2D-GE Spot Detection Combining Multidirectional Texture and Spatial Intensity Cues

Eleni Zacharia, Eirini Kostopoulou, Dimitris Maroulis, Nicholas P. Anagnou, and Kalliopi I. Pappa

***Abstract*—Spot detection is a challenging task of 2D Gel Electrophoresis image analysis. The available software packages and techniques miss some of the protein spots while they detect a high number of spurious spots. This paper introduces a novel approach for the detection of protein spots on 2D gel images which is based on multidirectional texture and spatial intensity information. The proposed approach is compared with two commercial software packages using real 2D-GE images. The outcome demonstrates that the proposed approach outperforms the two software packages; it detects almost all of real protein spots and a low number of spurious spots.**

I. Introduction

TWO dimensional gel electrophoresis (2D-GE) has been established as a mature technique for large-scale proteomic analysis as it separates protein samples based on protein net charge and molecular mass [1-2]. Due to this technique, many fields of research (i.e. cancer and infectious disease diagnosis and treatment, drug production, diagnostic markers development etc.) have been facilitated. 2D-GE technique results in a digital grayscale image containing thousands of white spots on a dark background. Each spot reflects the presence of a certain amount of an individual protein.

Detection of protein spots on a 2D-GE image is a crucial and challenging task which is impeded by the poor quality of images [3]: 2D-GE images contain inhomogeneous background and are contaminated with noise and artifacts. Furthermore, such images contain thousands of spots of various sizes and shapes. In many cases spots are so poorly contrasted that they are not clearly visible. Moreover, adjacent spots are often highly overlapped.

2D-GE image analysis software packages such as Melanie 7 (Bio-Rad) [4] and Delta 2D (Decodon) [5] are widely used in biologic laboratories. Despite their respective merits, they require human intervention for selection of mandatory input parameters as well as correction of the results. Accurate results stemming from the existing software packages have not yet been achieved [6]. Other well-known approaches such as watersheds [7] and inner-marked watershed methods [8] attempt to cope with overlapping spots; nevertheless, they lead to over-segmentation and require user intervention in order to refine the results. The active contour-based method [9] and Anjos et al. watershed-based method [10] aims to confront most challenges; however, they cannot segment highly overlapping spots. Additionally, Anjos et al. method requires human intervention which is a time-consuming procedure and adds subjectivity to the results. Our previous work for spot detection and segmentation [11] was based on 2D histograms and 3D spots morphology. Although it was automatic and proved to be very effective even when applied to images produced by different acquisition devices, it presents a high computational complexity.

In this paper, an original approach for spot detection on 2D-GE images is presented. The proposed approach detects the regional intensity maxima of the 2D-GE image and subsequently chooses those which correspond to spot centers by combining multidirectional texture (obtained through contourlet transform [12] decomposition), and spatial intensity information. Multiple regional intensity maxima located in the same spot are merged. Experiments were conducted on real 2D-GE images in order to evaluate the proposed approach and compare it with Melanie 7 and Delta 2D commercial software packages. The results demonstrate that the proposed approach is effective and detects almost all of real protein spots as well as a low number of spurious spots. Furthermore, it outperforms the two aforementioned software packages.

The remainder of this paper is organized in four sections. Section II provides an overview of the contourlet transform. Section III describes the proposed approach for spot detection on 2D-PAGE images. Section IV presents the experimental results obtained by Delta2D, Melanie 7 and the proposed approach. Finally, conclusions are summarized in Section V.

II. Background

The Contourlet Transform (*CT*) is a multiscale image representation scheme that can effectively represent smooth contours in different directions of an image [12]. Fig. 1 shows a flowchart of the contourlet transform. The method

Manuscript received July 30, 2013. This research has been financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) – Research Funding Program: THALIS, UOA, CERVI-CAN-PROT.

E. Zacharia, E. Kostopoulou and D. Maroulis are with Real Time Systems & Image Analysis Group, Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimiopolis, Ilisia, 15784, Greece (phone: +30-210-7275317; e-mail: eezacharia@gmail.com, ikostop@di.uoa.gr, dmaroulis@di.uoa.gr).

N. P. Anagnou and K. I Pappa are with Department of Basic Medical Sciences, Laboratory of Biology, University of Athens School of Medicine and Laboratory of Cell and Gene Therapy, Biomedical Research Foundation of the Academy of Athens, Greece. K. I Pappa is also with First Department of Obstetrics and Gynecology, University of Athens School of Medicine, Greece (email: anagnou@med.uoa.gr, kalliopib@yahoo.com).

