

DETECTION AND SEGMENTATION IN 2D GEL ELECTROPHORESIS IMAGES

Eirini Kostopoulou, Eleni Zacharia, and Dimitris Maroulis

Department of Informatics and Telecommunications, University of Athens, Greece

Emails: ikostop@di.uoa.gr, eezacharia@gmail.com, dmaroulis@di.uoa.gr

ABSTRACT

Analyzing 2D-gel electrophoresis images remains a challenging task. Amongst the respective stages of the analysis, detection and segmentation of each individual spot are the most crucial. The commercial availability of several software programs and techniques notwithstanding, spot detection and segmentation are extremely resource intensive and frustratingly time-consuming; without human intervention, they fail to detect several protein spots, such as low intensity spots or overlapping protein, while they detect various spurious spots. This paper presents an original approach to detecting and segmenting spots in 2D-gel electrophoresis images. The conducted experiments in a set of 16-bit images demonstrate that the proposed approach is very effective and it outperforms existing techniques even when it is applied to images containing several overlapping spots as well as to images containing spots of various intensities, sizes and shapes.

Index Terms— 2D-PAGE image analysis, spot detection, spot segmentation, proteomics.

1. INTRODUCTION

Two-Dimensional Polyacrylamide gel electrophoresis (2D-PAGE) is a well established technique for simultaneously analyzing and separating large collections and complex mixtures of proteins. The result is a digital image (2D-PAGE image) containing thousands of spots, each one corresponding to an individual protein.

The analysis of 2D-PAGE images remains a crucial stage in the analysis of the proteins. Amongst the stages of image analysis, detection and segmentation of each individual spot are the most challenging. These stages may reveal alterations in protein expression within a given biological system. However, incorrect detection and segmentation of 2D-PAGE images can detrimentally affect the expression profiles of proteins.

A variety of different methodologies and software programs have been proposed in the literature intended to solve the spot-detection or segmentation issue. Many of these are based on edge detection algorithms such as the Laplacian filtering, in conjunction with smoothing or morphological operators [1]-[5]. However, these methods

do not perform well in low quality images. Thresholding methods [6, 7] based on 1D histograms fail to detect and accurately segment the protein spots in 2D-PAGE images. The reason behind this apparent failure can be attributed to the fact that an 1D histogram does not provide information on the spatial correlation between the pixels in the image. Spot-segmentation techniques based on watershed methods [8-10] can cause over-segmentation. Although the selection of inner markers [11] is an approach used to address the over-segmentation problem, it requires human intervention. Last but not least, spot modeling techniques [12] using Gaussian fitting and diffusion models require the initial assumption that the spots have a specific common shape and size. Consequently, these well-established programs and techniques require human intervention in order to specify mandatory input parameters or to correct their results. It is worth pointing out, however, that human intervention does not only limit the throughput but it additionally brings the objectivity and reproducibility of results into question. Therefore, automating this part of the process is essential because: (i) it will allow rapid high-throughput analysis of the expression levels of thousands of proteins, and (ii) it will prevent variations in the protein expression profiles due to subjectivity.

Therefore, in spite of the commercial availability of several software programs and techniques, spot detection and segmentation continue to be challenging tasks. The main reason behind the challenges encountered in these stages lies in the nature of 2D-PAGE images. Indeed, the 2D-PAGE images contain thousands of spots of various intensities, sizes and shapes. Furthermore, adjacent spots are often not clearly separated but overlapping. Last but not least the 2D-PAGE images contain inhomogeneous background in addition to being contaminated with noise and artifacts. Due to the aforementioned nature of 2D-images, spot-detection and segmentation software: 1) fail to detect some low intensity spots, 2) detect spurious spots, 3) split a spot into two or more spots and 4) merge overlapping spots.

In this paper, an original approach to analyzing 2D-PAGE images is presented. The proposed approach can accurately detect and segment spots in 2D-PAGE images. Spot detection is implemented into two main stages. Firstly, a set of regions containing mostly spots is determined. Subsequently, the centers of spots existing in

each coherent region of the aforementioned set are specified. Spot segmentation is also implemented into two main stages. More precisely, each coherent region containing mostly spots is roughly segmented into regions each containing one individual spot, according to the detected centers of spots. Afterwards, the optimal contour of each spot is determined. The conducted experiments in a set of 16-bit 2D-PAGE images demonstrate that the proposed approach is fast, very effective and it outperforms existing techniques. Moreover, it can detect and segment very effectively overlapping spots, spots of various intensities, sizes and shapes as well as spots from images containing inhomogeneous background.

