Image and Vision Computing 28 (2010) 1267–1279

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Autonomous sub-image matching for two-dimensional electrophoresis gels using MaxRST algorithm

Daw-Tung Lin *

Department of Computer Science and Information Engineering, National Taipei University, No. 151, University Rd., Sanshia, Taipei County 237, Taiwan

ARTICLE INFO

Article history:
Received 11 October 2008
Received in revised form 22 December 2009
Accepted 19 January 2010

Keywords:
Sub-image matching
Two-dimensional electrophoresis
Features extraction
Gabriel graph
Relative neighborhood graph
Maximum relation spanning tree
Gaussian similarity measure

ABSTRACT

Matching two-dimensional electrophoresis (2-DE) gel images typically generates a bottleneck in the automated protein analysis, and image distortion and experimental variation, which reduce the matching accuracy. However, conventional matching schemes only compare two complete images, and landmark selection and registration procedures are rather time-consuming. This work presents a novel and robust Maximum Relation Spanning Tree (MaxRST) algorithm, in which an autonomous sub-image matching method does not require registering or manual selection of landmarks. The 2D gel images are represented graphically. Image features are then quantitatively extracted regardless of image size. Similarity between a sub-image and large image is then determined based on Gaussian similarity measurement inspired by fuzzy method, thereby increasing the accuracy of fractional matching. The proposed autonomous matching algorithm achieves an accuracy of up to 97.29% when matching 627 2-DE gel test images. In addition to accommodating image rotation, reversals, shape deformation and intensity changes, the proposed algorithm effectively addresses the sub-image mapping problem and was analyzed thoroughly using a large dataset containing 4629 images. The contributions of this work are twofold. First, this work presents a novel MaxRST strategy and autonomous matching method that does not require manual landmark selection. Second, the proposed method, which extends 2-DE gel matching to query sub-image and a database containing large sets of images, can be adopted for mapping and locating, and to compare small gel images with large gel images with robustness and efficiency.

1. Introduction

Two-dimensional electrophoresis (2-DE) is a very important means of analyzing protein characteristics. O'Farrel first introduced 2-DE protein separation in 1975 [28]. A 2-DE gel comprises two modes of electrophoretic separation performed sequentially based on molecular charges as the first direction and molecular weights as the second direction. Proteins are visualized following separation by staining or radiographic labeling [44]. The resultant 2-DE gel image typically contains thousands of protein spots. Comparing two or more gel images is of priority concern. Such analysis can yield considerable information concerning proteins associated with different diseases or for new unidentified proteins [7]. However, comparison by observation is a very labor-intensive task. Manually matching a sample sub-image with a reference gel image using only bare eyes is difficult and time-consuming, especially when either test image may be distorted. Fig. 1 presents a real example.

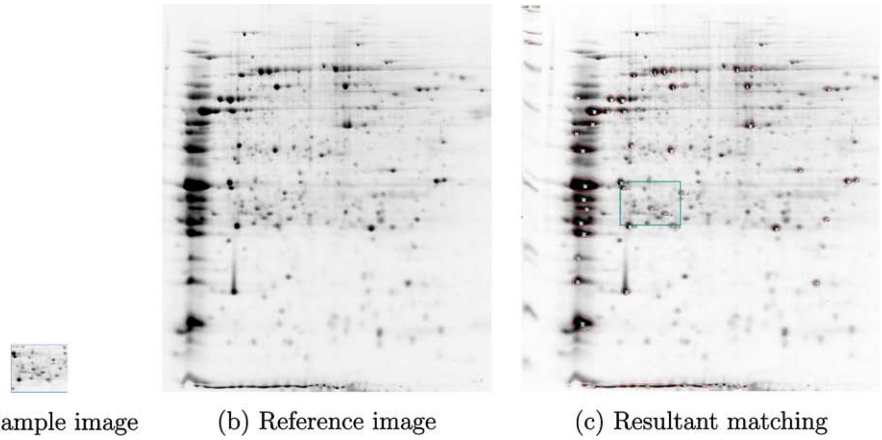
The 2-DE gel image analysis is the first step in a study for exploring differential patterns of qualitative protein expressions [35,41,43]. Moreover, a 2-DE gel image database should contain a large sets of sub-images [4,31]. Matching 2D electropherograms generally creates a bottleneck in automated protein analysis, image distortion and experimental variation, all of which reduce matching accuracy. However, traditional matching methods only compare two complete images, whereas landmark selection and registration procedures are time-consuming. An automated analysis of such images requires an efficient means of comparing multiple and various size of images simultaneously. Although a few commercial software programs exist for analyzing 2D gel images, most packages require intensive user interactions [18]. Thus, the inability to elucidate the role of pattern matching makes efficient matching impossible.

