A Spot Segmentation Approach for 2D Gel Electrophoresis Images Based on 2D Histograms

Eleni Zacharia¹, Eirini Kostopoulou¹, Dimitris Maroulis¹ and Sophia Kossida²

¹ Dept. of Informatics and Telecommunications, University of Athens, Greece

² Division of Biotechnology, Foundation of Biomedical Research of the Academy of Athens,

Greece

eezacharia@gmail.com, ikostop@di.uoa.gr, dmaroulis@di.uoa.gr, skossida@bioacademy.gr

Abstract

Spot-Segmentation, an essential stage of processing 2D gel electrophoresis images, remains a challenging process. The available software programs and techniques fail to separate overlapping protein spots correctly and cannot detect low intensity spots without human intervention. This paper presents an original approach to spot segmentation in 2D gel electrophoresis images. The proposed approach is based on 2D-histograms of the aforementioned images. The conducted experiments in a set of 16-bit 2D gel electrophoresis images demonstrate that the proposed method 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.

1. Introduction

Over the last decade, two-Dimensional Polyacrylamide Gel Electrophoresis (2D-PAGE) has become an important tool in biological and medical research. The reason for its popularity is that it enables scientists to analyze large collections and complex mixtures of proteins. The initial product of the 2D-PAGE is a digital image, reported as 2D-PAGE image, which contains thousands of spots. Segmentation of these spots may reveal alterations in protein expression within a given biological system. However, incorrect segmentation of 2D-PAGE images can detrimentally affect the expression profiles of proteins.

Spot-segmentation still remains a challenging process. The reason for this lies in the nature of 2D-PAGE images. Indeed, the 2D-PAGE images may contain up to 10,000 spots, each one with different intensity, size and shape. Furthermore, adjacent spots are not clearly separated but they overlap. Last but not least the 2D-PAGE images contain inhomogeneous background as well as they are contaminated with noise and artifacts. As a result, a remarkable number of techniques and software systems has been proposed and developed in order to implement the spot-segmentation of 2D-PAGE images.

Spot-Segmentation techniques based on watershed methods [1, 2] can cause over-segmentation. While selection of inner markets [3] is an approach used to address the over-segmentation problem, it requires human intervention which as already mentioned can affect the segmentation results. Morphology-based [4] and edge-detection methods [5, 6] do not perform well in the low quality images. Thresholding methods [7, 8] based on 1D histograms fail because the 1D histogram does not provide information about the spatial correlation between the pixels in the image. Therefore, low quality 2D-PAGE images can not be accurately segmented by this technique.

In this paper, an original approach to analyzing 2D-PAGE images is presented. The proposed approach is divided into three main stages: 1) Rough segmentation of the 2D-PAGE image into a set of regions containing mostly background and into a set of regions containing mostly spots, 2) Each coherent region containing mostly spots is segmented in order to separate adjacent spots which overlap, 3) Refinement procedure for the determination of the optimal contour of each spot. The conducted experiments in a set of 16-bit 2D-PAGE images demonstrate that the proposed method is fast, very effective and it outperforms existing techniques even when it is applied to images containing several overlapping spots or spots of various intensities, sizes and shapes. Moreover, the results have showed that the proposed method can be applied to images containing inhomogeneous background.

The remainder of this paper is structured in four sections. In section 2, a brief description of the thresholding technique using 2D Histograms is provided. In Section 3 the proposed spot segmentation approach is presented. In Section 4 experiments evaluate the proposed method and compare it to an existing software package for 2D-PAGE image analysis. Finally, our conclusions are apposed in section 5.

2. Threshold using 2D Histograms

Let *I* be a 2D-PAGE image and I_{mean} its mean image [9]. The 2D histogram is depicted in Fig. 1. Its x-axis and y-axis depict the gray levels of *I* and I_{mean} images respectively. Its z-axis represents the joint probability mass function $P(I_r, I_m)$:

$$P(I_r, I_m) = \frac{r(I_r, I_m)}{M \cdot N}$$
(1)

 I_r and I_m denote the intensity values of two corresponding pixels belonging to I and I_{mean} images respectively. More precisely, I_m intensity value of a pixel (x,y) is defined as:

$$I_m(x,y) = \frac{1}{9} \sum_{k=1}^{1} \sum_{l=-1}^{1} I(x+k,y+l)$$
(2)

and $r(I_r, I_m)$ denotes the occurrence of the intensities' pair (I_r, I_m) . *M*, *N* are the dimensions of both *I* and I_{mean} images.



Figure 1. 2D histogram

According to 2D Otsu recursive method [9, 10], an optimal vector (S,T) which divides the 2D histogram into four quadrilaterals (Fig. 1) can be determined. The quadrilaterals 1 and 2 contain the distributions of background and spots respectively, whereas the quadrilaterals 3 and 4 contain the distributions of pixels near edges and noises [9, 10].

3. Proposed Method

The proposed approach is divided into the following three stages:

1st Stage: The regions of the image where proteins exist are determined. More precisely, the 2D-PAGE image is roughly segmented into a set of regions containing mostly background B and into a set of regions containing mostly spots S. In order to segment spots of low intensities which are not clearly visible, the recursive 2D Otsu thresholding technique is applied twice.

Firstly, the recursive 2D Otsu thresholding technique is applied in the total 2D-PAGE image (Fig. 2a). As a result, the image is divided into: a) a set of regions F_0 which contain mostly spots (Fig. 2b white areas) and b) a set of regions B_0 which contain mostly the background (Fig. 2b black area).

