A Fuzzy-Watershed Image Segmentation of two-dimensional gel images—Survey Paper

Shaheera Rashwan 1, Amany Sarhan 2, Talaat Faheem 2, Bayoumy Abdel Rahman Bayoumy 3

1 Knowledge-Based Systems and Robotics Department, Informatics Research Institute, Mubarak City for Scientific Research, Borg ElArab, Alexandria, Egypt srashwan@mucsat.sci.eg http://www.mucsat.sci.eg/

2 Computer Science Department, Faculty of Engineering, Tanta University, Tanta, Egypt Amany_m_Sarhan@yahoo.com

3 Department of Computer-Based Applications, Institute of Informatics Mubarak City for Scientific Research, Borg ElArab, Alexandria, Egypt

Abstract

2D gel electrophoresis (2DGE) plays an important role in proteomics. It can separate proteins effectively with their pI values and molecular weights. Proteomics researchers needed to identify interested protein spots by examining the gel. This is time-consuming and labor extensive. It is desired that the computer can analyze the proteins automatically by first detecting and quantifying the protein spots in the digitized 2DGE images. In our work, we will investigate the use of the watershed algorithm in segmenting the protein spots from the varying background. However, the watershed algorithm often produces an over-segmented result. So, we will introduce the notion of fuzzy relations to improve the segmentation result.

Keywords: 2D gel electrophoresis, protein spots, watershed algorithm, over-segmentation, fuzzy relations

Background

Two-Dimensional Gel Electrophoresis (2-DE) is one of the most widely used techniques in molecular biology, used in a plethora of applications. The methods used for image processing on 2-DE data are primarily of two types. These can roughly be characterized as derivative-based([1],[2]), and watershed-based([3],[4]).

Derivative-based methods are generally applied when the foreground - in our case, the spots – is of a significantly different intensity than the background - the gel – and when in areas far away from the border of foreground and background, the intensity is relatively uniform.

If such conditions are satisfied, a gradient-based approach will produce a modified image which consists of a uniform, low intensity everywhere but the border regions. Such filters are often represented by matrices which indicate what the modified value of a given point will be.

Examples of filters used are horizontal gradient operator, such as a second-derivative filter or a Kalman filter[2] Note that we have not addressed how to interpret the modified image, which will be covered below. In our case, however, the images require some preprocessing before such a filter can be applied, as the image is noisy (not sufficiently smooth). Image segmentation/classification and topographic object extraction are critical
for subsequent image analysis and further image understanding of remote sensing data, as they have to conform to the following facts:

1. Remotely sensed imagery has multispectral and multiscale nature;
2. In contrast to other image modalities, remote sensing images contain various objects with heterogeneous properties with respect to size, form, spectral behavior etc, so meaningful objects should be extracted at the appropriate scale;
3. Model-based interpretation of remote sensing imagery is more difficult due to the heterogeneity of inherent object classes;
4. Suboptimal solutions will probably not be considered for remote sensing applications because there is no need for real-time applications.

The methods used for image processing on 2DGEL data in our work are the watershed-based methods. The watershed approach was first proposed by Beucher and Lantuejoul in 1979 as a method for contour detection, and is a beautiful, simple approach to the segmentation of an image. The model is that of a topographical map, with the more intense points lower down and less intense points higher up. Then "holes are pierced in each local minimum of the topographic surface [and] the surface is slowly immersed into a 'lake' thereby filling up all the catchment basins ... as soon as two catchment basins tend to merge, a dam is built."[4] The result of this is that the image is partitioned into some number of such catchment basins, which then form the units which will be examined for spots. In general a threshold will be defined so that basins have a certain minimum depth and/or size.

The advantage of Watershed-Based over the Derivative-Based approaches is that it is less sensitive to background noise. As long as background noise does not exceed the minimum depth for catchments basins, it will be properly ignored by this technique.

Image processing analysis generally consists of the following five steps which are
1. Image acquisition operations to convert images into digital form;
2. Pre-processing operations to obtain an improved image with the same dimensions as the original image;
3. Image segmentation operations to partition a digital image into disjoint and non-overlapping regions;
4. Object measurement operations to measure the characteristics of objects, such as size, shape, color and texture;
5. Classification operations to identify objects by classifying them into different groups.

