A Spot Matching Method using Topological Patterns of Neighbor Spots in 2-DE

Chan-Myeong Han[1], Dae-Seong Jeoune[2], Hwi-Won Kim[3], Young-Woo Yoon[1]

[1] Department of Computer Engineering, Yeungnam University
#280 Daehak-ro, Gyeongsan, Gyeongbuk, 712-749, Republic of Korea
cmhan@yu.ac.kr, ywyoon@yu.ac.kr

[2] Department of Media Design, Daegu Future Colleage
#114 Mirae-ro, Gyeongsan, Gyeongbuk, 712-716, Republic of Korea
dsjeoune@dmc.ac.kr

[3] Education Center for IT, Kyeongbuk College
#77 Daehak-ro, Yeungju, Gyeongbuk, 750-712, Republic of Korea
khw@kbc.ac.kr

Abstract. The proposed method compares two topological patterns from two central spots to be matched. Similarity transform is applied to one of two patterns in order to correct global and local distortion before comparing topological patterns. And then, matching between neighbor spots from two patterns is performed and similarity is evaluated using normalized Hausdorff distance. Final matching of spots is determined by the number of neighboring matched pairs, the number of outlier spots and the normalized Hausdorff distance. The proposed method shows good results in case where relative topology is well preserved and there are no outliers.

Keywords: spot matching, topology, neighbor spot, Hausdorff distance

1   Introduction

Proteomics is the large-scale study of learning functions of proteins and the very basic process is to identify proteins included in cells. Two-dimensional electrophoresis (2-DE) is the most frequently used in Proteomics [1-3]. As for research on proteins, spot matching is a main bottleneck. So, the implementation of fast and precise spot matching algorithm with no intervention is the most essential part [4]. However, it is impossible to obtain the same gel image each time due to many experimental parameters even if the same sample is used for a couple of electrophoresis experiments. It makes spot matching more difficult.

This paper proposes and evaluates a novel method which mimics human's recognition process and treats spot matching as point pattern problems and graph problems. The proposed method determines matched pairs by comparing topological patterns of two central spots. The definition of neighbor spots and a method for calculating similarity of topological patterns are proposed to solve this problem.

2   Previous Research

The typical method of spot matching in 2-DE gel image is a method by landmarks which are manually defined. Spots around landmarks are matched in turn. Initial matching is performed with landmarks and next matching is performed with best matching of neighbor spots [5-6]. A majority of conventional software use landmarks manually defined. Nevertheless, a process of manually defining landmarks has high error rates. The iterative closest point method as an automated approach has been proposed, in which spot matching is performed according to distances between matched pairs from two sets of spots and parameters of non-linear transformation are acquired [7].

A method by hierarchical structure and minimization of energy has been proposed, which is an integration of the hierarchical-based and optimization-based methods [8]. Similarly some studies propose matching methods motivated by the preservation of topology. As a measure to compare similarity of topology patterns, distances and angles among neighbor spots are compared [9].

3   Proposed Spot Matching Scheme

3.1   Spot Matching using Topological Patterns

The definition of graph is tightly related to neighbor spots and affects the performance of spot matching. In this paper, 5-NNG is used to define neighbor spots based on the previous experiment A spot matching method by topological patterns of neighbor spots estimates correspondence with similarity of patterns of neighbor spots $N_{graph}(v)$ located around the central spots. If central spots, p_i and q_j from reference gel and target gel are given, two sets of neighbor spots, $N_{5\text{-}NNG}(p_i)$ and $N_{5\text{-}NNG}(q_j)$ are extracted. And then, similarity between two patterns is compared to decide whether two central spots are best matched or not.

For correspondence estimation, topological patterns for two central spots are compared in their best state after moving spots to their best positions. A series of transformation should be applied before comparison of topological patterns. Transform involved with three parameters of transposition, rotation and scale is called "similarity transform". After similarity transform, the deformed pattern becomes exactly the same with the original one and then, it is needed to confirm two pairs of spot matching through parameter evaluation.

Central spot and one of neighbor spots can be pivot pair, which is exactly superimposed and other spots moved around according to parameters derived from central pair and pivot pair. The problem is that these neighbor pairs are not known until pattern comparison ends. All possible combinations of pairs should be considered and the final pivot pair is derived from a neighbor pair according to similarity. Equation (1) represents how to calculate new coordinates of neighbor spots of q_j with scale parameters s and rotation parameters θ. The new coordinates make neighbor spots of q_j move to their best positions to be matched in similar geometric conditions.

$$\begin{bmatrix} x'_{qj} \\ y'_{qj} \end{bmatrix} = s \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_{qj} - x_{q(center)} \\ y_{qj} - y_{q(center)} \end{bmatrix} + \begin{bmatrix} x_{p(center)} \\ y_{p(center)} \end{bmatrix} \qquad (1)$$

$$where, \quad sacle \quad s = \frac{\sqrt{(x_{p(center)} - x_{p(pivot)})^2 + (y_{p(center)} - y_{p(pivot)})^2}}{\sqrt{(x_{q(center)} - x_{q(pivot)})^2 + (y_{q(center)} - y_{q(pivot)})^2}}$$

$$rotation \quad \theta = a\tan\left(\frac{y_{q(pivot)} - y_{q(center)}}{x_{q(pivot)} - x_{q(center)}}\right) - a\tan\left(\frac{y_{p(pivot)} - y_{p(center)}}{x_{p(pivot)} - x_{p(center)}}\right)$$

