International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 14 (2017) pp. 4320-4328

How to Determine Spot Matching Results by Probabilistic Reliability using Homogeneous Multiple Graphs in Two-Dimensional Polyacrylamide Gel Electrophoresis

Dae-Seong Jeoune
Associate Professor,
Dep't of Drone Photo & Videography, Daegu Mirae College
Gyeongsan, Gyeongbuk, Republic of Korea.
ORCID: 0000-0003-1669-5700

Chan-Myeong Han
CEO,
EO3 Co. Ltd.
Daegu Metropolitan City, Republic of Korea.
ORCID: 0000-0002-9492-6685

Yun-Kyoo Ryoo
Associate Professor,
Dep't of Medical Computer Science, Daegu Health College
Daegu Metropolitan City, Republic of Korea.
ORCID: 0000-0001-7846-2182

Jae Hak Sa
Director,
MIS Division, CRETEC
Daegu Metropolitan City, Republic of Korea.
ORCID: 0000-0003-4550-8417

Beomcheol Lee
Dep't of Computer Engineering
Yeungnam University
Gyeongsan, Gyeongbuk, Republic of Korea.

Wook Hyun Kim*
Professor,
Dep't of Computer Engineering, Yeungnam University
Gyeongsan, Gyeongbuk, Republic of Korea.

Abstract

In proteomics, two-dimensional polyacrylamide gel electrophoresis, namely 2D-PAGE or 2-DE, is widely used to separate and identify proteins, producing gel images that consist of a lot of spots. They often introduce automated algorithm in spot matching to identify proteins from images in order to overcome time and cost constrains. In this paper, we propose a straightforward method how to determine the spot matching results using probabilistic reliability in the final stage of experiment. In the algorithm, homogeneous multiple graphs of k-NNG are used and the results from each iteration by a single graph are summed together to produce reliability of each spot pair in the form of probability. Experiment has performed to confirm the effectiveness of the proposed algorithm. Consequently, the proposed method can be a good guide even though the final decision depends on medical scientists.

Keywords: 2D-PAGE, Spot Matching, Grassfire Algorithm, Homogeneous Multiple Graphs, k-NNG, Probabilistic Reliability

INTRODUCTION

The two-dimensional polyacrylamide gel electrophoresis, namely 2D-PAGE or 2-DE, is one of the widely used methods to analyze proteins in the field of *proteomics*. By the processes of IEF and SDS-PAGE, proteins in cellular tissue can be separated in a gel in the form of tens to hundreds of spots[1-3]. And then medical scientist(s) may identify proteins using gel image, causing huge amount of time, cost and efforts in manual process. Actually, it is very hard to carry out this process in manual because hundreds or thousands of proteins are contained in a gel. Therefore, it is necessary to automate the analytical process of 2D-PAGE[3,4].

So far, many algorithms have been proposed to make the spot matching process automated because it can significantly reduce time and cost in the most important identification process in 2D-PAGE. Most spot matching algorithms produce good result over than 99% in detection rate and accuracy. But no one can guarantee the matched pairs does not contain false-positive pair(s).

There have been proposed various matching algorithms in 2D-PAGE that take advantage of features such as landmark protein, graph, iterative closest point, neighborhood descriptor, etc.[5]. Nevertheless, most of research related to spot matching devotes itself to enhancing detection rate and matching accuracy. And some take into account outlier or missing spots in a gel. It is important that spot matching results does not contain false-positive ones. In recent, the research on the individual reliability measure for every matched pairs has been introduced [6].

In this paper, therefore, we propose a method to measure and assess of each matched pair from the automated spot matching algorithm in 2D-PAGE in terms of probabilistic reliability. The proposed method adopts the grassfire spot matching algorithm and homogeneous multiple graphs. Also, the experimental results with probabilities of matching pairs

are presented, which provides useful information to medical scientists for decision making in the protein identification phase.

DEFINITION AND MATERIAL

The definition of spot matching in 2D-PAGE using mathematical notion is described. And then, spot matching algorithm using topological patterns including the grassfire method are presented.

A. Definition of Spot Matching in 2D-PAGE

The typical process to analyze proteins in a tissue by medical scientists is so much time and cost consuming, which is shown in the left part in the Figure 1. As mentioned above, a number of proteins in a gel ranging from hundreds to thousands require automated way to identify them to significantly reduce efforts by manual tasks of medical scientists. It is possible to automate some process using the spot matching algorithms developed by computer scientists that utilize gel images, shown in the Figure 1.