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

is implemented via a double iterated filter bank, which is the combination of the Laplacian Pyramid, used to achieve multiscale decomposition and a Directional Filter Bank which reveals the directional details at each scale.

Specifically, let I^0 be an input image. At each level j, (j=1,…,J) the Laplacian Pyramid is firstly used to decompose I_L^{j-1} (or I^0 if j=1) into a low-pass image I_L^j and a high-pass image I_H^j. Subsequently, the high-pass image I_H^j is decomposed by the Directional Filter Bank into 2^{Dj} directional subband images C_d^j, d=0,…,2^{Dj}-1. The values of the directional subband images are called "*contourlet coefficients*" or shortly "*coefficients*".

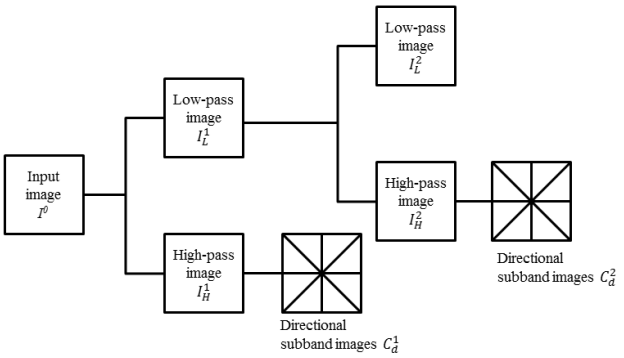

Fig. 1. Flowchart of the Contourlet Transform (CT)

III. Proposed Approach

According to Bettens et al. [13], protein spot intensity peaks at its central region and declines at regions further from the center. Based on this remark, spot centers correspond to regional intensity maxima of the 2D-GE image. However, not all of the regional intensity maxima correspond to spot centers as the 2D-GE image is contaminated with noise and artifacts.

The proposed approach detects regional intensity maxima $m \in M$ of a 2D-GE image I and subsequently chooses those which correspond to spot centers based on the following two proposed operators: (i) Filtering based on multidirectional texture information and (ii) Filtering based on spatial intensity information. However, multiple regional intensity maxima located in the same spot may remain after the filtering operators (i) and (ii). Therefore, a post-processing step is also proposed which merges the multiple regional intensity maxima of the same spot.

A. Filtering based on Multidirectional Texture Information

Once CT is applied to 2D-GE images, coefficients of high and low magnitude are produced mostly in background area and in spot areas, respectively. This differentiation in coefficient magnitudes becomes more noticeable when CT is applied to image texture such as the gradient direction of the 2D-GE image. Fig. 2 depicts the gradient direction of a 2D-GE sub-image as well as its four directional subband images.

As a result of the above observation, the gradient direction of the 2D-GE image I is first decomposed into its directional subband images C_d^j using CT. Each directional subband C_d^j is thresholded, locating the significant and insignificant coefficients of the corresponding subband (C_d^j). Each pixel p of I is assigned a label ($\lambda(p)$) based on the following criterion: If the corresponding coefficients of p in one or more subbands are significant, then p is labeled '1', otherwise p is labeled '0'. Fig. 3(a) illustrates a 2D-GE sub-image in which the pixels colored in red represent the pixels labeled '1'.

A regional intensity maximum m corresponds to a spot center if a square-shaped neighbourhood R_m around m contains a small percentage P_m of pixels labeled '1' (P_m<P_{max}, where P_{max} is a threshold). P_m is defined by the following equation:

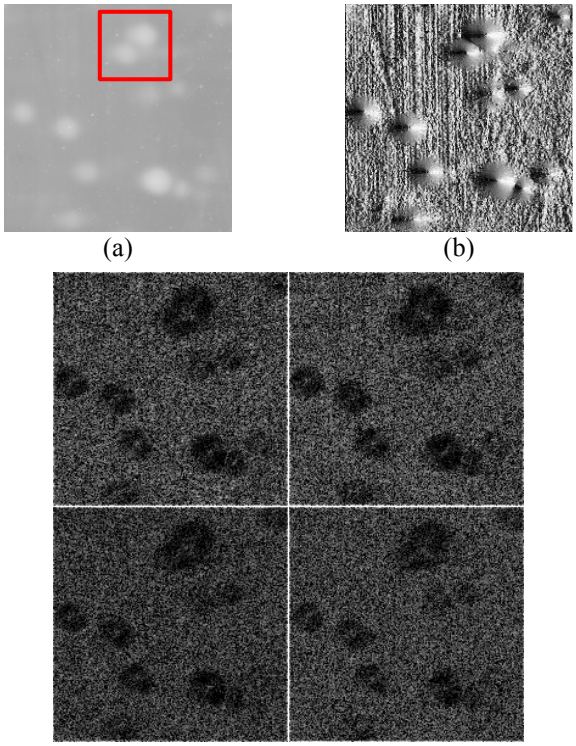

Fig. 2 (a) 2D-GE sub-image, (b) Gradient direction of (a), (c) contourlet coefficients in four directional subbands of (b).

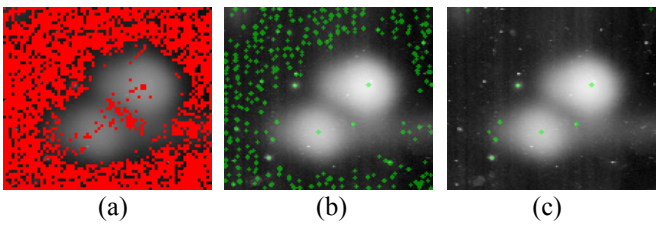

Fig. 3. Enlargement of the red square region of 2D-GE sub-image of Fig. 2(a). Pixels colored in red represent pixels labeled '1' (a). Pixels colored in green represent: (b) regional intensity maxima and (c) regional intensity maxima after applying the filtering based on multidirectional texture information.

$$P_m = \frac{\sum_{p \in R_m} \lambda(p)}{\sum_{p \in R_m} 1} \quad (1)$$

Fig. 3(b) depicts in green color the regional intensity maxima of the 2D-GE image, whereas Fig. 3(c) depicts in green color the regional intensity maxima after applying the filtering based on multidirectional texture information.

B. Filtering Based on Spatial Intensity Information

Assuming that R_m is a small region embedding the regional intensity maximum m of I, local thresholding is applied on R_m and a binary B_m containing many connected components is generated. If m is located near the boundaries of a connected component then m is not a spot center; otherwise it is. An example is illustrated in Fig. 4.

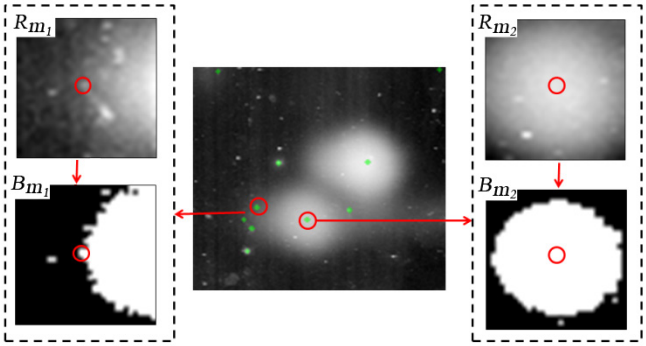

Fig. 4. (a) Enlargement of the 2D-GE sub-image of Fig. 3(c). The two regional intensity maxima m_1, m_2 embedded in red circles are examined for being spot centers. (a_1) R_m, B_m regions located around m_1 are depicted on the top and bottom respectively. m_1 does not correspond to a spot center. (a_2) R_m, B_m regions located around m_2 are depicted on the top and bottom respectively. m_2 corresponds to a spot center.

Based on this observation, optimal thresholding is applied on a square region R_m of I around each regional intensity maximum m and a binary *region* B_m, is created, in which the high and low intensity values are represented with one and zero respectively. Subsequently, the Euclidean distance d between the regional maximum m and its nearest zero pixel of the B_m region is computed. If d is higher than a threshold d_1, then m corresponds to a spot center.

C. Merging Multiple Regional Intensity Maxima Located Inside a Spot

Let M_1, M_2 be the sets of spot centers determined from the filtering based on multidirectional texture information and on spatial intensity information respectively. The final set of spot centers is defined as follows:

$$C = \{m : m \in M_1 \wedge m \in M_2\} \quad (2)$$

Multiple regional intensity maxima of the same spot are merged whether their Euclidean distance is less than a threshold d_2, and their centroids are considered as spot centers.

