Progress in Biological Sciences
Vol. 6, Number 2, Summer / Autumn 2016/129-137 - DOI: 10.22059/PBS.2016.590015

Neighborhood matrix: A new idea in matching of two dimensional gel images

Behrouz Alizadeh Savareh[1], Azadeh Bashiri[2], Mehrnaz Mostafavi[3]*

[1] *Student Research Committee, School of Allied Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.*

[2] *Department of Health Information Management, School of Allied Medicine, Tehran University of Medical Sciences, Tehran, Iran*

[3] *Faculty of Allied Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

Received: July 08, 2016; Accepted: December 9, 2016

ABSTRACT

Automated data analysis and pattern recognition techniques are the requirements of biological and proteomics research studies. The analysis of proteins consists of some stages among which the analysis of two dimensional electrophoresis (2-DE) images is crucial. The aim of image capturing is to generate a Photostat that can be used in future works such as image comparison. The researchers introduced a new method for matching two 2-DE gel images. In this method, a neighborhood circular region is defined to obtain information about spots' neighbors. In the present paper, the information obtained by this region is reordered into a matrix as a descriptor of the neighbors of each spot. The matrix is then used in matching the spots between two images. All conducted tests to evaluate the method's performance showed the power of the method in spot matching, even when the number of candidate matching spots in the second images increased. The proposed method provides a robust automatic comparison idea in gel images matching. Despite its low speed, its accuracy is excellent. The Novelty of the present study is the use of matrices as neighborhood descriptor. This idea is applicable in any other similar domain.

Keywords: Image matching; Gel electrophoresis; Recursion; Transformation

Introduction

Computerized analysis of data and meaningful pattern recognition techniques are required in many research fields like biology, which concerned with the study of life and living organisms (1). Genomics and proteomics are the common examples of biological sciences that cover important aspects of biology and biomedical engineering. Protein and chromosome mapping, listing the proteins coded by genes and examining gene expression are some researches done in these area (2).

In the field of proteomics, comparison of two

* **Corresponding author:** mz_mostafavi@sbmu.ac.ir

dimensional electrophoresis (2-DE) images is the selected method for proteins matching (3), which consists of applying the electric potential difference to the proteins stains and separating them on the gel to form 2-DE images. The formation of 2-DE images as the result of separation is a powerful resource for biomedical diagnosis (4). Spot by spot matching of 2-DE images of two or more proteins can lead to differentiate them. This technique is used to obtain a realistic and global view of the cellular genetic expression (1).

In a 2-DE image, proteins are shown as blob spots on gel. These spots can be revealed by staining techniques and captured by digital cameras. The aim of image capturing is to convert biological information to digital information in the form of an image that can be used in future works such as image comparison (5). The comparison of two images is done by spots matching. The results of 2-DE image comparison is the detection of the changes in the protein expression and identification of new proteins (6). Incorrect detection of the spots of 2-DE images or incorrect matching of them can detrimentally affect the expression profile of the proteins (7). Complex physical and chemical processes can cause some differences in the locations of the same protein spots in different gels images. Because of the flexibility and pliability of gels used in 2-DE, the extracted images were distorted. This distortion and deformation makes the 2-DE image registration complicated, and gel image matching becomes difficult (8). The difficulties are emerged in image registration, image distortion correction, spot detection, and spot matching (9), so adaptive registration methods are required for matching spots in two different gel images (10).

Two dimensional spot matching of two non-uniform images is a nondeterministic polynomial time (NP-hard) problem (11). Their computation is non-deterministic and is not solvable in polynomial time (12). Different algorithms have been developed for registration of 2-DE images. Such algorithms intensively use different information from the gel images, for example, spot intensities and spot shapes. The algorithms are hence complex, but the performance is still not superior. It seems that the problem has not been resolved yet; therefore, the aim of this study was to develop a new method for 2-DE gel images comparison.

Materials and methods

In this study, a set of 2D gel images taken from processing of differentiating human astrocytes from mesenchymal stem cells were used (13). Young and senescent human astrocytes on 25 mm^2 dishes were colored three times, and then 2-DE gels were fixed and scanned into a digital format. Sample of the dataset is shown in Fig. 1.