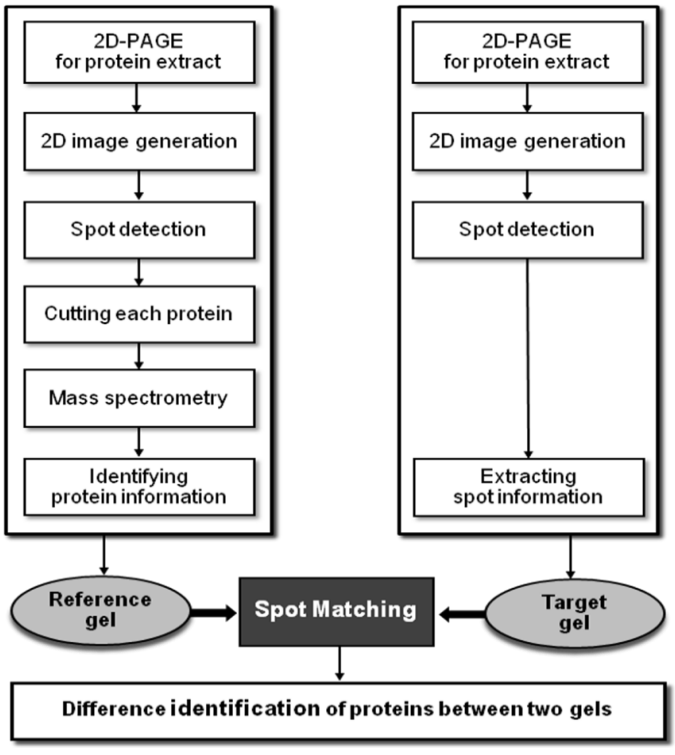
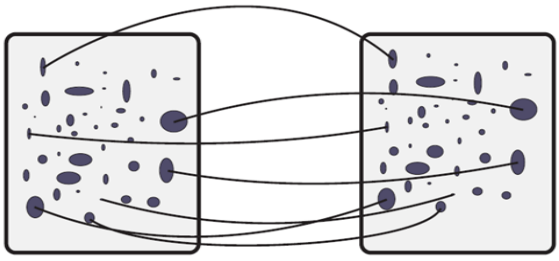


Figure 1. Spot Matching in 2D-PAGE

The definition of spot matching in 2D-PAGE is to identify the same proteins between two gel images: reference and target gels[6,7]. The concept of spot matching is depicted in the Figure 2.



(a) Reference gel ***P*** (b) Target gel ***Q***

Figure 2. Concept of Spot Matching in 2D-PAGE

It is assumed that a given reference gel and a target gel are 2-dimensional sets P and Q consisting of spots, respectively. Each of the center positions of spots in their images can be represented as coordinates $p_i=(x_i, y_i)$ in reference gel and $q_j=(x_j, y_j)$ in target gel, as shown in the equation (1) [6,8].

$$\boldsymbol{P} = \{p_1, p_2, \ldots, p_m\}, \text{ where } p_i = (x_i, y_i),\ 1 \le i \le m$$

$$\boldsymbol{Q} = \{q_1, q_2, \ldots, q_n\}, \text{ where } q_j = (x_j, y_j),\ 1 \le j \le n \tag{1}$$

Now spot matching problem with respect to each spots in two relevant 2D-PAGE images is to determine the maximum set M consisting of matching pairs between P and Q that satisfies the conditions given in the equation (2).

$$\boldsymbol{M} = \{(p_{i1}, q_{j1}), (p_{i2}, q_{j2}), \ldots, (p_{il}, q_{jl})\},$$

$$\text{where } p_{il} \in \boldsymbol{P},\ q_{jl} \in \boldsymbol{Q} \text{ and } l \le \min(m, n) \tag{2}$$

There might be missing spots in some cases. The word missing means that a spot really exists but it is not detected in the process of spot detection for many reasons. There might also be outlier spots in some cases. The word outlier means that the counterpart of a spot does not exist by bio-chemical reasons such as diseases or environmental conditions. However, the two words are used as the same meaning. In both cases, the spots have no counterpart.

B. Grassfire Spot Matching

Various spot matching methods have been proposed so far because the 2D-PAGE has characteristics of not only showing low repeatability but also including both global and local distortions. It is often classified into two categories. One is known as image matching approach that has been proposed in order to resolve spot matching problems involving nonlinear distortions[9,10]. And the other is point pattern matching method that utilizes geometric property of a graph. Here, the graph is formed as center positions of spots from

gel images. Again, the spot matching algorithm using the geometric property can be classified into several sub-categories such as using the landmark spots, graph theory, iterative closest point, similarity of the neighboring spots, etc.

