

## A Novel Approach for Protein Spots Quantification in Two-Dimensional Gel Images

Heshmat A.Rashwan<sup>1</sup>, Amany M. Sarhan<sup>2</sup>,  
Muhammed Talaat Faheem<sup>2</sup>, Bayumy.A.Youssef<sup>3</sup>

<sup>1</sup> *knowledge-Based Systems Department, Informatics Research Institute, Mubarak City for Science and Technology, Borg ElArab, Alexandria, Egypt*

<sup>2</sup> *Computers and Automatic Control Engineering Department, Faculty of Engineering, University of Tanta, Tanta, Egypt*

<sup>3</sup> *Computer-Based Applications Department, Informatics Research Institute, Mubarak City for Science and Technology, Borg ElArab, Alexandria, Egypt*  
*rashwan.shaheera@gmail.com*

### **Abstract**

*Two-dimensional polyacrylamide gel electrophoresis of proteins is a robust and reproducible technique. It is the most widely used separation tool in proteomics. Current efforts in the field are directed at development of tools for expanding the range of proteins accessible with two-dimensional gels. Proteomics was built around the two-dimensional gel. The idea that multiple proteins can be analyzed in parallel grew from two-dimensional gel maps. Proteomics researchers needed to identify interested protein spots by examining the gel. This is time consuming, labor extensive and error prone. It is desired that the computer can analyze the proteins automatically by first detecting then quantifying the protein spots in the 2-D gel images. In our previous work, we presented a new technique for segmentation of 2-D gel images using the fuzzy c-means algorithm using the notion of fuzzy relations. In this paper, we will describe the new relational fuzzy c-means algorithm (RFCM) and use it for automatic protein spots quantification. We will also use two methods to evaluate its performance: the unsupervised evaluation method and comparison with the expert spots quantification.*

**Keywords:** *2D gel images, Protein Spot Detection, Protein Spot Quantification, Fuzzy c-means algorithm, Fuzzy relations, and Computer vision.*

### **1. Introduction**

The last decade in life sciences was deeply influenced by the development of the “Omics” technologies (genomics, transcriptomics, proteomics, and metabolomics), which aim for a global view on biological systems. With these tools at hand, the scientific community is striving to build functional models to develop a global understanding of the living cell [1-3]. The analysis of the proteome as the final level of gene expression started out with techniques based on 2-D gel electrophoresis [4,5] and extended its reach with semi-gel-free and shot gun gel-free liquid chromatography–mass spectrometry (LC–MS)-based techniques in recent years.

Quantitative analysis based on LC-MS techniques is still in an early stage when considering available software and algorithms. Here, we focus on the computerized analysis of 2-D gels which are widely used in the scientific community. 2-D gels may separate up to 10,000 protein spots on one gel [6]. In a suitably equipped and experienced lab environment, 2-D gels are easy to handle, and they can be produced in a highly parallelized way.

On a proteome map, one can detect all spots of a whole experiment in a single gel image, whereas the average images proposed earlier suffer from dilution effects for weak and rare spots. The spots detected there can serve as a spot consensus pattern that is valid for the whole gel set of the experiment. The consensus spot pattern is then transferred according to the warping transform and used on all gels. This allows for 100% matching spots and, in turn, complete expression profiles for reliable statistical analysis [7, 8].

The goals of this step-Protein Spot detection- are to find the spot positions, find their surrounding boundary, and determine their quantities. There are two basic approaches that are used in current software: image segmentation and model-based quantitation.

The segmentation approach partitions the image into nonoverlapping segments, essentially classifying each pixel as belonging to a certain spot, or as being part of the background between spots. Spot boundaries and quantities are then derived from the spot's pixels.

The segmentation of the image can take various characteristics of the image into account: raw intensity, slope, and classification of pixels in the surrounding region. The advantage of this approach is that the image is clearly separated into spots and "non-spot" areas which are easy to assess by a user. If the software allows editing of spot boundaries, then any desired spot shape can, in principle, be obtained. Model-based approaches try to model a spot's intensity as a Gaussian normal distribution or some variant thereof. A spot's quantity and boundaries are then derived from the model. This paper is organized as follows: section 2 presents the previous developments in this area of research. In section 3, the original Fuzzy c-means algorithm was presented. In section 4, the new relational Fuzzy c-means algorithm was described. Section 5 shows the proposed protein spots quantification. Sections 6 and 7 present the experimental results and its discussion.

## 2. Previous developments

Previous developments in this area have employed a wide range of techniques, and a majority of applications address the matching of 2D gel electrophoresis images for proteins as opposed to DNA. This is because 2D gel electrophoresis images for proteins have a relatively more uniform background and are somewhat easier to work with than the 2D gel electrophoresis images for DNA.