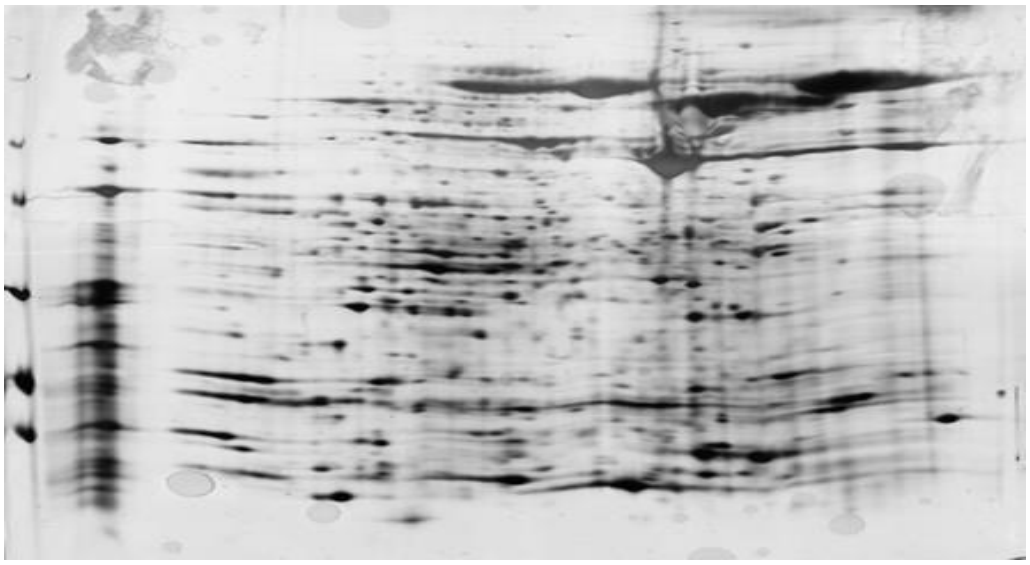

Figure 1. Sample of 2-DE images used for processing in the present study.

A two-step process was conducted to address 2-DE gel images matching:

1) The first step is the protein spots detection from 2-DE gel images: A modified version of the level set algorithm was applied to overcome the challenges involved.

2) The second step is protein spots matching: A novel idea is introduced by employing matrices.

The neighborhood area of the protein spot was searched to locate neighbor spots and a neighborhood matrix was formed, which was composed of neighbor spots structural properties. The matching step was addressed by using pair matrices for any spot in the first image (reference spots) and compare to the pairs of any potential candidate spot (any spot located in specific distance from reference spot) in the second image. All implantations were done by Matlab 2014 (14).

Spot detection

The spots were detected by a customized version of the level set method (15). This method is capable of segmentation in images with intensity inhomogeneity. By iterative manner, the method separates protein spots based on the local intensity, i.e. clustering property of the image intensities, and defines a local clustering criterion function for the image intensity in the neighborhood of each point. The method defines the energy by the level set functions that represent a part of the image domain and a bias field, which accounts for the intensity inhomogeneity of the image. Therefore, by minimizing this energy, the method is able to segment the image and estimate the bias field that can be used for intensity inhomogeneity correction. This approach does not work properly in situation where multiple neighboring protein spots are located very closely. The separated areas may have more than one protein as shown in Figure 2.

