Bioinformatics

In terms of research, for the past ten years the analysis of biomedical images – such as cDNA microarray images as well as 2D gels images that are obtained from two-dimensional electrophoresis of proteins – has been at the forefront of biomedical science. Indeed, these images are being increasingly applied in numerous fields of biomedical research such as cancer research, pharmaceutical research, toxicological research, infectious disease diagnosis and treatment, and agricultural development. The reason behind their broad use and success can be found in their main revolutionary feature: the ability to analyze the expression levels of thousands of genes over different samples simultaneously. The end product of either the microarray experiment or the two-dimensional electrophoresis of proteins is a high resolution digital image, containing thousands of spots, the intensities of which are proportional to the expression levels of specific genes. The major related research areas include image analysis and pattern recognition.

Proteomics images

Image analysis is necessary for the detection of spots' boundaries and the calculation of their intensity in proteomics images. The process of analyzing a 2D-PAGE image can be divided into three main phases namely: Spot-Detection, Spot-Segmentation and Spot-Quantification. During the 1 st phase, the protein spots located in the 2D-PAGE image are detected. During the 2nd phase, the entire area of each detected spot is determined, while during the 3rd phase, the brightness of each spot is measured.

Amongst the aforementioned phases, the detection and the segmentation phase are the most challenging ones beyond reasonable doubt. The main reason behind this fact lies in numerous factors such as: a) the inhomogeneity of the background, b) the presence of noise, c) the existence of streaks and artifacts, d) the existence of various intensities, sizes and shapes of the spots, e) the presence of complex regions containing overlapping spots.

State-of-the-art software packages require human intervention either for specification of input parameters and/or correction of the output. Hence, the results emerged from these packages lack objectivity and reproducibility. Methods that have been applied for proteomics image analysis include mathematical morphology and level sets.

RTS Image lab has developed an unsupervised proteomics image segmentation approach based on active contours. The proposed approach incorporates a detection process aiming to identify boundaries of spot overlap in complex regions, histogram adaptation in order to avoid unwanted noise amplifications and morphological reconstruction so as to identify boundaries of numerous faint spots. In addition, a spot-targeted level-set surface is formed, aiming to guide contour initialization whereas contour evolution is guided by region-based as well as morphologically-derived energy terms. Experiments have been conducted on datasets of both real and synthetic proteomics images in order to evaluate the segmentation accuracy of the proposed approach. The segmentation results demonstrate that the proposed approach is capable of identifying spot boundaries in multiplets, as well as boundaries of faint spots. Moreover, it outperforms state-of-the-art proteomics image analysis software packages in terms of segmentation quality. Furthermore, it is unsupervised as opposed to state-of-the-art software packages, in which the cumbersome and tedious process of manual editing by experienced biologists is required. Finally, it facilitates quicker convergence when compared with state-of-the-art methods.

Microarray images

Image analysis is necessary for the detection of spots' boundaries and the calculation of their intensity . The process of analyzing a microarray image can be divided into three main phases namely: Gridding, Spot-Segmentation and Spot-Intensity Extraction. During the 1st phase, the microarray image is segmented into numerous compartments, each containing one individual spot and background. During the 2nd phase each compartment is individually segmented into a spot area and a background area, while during the 3rd phase the brightness of each spot is calculated. The expression-levels of the genes in these spots result from their individual brightness. The analysis of the 2D gels includes only the latter two phases. This is a challenging task mainly due to the poor quality of microarray images. Indeed, these images are contaminated with noise, and artifacts. Moreover, real spots vary significantly from the ideal ones; they are not always circular in shape and their intensity is not always high enough to be clearly visible. Human intervention is therefore necessary either for the initialization of their input parameters, or for the rectification of their incorrect results. Consequently, the analysis and processing of the aforementioned images becomes on the one hand time-consuming, since the users have to choose the appropriate values for their input parameters and rectify their results, and on the other hand subjective, since the users initialize and correct the software programs in an individual manner. This subjectivity can in turn affect the biological results. As a matter of fact, the biological results often differ from the real ones. Methods that have been applied for microarray measurements include genetic algorithms and Support Vector Machines.

Pattern recognition

Pattern recognition is divided into the following stages:

1) The detection of differential expression

2) Pattern discovery

3) Class prediction

4) Inference of regulatory pathways and networks

Class prediction methods involve supervised machine learning techniques for diseases' diagnosis or prediction. This is a challenging task mainly due to three reasons: a) microarray data consist of a large number of gene expression measurements, while the number of samples is disproportionally small, b) a significant percentage of genes is usually not associated with the problem under investigation and c) the biochemical procedure used to produce microarrays, adds a lot of noise to the measurements. Methods that have been applied for microarray measurements include linear discriminant analysis, k-nearest neighbors, parzen windows, decision trees, Neural Networks and Support Vector Machines.