Image-Based Visualization
of Microarray Features and Classification Results

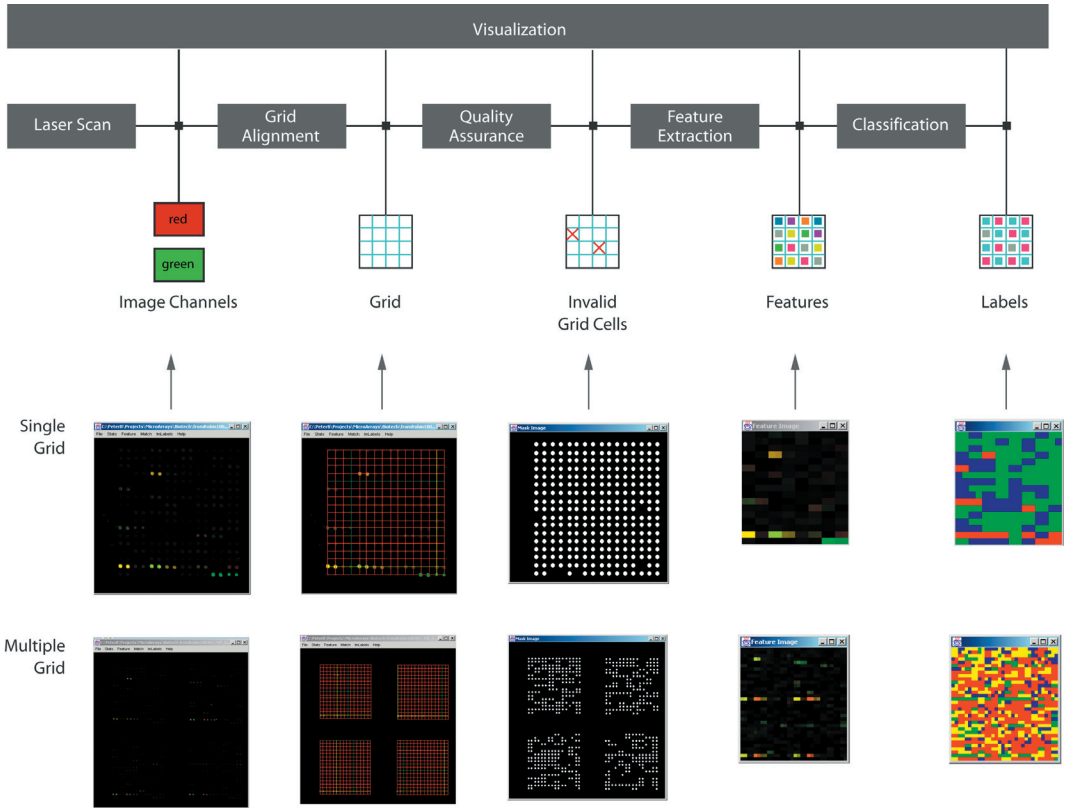
Peter Bajcsy, Research Scientist
Automated Learning Group
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
pbajcsy@ncsa.uiuc.edu

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

Introduction

In this work, we present a novel visualization approach for fast screening and inspection of DNA microarray data. The novelty lies in viewing the DNA microarray data as high-dimensional images, including laser scanned imagery, extracted features and labeled classification results. From a data analysis viewpoint, this type of display is very suitable for visual screening of errors and inspection of analyzed data because an image format is maintained throughout the entire analysis process. The microarray dots form a grid pattern, creating intrinsically grid-based information. Thus, it is natural to extract features and maintain the grid. Each extracted feature forms a point in the feature image which is then used for classification and visual inspection.

System Design



Overview of Image-Based Visualization

We have addressed the issue of fast screening and inspection of DNA microarray data by using novel image-based visualization approaches. These visualizations include input image channels, grid alignment results, screening results, high dimensional features, and classification labels.

Although extracted features from DNA microarray scanned imagery might not be spatially related, it could be beneficial to the experimentalists to introduce a spatial pattern of expected up-regulated and down-regulated genes into the design arrangement on a microarray glass slide. An introduction of a simple spatial pattern into the glass slide design could be used as a good example for performing a fast calibration. We present visual results that demonstrate the benefits of spatially related microarray grid information.

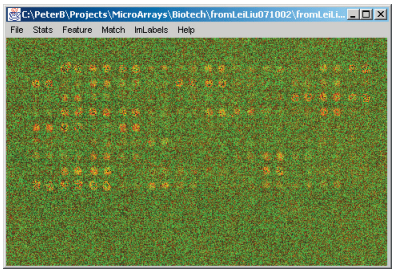
Another benefit of image-based visualization is the display of features. Extracted features are usually presented in table form with multiple variables associated with each grid location; for instance, mean, median, standard deviation, and ratios. A high-dimensional image (x, y, feature value) provides a 3D data cube with each image band (frame or 2D slice) ready to illustrate variations of features over the entire set of extracted features. For example, a fast detection of systematic errors can be conducted by visual inspection of several feature bands.