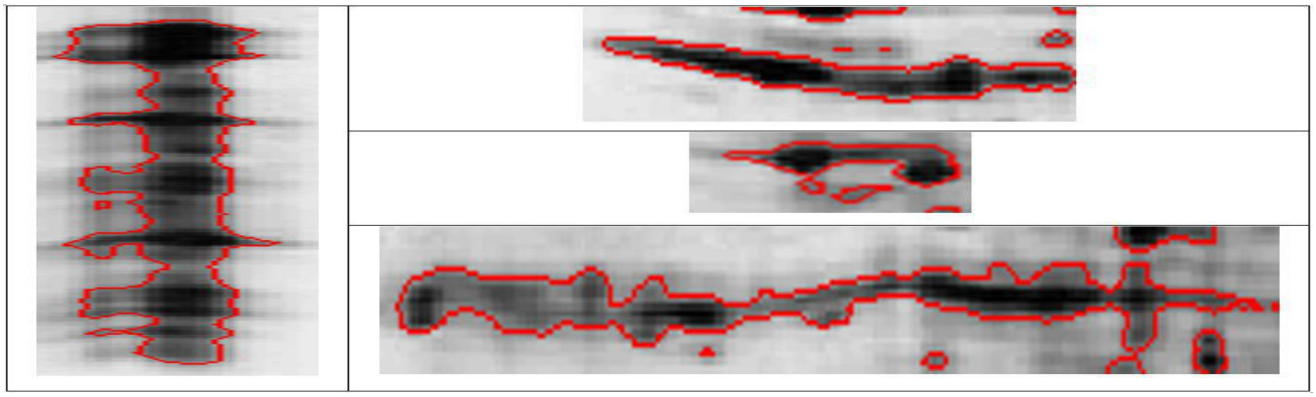

Figure 2. A 2-DE gel showing some areas with more than one protein.

To address the problem of separation areas that have more than one protein image, another step was applied based on the adaptive thresholding algorithms (16). For each pixel in the image, a threshold has been calculated, if the pixel value is below the threshold, it was set to the background value; otherwise, it was assumed as foreground value. Adaptive thresholding method was used as a complementary part to the level set method for image segmentation to slab the areas with more than one protein into areas with only one protein in a recursive manner (Figure 3).

This method is a recursive combination of the level set method segmentation and adaptive thresholding (Fig. 3). Although the application of the level set method improved the precision of the spots detection (Fig. 2); some areas that are greater than bound and suspicious of containing more than one spot were double-checked. These areas were returned into the flowchart to pass through the adaptive thresholding method phase and enter the level set again. This process is performed recursively until all possible spots are found.

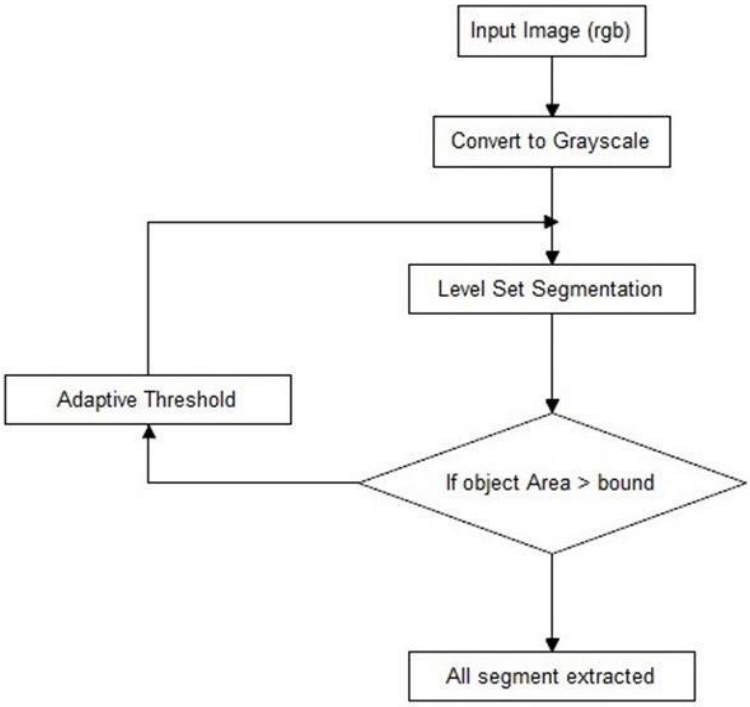

Figure 3. The algorithm of the level set method combined with adaptive threshold technique used in 2-DE gel analysis.

Spot matching

To overcome the matching problem, two matrices are defined for each spot in both images:

1) A property matrix that describes the spot morphology and regional information (Table 1).

2) A neighborhood matrix generated by sensing the neighborhood spot area for each spot. The formation of the neighborhood matrix from the neighboring area is illustrated in Fig. 4.

