2011 International Conference on Environment and BioScience
IPCBEE vol.21 (2011) © (2011) IACSIT Press, Singapore

2-D Electrophoresis Gel Matching Using Proximity Graphs

Faroq AL-Tam[+], António dos Anjos and Hamid R. Shahbazkia

D. E . Informatics Engineering Faculty of Sciences and Technology
University of Algarve - Portugal

Abstract. This work presents a method for matching 2-D Electrophoresis (2-DE) gel images. A simple registration algorithm based on automatic detection of landmarks and the Thin Plate Spline (TPS) is proposed. The matching method defines a scoring technique based on spots neighbors' similarity. Proximity graphs, Relative Neighborhood Graph (RNG) and Gabriel Graph (GG) are used to represent the gels. The final matching is carried out using a Bipartite Graph Matching (BGM) method. Experimental results show a good performance when comparing to state-of-the-art methods.

Keywords: matching, 2-D electrophoresis, registration, RNG, GG, BGM.

1. Introduction

In bioinformatics, Two-Dimensional Gel Electrophoresis (2-DE) is an electrochemical process used in Proteomics to display the protein expression of a sample under analysis. It consists essentially of two steps. First, iso-electric focusing is applied on a strip of gel, separating the proteins in the sample by electrical charge. In the second step, the gel strip is placed on a wider gel, and proteins are further separated by molecular weight [1]. After staining, the result is a transparent gel with as many dark spots as the number of detected proteins. The resulting gel can be compared with another gel in order to find differences between the respective protein expressions. These differences may indicate a response to a drug treatment, a disease state or any other biological disfunction. This kind of analysis is known as differential analysis. This work starts by a brief review of the related work in Section 2. Registration and matching methods are described in Section 3. Experimental results are discussed in Section 4.

2. Related Work

2.1. Matching Using Proximity Graphs

Proximity graphs are widely used in point pattern matching. In [2] an incremental Delaunay Network (DN) is used. Spots intensities are ranked in a discrete way and then, an algorithm uses them to create the network. It shares the same idea used in [3] to solve a point pattern matching problem in Astronautics. In [4] an RNG graph is used with a DN to find the best matches for DNA spots. RNG is created for the reference image and DN is created for the sample image. A breadth-first search algorithm is then used to find the best match for the RNG of the reference image in the DN of the sample. In [5] a sub-image matching algorithm using RNG and GG graphs is proposed. The features are extracted from the graphs then a Gaussian-like measure using fuzzy logic is used to find the matches. The algorithm does not consider any additional features that can be collected from the spots which can increase the matching process accuracy. Inexact graph matching algorithm is used in [6]. The reference (model) and the input (sample) images are segmented using [7]. The neighbours of a given point are represented using the Shape Context (SC) [8] metric. The

[+] Corresponding author. Tel.: (+351939014047);
E-mail address: faroq.al.tam@gmail.com.

structural information of the edges is used besides the SC. Furthermore, a deformed graph is used to evaluate existing non-linear deformations.

2.2. Point Pattern Matching

Here some related work in point pattern matching is highlighted. In [9] a relative shape context is proposed and relaxation labelling is used to find the final point matching. A similar method is also found in [10]. In [11] a graph representation is combined with the SC metric. Although this method shows a good performance when compared to some state-of-the-art methods like Robust Point Matching (RPM) [12], there is no guarantee to have an optimal assignment solution when notable deformations and numerous outliers exist. In [13] a similar approach is used but a weighed sum of points' features, preserving the topology of a given point set, is proposed. In practice, applying these methods to match 2-DE gel images is not a good choice, but they can help in the final stages or when combined with other features from the gel spots (i.e. not only the position and the geometrical features).

3. The Proposed Method

The proposed method is based on spot information and RNG as well as GG graph properties. Information including the coordinates and the Volume of each spot in the static and deformed images are extracted using [14]. From now on, we call the static image S, and the deformed image D. The matching consists of two steps. The first corrects the geometrical distortions of D using automatically detected landmarks and TPS [15] registration. The second includes the representation of spots neighbours, and matching using a BGM method.