IV. RESULTS

Experiments were conducted in order to evaluate the performance of the proposed approach on a set of real 2D-GE images containing a total of ~2000 spots. The dataset was provided by the Biomedical Research Foundation of the Academy of Athens. Each image was accompanied by three annotated images depicting real protein spots with blue crosses. The ground truth images, which were used in the experiments, resulted from the majority rule applied on each triad of annotated images. The performance of the proposed approach has been evaluated and compared with the performance of two commercial packages; Delta2D and Melanie 7. For CT decomposition, the "9-7" biorthogonal filters were used for the LP and these filters were mapped to their corresponding 2D filters for the DFB, using the McClellan transform as proposed in [12]. Decomposition of one scale was selected for the LP and of four directions with the DFB. In all the experiments, P_{max}, d_1, and d_2 were experimentally set to 0.5, 4 and 6.

Fig. 5 illustrates the detection results obtained by Melanie 7, Delta2D, and the proposed approach on a real 2D-GE sub-image. On this image, one may observe that all methods miss faint spots and detect spurious spots. However, the proposed approach has missed and falsely detected a clearly lower number of spots. In particular, the proposed approach detects 97 real spots out of 107 in total, misses 10, and detects 3 spurious spots. On the same image, Delta 2D detects 85 real spots, misses 22, and detects 12 spurious spots, whereas Melanie 7 detects 76 real spots, misses 31, and detects 1 spurious spot.

The detection results were evaluated using *Sensitivity* (S), *Precision* (P) and *F-measure* (F) defined as follows:

$$S = \frac{TP}{TP + FN}, \; P = \frac{TP}{TP + FP}, \; F = 2 \cdot \frac{S \cdot P}{S + P} \quad (3)$$

where TP, FN, FP denote the number of spots that are correctly detected, missed, and falsely detected respectively. It is worth pointing out that F (the weighted harmonic mean of S and P) is a more reliable measure than just S and P as it takes into account both the number of detected protein spots and the number of spurious spots.

In a total of approximately 2.000 real protein spots, the proposed approach achieved an S value of 78.1%, whereas Delta2D and Melanie 7 obtained 74% and 69% respectively. Furthermore, the proposed approach achieved a P value of 92% whereas Delta2D and Melanie obtained 76.7% and 95.3% respectively. Last but not least, the proposed approach achieved an F value of 84.4%, whilst Delta2D and Melanie 7 achieved a value of 75.4% and 80.1% respectively. Since F is a more reliable measure than S and P, it can be derived that the proposed approach outperforms

