

Protein Spot Detection in 2D-GE Images using Morphological Operators

E.A. Mylona^a, M.A. Savelonas^a, D. Maroulis^a, A. Vlahou^b, M. Makridakis^b
^a*Dept. of Informatics and Telecommunications, University of Athens, Greece*
^b*Biomedical Research Foundation of the Academy of Athens, Greece*
^a*{emylona,m.savelonas,d.maroulis}@di.uoa.gr*
^b*{vlahoua,mmakrid}@bioacademy.gr*

Abstract

This work addresses the detection of protein spots in 2D gel electrophoresis images. A novel morphology-based approach is proposed, which utilizes the dilation operator for the localization of regional intensity maxima associated with protein spots. A disk-shaped structuring element (SE) is selected in agreement with the prevalent roundish shape of the majority of protein spots. Thus, spots within rectangular-shaped streaks are correctly detected. SE size is set considering that a certain radius value allows the discrimination of individual spots located in complex spot regions as well as small-sized spots. Moreover, spurious intensity maxima associated with noise are ignored by applying regional intensity constraints. The results of the experimental evaluation lead to the conclusion that the proposed approach detects more actual protein spots and less false spots than a renowned 2D-GE image analysis software package. Furthermore, it does not require user intervention.

1. Introduction

Two-dimensional gel electrophoresis (2D-GE) is the leading proteomics technique, being capable to reflect the molecular anatomy of any biological sample. This technique radically changed the field of protein research [1], facilitating applications which range from disease prognosis, medical diagnosis and drug development to the monitoring of environmental pollution.

The process of 2D-GE involves the placement of a biological sample on a strip of polyacrylamide gel, which consists of a pH gradient. This gradient imposes the so called isoelectric focusing (IEF), which separates proteins according to their isoelectric points. The strip is subsequently placed on a sodium dodecyl sulfate polyacrylamide gel (SDS PAGE) and an

electric field is applied. At this stage, the proteins migrate until being immobilized at a position associated with the protein molecular weight.

2D-GE results in a digital grayscale image containing a few hundred up to several thousands of dark protein spots on a bright background or vice versa. Each spot reflects the presence of a certain amount of an individual protein. The analysis of 2D-GE images is either performed manually by biologists or with the assistance of software packages. However, manual analysis is tedious, time consuming and subjective [2-3], whereas most state-of-the-art software packages require manual editing and correction of software-generated results. These issues pose the need for robust, efficient and objective computer-assisted methods to enhance this key technology.

Additional issues arising in 2D-GE image analysis include the inhomogeneity of the background, the presence of rectangular-shaped streaks as well as the abundance of overlapping protein spots. Methods based on stepwise thresholding [4] and watersheds [5] aim to cope with these issues. However, the former fail in the presence of the inhomogeneous background [6] whereas the latter leads to the detection of false spots [7]. An active contour-based 2D-GE image analysis approach has been proposed by the authors in [8].

Morphology-based methods have been successfully employed in image analysis [9], including applications on a wide range of biomedical images such as: electron micrographs of muscle cross-sections [10], gastric tumor pathologic cell images [11] and digital color eye fundus images [12]. These methods allow the extraction of image components, useful in the representation of region shape such as boundaries, skeletons and convex hulls, by probing the image with an appropriately defined structuring element (SE).

In this work, a novel morphology-based approach is developed for the detection of protein spots in 2D-GE images. The dilation image operator is utilized for the detection of regional intensity maxima, which are associated with protein spots. This morphological

operator, although well known in image analysis literature, has not previously been employed in this application. A disk-shaped SE is applied on the image, adjusted to the dominant roundish shape of protein spots. Thus, protein spots located within rectangular streaks are correctly detected. SE is selected small-sized in order to allow the detection of local maxima associated with small spots, even in cases where such small spots overlap with larger spots. Moreover, regional intensity constraints are imposed in order to avoid spurious intensity maxima associated with noise. It should be noted that the proposed approach does not require user intervention.

The remainder of this paper is organized in four sections. Section 2 provides a brief background of the utilized aspects of mathematical morphology whereas Section 3 describes the main components of the proposed approach. Section 4 demonstrates the experimental results obtained, including quantitative comparisons with PDQuest (Bio-Rad), a well-known 2D-GE image analysis software package [13]. Finally, Section 5 discusses the conclusions of this work.