3.1. Step1: Gel Registration

In this step, a set of landmarks (control points) is detected in S and D, and then TPS-based warping is applied to D. TPS is sensitive to mismatches but it gives an elastic warp using only a few landmarks. Keeping this in mind, we set some constraints for the candidate spots to be considered as landmarks. Before explaining the constraints used, we first explain the features considered. Given two gels represented by two sets of vertices $G_S = \{u_1(x,y,I),\dots,u_n(x,y,I)\}$ and $G_D = \{v_1(x,y,I),\dots,v_m(x,y,I)\}$, where u_1 is the first spot in the static gel G_S, and the same for v_1 and the deformed gel G_D. Components x and y represent the coordinates of the centre of the spot, and $I \in [0,1]$ is the Normalized Volume (NV) of the spot. n and m are the number of spots in G_S and G_D respectively. We would like to note that the terms point, spot and vertex are used interchangeably. Volume Normalization is also used in [2] but in a discrete way. Having two RNG graphs RNG_S and RNG_D, their vertices G_S and G_D respectively, consider the following features for each vertex in each graph:

- Degree: the number of edges connected to a vertex $R(.)$.
- Normalized Volume $I(.)$.
- Vertex position $\Gamma(.)$.
- Relation with neighbours' I values.

Degree is also used in [5]. We want to highlight here that, the RNG graph is insensitive to transformations, so it yields good stability when considering the degree as a feature. Considering these features and given two vertices u_i in RNG_S and v_j in RNG_D they are considered as landmarks if they have:

- $|R(u_i) - R(v_j)| \leq \lambda$
- $I(u_i) \geq \rho \ and \ I(v_j) \geq \rho$
- $|I(u_i) - I(v_j)| \leq \varepsilon$
- $\sqrt{(\Gamma(u_i) - \Gamma(v_j))^2} < \sqrt{(\Gamma(u_k) - \Gamma(v_j))^2} \ \ \forall u_k \in RNG_S \ and \ i \neq k$
- $I(u_i) > I(u_k) \ and \ I(v_j) > I(v_l) \ \ \forall u_k \in RNG_S$ and u_k is adjacent to u_i, and $\forall v_l \in RNG_D$ and v_l is adjacent to v_j.

Where λ, ρ and ε are parameters. All points that verify all of these conditions are used as landmarks for the TPS. The first constraint measures how similar two given vertices are in number of neighbours. The second one shrinks the search space to spots that only have I greater than a given threshold. The third constraint defines the maximum difference between two vertices considering the size and intensity (i.e. I). It is unlikely that two corresponding spots have a big I difference. The fourth constraint measures the Euclidean distance between two given spots. The last constraint is very important because it only accepts spots that are *master* to be selected. The idea of "master spot" is illustrated in Fig. 1. In practice, we found it hard to set up ρ and ε to fixed constants due to the variations from one image to another. Therefore, we empirically set $\rho \in [0.1,0.6]$, $\varepsilon \in [0,0.3]$ and propose a greedy algorithm (Fig. 2) to find the best parameters.

Fig. 1: Master Spot. Left: spots identified, g and r are the master spots. Right: Spots' I values.

Algorithm 1: FindBestParameters(S, D, RNG_S, RNG_D)

$\varepsilon \leftarrow 0.3$
$best_\rho \leftarrow 0$
$best_\varepsilon \leftarrow \varepsilon$
$max_{NMI} \leftarrow 0$
$C_{NMI} \leftarrow 0$
while $\varepsilon > 0$
  do
    $\rho \leftarrow 0.6$
    while $\rho \geq 0.1$
      do
        $landmarks \leftarrow findmatches(RNG_S, RNG_D, \rho, \varepsilon)$
        if $length(landmarks) > 2$
          $D_w \leftarrow TPS(D, landmarks)$
          $C_{NMI} \leftarrow NMI(D_w, S)$
          if$(max_{NMI} \leq C_{NMI})$
            then
              $max_{NMI} \leftarrow C_{NMI}$
              $best_\varepsilon \leftarrow \varepsilon$
              $best_\rho \leftarrow \rho$
        $\rho \leftarrow \rho - 0.05$
    $\varepsilon \leftarrow \varepsilon - 0.05$
return$(best_\varepsilon, best_\rho)$

Fig. 2: Finding the best parameters algorithm

As for parameter λ, we set it equal to 1 because it is related to the number of neighbours of a given vertex. In Fig. 2, $findmatches(RNG_S, RNG_D, \rho, \varepsilon)$ function takes two RNG graphs, current ρ and ε, applies the mentioned constraints and returns two point sets comprising the landmarks. $findmatches(.)$ iterates and tests all vertices in RNG_S over all vertices of RNG_D, which obey the constraints, using the current parameters. $TPS(D, landmarks)$ uses a TPS to warp the deformed image D using the selected landmarks, and returns a registered version D_w. D_w is compared to S using the Normalized Mutual Information (NMI) measure [16] and the similarity value is stored in C_{NMI}. The number of bins used for the NMI histogram is 256.