Spot matching, a challenging task in 2-DE gel comparison, has been discussed in numerous studies. Existing matching methods are generally feature-based or image-based approaches [5,19]. Both schemes have advantages and disadvantages. Feature-based methods require extraction of image information and quantification of protein spots in gel images [3,10,14,24,27,29,39]. Spot matching is performed by measuring similarities in corresponding

---

* Tel.: +886 913 061723; fax: +886 226 744448.
E-mail address: dalton@mail.ntpu.edu.tw

0262-8856/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2010.01.004



(a) Sample image (b) Reference image (c) Resultant matching

Fig. 1. Manually matching a sample sub-image: (a) with a reference gel image; (b) using only bare eyes is difficult and time-consuming, especially when either test image may be distorted; (c) resultant matching.

feature data such as spot coordinates, boundaries, areas, and intensities. The effectiveness of this approach requires selecting appropriate landmarks and features. Image-based approaches must generally apply image warping and registration techniques to align two gels according to the intensity distribution in the overall image [12,34,37,38,41]. This method can deal with varying migration positions of spots between gels. However, the succeed of alignment relies on the geometric transformation function and optimization algorithm. The relevant literature includes the following studies:

- *Feature-based methods:* Variations of the feature-based method have been studied extensively. Lemkin and Lipkin [24] considered an electropherogram as a point pattern and matched spots in the GELLAB system. The point matching problem was simplified by partitioning spots into local landmark regions. However, this method requires manual selection of initial landmark spots. Anderson et al. [3] developed a TYCHO system for image analysis of 2D gel. The pattern matching and comparison were performed by adopting coordinate transformation of a matched gel to fit a reference gel with user-defined landmarks. Garrels [10] presented a highly accurate spot-matching method based on graph conversion. The manual intervention of landmark setting was reduced but not eliminated. The Melanie system [27] also used the graph-based approach but compared on-spot clusters instead of individual spots. This approach requires that users allocate landmarks manually. Pánek and Vohradský [29] proposed a point matching algorithm for comparing spot neighborhoods characterized by a syntactic descriptor, which minimized the influence of spot displacements. Matching accuracy was reported as high as 98%. Nevertheless, users must select landmarks that are evenly distributed throughout the gel. Hoffmann et al. [14] introduced a 2D point pattern matching technique that incorporated several features of the incremental Delaunary triangulation of point set. Although this method does not require interactive landmark setting, its dependency on noise and the random incremental mechanism of Delaunary triangulation are problematic. Srinark and Kambhamettu [39] designed a new feature-matching procedure for spot matching by integrating hierarchical- and energy-minimization-based methods. This approach is robust in the presence of image distortion inherent to the electrophoresis process. The energy minimization strategy is based on a greedy search algorithm that may be time-consuming.
- *Image-based methods:* Smilansky [37] compared the intensity of entire 2-DE gel images by using piecewise-bilinear mapping with a grid generated by a Delaunay triangulation. The accuracy of this approach, which is the first raw-image-based registration technology for 2-D gel analysis, depends on the choice of transformation function and landmarks. Veeser et al. [41] utilized piecewise bilinear mapping to represent the associated deformation of multiresolution 2D gel images and measured their similarity by cross-correlation between two intensity distributions. Although this method does not require landmarks or *a priori* alignment, the computational complexity of the algorithm seems high. Gustafsson et al. [12] applied the current leakage correction model to warp the distorted image and then aligned the image by maximizing a penalized likelihood criterion. Matching accuracy was improved by using the proposed warping method for image preprocessing. The drawback of this method is the parameter estimation of current leakage warping function and the map between the gel image and model image. The estimation was done manually by specifying various points on the cathode gel edge and on the gel front. Sorzano et al. [38] presented an elastic registration algorithm by expressing the deformation field as a linear combination of B-splines and using a dissimilarity measure, soft landmark constraint and multiresolution strategy to minimize matching energy function. Although this method can handle varying deformations, its computational time is very high (between 45 s and 150 s) when processing a 512 × 512 pixel image on a Pentium IV PC. Rogers and Graham [34] combined image and feature contexts to develop a gel registration algorithm based on iterated closest point (ICP) matching and distance metric combining Euclidean and shape-context features. The proposed methodology provides robustness against large distortions and presence of outliers.

This work presents a novel and robust Gaussian similarity measure inspired by the Fuzzy method and a Maximum Relation Spanning Tree (MaxRST) strategy, in which an autonomous matching method does not require manual registration or manual selection of landmarks. The 2-DE gel images are represented graphically. Images features are then determined quantitatively regardless of image size, distortion, rotation and variations in brightness. The similarity between a sub-image and full image is then determined, thus increasing the matching accuracy. Extended from our preliminary investigation [26], additional features have been adopted, the MaxRST algorithm has been modified, the global matching method has been modified with normalized feature values, a large dataset containing 4629 2-DE gel images has been created, the merits of combining the Gabriel graph (GG) model and relative neighborhood graph (RNG) model has been justified, and most importantly, matching accuracy has been improved and evaluated