notable problem; while many missed locations contain no obvious illumination and therefore no edges, some spots produce edge responses but are subsequently missed by the circle detection algorithm. The cause of these false negatives will require further investigation.

Figure 3. Minimal False Positive Spots with Some False Negatives (Missed Spots)

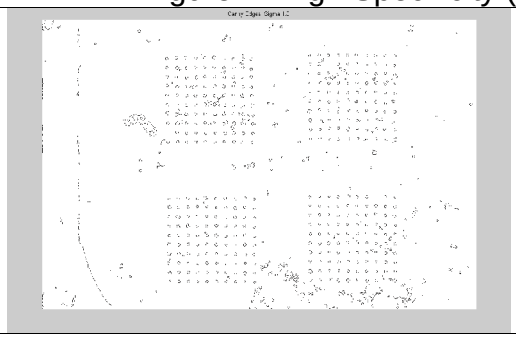


Edge Map

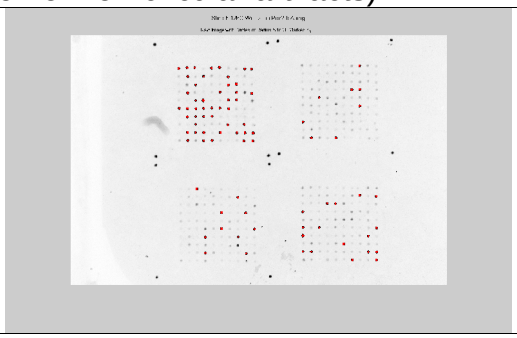


Detected DNA Spots In Red

Figure 4. High Specificity (Rejection of non-circular artifacts)



Edge Map



Detected DNA Spots In Red

4 Conclusion and Discussion of Future Work

Three task areas need further investigation: two focus on the parameters and detection algorithms while the other must develop a reference dataset to serve as a benchmark for testing. The initial work has been assessed visually to evaluate the how many visual spots were properly detected or missed and how many spurious locations were falsely detected. A labeled reference dataset will permit an automated measurement of the success of the algorithm. This reference model dataset will facilitate the other anticipated work on parameters and other detection and refinement algorithms. The approach contains numerous parameters such as the voting threshold that decides what constitutes a circle as well as the smoothing and other parameters in the edge detection step. A sensitivity analysis of these parameters should be performed to determine an appropriate setting for these parameters which provides the most robust results.

Another important task will consider the larger context of the entire microarray slide and determine how reliable can the approach detect the control points on the slide. With a suitable number and arrangement of these control points, the orientation and scale of the slide can be measured and masks can be created to identify the areas of interest where the DNA probes are located (known as wells). This masking can permit a relaxed approach to circle detection since any false positives in the non-well regions can be ignored.

References

[1] Chandler DP, Alferov O, Chernov B, Daly DS, Golova J, Perov A, Protic M, Robison R, Schipma M, White A, Willse A. "Diagnostic oligonucleotide microarray fingerprinting of Bacillus isolates," Journal Clinical Microbiology. 2006 Jan; 44(1):244-50.

[2] Y. H. Yang, M. J. Buckley and T. P. Speed. "Analysis of cDNA microarray images," Briefings in Bioinformatics, 2001 Vol. 2, pp. 341-349.

[3] Amin Sarafraz "Circle Detection via Standard Hough Transform," Matlab Central, 2004