3.2. Step2: Gel Matching

We denote RNG_S for Relative Neighbourhood graph for the static gel, and GG_D for Gabriel graph for the deformed gel. GG and RNG define the neighbourhood of a vertex by the distance to other vertices.

Furthermore, RNG is a sub-graph of GG having less (or equal) number of neighbours than GG, for a given vertex. Besides graph representation, we define these features: Θ: angle between a point and the center-of-mass point of the gel; Ω: distance between a spot and the center-of-mass point of the gel; V: spot volume; Γ: spot position in the gel. Given two graphs RNG_S and GG_D (see Fig. 3g and Fig. 3h) and two vertices $i \in RNG_S$ and $j \in GG_D$. We define a cost function for each feature as:

$$\mathrm{C}_{\Psi}(i,j) = \Delta_{\Psi}(\mathrm{i},\mathrm{j})^{\log\left(\sqrt{10}+\sum_{\mathrm{k}\in \mathrm{i_N}} \min_{\mathrm{l}\in \mathrm{j_N}} \Delta\Psi(\mathrm{k},\mathrm{l})\right)} \quad (1)$$

Where Δ_{Ψ} is the Squared Difference between two given spots regarding feature Ψ, where Ψ can be any one of Θ, Ω, V or Γ. i_N and j_N refer to all neighbours of vertices i and j respectively. The constant $\sqrt{10}$ in Eq. (1) makes the power equal to 0.5 in the best case, i.e. $\log\left(\sqrt{10}+0\right) = 0.5$. The best case scenario is when the features of i_N and j_N are exactly the same. In Eq. (1) we reward and penalize vertices i and j according to their neighbours' similarity. In order to combine all the features, we first normalize them and weight them in a single multi-objective function:

$$C(i,j) = a_1 \times \mathrm{C}_{\Theta}(\mathrm{i},\mathrm{j}) + a_2 \times \mathrm{C}_{\Omega}(\mathrm{i},\mathrm{j}) + a_3 \times \mathrm{C}_{\mathrm{V}}(\mathrm{i},\mathrm{j}) + a_4 \times \mathrm{C}_{\Gamma}(\mathrm{i},\mathrm{j}) \quad (2)$$

Where as are the weights. We set a_1, a_2 *and* a_4 equal to 1 and a_3 equal to 0.2. For finding the optimal matching and the one-to-one mapping, a BGM (The Hungarian) algorithm is used. To make gels of G_S and G_D equal in number of spots, dummy spots are appended to the gel that has the lower number of spots. Finally, Eq. (2) is applied to all possible permutations, resulting in a square matrix to be used by BGM, and the optimal assignment is to find the permutation $P(i)$ that minimizes the following function [8]:

$$\sum_{i} C_{i,P(i)} \quad (3)$$

Where $C_{i,P(i)}$ is the cost value of matching spot i to the correspondent spot $P(i)$, and is calculated using Eq. (2). $P(i)$ is the permutation for vertex i (i.e. j). The cost of matching a $dummy$ spot is ∞, so the outliers are supposed to match those dummy spots. After the initial matching, to reduce the number of false matches, we propose the following iterative post-processing procedure:

1- Set a threshold (0.002 after the coordinates of the spots are normalized) to reject the matches that are too far, and use the remaining as landmarks.

2- Transform GG_D's vertices by using TPS and the landmarks from 1.

3- Calculate Eq. (2) for all spots in RNG_S and GG_D and apply the BGM as mentioned above.

4- Go to 1 and repeat until the matches become stable (i.e. no change in number of matched spots and their positions between iteration n and iteration $n-1$).

In practice, after a few number of iterations (in this work ≤ 20 iterations) the results improve a lot. This iterative procedure is similar to the one used in some point pattern matching methods like [11].

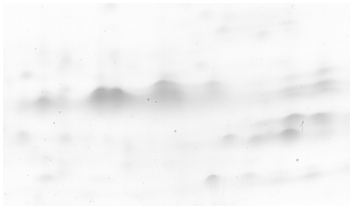
4. Results

We run the proposed method on 8 different pairs of gel images and compare it to SC [8] and RPM [12] methods. Table 1 shows the results in terms of true and false matches.