Kim et. al.[9] proposed a hierarchical segmentation based on thresholding and the detection of watersheds. They first pre-process the images to remove noise and enhance contrast, then thresholding is applied which produces large regions. A watershed detection algorithm is then applied recursively on these regions until only a single blob is detected which is considered to be a spot. Their method relies on setting several parameters and is sensitive to noise, and 2D gel electrophoresis images typically contain noise.

Sugahara et. al.[10] smoothed image regions by averaging pixel intensities using an  $m \times m$  window and performed a thresholding operation which ultimately subtracted the background, and then created a binary image for spot detection. This method relies heavily on the selection of a proper threshold value which can cause either an over-segmentation of spots in some regions as well as an under-segmentation of spots in other regions.

Takahashi et. al.[11] performed image enhancement and smoothing before defining local maxima in order to label the spots. This method also relies on the definition of threshold values in order to function properly.

More recently, Morris et. al.[12] developed a very accurate and robust method of detecting spots in 2D gel electrophoresis images. Their process involves an "average gel" which is created by first using registration software to create an alignment of all gels being used. The pixel intensities are then averaged across the aligned gels. The gels are each denoised using the average gel and pinnacles (regions that are a local maximum in both the horizontal and vertical directions and above a certain threshold) are detected which denote the spot locations. A disadvantage of this method is the need to perform image registration as a pre-processing step and the need to define a threshold in order to determine which regions are pinnacles.

In their paper [13], Umer Z. Ijaz et al. presented a technique that uses the clustering techniques like K-mean and fuzzy C-mean to distinguish between different types of protein spots and unwanted artifacts.

Christopher S. Hoeflich et al [14] presented a new technique using the labeling of each image pixel as either a spot or non-spot and use a Markov Random Field (MRF) model and simulated annealing for inference. Neighboring spot labels are then connected to form spot regions.

Dimitris K. Iakovidis,et al [15] presented a novel approach to unsupervised protein spot detection in 2D-PAGE images based on a genetic algorithm. This algorithm searches within a multidimensional parameter space to determine, in parallel, the parameters of multiple diffusion models that optimally fit the characteristics of possible spots. The detection and quantification of the spots is achieved by superposition of diffusion functions modeling adjacent spots

In their paper [16] P. Tsakanikas et al. introduce the use of Active Contours without Edges coupled with Contour Transform - based image enhancement for extracting accurately the gel image foreground (regions with spots) from the background. They demonstrate, using both synthetic and real gel images, that the proposed approach extracts tight spot regions which do not include background areas but include almost all spots detected by PDQuest a popular commercial 2DGE image analysis package. Furthermore, their method does not require manual calibration for every new image in order to detect weak but often important "faint" spots.

In [17], Yoon et al, to find protein spots more accurately and reliably from gel images, propose Reversible Jump Markov Chain Monte Carlo method (RJCMCMC) to search for underlying spots which are assumed to have Gaussian-distribution shape. Their statistical method identifies very weak spots, restores noisy spots, and separates mixed spots into several meaningful spots which are likely to be ignored and missed. Their proposed approach estimates the proper number, centre-position, width, and amplitude of the spots and has been successfully applied to the field of projection reconstruction NMR PR-NMR) processing.

In this work, we intend to present a novel algorithm that uses the Fuzzy C-means algorithm as a primary step in the segmentation process.

### **3. The Fuzzy C-Means Segmentation Algorithm**

Fuzzy C-means method, also known as Fuzzy ISODATA, which was originally introduced by Bezdek in 1981 as an extension to Dunn's algorithm [18] is the most widely used fuzzy clustering algorithm in practice.

Fuzzy C-Means is a data clustering technique based on optimizing the objective function:

$$J(U, V) = \sum_{j=1}^c \sum_{i=1}^N (\mu_{ij})^m \|x_i - v_j\|^2 \quad (1)$$

It requires every data point in the data set to belong to a cluster to some membership degree. The purpose of the FCM is to group data points into different specific clusters. Let  $X = \{x_1, x_2, \dots, x_N\}$  be a collection of data. By minimizing the objective function (1),  $X$  is classified into  $c$  homogeneous clusters where  $\mu_{ij}$  is the membership degree of data  $x_i$  to a fuzzy cluster set  $v_j$ ,  $V = \{v_1, v_2, \dots, v_c\}$  are the cluster centers.  $U = (\mu_{ij})_{N \times c}$  is a fuzzy partition matrix, in which each  $\mu_{ij}$  indicates the membership degree for each data point in the data set to the cluster  $j$ .

