Quality Assurance Methods
for Processing Microarray Imagery

Peter Bajcsy, Research Scientist
Automated Learning Group
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
pbajcsy@ncsa.uiuc.edu

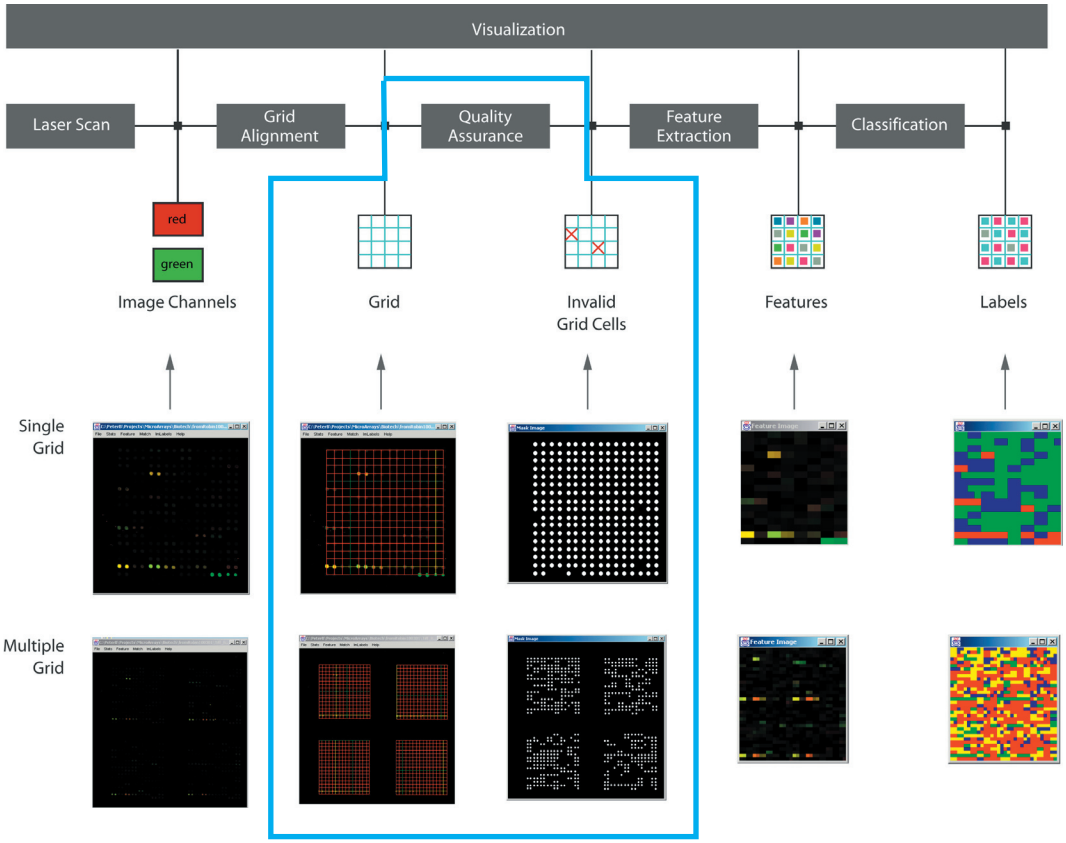
Zonglin L. Liu, Research Associate Professor
W.M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
z-liu@uiuc.edu

Lei Liu, Director of Bioinformatics
W.M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
leiliu@uiuc.edu

Introduction

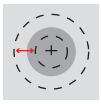
We present a set of novel quality assurance (QA) control methods for processing DNA microarray laser scanned imagery. These methods have been designed to detect systematic errors in microarray images and remove any unreliable information from further data analysis. The QA methods are applied once grid cells with dots have been identified. Each grid cell is screened for errors due to location and size, small signal-to-noise ratio (SNR), shape irregularity (topology) of a dot, and inconsistent intensity probability distributions. The goal of screening is to eliminate grid cells with spotting or hybridization errors before features are extracted and any data mining is applied.