2. Mathematical Morphology

2.1. Dilation Operator

Let E be an Euclidean space, A a grayscale image in E and B a SE. The dilation image operator, denoted as $A \oplus B$, is defined, in the general case, by:

$$D(A, B) = (A \oplus B)(x) = \sup_{y \in E} [A(y) + B(x - y)] \quad (1)$$

where \sup denotes the supremum and x, y are elements of E (pixels in the case of a digital image) [14]. The SE plays the exact same role with the convolution kernel in the linear filter theory. Figure 1 illustrates the dilation image operator on an image A with a SE B .

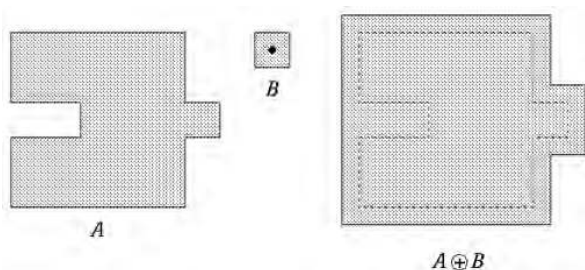


Figure 1. The dilation image operator on image A with SE B .

The shape and size of the SE are selected according to the shape and size of the objects of interest, respectively. For roundish-shaped small objects the SE is selected as disk-shaped with a radius smaller than the one of the object, aiming at detecting a large amount of regional intensity maxima [15]. Figure 2 illustrates a disk-shaped SE of radius 3.

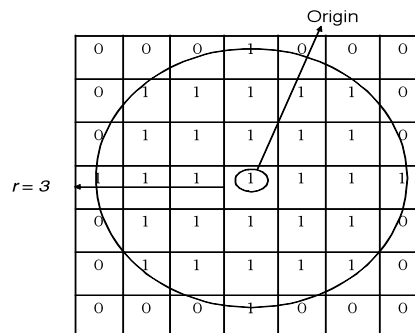


Figure 2. A disk-shaped SE of radius 3.

2.2. Regional Intensity Maxima

Regional intensity maxima are connected components of pixels with a constant intensity value, bounded from pixels of lower intensity value. A regional intensity maximum M at elevation t is defined by:

$$\left\{ \begin{array}{l} \forall p \in M, I(p) = t \\ \forall p \in \delta_{SE}(M) \setminus M, I(p) < t \end{array} \right\} \quad (2)$$

where p is the pixel location, $I(p)$ is the intensity of p and δ_{SE} is the region generated by the dilation of M according to the SE.

In the case of a 2D-GE image a regional intensity maximum is associated with a bright protein spot on a dark background, often corresponding to saturated spot region.

3. Proposed Morphology-based Approach

Based on the observation that the majority of the regional intensity maxima is associated with protein spots, the proposed approach aims to detect such maxima. The algorithm comprises of the following steps:

1. the four nearest neighbors (4NN) median filter is applied on the input 2D-GE image [16], so as to minimize the detection of background intensity peaks,

2. the image is dilated with a disk-shaped SE approximating the dominant roundish shape of protein spots. The typical size of a protein spot ranges from 20 to 300 pixels. Therefore, the radius of the disk r is selected smaller than the smallest protein spot radius, as determined by observation.
3. the image is scanned so as to detect regional intensity maxima associated with protein spots. The selected maxima should satisfy the following constraints:
 - their intensity should be equal to or higher than the intensity over a $q \times q$ adjacent region whose size should be equal to or higher than 20 pixels, so as to include salient maxima associated with spots and avoid extremely local maxima associated with noise,
 - pixel intensity over a $w \times w$ neighborhood should exceed a threshold value T , so as to avoid spurious maxima associated with background peaks and artifacts. The threshold value is selected equal to or higher than the faintest spot's intensity.

Figure 3 depicts: (a) a part of a real 2D-GE image, (b) the dilated image of Fig. 3(a) with a disk-shaped SE and (c) the binary image identifying the locations of the selected regional intensity maxima.

4. Results

The Biomedical Research Foundation of the Academy of Athens provided real 2D-GE images. Each such image is accompanied by three annotated images, which were manually edited by expert biologists, in order to allow quantitative evaluations of the proposed approach. The ground truth images, which were used in the experiments, resulted from the majority rule applied on each triad of annotated images. The proposed algorithm has been implemented in Matlab R2009b and executed on a 3.2 GHz Intel Pentium workstation. Parameters r , q , w and T of the proposed approach were experimentally determined as 4, 4, 3 and 75 respectively for each image that was utilized. The values of the parameters did not need to change since the 2D-GE images used for the experiments were produced from the same 2D-GE experiment.