The grassfire algorithm is adopted in this paper to perform basic spot matching with 2D-PAGE images. It has ability to produce accurate spot matching result based on the topological pattern of neighboring spots in two graphs. One is converted from the given reference gel and the other from the corresponding target gel. The grassfire algorithm is characterized by three factors such as type of graph, determination of seed spot pair for starting execution at the initial point pair and the spreading direction of consecutive spot matching. And it shows different spot matching result according to the type of a graph employed because the algorithm takes advantage of topological pattern of neighbor spots in its execution and it is obvious that the configuration of neighbor spots for each spot in a graph may be different. In general, there are commonly used types of graph in spot matching such as the Delaunay triangulation graph, the Gabriel graph, the relative neighbor graph, the k-nearest neighbor graph(k-NNG), etc. Among these, the grassfire algorithm takes advantage of k-NNG as the graph for building reference and target gel.

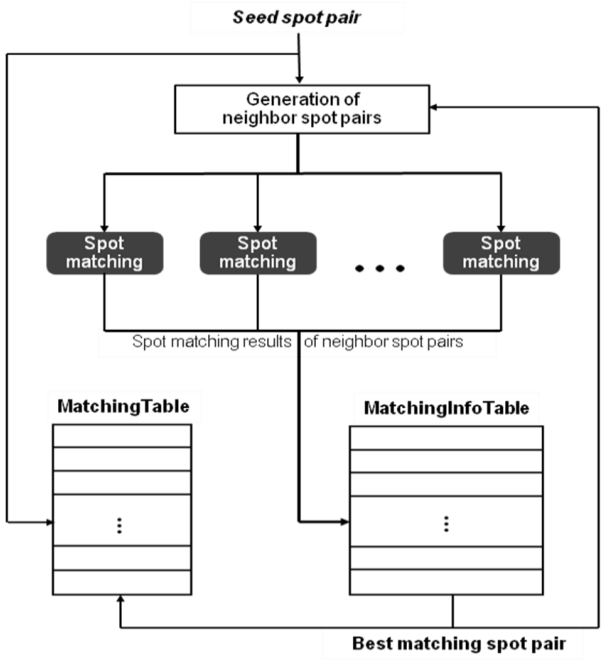


Figure 3. Organization of Grassfire Spot Matching

The grassfire algorithm starts spot matching process at a single distinct pair of spots, which is called the seed spot pair[5,7]. Clearly, a true-positive matching pair should be determined as a seed spot pair to prevent the algorithm from producing unreliable spot matching result due to input of wrong pair. In general, it is determined either in manual or automatic selection algorithm using landmark spots. Next, a single spot pair with the highest similarity among the neighbors of the seed spot is selected to be compared for next spot matching. The organization of the grassfire spot matching is depicted in the Figure 3. The spot matching results should be checked doubly with the help of the 'MatchingTable'. If a certain matched pair of neighbors is already exists in the 'MatchingTable', it is discarded and not stored in the 'MatchingInfoTable'. In this manner, the same process is iterated until all the spots are exploited and the 'MatchingInfoTable' becomes empty.

The grassfire algorithm takes advantage of matching information of the previous stage to the next matching stage, which results in the good performance in detection rate and matching accuracy as well as speed. And spot matching is spread out toward the direction of showing the best matching accuracy among neighbor spot pairs because the next matching spot is determined as a single spot pair of producing the highest topological similarity among candidates of neighbor spot pairs. Moreover, the position of the seed spot has strong influence on the order of matching, hence closely related to the direction. As a consequence, the grassfire scheme produces the same matching result regardless of the position of the seed spot pair. This is why a spot pair with the best matching value of similarity among candidates of neighbor spot pairs in the previous stage is chosen in the next matching stage of algorithm execution.

PROPOSED METHOD

The relationship between spot matching accuracy and reliability of the matched spot pairs are explained. And the proposed probabilistic reliability measuring scheme for each pairs of spot matching results using homogeneous multiple graphs are described in detail with mathematical notations. Also, the algorithm for the proposed scheme is explained with a simplified pseudo code in this paper.

A. Matching Accuracy vs. Reliability

Matching accuracy is the most important performance measure in spot matching algorithm and detection rate as well. Actually, the accuracy of 100% is not guaranteed among the detected spot pairs because the false-positive matching pairs can be contained as matched spot pair(s) in the result. Therefore, it is necessary to verify spot pairs in the matching result whether each of them definitely belongs to true- or false-positive matching.

The verification for confirming the spot matching results has been carried out by human with naked eyes traditionally, especially by medical scientists, which is not a trivial tasks because hundreds to thousands of pairs should be examined. It requires huge amount of time and cost in case all of the matched spot pairs are examined. On the other hand,

verification of spot matching results at random does not guarantee that all the false-positive matched spot pairs are selected for the candidates to be examined.