Table 1. The properties of the matrix describing the spot morphology in a 2-DE gel

Area	Eccentricity	Major axis length	Minor axis length
Equivalence diameter	Extent	Mean intensity	Bounding box length/Bounding box width

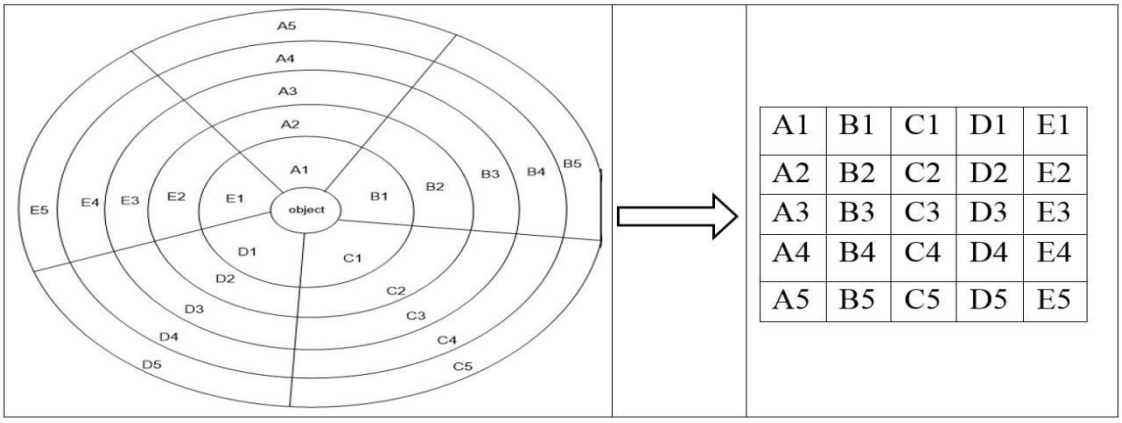

Figure 4. The neighborhood matrix formation from the area of a protein spot in a sample 2-DE gel.

As seen above, the neighborhood area of a spot was segmented into some areas as A1-A5, B1-B5, C1-C5, D1-D5, and E1-E5. Each cell value of the matrix was equal to the sum of the spots area located in the areas with the same name. Then the candidate pair spots in the first image were searched with the area by the radius "R" in the second image. For example, if 15 candidate pair spots are found in second image, a similarity value is calculated through the following formula:

Similarity value = $ssim_p \times ssim_n$; where $ssim_p$ indicates structural similarity of property matrix of spot in the first image and property matrix of spot in the second image and $ssim_n$ stands for structural similarity of neighborhood matrix of spot in the first image and neighborhood matrix of spot in the second image.

The similarity of the two matrices was confirmed by calculation of the structural similarity index measure (SSIM) between the two matrices (17). After the calculation of the similarity value for the spots of interest, the values are sorted in a descending order, and the highest value is selected as the paired spot for reference spot. This process is repeated for all spots in the first image.

Results

Detecting spots from 2-DE gel images is a basic step that must be carried out as accurately as possible, which, affects the performance of those foregoing steps, including matching step, before the matching step is started. It is obvious that the accuracy of all the following steps depends on its correctness.

The result of spot detection in pure level set mode and modified one are shown in Figure 5. The result of separating protein spots from the background using a modified version of the LSM is appropriate and by using this method, areas with multiple spots can be separated successfully.