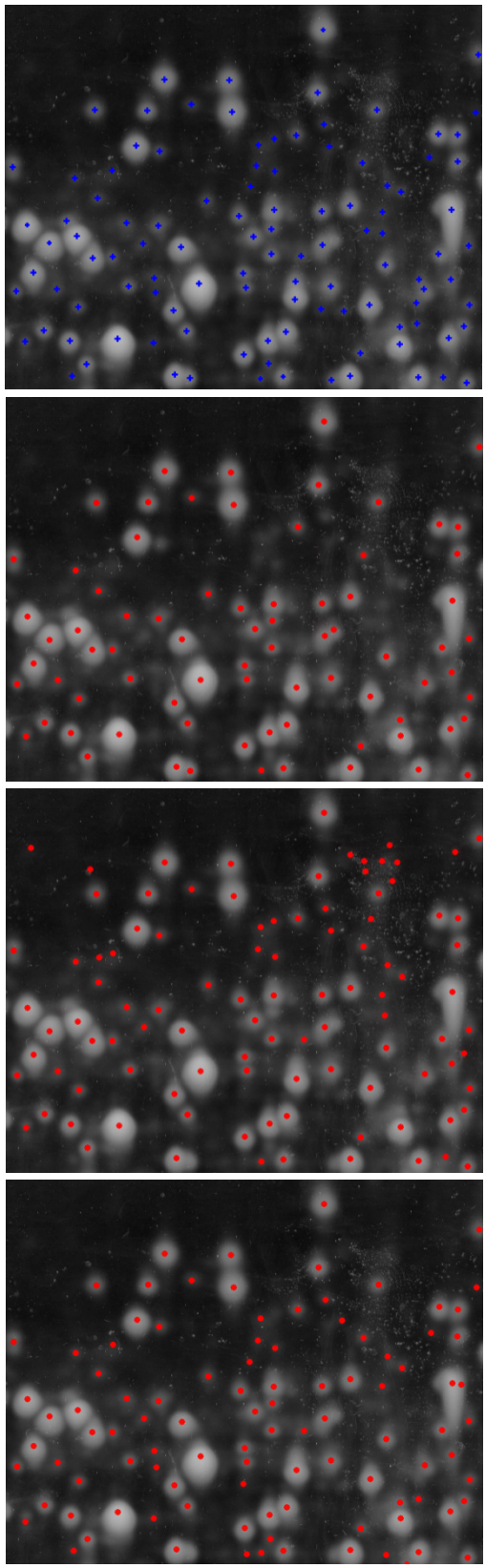

Fig. 5. (a) Ground truth image and detection results obtained by: (b) Melanie 7, (c) Delta2D, and (d) the proposed approach.

in overall the two compared software packages.

V. CONCLUSION

In this paper, an approach for the detection of protein spots on 2D-GE images is proposed. The proposed approach detects regional intensity maxima of the 2D-GE image and subsequently chooses those which correspond to spot centers based on two filtering operators. Multiple regional intensity maxima located in the same spot are merged. Experiments were conducted on real 2D-GE images in order to evaluate the proposed approach and compare it with two widely used software packages. The results have shown that the proposed approach detects spots in a more effective manner and therefore it outperforms Melanie 7 and Delta 2D software packages. Furthermore, it detects almost all of real protein spots and a low number of spurious spots. Future perspectives of this work involve the segmentation of protein spots and the development of an integrated image analysis system for the detection and segmentation of protein spots on 2D-GE images.

REFERENCES

[1] J.M.Lee, J.J. Han, G. Altwerger, and E.C. Kohn, "Proteomics and biomarkers in clinical trials for drug development", *Proteomics J.*, vol. 74, no. 12, pp.2632-2641, November 2011.

[2] A.W. Dowsey, M.J. Dunn, and G.Z. Yang, "The role of bioinformatics in two-dimensional gel electrophoresis," *Proteomics J.*, vol. 3, no. 8, pp. 1567–1596, August 2003.

[3] A. Görg, W.Weiss, and M. Dunn, "Current two-dimensional electrophoresis technology for proteomics," *Proteomics*, vol. 4, no. 12, pp. 3665–3685, December 2004.

[4] Delta2D [www.decodon.com]

[5] Melanie 7 [www.genebio.com]

[6] B.N. Clark, and H.B. Gutsein, "The myth of automated, high-throughput two-dimensional gel analysis," *Proteomics*, vol. 8, no.6, pp. 1197-1203, March 2008.

[7] K.P. Pleissner, F. Hoffman, K. Kriegel, C. Wenk, S. Wegner, A. Sahlstrom, H. Oswald, H. Alt and E. Fleck, "New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis databases," *Electrophoresis*, vol. 20, pp. 755-765, January 1999.

[8] Y. Kim, J. Kim, Y. Won, and Y. In, "Segmentation of protein spots in 2-D gel electrophoresis images with watershed using hierarchical threshold," *in Proc. 18th International Symposium on Computer and Information Sciences*, 2003, vol. 2869, pp. 389-396.

[9] M. Savelonas, E. Mylona, and D. Maroulis, "Unsupervised 2D Gel Electrophoresis Image Segmentation based on Active Contours," *Pattern Recognition*, vol. 45, no. 12, pp. 720-731, 2012.

[10] A dos Anjos, AL Møller, BK Ersbøll, C. Finnie, and H.R. Shahbazkia, "New approach for Segmentation and Quantification of Two-Dimensional Gel Electrophoresis Images," *Bioinformatics*, vol. 27, no.3, pp. 368-375, 2010.

[11] E. Kostopoulou, E. Zacharia, and D. Maroulis, "An Effective Approach for Detection and Segmentation of Protein Spots on 2D-Gel Images," *IEEE Journal of Biomedical and Health Informatics*, to be published. Available: http://dx.doi.org/10.1109/JBHI.2013.2259208.

[12] M.N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. on Image Processing*, vol. 14, no. 12, pp. 2091-2106, 2005.

[13] E. Bettens, P. Scheunders, D. Van Dyck, L. Moens, and P. Van Osta, "Computer analysis of two dimensional electrophoresis gels: A new segmentation and modeling algorithm," *Electrophoresis*, vol. 18, no. 5, pp. 792-798, 1997.