The value of  $U$  should satisfy the following conditions:

$$\mu_{ij} \in [0,1], \quad \forall i = 1, \dots, N, \forall j = 1, \dots, C \quad (2)$$

$$\sum_{j=1}^c \mu_{ij} = 1, \quad \forall i = 1, \dots, N \quad (3)$$

The  $\|x_i - v_j\|$  is the Euclidean distance between  $x_i$  and  $v_j$ . The parameter  $m$  is called fuzziness index, which control the fuzziness of membership of each datum. The goal is to iteratively minimize the aggregate distance between each data point in the data set and cluster centers until no further minimization is possible.

The whole Fuzzy C-Means process can be described in the following steps.

**Step 1:** Initialize the membership matrix  $U$  with random values, subject to satisfying conditions (2) and (3).

**Step 2:** Calculate the cluster centre  $V$  by using following equation

$$v_j = \frac{\sum_{i=1}^N (\mu_{ij})^m x_i}{\sum_{i=1}^N (\mu_{ij})^m}, \quad \forall j = 1, \dots, C \quad (4)$$

**Step 3:** Get the new distance:

$$d_{ij} = \|x_i - v_j\|, \quad \forall i = 1, \dots, N, \forall j = 1, \dots, C \quad (5)$$

**Step 4:** Update the Fuzzy partition matrix  $U$ :

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}}, \quad \text{if } d_{ij} \neq 0 \quad (6)$$

Else  $\mu_{ij} = 1$

**Step 5:** If the termination criteria have been reached, then stop. Else go back to step 2.

The suitable termination criteria can be set by checking whether the objective function is below a certain tolerance value or if its improvement compared to the previous iteration is below a certain threshold. Moreover, the maximum number of iteration cycles can be used as a termination criterion as well.

#### 4. Protein Spot Detection Utilizing the Relational Fuzzy C-Means Algorithm

In this section, we present our new algorithm for matching protein spots in the 2-D gel images. We called the new algorithm the Relational Fuzzy C-means as it builds on the

traditional Fuzzy C-means algorithm but is modified by introducing the notion of fuzzy relations in order to differentiate spot pixels from the varying background.

The algorithm is composed of 4 steps where the first step of it to apply the FCM to the image to produce preliminary clusters. Then these clusters are then internally refined to identify the inner spots by separating the background pixel from the contained pixel in the cluster by applying steps 2 to 4 on the clusters.

A summary of the steps of the proposed algorithm is given below.

**Step1:** *Apply the FCM algorithm presented in section 2 with  $C$  more than 2. The output is the partitioning of pixels in the image to different clusters each having a center value  $v$ .*

**Step 2:** *For each two pixels  $x, y$  belonging to two different clusters, Create a Fuzzy Relation between  $x$  and  $y$  named  $I(x,y)$  where  $I(x,y)$  define the degree of closeness between intensities of pixels  $x$  and  $y$ .*

**Step 3:** *Compare pixels  $x, y$*

*- if pixel  $x$  is much more darker than pixel  $y$*

*then pixel  $x$  is a spot pixel*

*else if pixel  $x$  is much more lighter than pixel  $y$*

*then pixel  $y$  is a spot pixel*

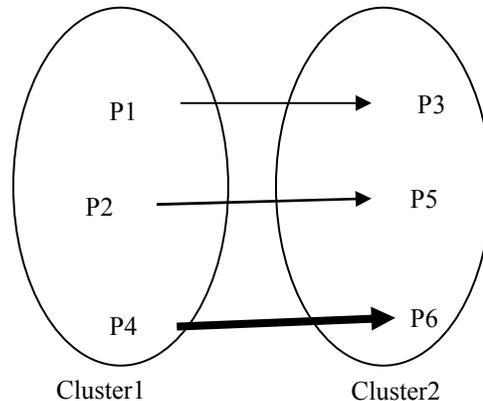
*else if difference between intensities is low and one of the pixels is a spot pixel*

*then the other is a spot pixel also*

*end if*

**Step 4:** *Mark spot pixels and differentiate them from the background. This is done by assigning the spot pixels to the maximum value of centers of clusters and the background-non spot pixels- to the minimum value of centers of clusters.*

In figure 1, a representation of a fuzzy relation between two points in two different clusters: cluster1 and cluster2,  $I(x,y)$ , is shown where  $x$  and  $y$  are any two points each in one different cluster. Here the arrow represents the degree of closeness between two pixels from different clusters.



**Figure 1: Representation of the fuzzy relation between pixels in two clusters.**

The dark arrow represents the strong relation between the two pixels in two different clusters.