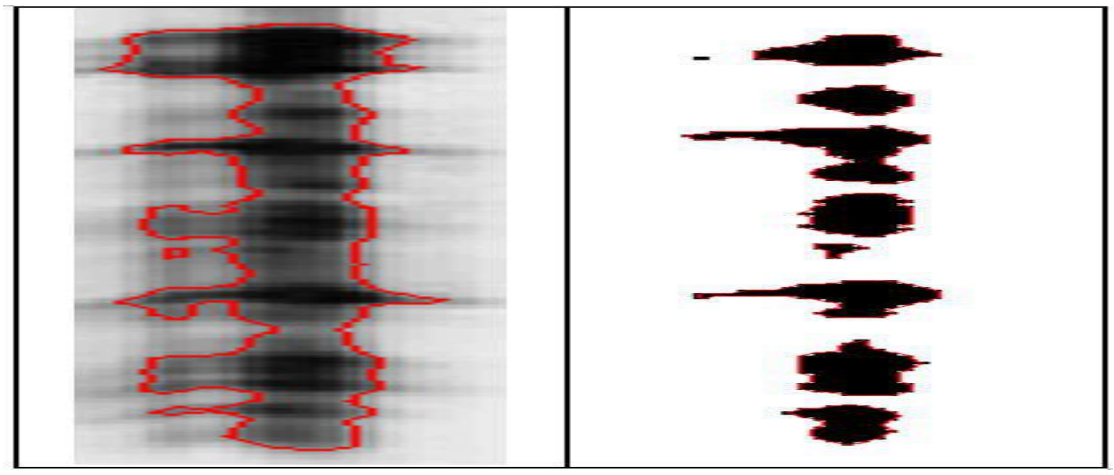

Figure 5. The result of modified level set method applied to a big area of a 2-DE gel. The image at left side shows the spot detection in pure level set mode, while that at the right side indicates the modified level.

In order to test the performance of the methods, a sample image analyzed by our method and the same image detected by some common spot detector algorithms: SURF (a)(18), BRISK (b)(19), FAST(c)(20), MIN EIGEN (d)(21), MSER (e)(22) and our method (f) were compared. As shown in Figure 6 other algorithms than our method used for spot detection were not able to provide acceptable results. Some of the compared algorithms mis-spots and some take more than one spot in one spot location.

The result of customized level set method was notably better than other algorithms in spot detection. Also, noisy area (yellow highlighted) rejected by our method. Regarding the matching points, our method was compared with two famous methods in the spot matching: Harris matching (23) and progressive graph matching (24). The result of matching comparisons is shown in Figure 7.

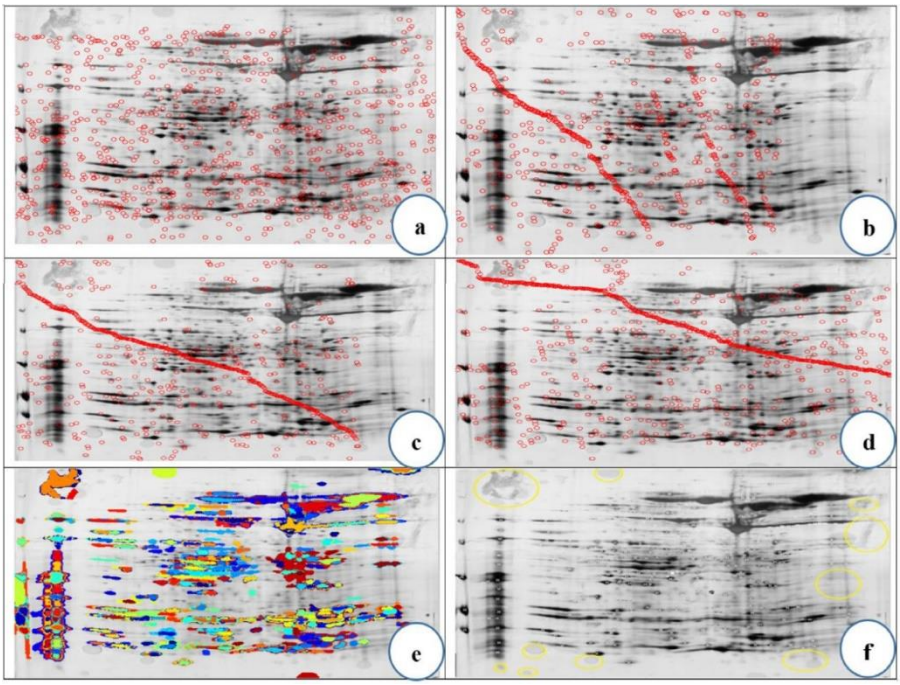

Figure 6. Spot detection comparison using some common spot detector algorithms: SURF (a), BRISK (b), FAST(c), MIN EIGEN (d), MSER (e) and our method (f). References are given in the text.