System Design



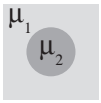
Quality Assurance Methods

The first step in processing microarray images is to conduct grid alignment. This step is accomplished by one of our Image To Knowledge (I2K) tools. The resulting grid is passed to the quality assurance tools (or screening tool) and processed according to a user's selection.



Location and Size Method

This screening method is designed to examine the location and size of dots. The method eliminates grid cells that do not satisfy radius +/- delta of the dot size where the location of a dot is computed based on local spatial statistics of a grid cell. A mask is generated as a result of the screening methods selected.

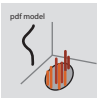


SNR Method

The QA method based on SNR analysis consists of two steps. First, analyzing the difference between minimum and maximum intensity values inside of a grid cell enables them to be marked with no signal compared to the background. Second, a SNR value computed over a dot template in a red or green image band is used for grid cell removal with unreliable information (small SNR). This type of screening can reveal problems with spills during spotting or errors during the hybridization process.

Topology Method

The goal of a QA method based on topology analysis is to eliminate disconnected or partial dots. The method is based on performing connectivity analysis inside each grid cell and eliminating those grid cells that contain the largest connected signal area outside of the allowed interval of size variation (dot area +/- delta). In other words, topologically deviating dots will be eliminated from further processing. This type of analysis is also intended to screen errors arising from spotting and hybridization.

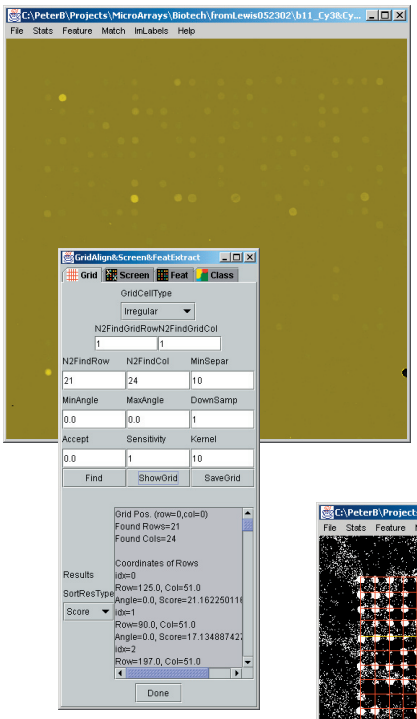


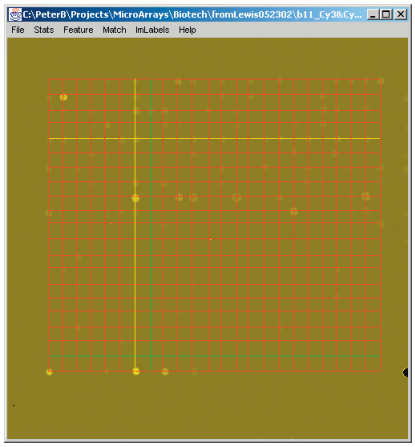
Statistical Probablility Distribution Function Method

The last QA method is based on statistical probability distribution functions (PDFs) of foreground (signal) pixel intensities inside of a grid cell. The type of PDF that models foreground is estimated by using higher order central moments and a probability distribution plane. A grid cell passed the screening test if an estimated PDF type of foreground matches the expected PDF types, for instance, a uniform PDF for highly up-regulated foreground pixels (due to saturated dynamic range). The goal of this screening is to eliminate grid cells with inconsistent intensity probability distributions.