3.2 Neighbor Spot Matching and Similarity Measure

The matched pairs of neighbor spots between $N_{5\text{-}NNG}(p_{center})$ and $N_{5\text{-}NNG}(q_{center})$ must be determined before similarity of two patterns is calculated for matched pairs that are used in evaluation of similarity. Distances for all possible combinations of neighbor spots are calculated and one-to-one matched pairs are formed in the order of shorter distance. As for the matched neighbor pairs, Euclidean distances of two spots are the shortest. Two spots are not accepted as a matched pair if distance between them is greater than the threshold.

Similarity can be calculated using the matched neighbor matched pairs. Hausdorff distance is the most popular when getting similarity of patterns as shown in equation (2), where p_i and q_j are spots from patterns of p_c and q_c, respectively, and $d(p_i, q_j)$ is Euclidean distance function between the points.

$$h(P,Q) = \max_{p_i \in N_{5-NNG}(p_c)}\left(\min_{q_j \in N_{5-NNG}(q_c)}(d(p_i, q_j))\right) \qquad (2)$$

Hausdorff distance does not consider a scale parameter. So it must be normalized by dividing with pivot distance, where pivot distance(pd) is a distance between central spot p_c and pivot spot. The normalized Hausdorff distance(NHD) is defined as equation (3). NHD is not enough as a criterion for decision of matching. In this paper, three criteria with priority are introduced for better matching result; (1)the more number of neighbor matched pairs, (2)the fewer number of outlier spots and (3)NHD.

$$NHD = h(p_i, q_j) / pd \qquad (3)$$

4 Experiment and Result

Experiment using data set called "Human leukemias" from the web site (*www.lecb.ncnifcrf.gov/2DgelDataSets*) is performed. It has 128 pairs of gels and each gel has approximately 22 manually matched pairs of spots. The file, "landmark.tbl" has information in one piece on 128 pairs of gels and it is separated into 128 files. Each gel pair has one-to-one matched pairs and there is no outlier. The program language *perl* is used to implement the proposed algorithm and *python* with turtle graph library is used to visualize the matching results.

Table 1. Result of Experiment

Measures	Values
Total number of gel pairs	128
Total number of spot pairs	2,763
Number of detected spot pairs	2,716
Detection rate (%)	98.2
Matching accuracy (%)	99.3

5   Conclusion

In this paper, a new spot matching algorithm for 2-DE is proposed based on neighbor spot and the similarity of topological patterns. It presents neighbor spots as a graph and 5-NNG was proved to be the most proper for spot matching. The proposed method can robustly match all of spots under large global/local distortion. It shows excellent results especially under circumstances where there are no outliers and the topology of spots are well preserved.

Future research is required to solve problems with outliers. The proposed algorithm can be used with slight change by inserting virtual missing spots corresponding to outliers. And, the proposed method can be enhanced to be applied to the problems with outliers using various graphs and new comparing methods of similarity.

References

1. J. L. Harry, M. R. Wilkins, B. R. Herbert : Proteomics: Capacity versus Utility, Electrophoresis, vol. 21, pp. 1071-1081 (2000)

2. P. H. O′Farell : High Resolution Two-dimensional Electrophoresis of Proteins, Biological Chemistry, vol. 250, pp. 4007-4021 (1975)

3. R. Westermeier : Electrophoresis in Practice: A Guide to Theory and Practice (1993)

4. M. Daszykowski, E. Mosleth Faeregestad, H. Grove, H. Martens, B. Walczak : Matching 2D Gel Electrophoresis Images with Matlab 'Image Processing Toolbox', Chemometrics and Intelligent Laboratory Systems, vol. 96, pp.188-195 (2009)

5. S. Veeser, M. J. Dunn, G. Yang : Multiresolution Image Registration for Two-dimensional Gel Electrophoresis, Proteomics, vol. 1, pp.856-870 (2001)

6. J. Salmi, T. Aittokallio, J. Westerholm, M. Griese, A. Rosengren, T. A. Nyman, R. Lahesmaa, O. Nevalainen : Hierarchical Grid Transformation for Image Warping in the Analysis of Two Dimensional Electrophoresis Gels, Proteomics, vol. 2, pp. 1504-1515 (2002)

7. Mike Rogers, Jim Graham : Robust and Accurate Registration of 2-D Electrophoresis Gels Using Point-Matching, IEEE Transactions on Image Processing, vol. 16, no. 3, pp. 624-635, (2007)

8. Thitiwan Srinark, Chandra Kambhamettu : An Image Analysis Suite for Spot Detection and Spot Matching in Two-dimensional Electrophoresis Gels, Electrophoresis, vol. 29, pp. 706-715 (2008)

9. Ankhbayar Hukhuu, Jeong-Bae Lee, Young-Sup Hwang : Automatic Matching of Protein Spots by Reflecting Their Topology (in Korean)., The KIPS Transactions: Part B, vol. 17-B, no. 1, pp. 79-84 (2010)