Once certain information for the matched pairs by an automated manner are given, one can easily make decision in the spot matching verification phase. In this paper, we present the information in the as a form of probabilistic reliability for each pair of the matching result. Then, spot pairs belonging to a certain range of probabilistic reliability are to be verified.

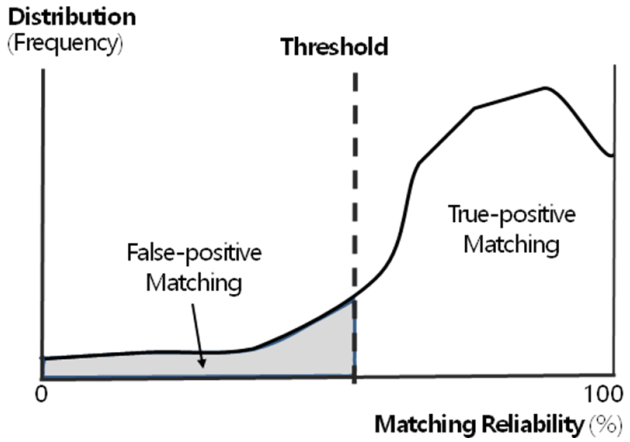


Figure 4. Thresholding for Decision Making from Probabilistic Matching Reliability Distribution

For example, let us suppose that the matched spot pairs whose probabilistic reliability is less than 50% are candidates for manual verification. The shaded area in the Figure 4 denote the candidates for verification. In case all of them turn out to be correct as true-positive matched pairs, the other spot pairs with probabilistic reliability of higher than 50% are regarded as correct ones. In case, otherwise, there exist some incorrect false-positive spot pairs among the candidates, the threshold for choosing candidates should be adjusted to higher probabilistic reliability from 50% up to 70%, for instance. In this manner, spot matching result verification in manual is accomplished with the minimum cost and time by greatly reducing the number of spot pairs to be examined. Moreover, the initial threshold to group the candidate spot pairs should be determined appropriately by medical scientists as a result of experience, discipline and/or empirical study according to the purpose of experiment.

B. Probabilistic Reliability Measuring Scheme using Homogeneous Multiple Graphs

In order to provide useful information, or - the probabilistic reliability of every matched spot pairs, the proposed method employs multiple graphs of homogeneity. The basic idea starts from the fact that different graph produce different matching result in the grassfire spot matching algorithm. This is obvious because it takes advantage of topological pattern of neighbor spots. In the proposed method, therefore, individual execution of spot matching by different but homogeneous graph produce result and it is accumulated by iteration. Then, the matching probability of every matched pair is calculated across multiple graphs as reliability for verification.

Now, the proposed method is described in detail as the normalized form with mathematical notation. Let us assume G a set of graphs containing multiple graphs as its element, as shown in the equation (3). The notation $|G|$ in the equation (4) is defined as the size of G. It takes the value corresponding to the number of individual graph as its elements, meaning the number of iteration of spot matching in the proposed method.

$$G = \{G_k \mid k = 1 \text{ to } N\} = \{G_1, G_2, \ldots, G_N\} \quad (3)$$

$$|G| = N \quad (4)$$

With the use of graph G, the matching result M for each spot pair (p_i, q_j) by an algorithm execution for a certain graph G_k can be expressed in the form of a tuple (G_k, p_i, q_j) as shown in the equation (5). Every result for all spot pairs with respect to individual G_k are computed in turn and stored in the matrix. And then, the corresponding results of spot pair (p_i, q_j) across G, from G_1 to G_N, is summed up to get the accumulated spot matching frequency $F_G(p_i, q_j)$ as shown in the equation (6).

$$M(G_k, p_i, q_j) = \begin{cases} 1, & \text{if matched} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$F_G(p_i, q_j) = \sum_{k=1}^{N} M(G_k, p_i, q_j) \quad (6)$$

$$R_G(p_i, q_j) = \frac{1}{|G|} \cdot F_G(p_i, q_j) = \frac{1}{N} \cdot F_G(p_i, q_j) \quad (7)$$

Finally, the reliability of every spot pair based on probability with respect to the homogeneous multiple graphs can be obtained. The probabilistic matching reliability R_G of each spot pair (p_i, q_j) represents the average value across multiple graphs, which is calculated by dividing the accumulated spot matching frequency with the number of graphs wholly used in the algorithm, as shown in the equation (7). After all, the medical scientists would take advantage of probabilistic reliability as useful criteria to determine whether each spot pair is actually matched or not.