The remainder of this paper is structured in three sections. In section 2, the proposed approach to detecting and segmenting the spots in 2D-PAGE images is presented. In Section 3 experiments evaluate the proposed approach and compare it to a widely used software package for 2D-PAGE image analysis. Finally, our conclusions are apposed in section 4.

2. PROPOSED APPROACH

2.1. Spot Detection

Spot detection is implemented into two main stages: Firstly, a set of regions containing mostly spots (S) is determined. Thereafter, the centers of spots that exist in each coherent region of S are specified.

Determination of the S set: The determination of the S set is implemented using the recursive 2D Otsu thresholding technique in two phases [13],[14]. For better comprehension of the two phases of this stage, an imitation of a synthetic image is illustrated in Fig. 1. In this figure, each spot is colored uniformly with a grey value, instead of a range of grey values.

Firstly, the recursive 2D Otsu thresholding technique is applied in the initial 2D-PAGE image (Fig. 1a). As a result, a set of regions S_1 which contain mostly spots is determined (Fig. 1b white areas). As it can be observed, the S_1 set does not contain the total number of spots. In order to detect low intensity spots which are not clearly visible the 2D Otsu thresholding technique is subsequently applied to the area of 2D-PAGE image which corresponds to the black area of Fig. 1b. Likewise, a set of regions S_2 which contains mostly spots is determined from the aforementioned area (Fig. 1c white areas). Finally, the set of regions S containing mostly spots (Fig. 1d white areas) is defined as:

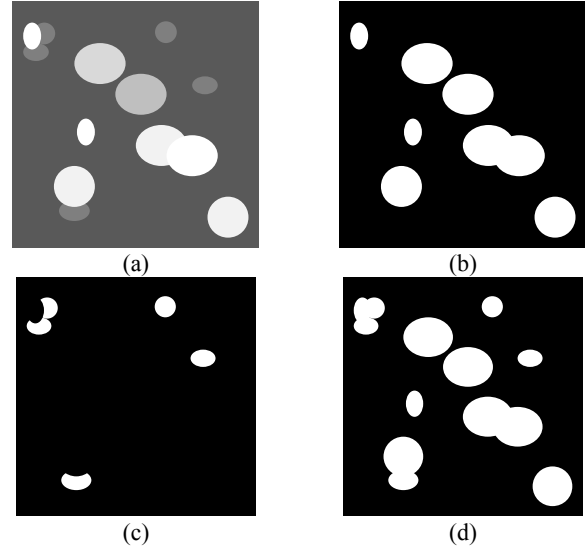


Fig. 1. Determination of the S set: (a) imitation of a synthetic 2D-PAGE image, (b) S_1 set, (c) S_2 set, (d) S set.

$$S = \{p \mid p \in S_1 \cup S_2\} \quad (1)$$

More precisely it contains the pixels p belonging either to S_1 or to S_2 sets of regions.

The 2D Otsu thresholding technique can be advantageous for the determination of the two aforementioned sets over 1D thresholding techniques. This is due to the fact that it is based on 2D-Histograms and thus it provides information on the spatial correlation between the pixels in the image.

Determination of the centers of each spot: According to Bettens et al [15] protein spots have an intensity that peaks at their central region and declines at regions further from their centre (Fig. 2). In the case when the peak is thin, spots resemble a 3D-Gaussian function, whereas in the case when the peak is wide, spots resemble a plateau. Consequently, the most probable candidates for spots' centers are the pixels of each coherent region of S that have intensity value higher than their adjacent pixels.

However, spots may have more than one peak since the 2D-PAGE images are contaminated with noise. Thus,

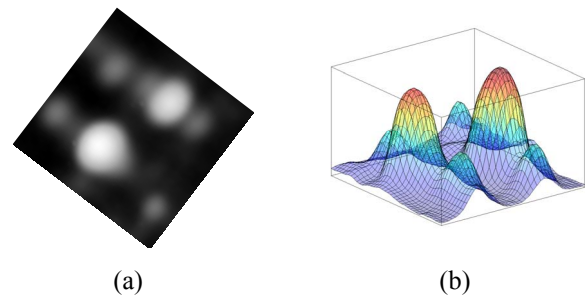


Fig. 2. 2D and 3D representation of an area of a real 2D-PAGE Image.