## 5. Our proposed Protein Spots Quantification

The previous work concerning Protein Spots quantification was based on 2D gel matching and identification [27, 28] after the detection step. In this section, we propose a novel technique for protein spot quantification based on the assumption that all spots have approximately same size. Area of Protein spots can be identified from the sample preparation step. The SDS-PAGE (sodium dodecyl sulfate) in the 2nd Dimension separates proteins by their size (molecular weight, MW) and no other physical feature [24].

This means that the average protein size in a 2D gel image is well known by the expert user who had prepared the gel. Having this knowledge, we can estimate the number of pixels per protein spot- Drawing a circle having the same area and counting number of pixels inside the circle will yield to an appropriate pixels count of a protein spot. From the relational Fuzzy C-means segmentation step, the number of spot pixels in a 2D gel image is defined- refer to step 3 in the algorithm.

Now the quantification of protein spots in an image is computed as follows:

$$\text{Number of Protein Spots} = \frac{\text{Total no of Spot pixels}}{\text{no of pixels per spot}} \quad (7)$$

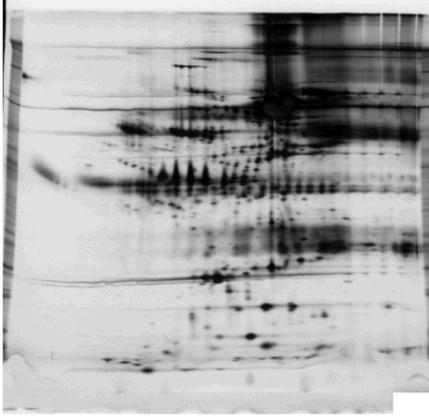
This method is computational inexpensive, however it will be proven to be accurate.

## 6. Experimental Results

The LECB 2-D PAGE gel images database is available for public use. It contains data sets from four types of experiments with over 300 gif images with annotation and landmark data in html, tab-delimited and xml formats. It could be used for samples of several types of biological materials and for test data for 2D gel analysis software development and comparison with other similar samples.

PAGE is polyacrylamide gel electrophoresis. The LECB was the U.S. National Cancer Institute's Laboratory of Experimental and Computational Biology. The database is available at two Web sites [20,21].

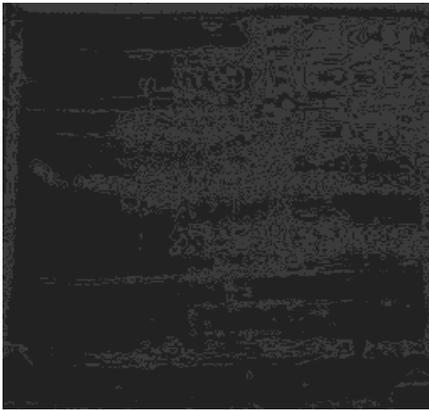
In our work we used these data and applied our algorithm to 2D gel images of Fetal Alcohol Syndrome cell lines. The results are shown in the following figures. Figures 2 and 6 show four test cases 2-D gel electrophoresis images of a Patient- Fetal Alcohol Syndrome. Figures 3 and 7 show the gradient images of the 2-D gel images presented in Figures 2 and 6. The results of applying the current Fuzzy C-means Segmentation algorithm on these images at  $C=2$ , are shown in figures 4 and 8. The results of applying the proposed Relational Fuzzy C-means Segmentation algorithm on these images at  $C = 6$  and  $\beta = 20$ , are shown in figures 5 and 9. Experimentally,  $\beta = 20$  was shown to be the best parameter estimation to reduce error since when we increased the parameter to 22 or decreased to 18, image quality was poor and error increased so much.



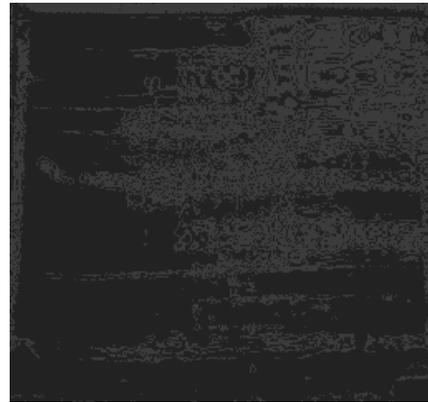
**Figure 2.** 2-D gel electrophoresis image of patient Fetal Alcohol Syndrome