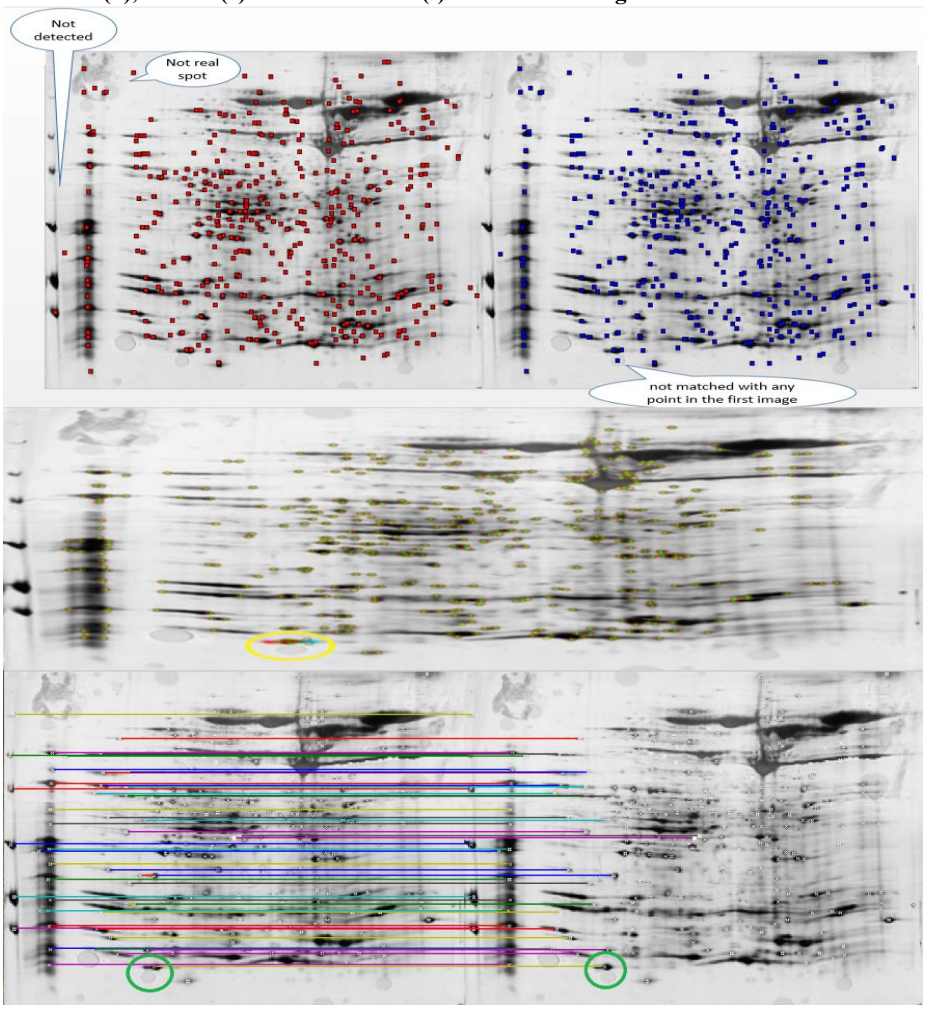

Figure 7. Spot matching comparison. Upper image: progressive graph matching, middle: Harris matching, lower: our method.

In addition, to evaluate the performance of algorithm in different elastic distortion, total similarity matrices were generated by increasing the radius of searching in a 3D form. In Figure 8, the vertical direction indicates the index of spots in the first image and the horizontal direction is the index of spots in the second image.