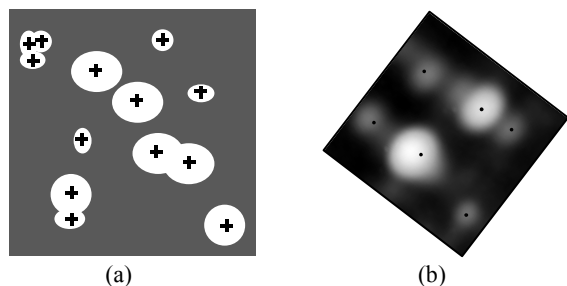


Fig. 3. Determination of the centers of spots: (a) of Fig. 1a, (b) of Fig. 2a

the determination of the spots' centers is achieved by obtaining the local intensity maxima in 6×6 neighborhoods of the S coherent regions. The spots' centers of the Fig. 1a are illustrated with crosses on Fig.3a, whereas the spot centers of the real image Fig 2a are illustrated with dots in Fig. 3b.

2.2. Spot Segmentation

Let S_c be a coherent region of the S set whose intensity range is $[I_{min}, I_{max}]$, where I_{min} denotes the minimum intensity value of a pixel contained in the S_c and I_{max} the maximum intensity value. The coherent region S_c of S is segmented into sub-regions around the local maxima as follows:

For each intensity value i from I_{max} to I_{min} , the proposed algorithm assigns a label to the pixels p of intensity i applying a majority voting criterion among the labels of its adjacent pixels within a 3×3 neighborhood. The pixel is assigned a new label when all its adjacent pixels are unlabeled.

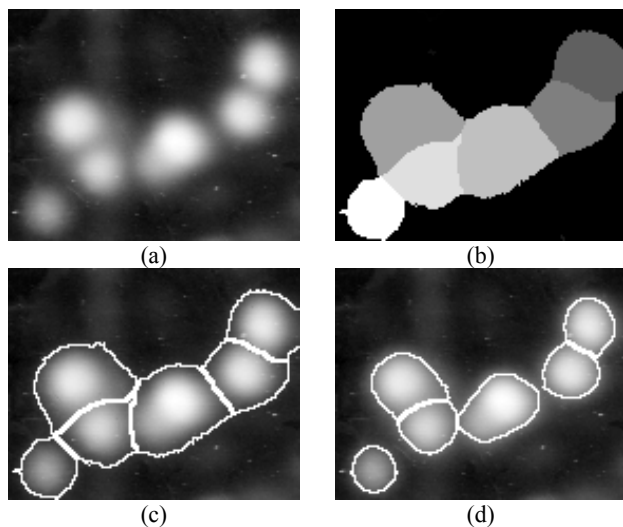


Fig. 4. Spot Segmentation: (a) an area of a real 2D-PAGE image, (b) mosaic of labeled regions, (c) not optimal contour, and (d) optimal contour.

Finally, the algorithm results in a mosaic of labeled regions S_L each containing a single protein spot. An example of an area of a real 2D-PAGE image as well as its mosaic of labeled regions is illustrated in Fig. 4a and Fig. 4b correspondingly.

Fig. 4c depicts the contours obtained using the labeled regions of the mosaic. As it can be observed, these contours are not the optimal ones. Therefore, the optimal contour of each spot is subsequently determined. More precisely, the optimal thresholding [16] method is automatically applied to each labeled region S_L in order to remove the background area that surrounds the protein spot located in S_L (Fig. 4d).

3. RESULTS

Several experiments on a set of five real 2D-PAGE images each containing thousands of spots were conducted so as to evaluate the performance of our method in spot-detection and spot-segmentation.

All the 2D-PAGE images used in the experiments were provided by the Biomedical Research Foundation of the Academy of Athens. The 2D-PAGE images are of low quality since they have inhomogeneous background in addition to being contaminated with noise. Moreover, they contain numerous spots of various intensities, sizes and shapes. Furthermore, they contain numerous spots which overlap.

The accuracy of the proposed approach was analyzed by means of a statistical analysis, and its results were compared with those of the Melanie 7 software package [3]. It should be noted that the results were validated with the ground truth in spot-detection provided by expert biologists in the proteomics field.

The proposed approach has correctly detected 79,3% of real spots, while the Melanie 7 software package has detected only 65,9% of spots. Moreover, the percentage of spurious spots detected with the proposed approach was clearly lower, 3% for the proposed approach versus 5% for the Melanie software package.