In addition to visually exploring extracted features, image-based visualization provides a means for efficient inspection of labeled classification results. For instance, a result of K-means clustering can be easily displayed with color labels as one resulting image. The results of hierarchical clustering methods, e.g., single-link or complete-link clustering, can be shown as a cross section of a labeled 3D cube (x, y, clustering level) instead of a standard dendogram-based visualization.

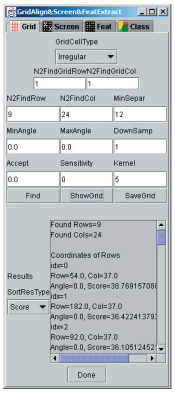
In this work, we also present the use of image-based visualization for exploration of experimental variables. In order to investigate a functional dependency of a gene expression level on a selected variable, one can stack classification results into a 3D data cube (x, y, variable value) and view a 2D cross section of the multi-grid data cube along the variable axis. This type of visualization serves as a good inspection tool while conducting gene correlation studies.

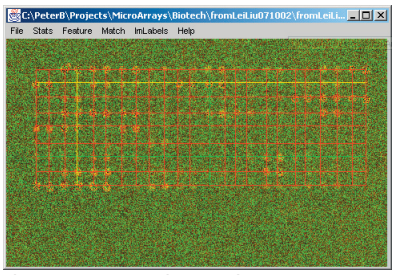
Example Results

Grid Alignment and Visualization



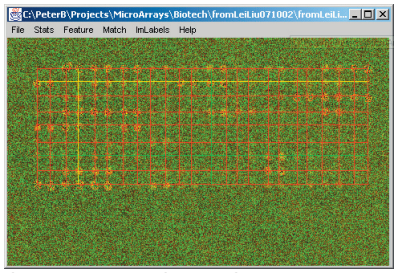
Original Image



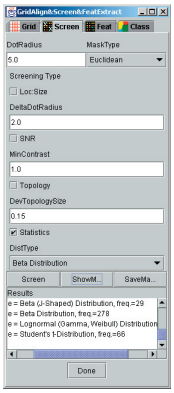


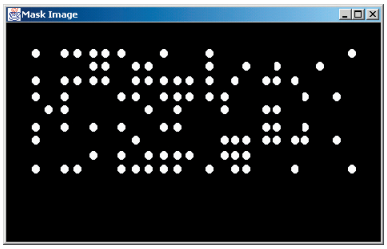
Original Image with Overlaid Grid

Screening and Visualization



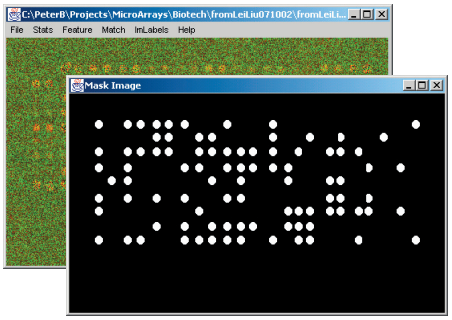
Original Image with Overlaid Grid



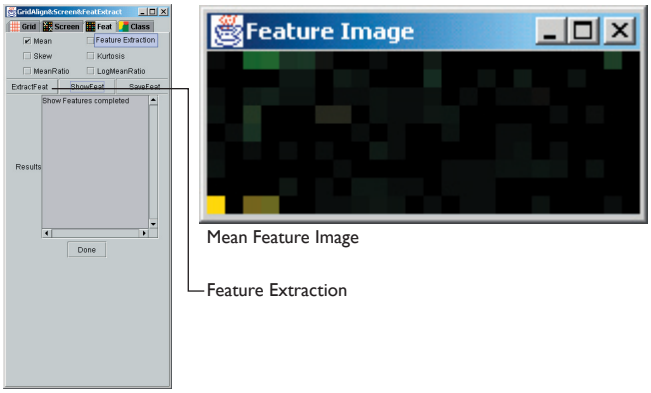


Mask Image

Feature Selection and Visualization



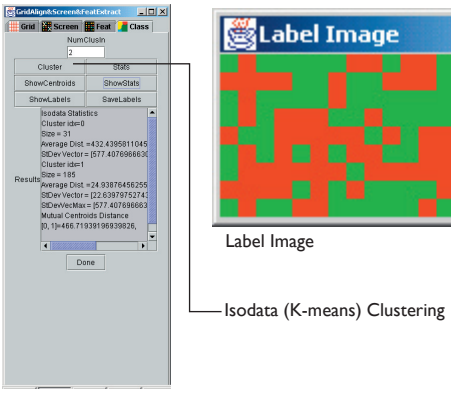
Original and Mask Image



Mean Feature Image

Class Labeling and Visualization

Mean Feature Image



Label Image