C. Algorithm Description

The algorithm to implement the proposed probabilistic reliability measuring scheme described above has built, which is illustrated with simplified pseudo code in the Figure 5. As for graphs, we adopted *k-NNG*(k-nearest neighbor graph) in the algorithm because the grassfire algorithm used it originally and the same kinds of k-NNG shows good homogeneity or similarity compared to the other kinds of graphs. Once again, it takes advantage of the grassfire spot matching algorithm by iterating as many times as multiple graphs with homogeneous property used in the proposed method. It is comprised of three principal steps as follows:

```
01: M_array(m, n) ← 0
02: P_match(m, n) ← 0
03: for each graph ∈ k-NNG do
04:   M_pair = grassfire(P, Q, k-NNG)
05:   for each spot pair (pi,qj)∈ M_pair do
06:    M_array(pi,qj) = M_array(pi,qj) + 1
07:   end for
08: end for
09: for each pi ∈ P do
10:   for each qj ∈ Q do
11:    P_match(pi,qj) = M_array(pi,qj)/|k-NNG|
12:   end for
13: end for
```

Figure 5. Pseudo Code for the Proposed Method

• (STEP 1) Initialization for the variables used in the algorithm is performed in the first step. Here, two kinds of arrays *M_array* and *P_match* are to be used, where *(m, n)* denotes the number of spots in reference gel *P* and target gel *Q*, respectively. The former *M_array* is to store matching account through grassfire execution by the whole multiple graphs. And the latter *P-match* is for storing the probabilistic matching reliability.

• (STEP 2) In the second step, actually spot matching between the reference and target gel is performed using the grassfire algorithm for every multiple graphs in turn. First, the grassfire algorithm configures the information of *k-NNG* such as neighbor spots from *P* and *Q*. And then, it returns result of the matched pairs to *M-pair* after execution. For each iteration in this step, the matching results of each spot pair for is accumulated to *M-array*.

• (STEP 3) In the final step, the pair-wise reliability of matching probability throughout whole graphs used in the algorithm is calculated. It is performed by dividing the accumulated sum with the number of graphs. The notation |*k-NNG*| in the line 11 denotes the number of graphs the same as |*G*| in the equation (4).

At this point, the problem in the proposed method arises on how many graphs are to be used for effective and efficient information in real environment by medical scientists. In conclusion, it is out of the question because the execution time by additional graphs would be ignored when compared to the time and cost of manual task. Therefore, one can introduce as many as graphs to obtain more precise measure of probabilistic matching reliability.

EXPERIMENT AND RESULT

The proposed method is implemented using PERL language to verify its validity. And the experiment is performed using a specific 2D-PAGE image data set. Here, the feature of data set is described. And the experiment results are also presented.

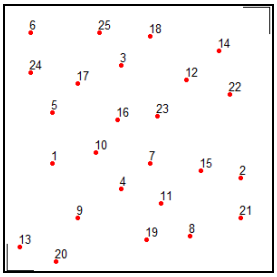
A. Data Set for Experiment

It requires a test data set of 2D-PAGE image to experiment spot matching procedure by the proposed method. Moreover, the individual matching spot pairs for a given data set - a reference gel and a corresponding target gel - should be known in advance. Actually, it is very difficult and time-/cost-consuming task to prepare gel image samples from tissues of human being while confirming every matching spot pairs in manual because a single image gel contains hundreds or thousands of spots. Therefore, a synthesized gel image pair is used for experiment in this paper. In the literature [11], the method to generate a pair of gels is introduced. It reflects inherent property of real gel images by maintaining topology patterns by following the normal distribution with the minimum distance between spots.

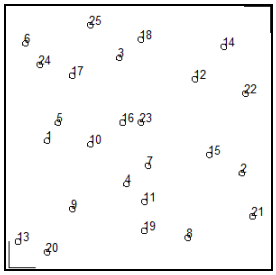
A data set of the synthesized gel images is shown in the Figure 6. It contains 25 spots in a gel of size 64×64 pixels with the minimum of 10 pixel distance. And there exists no outliers or missing spots for simplicity. Several parameters need to be set to generate a synthesized gel pairs such as size of a gel, the number of spots, the minimum distance between gels, the normal distribution, etc.

The target gel image is generated with a certain variation based on the reference gel image. First, a reference gel is randomly generated based on the given parameters. In the process of reference spot generation, the minimum distance between/among spots plays an important role because it prevents the generated spots from overlapping or placing too much close. Next, once all the spots of a reference gel is produced, their corresponding spots for a target gel should be generated in turn. The spots in the reference gel are transformed to form corresponding spots in the target gel

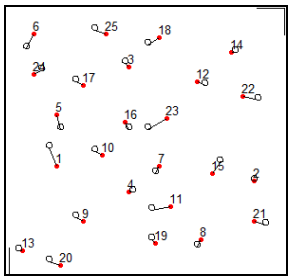
using the random number of the normal distribution. The displacements for every spots in the reference gel are calculated using the normal distribution along with X- and Y-axis, respectively. And then, they are added to the coordinates of the corresponding spots in the reference gel. The displacement information between reference and target gel in the Figure 6 is presented in the Table 1 together with spot coordinates. Finally, a single number is allocated to two corresponding spot pair one after another. This process provides an easy way to identify the detected pot pairs in the spot matching result by an automated spot matching algorithm whether each pair belongs to either false-positive or true-positive with the help of pre-allocated number of spots.