Two indicative examples of spot-detection and spot-segmentation results in two areas of 2D-PAGE images used in the experiments are illustrated in Figs. 5,6. In particular Figs. 5a,6a depict two areas of 2D-PAGE images. The ground truth in spot-detection of the aforementioned areas of images are depicted in Figs. 5b,6b with crosses.

Figs. 5c,6c illustrate spot detection results of the original images using the proposed approach while Figs. 5d,6d illustrate spot detection results using Melanie 7 software package. The real spots detected by the proposed method are depicted in Figs. 5c,6c with crosses outlined by squares, while the real spots detected by the Melanie 7 software are depicted in Figs. 5d,6d with crosses outlined by diamonds.

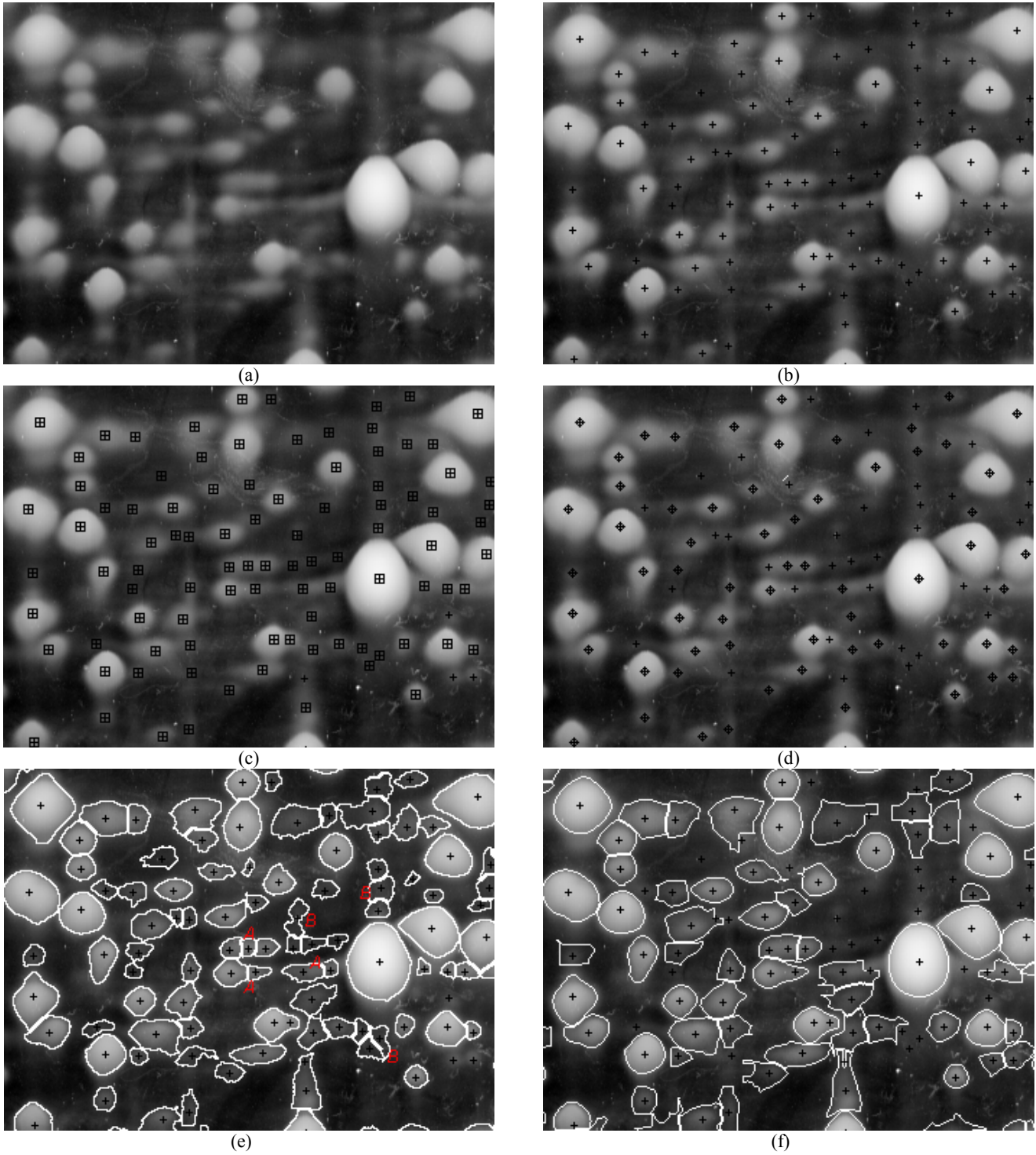


Fig. 5. Spot-detection and spot-segmentation results in an area of 2D-PAGE image, (a) area of 2D-PAGE image, (b) Ground truth in spot-detection, (c, d) Spot-detection results of the proposed approach (squares) and Melanie 7 software package (diamonds) respectively, (e, f) Spot-segmentation results of the proposed approach and Melanie 7 software package respectively.