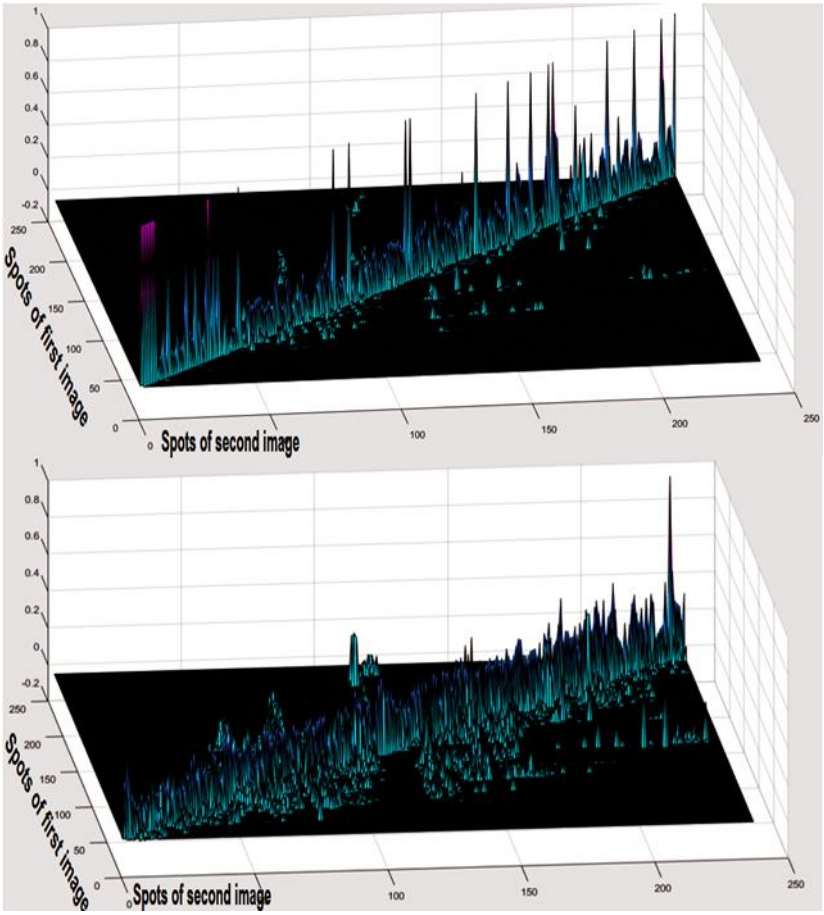

Figure 8. Similarity values between all spots in the first and second image by increasing the radius of search.

Discussion

Based on our results, the performance of the proposed algorithm in spot detection was superior to all other methods: SURF, BRISK, FAST and MIN EIGEN (18-21). These methods were not able to detect spots properly and there were no compatibilities between the real spots and their output. The performance of MSER (22) was relatively better. However, it could not be ignored that the noisy areas (yellow circles in figure 6f) and detecting them as spots is not acceptable. In such situations, the customized level set (15) could separate real spots from noisy areas properly. Although there were some undetected spots, they could be ignored due the acceptable performance of algorithm.

In the matching, the performance of the proposed algorithm was better than the others since a small shift in the location of a spot from one image to another caused both Harris matching algorithms (23) and progressive graph matching (24) go wrong. By adding some shifts in the artificial location of the spots, none of these algorithms could establish an acceptable match for the shifted spots. Although the execution time increased, the diagonal line of total similarity value (correct match) was not threatened by increasing the radius of search (Figure 8). Ignoring the execution time increased the tolerance of algorithm against more elastic distortion. This means that the algorithm introduced in this study is able to deal with a greater amount of distortion in the images.

Neighborhood matrix was used as descriptor for

spatial information about a spot neighbor in this paper. Positive results indicate the success of this method, a closer examination of the aspects of using the matrix contains important points in clarifying the new dimensions of the application of this method and its optimization. The creation of the matrix was done roughly where neighborhood circle area sectored sharply. The ability of the matrix can be increased in the spatial description of spots neighborhood using fuzzy bordering of the sectors. The fuzzy bordering sectors of neighborhood areas will increase the tolerance of the algorithm against more elastic distortions. However, this process requires a complex process to form the matrix in fuzzy form and can be the subject of future studies.