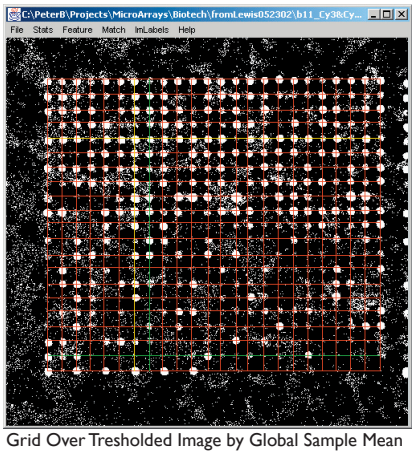
Example Results

Step 1: Grid Alignment



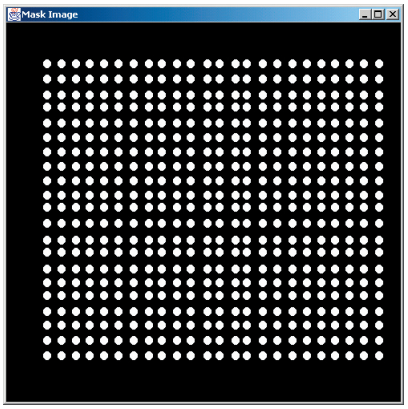


Grid Alignment Visualization

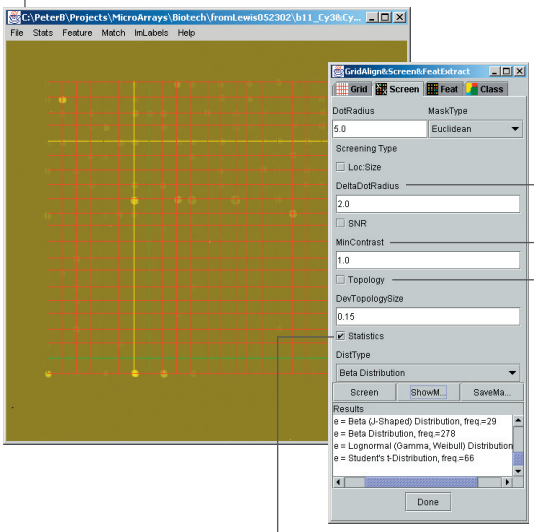


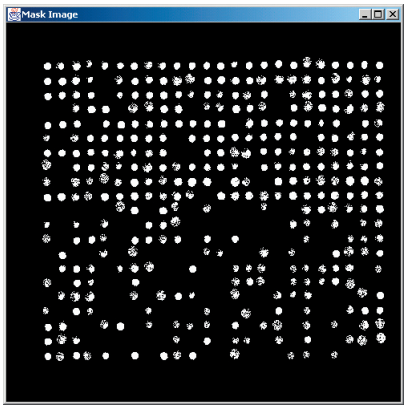
Grid Over Tresholded Image by Global Sample Mean

Step 2: Screening

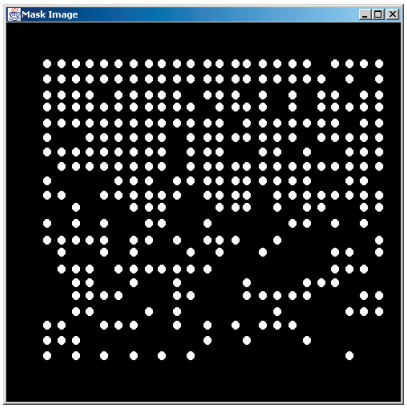


Mask Image: No Screening

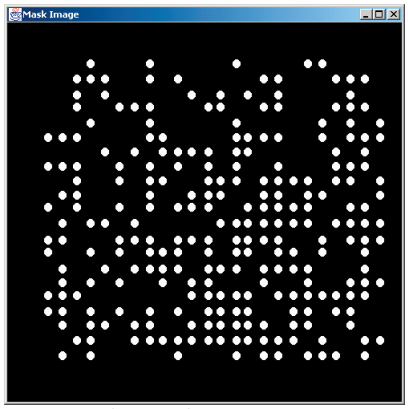




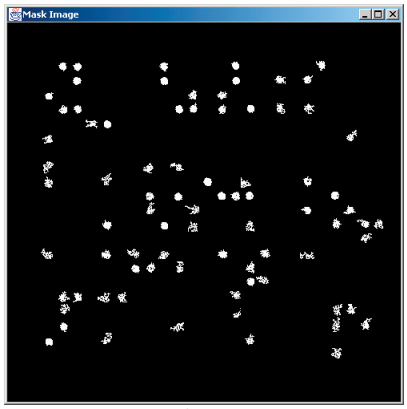
Mask Image: Location and Size Screening



Mask Image: SNR Screening



Mask Image: Statistical Screening



Mask Image: Topology Screening