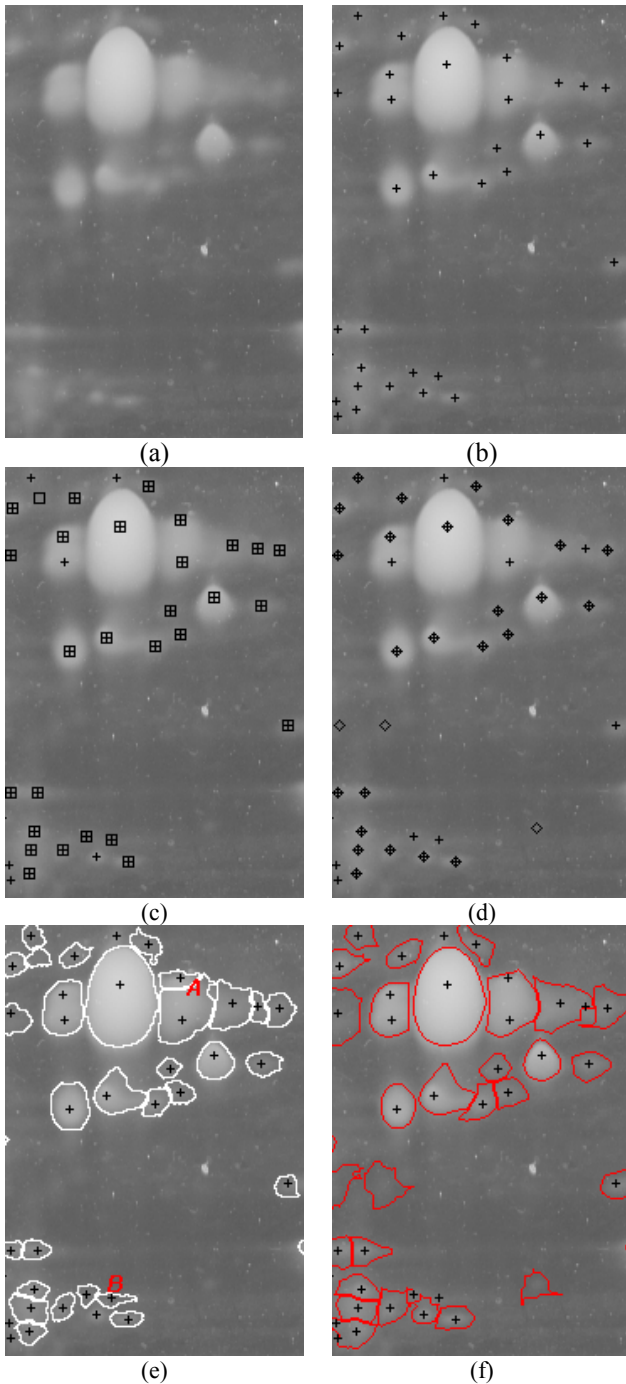


Fig. 6. Spot detection and segmentation results in an area of a 2D-PAGE image. (a) an area of 2D-PAGE image, (b) Ground truth in spot-detection, (c, d) Spot-detection results of the proposed approach (squares) and Melanie 7 software package (diamonds) respectively, (e, f) Spot-segmentation results of the proposed approach and Melanie 7 software package respectively.

The missed spots are depicted in Figs. 5c,6c,5d,6d with single crosses, while the spurious spots detected by the

proposed approach and Melanie 7 software package are depicted with squares and diamonds respectively, which do not contain crosses.

The proposed approach has detected almost all of the real spots in both images (95 spots out of 99 of Fig. 5a and 28 spots out of 34 of Fig. 6a), while Melanie 7 software package detected only 64 real spots on Fig. 5a and 25 on Fig. 6a.

