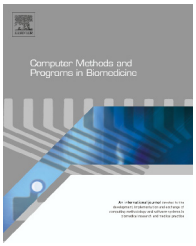
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 122 (2015) 26–39

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



2D-gel spot detection and segmentation based on modified image-aware grow-cut and regional intensity information

E. Kostopoulou, S. Katsigiannis, D. Maroulis*

Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 15784 Panepistimioupolis, Athens, Greece

ARTICLE INFO

Article history:
Received 7 January 2015
Received in revised form
15 June 2015
Accepted 16 June 2015

Keywords:
Segmentation
Proteomicss
2D-gel electrophoresis

ABSTRACT

Background: Proteomics, the study of proteomes, has been increasingly utilized in a wide variety of biological problems. The Two-Dimensional Gel Electrophoresis (2D-PAGE) technique is a powerful proteomics technique aiming at separation of the complex protein mixtures. Spot detection and segmentation are fundamental components of 2D-gel image analysis but remain arduous and difficult tasks. Several software packages and academic approaches are available for 2D-gel image spot detection and segmentation. Each one has its respective advantages and disadvantages and achieves a different level of success in dealing with the challenges of 2D-gel image analysis. A common characteristic of the available methods is their dependency on user intervention in order to achieve optimal results, a process that can lead to subjective and non-reproducible results. In this work, the authors propose a novel spot detection and segmentation methodology for 2D-gel images.

Methods: This work introduces a novel spot detection and spot segmentation methodology that is based on a multi-thresholding scheme applied on overlapping regions of the image, a custom grow-cut algorithm, a region growing scheme and morphological operators. The performance of the proposed methodology is evaluated on real as well as synthetic 2D-gel images using well established statistical measures, including precision, sensitivity, and their weighted measure, *F*-measure, as well as volumetric overlap, volumetric error and volumetric overlap error.

Results: Experimental results show that the proposed methodology outperforms state-of-the-art software packages and methods proposed in the literature and results in more plausible spot boundaries and more accurate segmentation. The proposed method achieved the highest *F*-measure (94.8%) for spot detection and the lowest volumetric overlap error (8.3%) for the segmentation process.

Conclusions: Evaluation against state-of-the-art 2D-gel image analysis software packages and techniques proposed in the literature, including Melanie 7, Delta2D, PDQuest and Scimo, demonstrates that the proposed approach outperforms the other methods evaluated in this work and constitutes an advantageous and reliable solution for 2D-gel image analysis.

* *Corresponding author*. Tel.: +30 210 7275317; fax: +30 210 72753333.
E-mail addresses: ikostop@di.uoa.gr (E. Kostopoulou), stamos@di.uoa.gr (S. Katsigiannis), dmaroulis@di.uoa.gr (D. Maroulis).

http://dx.doi.org/10.1016/j.cmpb.2015.06.007
0169-2607/

1. Introduction

In recent years, proteomics, i.e. the study of proteomes under different conditions, has been increasingly utilized for revealing the complex processes of cells [1]. Some key opportunities offered by the field of proteomics are the evaluation of new drugs and the exploration of biological events [2–6]. The *Two-Dimensional Polyacrylamide Gel Electrophoresis* (2D-PAGE) technique is a powerful technique in proteomics aiming at protein separation and identification and has been widely used due to its ability to separate thousands of proteins on polyacrylamide gels. This is achieved using isoelectric focusing and sodium dodecyl sulfate polyacrylamide gel electrophoresis that allows for protein separation according to the differences in their net charge and their molecular mass [7–10]. Results are then visualized into a digital image that can contain thousands of protein spots. In a 2D-gel electrophoresis experiment the aim of 2D-gel image analysis is the rapid identification of: (a) proteins located on a single gel and (b) differentially expressed proteins between samples from a series of 2D-gels.