REFERENCES

1. Han, C.M., Park, J.H., Chang, C.S., and Ryoo, M.C. (2014) A spot-matching method using cumulative frequency matrix in 2D gel images. *Biotechnol. Biotechnol. Equip.*., **28**, 37-43.
2. Polášková, L., (2014) *Dynamic spot analysis in the 2D electrophoresis gels images.* Fakulta elektrotechniky a komunikačních technologií.
3. Wörz, S., Winz, M.L., and Rohr, K. (2008) *Geometric alignment of 2D gel electrophoresis images using physics-based elastic registration*. In: Biomedical Imaging: From Nano to Macro, 2008. 5th IEEE International Symposium.
4. Freire, J.A., Seoane, Rodriguez, A., Ruiz Romero, C., López-Camos, G., and Dorado, J. (2010) A Block Matching based technique for the analysis of 2D gel images. *Stud. Health Techn. Inf.*, **160**, 1282-1286.
5. Natale, M., Mareska, B., Abrescia, P., Bucci, E.M. (2011) Image analysis workflow for 2-D electrophoresis gels based on Image J. *Proteom. Ins.*, **4**, 37-49.
6. Xin, H., Zhu, Y., and Wu, L. (2007) *A comparative study of adaptive registration methods for two dimensional gel electrophoresis images*. In: 29th Annual International Conference of the IEEE: Engineering in Medicine and Biology Society, 2007. EMBS 2007.
7. Ceccarelli, M. and Carstensen, J. (2010) *Bayesian grid matching for 2D gel registration.* In: IEEE International Conference on Imaging Systems and Techniques (IST 2010). Thessaloniki, Greece, 1-2 July 2010. Pp. 279-283.
8. Li, C., Huang, R., Ding, Z., Gatenby, J.C., Metaxas, D.N., and Gore, J.C. (2011) A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE *Trans. Image Proc.*, **20**, 2007-2016.
9. Fisher, B., Perkins, S., Walker, A., and Wolfart, E. (1996) *Hypermedia image processing reference*. Wiley Chichester, UK.
10. Lin, D.T., Kuo, J.L., Lin, E.c., and Huang, S.Y. (2005) Fuzzy similarity measure and fractional image query for large scale protein 2d gel electrophoresis. In: Khosla, R. (ed.), *Knowledge-Based Intelligent Information and Engineering Systems*. Springer, Stuttgart.
11. Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P. (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Proc.*, **13**, 600-612.
12. Davies, E.R. (2004) *Machine vision: theory, algorithms, practicalities*. Elsevier, Amsterdam.
13. Zali, H., Tavirani, Moh.R., and Tavirani, M.R. (2015) Protein protein interaction network of Alzheimer's disease from human hippocampal proteome. IJAPBS, **4**, 130-144.
14. Gonzalez, R.C., Woods, R.E., and Eddins, S.L. (2004) *Digital image processing using MATLAB*. Pearson Education, India.
15. Li, C., Huang, R., Ding, Z., Gatenby, J.C., Metaxas, D.N. and Gore, J.C. (2011) A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE Transactions on Image Processing, **20**, 2007-2016.

16. Saikumar, T., Nagarani, M., Yojana, K., and Shashidhar, B. (2012) *Image segmentation of an adaptive threshold algorithm using watershed transform and fuzzy c-means clustering on level set method*. In: IEEE International Conference On Advances in Engineering, Science and Management (ICAESM -2012), Tamil Nadu, India.
17. Kaur, E.S., and Bawa, E.N. (2017) Performance evaluation of compressed images by using Gradient based optimization and Edge restoration method. *Int. Res. J. Eng. Technol*., **4**, 1620-1624.
18. Bay, H., Tuytelaars, T. and van Gool, L. (2006) *Surf: Speeded up robust features.* In: IEEE International Conference on Computer Vision. Pp. 404-417.
19. Leutenegger, S., Chli, M. and Siegwart R.Y. (2011) *BRISK: Binary robust invariant scalable keypoints*. In: IEEE International Conference on Computer Vision.
20. Rosten, E. and Drummond T. (2005) *Fusing points and lines for high performance tracking*. In: Tenth IEEE International Conference on Computer Vision.
21. Shi, J. (1994) *Good features to track.* In: IEEE Conference on Computer Vision and Pattern Recognition.
22. Obdržálek, D., Basovník, S., Mach, L. and Mikulík, A. (2009) *Detecting scene elements using maximally stable colour regions*. In: Gottscheber, A., Obdržálek, D., and Schmidt, C. (eds.), International Conference on Research and Education in Robotics. Springer, Stuttgart.
23. Muja, M. and Lowe, D.G. (2012) *Fast matching of binary features*. In: Ninth IEEE Conference on Computer and Robot Vision (CRV).
24. Cho, M. and Lee, K.M. (2012) Progressive graph matching: Making a move of graphs via probabilistic voting. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.