The proposed approach has efficiently segmented the real spots. Figs. 5e,6e and Figs. 5f,6f illustrate spot segmentation results using the proposed approach and Melanie 7 software package respectively. In cases where "A" is depicted, the proposed approach has efficiently segmented real overlapping protein spots while Melanie 7 did not. Moreover, in cases where "B" is depicted, the proposed approach has segmented real protein spots very efficiently while Melanie 7 did not.

4. CONCLUSION

In this paper, an original method for spot detection and spot-segmentation in 2D-PAGE images is presented. The proposed approach is based on the concept of the 2D Otsu thresholding technique as well as on the concept of a mosaic of labeled regions. The experimental results over real 2D-PAGE images demonstrate that it is very efficient and it outperforms state-of-the-art software program, Melanie 7. Moreover, it can be applied to low quality images (images having inhomogeneous background, being contaminated with noise and/or containing spots of various intensities, sizes and shapes as well as overlapping spots), yielding excellent results.

5. ACKNOWLEDGMENT

The authors would like to thank the Biomedical Research Foundation of the Academy of Athens for the provision of 2-D page images as well as for the provision of the corresponding ground truth. This research was funded by the Greek State Scholarship Foundation (IKY) and partially by the special account of research grants of the University of Athens, Greece.

6. REFERENCES

- [1] L.Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," IEEE Trans. on Image Processing, vol. 2, no. 2, pp. 176-201, 1993.
- [2] Pf. Lemkin and Le. Lipkin, "2D electrophoresis gel data base analysis: aspects of data structures and search strategies in GELLAB," Electrophoresis, vol. 4, no. 1, pp. 71-81, 1983.

- [3] Rd. Appel et al, "Melanie II-a third generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms," *Electrophoresis*, vol. 8, no. 15, pp. 2735-2748, 1997.
- [4] N.L. Anderson, J. Taylor, Ae. Scandora, P. Coulter, and N.G. Anderson, "The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns," *Clinical Chemistry*, vol 27, pp. 1807-1820, 1981.
- [5] D. Olson and M.J. Miller, "Elsie 4: Quantitative analysis of sets of two dimensional gel electrophoretograms," *Analytical Biochemistry*, vol. 169, pp. 49-70, 1988.
- [6] P. Cutler, G. Heald, I.R. White, and J. Ruan, "A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection," *Proteomics*, vol. 3, pp. 392-401, 2003.
- [7] P. Peer, and L.G. Corzo, "Local Pixel Value Collection Algorithm for Spot Segmentation in Two-Dimensional Gel Electrophoresis Research," *Hindawi Publishing Corporation Comparative and Functional Genomics*, vol. 2007, 2007.
- [8] K.P. Pleissner et al, "New algorithmic approaches to protein spot detection and pattern matching in two dimensional electrophoresis databases," *Electrophoresis*, vol. 20, pp. 755- 765, 1999.
- [9] L. Vincent, and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583-598, 1991.
- [10] Jb.T.M. Roerdink, and A. Meijster, "The Watershed Transform: Denitions, Algorithms and Parallelization Strategies," *Fundamenta Informaticae*, vol. 41, pp. 187-228, 2001.
- [11] Y. Kim, J. Kim, Y. Won, and Y. In, "Segmentation of protein spots in 2-D gel electrophoresis images with watershed using hierarchical threshold," *LNCS*, vol. 2869, pp. 389-396, 2003.
- [12] M Rogers, J Graham and R P Tonge, "Using Statistical Image Models for Objective Evaluation of Spot Detection in 2D Gels", *Proteomics*, vol 3, pp 879-886, 2003.
- [13] J. Zhang, and J. Hu, "Image Segmentation Based on 2D Otsu Method with Histogram Analysis," *International Conference Science and Software Engineering*, 2008.
- [14] J. Gong, L. Li, and W. Chen, "Fast recursive algorithms for two-dimensional thresholding," *Pattern Recognition*, vol. 31, no. 3, pp. 295-300, 1998.
- [15] E. Bettens et al, "Computer analysis of two-dimensional electrophoresis gels: a new segmentation and modeling algorithm," *Electrophoresis*, vol. 18, pp. 792-798, 1997.
- [16] R. Gonzalez, R. Woods, "Digital Image Processing," Prentice Hall, Second Edition.