Figure 4 illustrates the detection results obtained by the proposed approach on a real 2D-GE image, whereas Fig. 5 illustrates the detection results obtained by PDQuest image analysis software package on the same image. PDQuest parameters were set by an expert biologist, following his experience.



Figure 3. (a) a real region of a 2D-GE image, (b) the dilated image and (c) the binary image identifying the locations of selected regional intensity maxima.

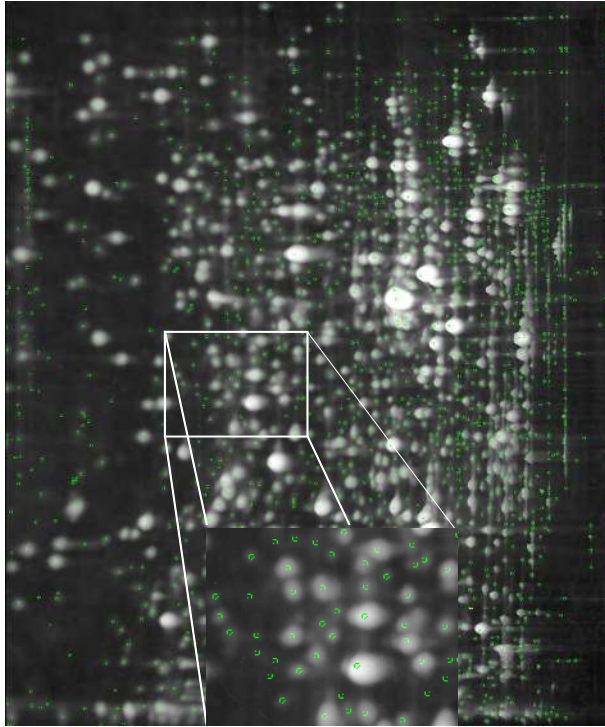


Figure 4. Detection results obtained by the proposed approach on a real 2D-GE image.

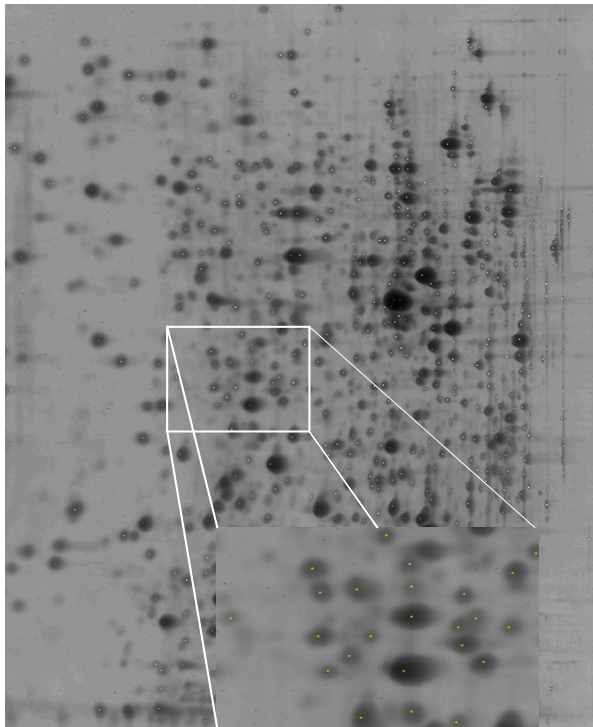
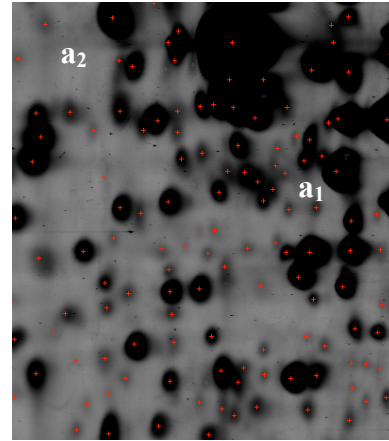
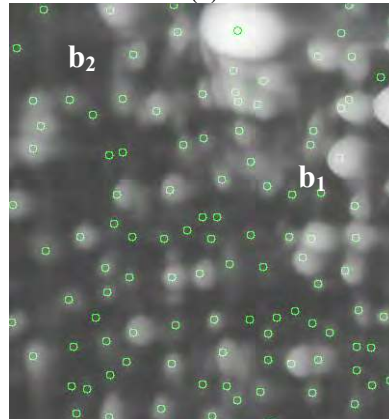


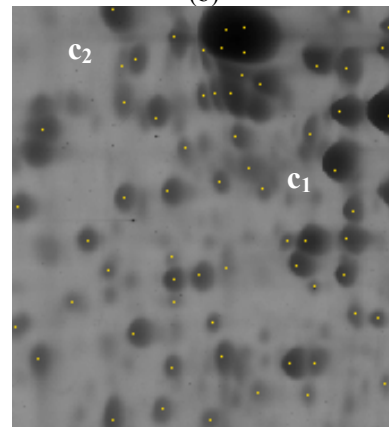
Figure 5. Detection results obtained by PDQuest software package on the 2D-GE image of Fig. (4).



