Pattern Recognition Letters 32 (2011) 3–11

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Structural matching of 2D electrophoresis gels using deformed graphs

Alexandre Noma [a,*], Alvaro Pardo [b], Roberto M. Cesar Jr. [a]

[a] Institute of Mathematics and Statistics – IME, University of São Paulo – USP, Rua do Matão, 1010, 05508-090 São Paulo, Brazil
[b] Department of Electrical Engineering, Faculty of Engineering and Technologies, Catholic University of Uruguay, Av. 8 de octubre 2801, CP11600 Montevideo, Uruguay

ARTICLE INFO

Article history:
Available online 1 March 2010

Keywords:
Deformed graphs
Electrophoresis gels
Graph matching
Maximum common subgraph
Point matching
Structural pattern recognition

ABSTRACT

2D electrophoresis is a well-known method for protein separation which is extremely useful in the field of proteomics. Each spot in the image represents a protein accumulation and the goal is to perform a differential analysis between pairs of images to study changes in protein content. It is thus necessary to register two images by finding spot correspondences. Although it may seem a simple task, generally, the manual processing of this kind of images is very cumbersome, especially when strong variations between corresponding sets of spots are expected (e.g. strong non-linear deformations and outliers). In order to solve this problem, this paper proposes a new quadratic assignment formulation together with a correspondence estimation algorithm based on graph matching which takes into account the structural information between the detected spots. Each image is represented by a graph and the task is to find a maximum common subgraph. Successful experimental results using real data are presented, including an extensive comparative performance evaluation with ground-truth data.

1. Introduction

In this paper, the problem of 2D electrophoresis gel matching is addressed. Two-dimensional electrophoresis (2DE) is a well-known method for protein separation which is extremely useful in the field of proteomics. The basic idea is to separate proteins contained in a sample using two independent properties such as isoelectric point and mass. An example of images that are obtained is given in Fig. 1 together with the corresponding spot matching.

Each spot in the image represents a protein accumulation and its size depends on the amount of protein present in the sample. A grayscale is placed on the top of each image to allow grayscale calibration. Although it may seem a simple task, the manual processing of this kind of images is very cumbersome. Furthermore, since gel electrophoresis is generally used to compare samples, several pairs of images must be compared during a single experiment. For this kind of differential analysis, it is necessary to register two images by finding spot correspondences (Fig. 1).

One of the reasons for the popularity of 2DE is its simplicity. As a counterpart, the experimental setting and the materials used do not allow a highly controlled experiment. This means that, in general, strong variations between corresponding sets of spots are expected. All these elements show that, although 2DE gel images may seem simple, the complete task of individual spot matching and gel registration is a complex and time consuming process.

In the present work, we extend the algorithm described in (Noma et al., 2008) for the 2DE gel matching problem. For complex pairs of gels, in order to achieve translation invariance, we apply a strategy based on iterative closest point (ICP) (Besl and McKay, 1992). In the experiments, we illustrate more complex cases than those shown in (Noma et al., 2008), comparing the extended approach against two well-known methods described in the literature, the graduated assignment algorithm, described in (Chui and Rangarajan, 2000), which uses the Euclidean distance, and the bipartite graph matching combined with shape context, proposed in (Belongie et al., 2002), which is well-known in the literature to be very robust to outliers. Both methods are combined with transform estimation (e.g. thin-plate splines) in order to solve the point matching problem. While these two rely on transform estimation, our approach uses deformed graphs to constrain the structural relations between points. All of them are closely related to the ICP, which is attractive because of its simplicity and good convergence properties.

The proposed method is asymptotically faster than the ones proposed in the literature and can interact with the user in order to find the correspondences for complex pairs of gels, taking advantage of previous knowledge provided by the user by selecting or validating a couple of corresponding spots.

The remainder of this paper follows in Section 2 with a brief overview of related works. Section 3 formulates the 2DE gel matching task as a quadratic assignment problem via graph matching,

---

* Corresponding author. Tel.: +55 11 3814 4135.
E-mail addresses: alex.noma@gmail.com (A. Noma), apardo@ucu.edu.uy (A. Pardo), cesar@ime.usp.br (R.M. Cesar Jr.).