(a) Reference Gel ***P***



(b) Target Gel ***Q***



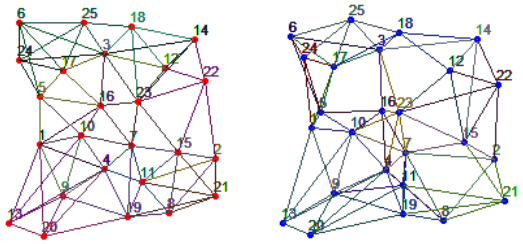
(c) Overlapped Image

Figure 6. Synthesized Gel Image Data Set for Experiment

Table 1. Coordinates(location information) of Spots in the Gel Images of Figure 7

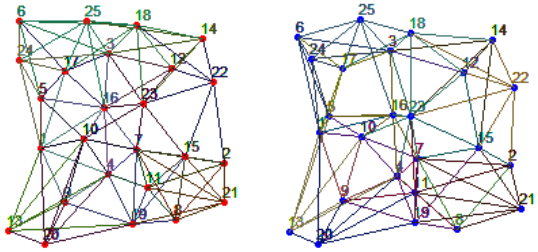
Number of Spot	Reference Gel $p_i=(x_i, y_i)$	Target Gel $q_j=(x_j, y_j)$	Displacement* (p_i - q_j)
1	(38, 78)	(33, 94)	(-5, 16)
2	(174, 67)	(173, 70)	(1, 3)
3	(88, 148)	(85, 154)	(-3, 6)
4	(88, 59)	(90, 62)	(2, 3)
⋮	⋮	⋮	⋮
24	(22, 143)	(27, 148)	(5, 5)
25	(72, 172)	(63, 177)	(-9, 5)

*Note: the minus sign(-) denotes the direction of displacement.



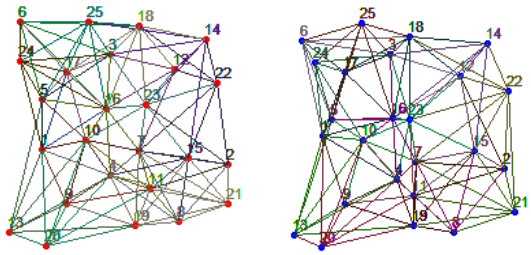
Reference Gel Target Gel

(a) 5-NNG



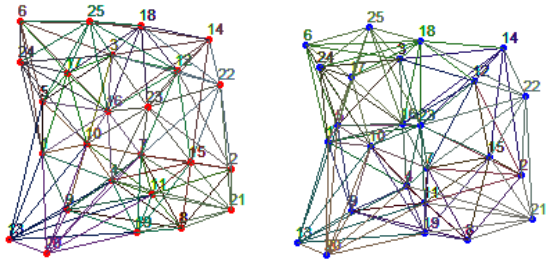
Reference Gel Target Gel

(b) 6-NNG



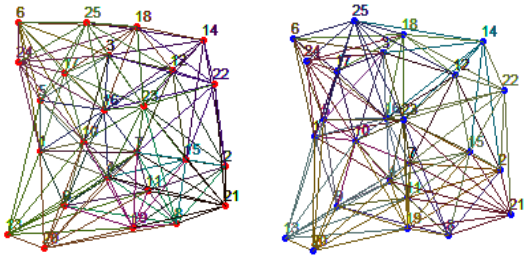
Reference Gel Target Gel

(c) 7-NNG



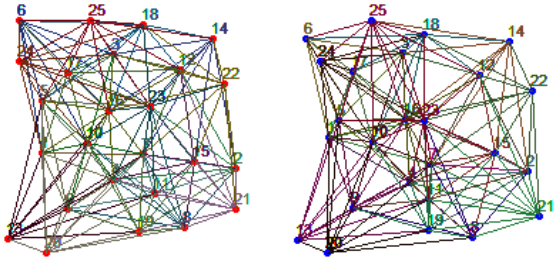
Reference Gel Target Gel

(d) 8-NNG



Reference Gel Target Gel

(e) 9-NNG



Reference Gel Target Gel

(f) 10-NNG

Figure 7. Graph Configuration for the Gel Image Data Set.