(a)



(b)



(c)

Figure 6. (a) ground truth image of Fig. 3(a), (b) detection results obtained by the proposed approach and (c) detection results obtained by PDQuest software package.

Figure 6(a) illustrates the ground truth image of Fig. 3(a), whereas Fig. 6(b) and Fig. 6(c) illustrate the detection results obtained by the proposed approach and PDQuest image analysis software package, respectively. It is evident that considering the ground

truth, the proposed approach manages to detect more actual protein spots than PDQuest.

The detection results were quantified by means of the detection accuracy (DA) [17], which is defined as:

$$DA = \frac{TP}{TP + FN} \quad (3)$$

where true positives (TP) and false negatives (FN) are defined as correctly detected and missed actual spots, respectively. TP and FN were calculated on the basis of the ground truth images.

Considering the detection results illustrated in Fig. 6, it can be derived that the proposed approach detects 105 true positive (TP) and 25 false negative (FN) protein spots, whereas PDQuest detects 67 TP and 68 FN protein spots, respectively. These quantitative results verify that the proposed approach: a) correctly detects much more actual spots than PDQuest, b) detects much less false spots than PDQuest. Table I presents the calculated TP and FN spots, as detected by the proposed approach (Fig. 6(b)) and PDQuest (Fig. 6(c)).

TABLE I
Detection Results

	TP	FN	DA
Proposed Approach	105	25	81%
PDQuest	67	68	50%

In a total of approximately 20,000 actual protein spots appearing in 10 2D-GE images provided, the proposed approach achieves a DA of $85.4 \pm 4.7\%$, whereas PDQuest obtains a DA of $55 \pm 8.9\%$.

Figure 7 illustrates: detailed ground truth (a_1 - a_2), detection results of the proposed approach (b_1 - b_2) and of PDQuest (c_1 - c_2), over detailed sub-images of Fig. 6, containing overlapping spots (a_1 - c_1) and vertical streaks (a_2 - c_2). In cases of complex regions consisting of multiple overlapping spots, unlike PDQuest (Fig. c_1), the proposed approach is capable of detecting each individual spot (Fig. b_1), as allowed by the selection of the small size of r of the disk-shaped SE. Moreover, it efficaciously locates spots within vertical streaks (Fig. b_2), which is not the case with PDQuest (Fig. c_2).

5. Conclusions

In this work, a novel morphology-based approach is proposed, aiming at the detection of protein spots in 2D-GE images. Several issues arise in the analysis of

such images, including the inhomogeneity of the background, the presence of noise and complex regions containing overlapping spots, as well as the presence of rectangular-shaped streaks.

Aiming to cope with these issues, the proposed approach utilizes the dilation image operator for the location of regional intensity maxima associated with protein spots. The SE is selected disk-shaped in agreement with the dominant roundish shape of the majority of spots. This allows the correct detection of spots located within streaks. Additionally, the size of the disk-shaped SE is set to a relatively small value, so as to allow the detection of local maxima associated with small spots, even in cases where such small spots overlap with larger spots. Furthermore, intensity constraints are imposed on the selection of regional maxima, in order to select salient maxima and avoid spurious ones, which are associated with noise.

The detection results obtained by the proposed approach and PDQuest image analysis software package were quantitatively compared considering ground truth images provided by expert biologists. The proposed approach evidently outperforms PDQuest in terms of detection accuracy, as it correctly detects much more TP and misses much less FN . Moreover, unlike PDQuest, it does not require user-intervention. Future perspectives of this work involve the development of an integrated 2D-GE image analysis system for the identification of protein spots.

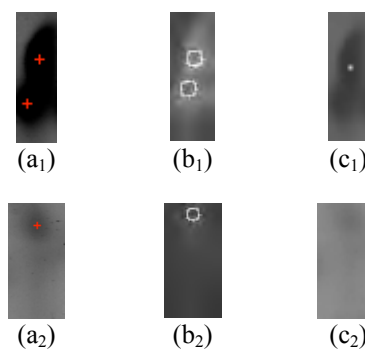


Figure 7. (a_1 - a_2) Detailed ground truth, (b_1 - b_2) detection results of the proposed approach and (c_1 - c_2) detection results of PDQuest, containing overlapping spots (a_1 - c_1) and vertical streaks (a_2 - c_2).