Image acquisition

Image acquisition, that is capture
of an image in digital form, is obviously the first step in any image processing system. Illumination is an important prerequisite of image acquisition for food quality evaluation. The quality of captured image can be greatly affected by the lighting condition. A high quality image can help to reduce the time and complexity of the subsequent image processing steps, which can decrease the cost of an image processing system.

Image pre-processing

2DGEL images are subject to various types of noises. These noises may degrade the quality of an image and subsequently it cannot provide correct information for subsequent image processing. In order to improve the quality of an image, operations need to be performed on it to remove or decrease degradations suffered by the image during its acquisition. The purpose of preprocessing is an improvement of the image data, which suppresses unwilling distortions or enhances some image features that are important for further processing and creates a more suitable image than the original for a specific application. Two different types of image preprocessing approaches can be identified for quality evaluation: pixel pre-processing and local pre-processing, according to the size of the pixel neighborhood that is used for the calculation of a new pixel.

Image Segmentation

In computer vision, segmentation refers to the process of partitioning a digital image into multiple regions (sets of pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. The result of image segmentation is a set of regions that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s).

Several general-purpose algorithms and techniques have been developed for image segmentation. Since there is no general solution to the image segmentation problem, these techniques often have to be combined with domain knowledge in order to effectively solve an image segmentation problem for a problem domain.

Examples of these techniques are: Clustering Methods, Histogram-Based Methods, Region Growing Methods, Level Set Methods, Graph Partitioning Methods, Watershed Transformation, Model based Segmentation, Multi-scale Segmentation, Semi-automatic Segmentation, Neural Networks.
Segmentation

Synopsis of the study

Even several algorithms have been designed for spot detection and quantification like Gaussian fitting, Laplacian of Gaussian spot detection (LOG), line analysis and watershed transformation (WST), spot detection is still a challenging problem because the background intensity varies from different regions in the image and global or local background subtraction is still a problem. The image analysis software must be able to detect spots fully automated to be able to run as a component in a batch-process.

Regarding this is a new version of the watershed transformation was developed which reduces over-segmentation. Our work provides an improved solution to the problem of protein spot detection by applying fuzzy watershed transform algorithm. The over-segmentation draw back would be overcome by applying fuzzy relations in watershed transform algorithm.

The objectives of this study:

1- Our aim in this research is the development of a new watershed algorithm efficient for 2Dgel image analysis.

2- Building a new system which can perform the automatic segmentation of Protein spots. In this system, the problem of over-segmentation will be overcome by introducing the notion of fuzzy relations.

Technical approach

Watershed algorithm is sensitive to noise and often produces over-segmentation. To overcome the problem of noise sensitivity, the image is transformed into fuzzy domain by using the s-function and maximum fuzzy entropy principle, and then enhanced by the fuzzy intensifying function. The watershed method is applied in the fuzzy domain to get segmentation which can segment the images well.

To overcome the problem of over segmentation, we present an efficient algorithm to merge similar catchments and effectively diminish over-segmentation. The proposed approach is based on the application of the Fuzzy C-means algorithm together with composition of Fuzzy relations.

The Fuzzy C-means algorithm is one of the most widely used clustering algorithms. It was initially proposed by Bezdek [5,6]. It is an unsupervised algorithm, based on the minimization of a fuzzy objective function, which is based on the intra-class scatter of the given data. The algorithm performs a partition of the data into c clusters and c centers, one for each cluster, are generated.
The Fuzzy C-means algorithm on its own could be, for instance, used to group together pixels having similar grey-level value. This will without doubt diminish the number of regions but because no information is taken into account about the connectivity between regions, this simple approach is more prone to errors. In order to construct a more robust technique, the use of Fuzzy Relations is introduced.

The most common form of the fuzzy relation composition is the max–min composition; the max–min composition for fuzzy relations can be interpreted as indicating the strength of such a relational chain. The strength of the relation between elements x and z is then the strength of the strongest chain between them. The concepts of fuzzy relations can be applied to the problem of watershed over-segmentation.

In our work we will choose the best fuzzy relation which will be able to improve the watershed algorithm for spot detection in 2D gel images. The proposed spot detection process will be implemented and tested on 15 protein gel profiles (image size: 1498 x 1544) of porcine testis. The detection results will be compared by that generated by a commercial software tool: ImageMaster.

References