Table 1 Expermental results: S: Static, D: Deformed, T: Ture and F: False matches.

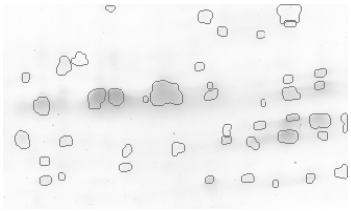
	Detected		Proposed				RPM				Shape Context			
#	S	D	T	T%	F	F%	T	T%	F	F%	T	T%	F	F%
1	79	114	67	**84.81%**	3	**3.80%**	60	75.95%	15	18.99%	56	70.89%	18	22.78%
2	44	49	34	**77.27%**	0	**0.00%**	22	50.00%	16	36.36%	16	36.36%	10	22.73%
3	341	347	295	**86.51%**	5	**1.47%**	292	85.63%	43	12.61%	242	70.97%	71	20.82%
4	231	233	215	**93.07%**	1	**0.43%**	215	**93.07%**	14	6.06%	189	81.82%	35	15.15%
5	160	278	133	**83.13%**	1	**0.63%**	126	78.75%	17	10.63%	103	64.38%	43	26.88%
6	160	260	141	88.13%	3	**1.88%**	149	**93.13%**	11	6.88%	116	72.50%	30	18.75%
7	160	207	123	76.88%	9	**5.63%**	129	**80.63%**	31	19.38%	102	63.75%	40	25.00%
8	160	185	68	42.50%	17	**10.63%**	79	**49.38%**	56	35.00%	78	48.75%	41	25.63%
	1335	1673	1076	**80.60%**	39	**2.92%**	1072	80.30%	203	15.21%	902	67.57%	288	21.57%

Although RPM and the proposed method are equivalent in terms of true matches, our method outperforms the others in terms of false matches' rate. The one-to-one matching for RPM in Table 1 is taken by applying the Hungarian algorithm after the RPM has finished. A notable problem in SC is the swapped-match, in which two points are "correct" but in swapped positions. An example of gel matching for our method is depicted in Fig. 3.

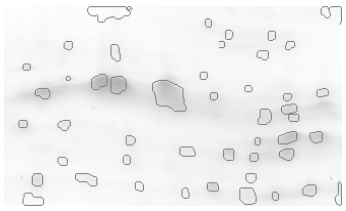


a- Static image S.

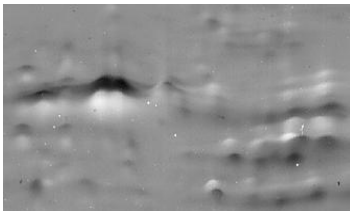
b- Deformed image D.



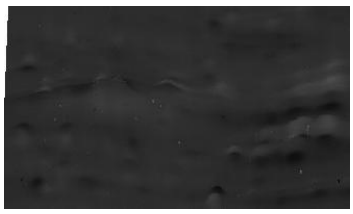
c- Spots detected on S.



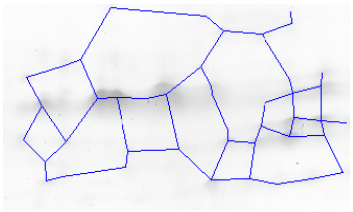
d- Spots detected on D.



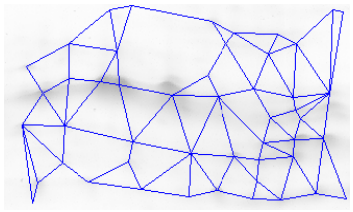
e- Difference between S and D images.



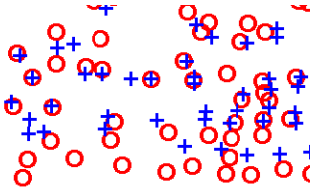
f- Difference between S and D after registration.



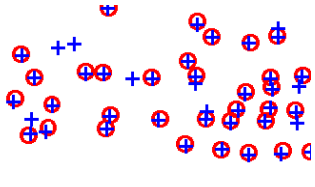
g- RNG graph of S.



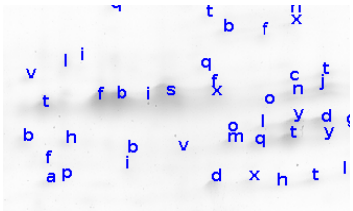
h- GG graph of D.



