Towards a Fully Automated Analysis of Hybridization Spot Array Images

Norbert Brändle Horst Bischof
Vienna University of Technology
Pattern Recog. and Image Proc. Group
Treitlstr. 3/1832, A-1040 Vienna
nob@prip.tuwien.ac.at

Hilmar Lapp
Novartis Forschungsinstitut
Genetics
Brunner Str. 59, A-1230 Vienna
hilmar.lapp@pharma.novartis.com

Abstract

Digital image analysis of high-density spot array images is a crucial part of high-throughput gene expression and transcript frequency profiling experiments in array technology. With increasing density of the spot arrays and large numbers of images a fully automated approach becomes mandatory in order to be capable of analyzing the masses of data. Fully automated means that no user interaction is necessary for images of reasonable quality, and that images of bad quality are detected automatically. When hybridization signals represent gene expression levels (e.g. hybridizations of arrays with probes derived from total PolyA+ RNA of cell-lines or tissues) the accuracy of signal quantification is essential for the reliability and significance of the gained expression information.

Our image analysis approach is divided into the two main parts *grid fitting* and *spot fitting*. Grid fitting finds the grid of spot locations in the digital image and represents the array distortions that occur in the course of the experiment. Grid fitting is based on simple and robust image processing operators (e.g. matched filter, Radon transform). Special emphasis was devoted to taking into account prior knowledge inherent in the imaging process. As a consequence there is no need for generic processes like Markov Random Fields or methods like Fuzzy Logic. The spot locations are used as initial estimates for the centers of *parametric spot models* modeling the shape of the spots. Fitting is performed by robust Maximum Likelihood estimators. The volume under the fitted spot model is used as the quantification of the hybridization signal. The main advantage of the parametric spot modeling approach is the possibility to cope with *spot-shape overlap*, a phenomenom frequently encountered in grids of high density.

We have tested our approach on a set of 543 oligonucleotide fingerprint images comprising 57600 spots each. The spot model was a bell-shaped two-dimensional Gaussian function. From this set of images 527 have been successfully analyzed. 16 images have been correctly detected as being of bad quality. The execution time for one image was 4 minutes on a Pentium II 400 MHz. We currently work on extending our approach to images arising from different hybridization experiments, e.g. complex probe hybridizations or primer oligo hybridizations.

Our approach is the first to our knowledge which is capable of analyzing high-density spot array images in a completely automated way (no theoretic grids have to be predefined or aligned) and which employs two-dimensional spot models for the quantification of the hybridization signal.

References

[1] N. Brändle, H. Lapp, and H. Bischof. Automatic Grid Fitting for Genetic Spot Array Images Containing Guide Spots. Accepted at: *"8th Intl. Conf. on Computer Analysis of Images and Patterns", Ljubljana, Slovenia, September 1–3*, 1999. URL: http://razor.fri.uni-lj.si/CAIP99