B. Experiment and Result

The synthesized 2D-PAGE data set containing 25 spots as mentioned above in the Figure 7 is used in the experiment, even though there exist hundreds to thousands of spots in a real gel image. This is why we aimed at showing the precise and intuitive results by simplicity. The procedure of experiment is very simple. First, the proposed algorithm is implemented in PERL programming language. And then, it is performed using the experiment data set, which produces fundamental spot matching results. Finally, they are summed and calculated to get outcomes in desired forms.

The spot matching performance is summarized in the Table 2, which denotes outcome by the grassfire spot matching algorithm contained in the proposed method using six kinds of graphs as its input from 5-NNG to 10-NNG individually. The detection rate and the matching accuracy are the performance measures. The latter is the ratio of the number of the true-positive matched spot pairs over that of the detected spot pairs and the former is the number of detected spot pairs as matched over that of overall spot pairs in two corresponding gels. In three cases of 5-NNG, 6-NNG and 7-NNG, the matching accuracy is less than 100 percent (shaded gray in the Table 2). This implies that false-positive matching pairs together with true-positive ones are contained among the detected matching spot pairs. Consequently, the true- and false-positive matching rates can be calculated from two measures of the detection rate and the matching accuracy in the Table 2, which is shown in the Table 3.

Table 2. Result Summary of the Grassfire Spot Matching using *k*-NNG

Measures	*k=5*	*k=6*	*k=7*	*k=8*	*k=9*	*k=10*
Detection Rate (%)	92.0	96.0	100	100	100	100
Matching Accuracy (%)	95.7	95.5	92.0	100	100	100

Table 3. True- and False-positive Matching Rate

Measures	*k=5*	*k=6*	*k=7*	*k=8*	*k=9*	*k=10*
True-positive Matching (%)	88.0	91.7	92.0	100	100	100
False-positive Matching (%)	4.0	4.3	8.0	0	0	0

The Table 4 denotes the matching results of each spot pair for six kinds of graphs including both true- and false-positive matching in detail. And the accumulated matching frequency(AF) and the probabilistic matching reliability(PR) across the multiple graphs are also shown. Here, four spot pairs shaded gray in the Table 3 such as (4, 11) of 6-NNG, (12,23) and (23, 12) of 7-NNG and (24,6) of 5-NNG are detected as matching pairs, but false-positive matching pairs by the grassfire spot matching method. Actually, they are matched by chance in the execution of the grassfire spot matching for a certain graph, showing relatively very low probabilistic matching reliability compared to the others.

Let us assume that a threshold value in the verification stage is determined to be no less than the PR of the spot pair (11, 11), namely greater than or equal to 66.7%. It is obviously clear that the detection rate and its true-positive matching accuracy would be 100% for the whole spot pairs finally determined by threshold. On the other hand, the threshold value is set to 90%, the detected matching pairs such as (4, 4), (6, 6), (11, 11), (12, 12), (23, 23) and (24, 24) are regarded as false-positive ones even though they are true-positive ones showing over than 80% of their PRs. When confirming the spot matching pairs by threshold, it should be determined carefully. In case it is very low, the probability to contain the false-positive matching pairs among results becomes high. On the contrary, when it is set too high, the true-positive matching pairs are excluded and a number of spot pairs would remain for manual examination with efforts of time and cost. In actual, medical scientists should find out the method to determine appropriate threshold based on their empirical knowledge and heuristic information according to the type of samples.

Table 4. Probabilistic Spot Matching Reliability across Homogeneous Multiple Graphs using *k*-NNG

Spot Pairs (p_i, q_j)	Spot Matching Results						AF*	PR**
	k=5	*k=6*	*k=7*	*k=8*	*k=9*	*k=10*		
(1,1)	√	√	√	√	√	√	6	100
(2,2)	√	√	√	√	√	√	6	100
(3,3)	√	√	√	√	√	√	6	100
(4,4)	√		√	√	√	√	5	83.3
(4,11)		√					1	16.7
(5,5)	√	√	√	√	√	√	6	100
(6,6)		√	√	√	√	√	5	83.3
(7,7)	√	√	√	√	√	√	6	100
(8,8)	√	√	√	√	√	√	6	100
(9,9)	√	√	√	√	√	√	6	100
(10,10)	√	√	√	√	√	√	6	100
(11,11)			√	√	√	√	4	66.7
(12,12)	√	√		√	√	√	5	83.3
(12,23)			√				1	16.7
(13,13)	√	√	√	√	√	√	6	100
⋮	√	√	√	√	√	√	6	100
(22,22)	√	√	√	√	√	√	6	100
(23,23)	√	√		√	√	√	5	83.3
(23,12)			√				1	16.7
(24,6)	√						1	16.7
(24,24)		√	√	√	√	√	5	83.3
(25,25)	√	√	√	√	√	√	6	100

* AF : accumulated matching frequency
** PR : probabilistic matching reliability across multiple graphs

The proposed method using homogeneous multiple graphs can be a new alternative to enhance accuracy of the spot matching algorithm. Consequently, the inaccurate matched spot pairs contained in the spot matching results can be excluded by calculating probabilistic matching reliability for all the matched spot pairs. Although the experiment in this paper shows a good example, one should take into account somewhat different aspects of results in the case of using real gel images caused by a number of spots in a gel, distance between/among spots, global and local distortions, and other properties.