6. Acknowledgements

We would like to thank the Biomedical Research Foundation of the Academy of Athens for the provision of 2D-GE images as well as ground truth of these images. This work was partially supported by the

European Social Fund and the Ministry of National Education of Greece through the HRAKLEITOS II program.

7. References

- [1] A.W. Dowsey, M.J. Dunn, and G.Z. Yang, "The Role of Bioinformatics in Two-Dimensional Gel Electrophoresis", *Proteomics*, vol. 3, 2003, pp. 1567-1596.
- [2] G. Mercurio, S. Maglio, A. Agrusti, G. DeNunzio, R. Cataldo, I. DeMitre, M. Favetta, A. Massafra, G. Marsella, D. Vergara and M. Maffia, "Network P2P for Exploring and Visualization of Proteomic Data Produced by Two Dimensional Electrophoresis", *Proc. IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 197-202, 2008.
- [3] N. Khan, S. Rahman and A.G. Stockman, "Integration of Biological Data Resources Using Image Object Keying," *Proc. IEEE Symposium on Computer-Based Medical Systems (CBMS)*, pp. 16, 2003.
- [4] 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, 2003, pp. 392-401.
- [5] L. Vincent and P. Soille, "Watersheds in Digital Spaces - an Efficient Algorithm Based on Immersion Simulations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 583-598.
- [6] Y. Kim, J. Kim, Y. Won and Y. In, "Segmentation of Protein Spots in 2D Gel Electrophoresis Images with Watersheds using Hierarchical Threshold", *Computer and Information Sciences*, vol. 2869, 2003, pp. 389-396.
- [7] A. Bleau and L.J. Leon, "Watershed-based Segmentation and Region Merging", *Computer Vision and Image Understanding*, vol. 77, 2000, pp. 317-370.
- [8] M. Savelonas, D. Maroulis, E. Mylona, "Segmentation of Two-Dimensional Gel Electrophoresis Images Containing Overlapping Spots", *Proc. IEEE International Conference on Information Technology Applications in Biomedicine (ITAB)*, Larnaca, Cyprus, 2009.
- [9] L. Vincent, "Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms", *IEEE Transactions on Image Processing*, vol. 2, 1993, pp. 176-201.
- [10] B. Bödvarsson, S. Klim, M. Mørkebjerg, S. Mortensen, C.H. Yoon, J. Chen, J.R. Maclaren, P.K. Luther, J.M. Squire, P.J. Bones and R.P. Millane, "A Morphological Image Processing Method for Locating Myosin Filaments in Muscle Electron Micrographs", *Image and Vision Computing*, vol. 26, pp. 1073-1080.
- [11] T.G. Li, S.P. Wang, and N. Zhao, "Gray-Scale Edge Detection for Gastric Tumor Pathologic Cell Images by Morphological Analysis", *Computers in Biology and Medicine*, vol. 39, 2009, pp. 947-952.
- [12] D. Welfer, J. Scharcanski, C.M. Kitamura, M.M Dal Pizzol, L.W.B. Ludwig and D.R. Marinho, "Segmentation of the Optic Disk in Color Eye Fundus Images using an Adaptive Morphological Approach", *Computers in Biology and Medicine*, vol. 40, 2010, pp. 124-137.
- [13] J.I. Garrels, "The Quest System for Quantitative Analysis of Two-Dimensional Gels", *Journal of Biological Chemistry*, vol. 264, 1989, pp. 5269-5282.
- [14] J. Serra, *Image Analysis and Mathematical Morphology*, 2nd ed. Academic, New York, 1982.
- [15] E.R. Urbach and M.H.F. Wilkinson, "Efficient 2-D Grayscale Morphological Transformations With Arbitrary Flat Structuring Elements", *IEEE Transactions on Image Processing*, vol. 17, 2008, pp. 1-8.
- [16] D.T. Lin, "Autonomous Sub-Image Matching for Two-Dimensional Electrophoresis Gels using MaxRST Algorithm", *Image and Vision Computing*, In Press, Corrected Proof, 2010.
- [17] C. Guda, E. Fahy and S. Subramaniam, "MITOPRED: a Genome-Scale Method for Prediction of Nucleus-encoded Mitochondrial Proteins", *Bioinformatics*, vol. 20, 2004, pp. 1785-1794.