2D-gel image analysis can be summarized into four main stages: (1) spot detection, (2) spot segmentation, (3) spot quantification and (4) image alignment. The objective of the first three stages is to detect the number of pixels that belong to protein spots in order to quantify the protein expression levels. The fourth stage is performed in order to match the corresponding protein spots from different images. Two different workflows for 2D-gel image analysis can be followed [11,12]. In the straightforward workflow, spot detection and segmentation are performed prior to image-alignment [13,14]. In the workflow followed by the Delta2D [15] and Progenesis Samespots [16] software tools, image-alignment is applied prior to spot detection and segmentation. The characteristics of 2D-gel images make spot detection and segmentation a very challenging task. 2D-gel images may contain thousands of spots that exhibit great variety in intensity, size and shape. Spots can be so poorly contrasted that they are not clearly visible and adjacent spots can often be highly overlapped. Furthermore, the overall quality of these images suffers due to artifacts, inhomogeneous background and high levels of noise [17].

Due to the importance of 2D-gel image analysis, many commercial software solutions have been developed [12]: e.g. PDQuest [18], DeCyder 2D [19], Melanie 7 [20], ImageMaster 2D [21], Delta2D [15] and Progenesis Samespots [16]. Each software has its respective advantages and disadvantages and achieves a different level of success in dealing with the challenges of 2D-gel image spot detection and segmentation [12,22]. An important observation is that all these software tools are dependent on manual parameter tuning, thus requiring human intervention that can lead to subjective results. Moreover, the most common problems in their results are that they may: (1) segment overlapping spots as one single spot, (2) over-segment a single spot into more, (3) fail to detect some spots, (4) mistake artifacts for spots and (5) incorrectly approximate the spot boundaries. The accumulation of all these errors alters the extracted protein expression levels, thus leading to erroneous biological conclusions. This problem is addressed by extensive manual editing of the results, which is a time-consuming process that requires an average of 1 to 4 man-hours per gel [23]. Human intervention leads to subjective and non-reproducible results and limits the throughput of the analysis process. As a result, the development of an automated 2D-gel image analysis method would be of utmost importance since it would allow for high throughput analysis of the expression levels of thousands of proteins, and would lead to safer, objective and reproducible biological conclusions.

Many researchers have tried to address the challenges of 2D-gel image analysis. Based on the assumption that the shape of protein spots follows a Gaussian distribution, Yoon et al. [24] proposed the Reversible Jump Markov Chain Monte Carlo (RJMCMC) method for separating overlapping spots and enhancing the weaker spots. RJMCMC can be very time consuming and due to the original assumption it underperforms in cases of images with many non-Gaussian shaped spots. Morris et al. [25] proposed “Pinnacle”, a method for spot detection and quantification. Pinnacle’s spot detection is performed on a denoised average image of a properly aligned 2D-gel image set and is achieved by detecting local minima (“pinnacles”) on the denoised average image and combining them within a defined proximity. Pinnacle’s main advantage lies on detecting overlapping spots [25]. Nevertheless, it occasionally results in detecting spurious spots, i.e. false positive spots [26]. Li et al. [27] proposed “RegStatGel”, a method that is also performed on an average image of a properly aligned 2D-gel set and incorporates the watershed algorithm [28] for spot segmentation. Its main drawback is that it underperforms in splitting overlapping spots [26]. Based on morphological operations, Mylona et al. [29] proposed a method for spot detection that achieves increased performance in detecting overlapping spots at the expense of occasionally missed spots. Dos Anjos et al. [14] proposed “Scimo”, a spot segmentation and quantification method that is also based on the watershed algorithm and achieves a more realistic estimation of close proximity spots, as well as partially overlapping non-saturated spots. However, the authors make the assumption that each basin of the watershed contains only one protein spot. This assumption is not always valid in case of major overlapping. An active contour-based method [13] has also been proposed for 2D-gel image segmentation but requires empirical adjustment for a large number of parameters for different image datasets, a process that is lengthy and time-consuming, as described in [13]. Kostopoulou et al. [30] proposed an effective spot detection and segmentation approach based on 2D histograms and 3D spot morphology. Although effective, this approach suffers from a high computational complexity. Zacharia et al. [31] proposed a spot detection approach based on multidirectional texture and spatial intensity information, but is not capable of segmentation.

In this paper, the authors present a novel approach for 2D-gel image spot detection and segmentation. The proposed detection approach is based on a multi-thresholding scheme applied on overlapping regions of the image. For the segmentation process, the proposed method combines a grow-cut algorithm [32] that utilizes a custom update rule, with a region growing approach and morphological operators. The proposed detection and segmentation methodology is evaluated against