CONCLUSION

The spot matching algorithm plays an important role to automate one or more phases in 2D-PAGE in the field of proteomics. Although it reduces time and cost significantly, the results cannot be guaranteed. Therefore, they should be carefully examined in manual task because the possibility of containing false-positive matching spot pairs always exists. The goal of this paper is to solve this problem by iterating the grassfire algorithm several times with homogeneous multiple graphs and summing up the every results to get probabilistic matching reliability for each spot matching pairs.

The proposed method has implemented and the experiment for six kinds of k-NNGs has performed to verify its effectiveness. In the experiment, the synthesized 2D-PAGE gel image data set is used, which contains 25 spots in a gel image size of 64×64 pixels. The experiment provides meaningful probabilistic reliability for the resulting spot matching pairs to decide whether each of them belongs to true-/false-positive matching by medical scientists.

The contribution of this paper will help one who engages in proteomics and related fields to utilize the automated spot matching method more efficiently in actual environment because it can minimize a lot of manual validation, hence reducing time and cost significantly. However, the effective determination for threshold still remains. The future research topic includes the issue for threshold and the diversity of graphs adopted for the algorithm execution.

ACKNOWLEDGEMENT

The corresponding author is Wook Hyun Kim.

REFERENCES

[1] T. Srinark and C. Kambhamettu, "An Image Analysis Suite for Spot Detection and Spot Matching in Two-Dimensional Electrophoresis Gels", Electrophoresis, Vol. 29, pp. 706-715, 2008

[2] J. L. Harry, M. R. Wilkins and B. R. Herbert, "Proteomics: Capacity versus Utility", Electrophoresis, Vol. 21, pp. 1071-1081, 2000

[3] A. Almansa, M. Gerschuni, A. Pardo and J. Preciozzi, "Processing of 2D Electrophoresis Gels", The 1st International Workshop on Computer Vision Applications for Developing Regions(ICCV), 2007

[4] M. Daszykowski, E. Mosleth Faregestad, 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] Yun-Kyoo Ryoo, Chan-Myeong Han, Ja-Hyo Ku, Dae-Seong Jeoune and Young-Woo Yoon, "Grassfire

Spot Matching Algorithm in 2-DE", International Journal of Bio-Science and Bio-Technology, Vol. 5, No. 4, pp. 167-174, SERSC, September 2013

[6] Dae-Seong Jeoune *et al.*, "Reliability Measure of 2D-PAGE Spot Matching using Multiple Graphs", Proceedings of 2014 International Conference on Image Processing, Computer Vision and Pattern Recognition(IPCV2014), pp. 311-315, CSREA, WORLDCOMP'14, Las Vegas, USA, July 21-24, 2014

[7] Dae-Seong Jeoune, Chan-Myeong Han and Wook Hyun Kim, "Fully Automated Detection of Landmark Spot Pairs using the Topology of Both the First and the Second Neighbor Spots in Two-Dimensional Electrophoresis", International Journal of Applied Engineering Research", Vol. 11, No. 18, pp. 9448-9454, 2016

[8] Chan-Myeong Han, Dae-Seong Jeoune, Hwi-Won Kim and Young-Woo Yoon, "A Spot Matching Algorithm using the Topology of Neighbor Spots in 2D-PAGE Images", International Journal of Software Engineering and Its Applications, Vol. 7, No. 5, pp. 87-98, SERSC, September 2013

[9] Jiann-Der Lee and Wei-Chun Chen, "A Novel Scheme for Registration of Two Dimensional Gel Electrophoresis Images", Biomedical Engineering: Applications, Basis and Communications, Vol. 18, No. 4, pp. 158-166, August 2006

[10] G. Shi, T. Jiang, W. Zhu. B. Liu and H. Kao, "Alignment of Two-Dimensional Electrophoresis Gels", Biochemical and Biophysical Research Communications, Vol. 357, pp. 427-432, 2007

[11] Dae-Seong Jeoune, Moon Sun Kim and Chan-Myeong Han, "Synthesis of 2D-PAGE Data Sets for Point-Matching Based Spot Matching Applications", International Journal of Information, International Information Institute, Vol. 18, No. 3, pp. 1113-1120, March 2015