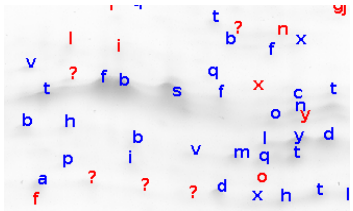
i- Initial matching between S and D.



j- Matching after 5 iterations of post-processing.



k- Final matching: Spots on S.



l- Matched spots on D.
Blue: Accepted. Red: Rejected

Fig. 3: Matching results of the proposed method for gel pair No. 2 in Table 1.

5. Conclusion

A method for matching 2-DE electrophoresis images has been presented. The distortions are corrected using a landmark-based method, and the final matching is done by using the RNG and GG graphs preserving the neighbourhood of a given point. A scoring method is used to reward or penalize two points according to their neighbour's similarity. Furthermore, an iterative post-processing technique is used based on TPS to reduce false matches. The method is compared to two widely-known robust methods for point matching. The results show a very good rate of true and false matches as well. Despite the proposed and RPM methods are equivalent in what concerns matching rate, manual intervention to correct the false matches will be less when using our method.

6. References

[1] P.H. O'Farrell, "High resolution two-dimensional electrophoresis of proteins.," *The Journal of biological chemistry*, vol. 250, May. 1975, pp. 4007–4021.

[2] F. Hoffmann, K. Kriegel, and C. Wenk, "An applied point pattern matching problem: comparing 2D patterns of protein spots," *Discrete Applied Mathematics*, vol. 93, 1999, pp. 75 - 88.

[3] G. Weber, L. Knipping, and H. Alt, "An Application of Point Pattern Matching in Astronautics," *Journal of Symbolic Computation*, vol. 17, 1994, pp. 321 - 340.

[4] Y. Watanabe, K. Takahashi, and M. Nakazawa, "Automated detection and matching of spots in autoradiogram images of two-dimensional electrophoresis for high-speed genome scanning," *Image Processing, 1997. Proceedings., International Conference on*, 1997, pp. 496 -499 vol.3.

[5] D.-T. Lin, "Autonomous sub-image matching for two-dimensional electrophoresis gels using MaxRST algorithm," *Image and Vision Computing*, vol. 28, 2010, pp. 1267 - 1279.

[6] A. Noma, A. Pardo, and R.M.C. Jr, "Structural matching of 2D electrophoresis gels using deformed graphs," *Pattern Recognition Letters*, vol. 32, 2011, pp. 3 - 11.

[7] A. Almansa, M. Gerschuni, A. Pardo, and J. Preciozzi, "Processing of 2D Electrophoresis Gels," *2007 ICCV International Workshop on Computer Vision for Developing Regions*, 2007.

[8] S. Belongie and J. Malik, "Matching with shape contexts," *Proc. IEEE Workshop Content-based Access of Image and Video Libraries*, 2000, pp. 20–26.

[9] J. Zhao, S. Zhou, J. Sun, and Z. Li, "Point pattern matching using Relative Shape Context and relaxation labeling," *Proc. 2nd Int Advanced Computer Control (ICACC) Conf*, 2010, pp. 516–520.

[10] J. Zhao, J. Sun, S. Zhou, Z. Li, and M. Chen, "Inexact point pattern matching algorithm based on Relative Shape Context and probabilistic relaxation labelling," *Proc. 3rd Int Computer Research and Development (ICCRD) Conf*, 2011, pp. 508–512.

[11] Y. Zheng and D. Doermann, "Robust point matching for nonrigid shapes by preserving local neighborhood structures.," *IEEE Trans Pattern Anal Mach Intell*, vol. 28, Apr. 2006, pp. 643–649.

[12] H. Chui and A. Rangarajan, "A new algorithm for non-rigid point matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2000, pp. 44–51.

[13] J.-H. Lee and C.-H. Won, "Topology preserving relaxation labeling for nonrigid point matching.," *IEEE Trans Pattern Anal Mach Intell*, vol. 33, Feb. 2011, pp. 427–432.

[14] A. dos Anjos, A.L.B. Møller, B.K. Ersbøll, C. Finnie, and H.R. Shahbazkia, "New approach for segmentation and quantification of two-dimensional gel electrophoresis images.," *Bioinformatics*, vol. 27, Feb. 2011, pp. 368–375.

[15] F.L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, 2002, pp. 567–585.

[16] C. Studholme, D.L.G. Hill, and D.J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, vol. 32, 1999, pp. 71 - 86.