Afterwards, the 2D Otsu thresholding technique is applied to each region of the set B_0 of the 2D-PAGE image (Fig. 2c, except black areas encircled by dashed curves). Likewise, B_0 is divided into F_{01} (Fig. 2d white areas) and B_{01} (Fig. 2d black area outside the dashed curves) sets of regions.

Finally, the set of regions containing mostly spots *S* (Fig. 2e white areas) is defined as:

$$S = \{ p \mid p \in F_0 \cup F_{01} \}$$
 (2)

More precisely it contains the pixels p belonging either to F_0 or to F_{01} sets of regions.



Figure 2. The two phases of the 1st stage based on 2D Otsu thresholding technique.

2nd stage: Each coherent region of S is in turn segmented in order to separate adjacent spots which overlap. According to Bettens et al [11] protein spots have an intensity that peaks at their central region and declines at regions further from their centre. In the case when the peak is thin, spots resemble a 3D-Gaussian function. In the case when the peak is wide, spots resemble a plateau. As a result, local maxima in the intensities contained in each coherent region of S are the most probable candidates for spot centers. Thus, spot detection is achieved with the determination of local intensity maxima of the S regions. Afterwards, each coherent region of S is segmented into subregions around the local maxima of image intensities by using a variant of the pixel value collection algorithm.

3rd stage: The optimal contour of each spot is determined. More precisely, the optimal thresholding [12] method is applied in order to remove the background area that surrounds each protein spot.

4. Results

Experiments were performed on a set of regions of real 2D-PAGE images in order to evaluate the performance of the proposed algorithm. The 2D-PAGES images were digitized at 2250×3000-pixels at 16-bit grey level depth. Each 2D-PAGE image contains approximately 1500 spots.

The accuracy of the proposed method was analyzed by means of a statistical analysis. The spot segmentation results of the proposed method were compared with the spot segmentation results of the Melanie 7 software package [13]. The results were validated with the ground truth provided by expert biologists in the proteomics field. It should be noted that the biologists used the Melanie 7 software package the results of which they then had to correct manually in order to reach the ground truth.

The percentage of the real spots segmented with the proposed method achieves the high value of 93.6%



(c)





Figure 3. Spot segmentation results. (a) A region of a 2D-PAGE image, (b) Ground truth of the segmentation provided by the experts, (c) Segmentation results using the proposed method, and (d) Segmentation results using the Melanie 7 software program.

thus outperforming the Melanie 7 software program. Moreover, in comparison with Melanie 7, the percentage of spurious spots detected with the proposed method is less than the half.

An indicative example of spot-segmentation results in a region of 2D-PAGE image used in the experiments is illustrated in Fig. 3. More precisely, Fig. 3a illustrates a region of a poor quality 2D-PAGE image. Even though this region of the image is small, it still contains numerous spots of various intensities, sizes and shapes. Furthermore, it contains numerous spots which overlap. Fig. 3b illustrates the ground truth of the segmentation provided by expert biologists. Fig 3c depicts the segmentation results provided by the proposed method, while Fig. 3d illustrates the segmentation results of Melanie 7.

It should be noted that we have increased the brightness of the image in order for the reader to be able to observe the results even in the cases of low intensity spots which in reality are not clearly visible.

In these images, it is evident that the proposed method has apparently segmented all the spots of the image. On the contrary, Melanie 7 has missed eight spots, indicated by arrows in Fig. 3d. Furthermore, the proposed method has efficiently separated all the overlapping spots, while Melanie 7 was unable to separate seven overlapping spots (see '2' in Fig. 3d). Last but not least, the proposed method has not detected any spurious spots, in comparison with Melanie 7 which has detected two spurious spot (see '1' in Fig. 3d).

5. Conclusions

In this paper, an original method for spotsegmenting 2D-PAGE images is presented. The proposed approach is based on the concept of 2D histograms. The experimental results over real 2D-PAGE images demonstrate that it is efficient and it outperforms competing state-of-the-art software programs. Moreover, it can be applied to images containing spots of various intensities, sizes and shapes as well as to images containing overlapping spots, giving excellent results.

Acknowledgments

The authors would like to thank Biomedical Research Foundation of the Academy of Athens for the provision of 2-D page images as well as for the provision of their ground truth.

This research was partially funded by the special account of research grants of the University of Athens,

Greece and by the Greek State Scholarship Foundation (IKY).

References

- K.P. Pleissner et al, "New algorithmic approaches to protein spot detection and pattern matching in twodimensional electrophoresis databases," *Electrophoresis*, vol. 20, pp. 755-765, 1999.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] P.F. Lemkin and L.E. 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.
- [6] R.D. 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.
- [7] 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.
- [8] 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.
- [9] J. Zhang, and J. Hu, "Image Segmentation Based on 2D Otsu Method with Histogram Analysis," *International Conference Science and Software Engineering*, 2008.
- [10] J. Gong, L. Li, and W. Chen, "Fast recursive algorithms for two-dimensional thresholding," *Pattern Recognition*, vol. 31, no. 3, pp. 295-300, 1998.
- [11] E. Bettens et al, "Computer analysis of two-dimensional electrophoresis gels: a new segmentation and modeling algorithm," *Electrophoresis*, vol. 18, pp. 792-798, 1997.
- [12] R. Gonzalez, R. Woods, "Digital Image Processing," Prentice Hall, Second Edition.
- [13] R.D. Appel, J.R. Vargas, P.M. Palagi, D. Walther and D.F. Hochstrasser, "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.