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.02.016

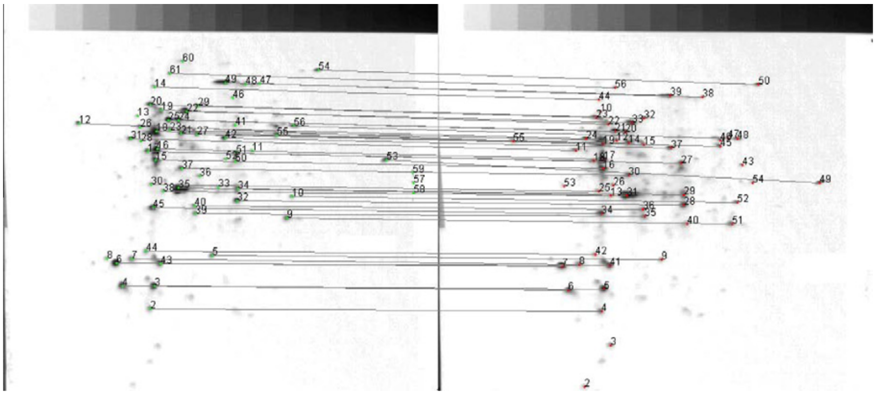

Fig. 1. A correspondence between a pair of gel images.

describing each term of the cost function and the optimization algorithm based on deformed graphs. Experimental results are illustrated in Section 4. Finally, the conclusions and final observations are given in Section 5.

2. Background

In this section, we describe the main classes of methods related to the 2DE gel matching task. This overview does not aim to be exhaustive, but just to exhibit the main trends in current research and to highlight some open issues we would like to address in this paper.

The 2DE gel matching problem is closely related to the 2D point pattern matching problem, in which there is an extensive literature with sophisticated methods, ranging from $O(n^6)$ complexity to $O(n^2)$ algorithms. As pointed out in Rogers and Graham (2007), in practice 3000–4000 spots can be visualized on a single gel image and many recent studies involve differential analysis of sets of up to 100 2DE gels, requiring efficient algorithms for this task. Here, we propose a simple $O(n^2)$ algorithm, exploring the 2DE gel matching problem requirements.

Most of the existing methods for point matching does not take into account the structural information between points to obtain the correspondence. Instead, they usually start by extracting point features (in our case, to represent the spots) which are then used for point and gel matching. In some cases, these point features can be used to establish point correspondence before obtaining the complete gel matching. Existing solutions are based on distances between pair of points (one from each image). For instance, the graduated assignment algorithm, described in Chui and Rangarajan (2000), uses the Euclidean distance. Another example is the bipartite graph matching (BGM) combined with shape context (SC), proposed in Belongie et al. (2002), which is well-known in the literature to be very robust to outliers. Both methods, combined with transform estimation (e.g. thin-plate splines), produce very good matches.

One of the best works, comparing different state-of-the-art methods for the 2DE gel matching problem, is described in Rogers and Graham (2007). The authors present a method to match sets of 2D points using an iterative algorithm that combines point correspondence and transform estimation. In order to establish point correspondences, they separated the procedure into distance computation and correspondence estimation. Finally, based on the point correspondence, they estimate a transformation between both sets of 2D points. These two steps may be iterated to refine the results, following the same strategy of the ICP algorithm. For the experiments, the authors present detailed evaluation for different distances between points and different point correspondence estimations. They evaluate the Euclidean distance and the SC distance for point matching. Regarding the correspondence estimation, they propose several methods: closest point, *k*-closest points, BGM and other graph based ones.

Following this trend, we propose an extension of our previous work described in Noma et al. (2008), in which it was proposed a quadratic assignment formulation, combining SC as the linear term and structural information as the quadratic term of a cost function, minimized by a correspondence algorithm based on deformed graphs in order to solve the gel matching problem. For complex pairs of gels, we combine an ICP strategy, alternating correspondence and translation estimation, exploring previous knowledge provided by the user, which is fundamental for complex cases, by selecting or validating a couple of corresponding spots. For the experiments, we use simulated pairs to evaluate the performance of our approach against well-known algorithms involving Euclidean distance, SC, BGM, transform estimation, and ICP, according to different degrees of degradation. Moreover, original pairs, similar to the ones obtained in practice, are also used to assess the results.

3. Framework based on deformed graphs

In this section, we present our graph matching framework based on deformed graphs applied to the 2DE gel matching problem.

3.1. Gel matching

Matching two gel images should be based on invariant features present on them, such as points representing each spot. The best methods for spot matching are based on point matching techniques. Given two sets of 2D points $x = \{x_1, \ldots, x_N\}$ and $y = \{y_1, \ldots, y_M\}$, and a distance $d_{pq} = d(x_p, y_q)$, the idea is to compute the correspondence between points, minimizing a cost function taking into account the distances between corresponding points.

The algorithm proposed in Almansa et al. (2007) is used for the detection of these points. This algorithm is based on the detection of meaningful spots determined by its contrast and shape. The point descriptor of the spots is the darkest point inside it, i.e. the peak of protein concentration. Fig. 2 illustrates the result of the meaningful boundaries detection algorithm applied to a real gel image.

3.2. Graph matching

We formulate the 2DE gel matching task as a quadratic assignment problem via graph matching. The linear term evaluates the 'appearance' while the quadratic term evaluates the 'structure'. Although graph matching is known to be computationally expen-