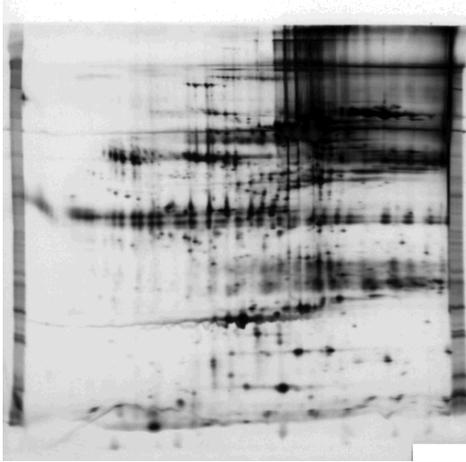
**Figure 3.** The gradient image of 2-D gel electrophoresis image of Patient Fetal Alcohol Syndrome in figure 2



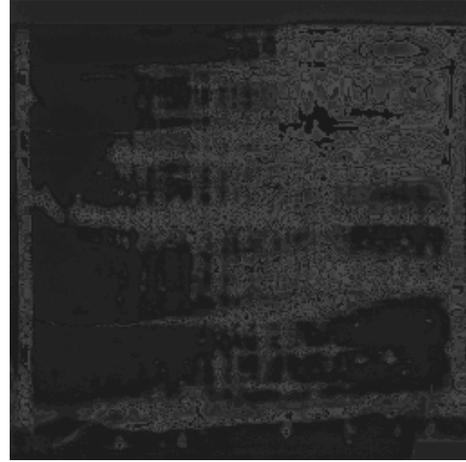
**Figure 4.** The gradient image of 2-D gel electrophoresis image of a Patient- human Fetal Alcohol Syndrome in figure 2 after applying the FCM, no of clusters =2



**Figure 5.** The gradient image of 2-D gel electrophoresis image of Patient Fetal Alcohol Syndrome in figure 2 after applying the RFCM, no of clusters =6,  $\beta = 20$



**Figure 6. 2-D gel electrophoresis image of patient Fetal Alcohol Syndrome**



**Figure 7. The gradient image of 2-D gel electrophoresis image of Patient Fetal Alcohol Syndrome in Figure 6**



**Figure 8. The gradient image of 2-D gel electrophoresis image of Patient Fetal Alcohol Syndrome in figure 6 after applying the FCM, no of clusters =2**



**Figure 9. The gradient image of 2-D gel electrophoresis image of Patient-human Fetal Alcohol Syndrome in figure 6 after applying the RFCM, no of clusters =6,  $\beta = 20$**

The new technique shows high performance and detects the protein spots precisely, as shown in Figures 5 and 9 even the less dark spots in the image appears (shown by squares) while in Figures 4 and 8 when applying the FCM algorithm those protein spots disappear totally which affects the spot quantization step in the whole process of 2-D gel image analysis.

### 6.1. The evaluation error $E_{CW}$

In this section, we will use the  $E_{CW}$  evaluation error [22] to evaluate the performance of the new algorithm the relational Fuzzy C-means algorithm (RFCM) versus the original Fuzzy C-means algorithm (FCM).  $E_{CW}$  is a composite evaluation method for color images. It uses intra-region visual error to evaluate the degree of under-segmentation, and uses inter-region region visual error to evaluate the degree of over-segmentation.

1)  $E_{in.tra}$  of  $E_{CW}$

$$E_{in.tra} = \frac{\sum_{p \in I} \mu(\|C_x^o(p) - C_x^s(p)\|_{L^*a*b} - TH)}{S_I} \quad (8)$$

where  $C_x^o(p)$  and  $C_x^s(p)$  are pixel feature value(color components in CIE  $L^*a^*b$  space) for pixel  $p$  on original and segmented image, respectively,  $TH$  is the threshold to judge significant difference, and  $\mu(t) = 1$  when  $t > 0$ , otherwise  $\mu(t) = 0$ .

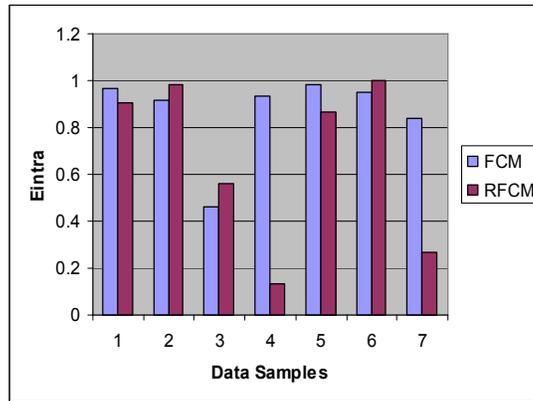
2)  $E_{in.ter}$  of  $E_{CW}$

$$E_{in.ter} = \sum_{i=1}^N \sum_{j=1, j \neq i}^N [\mu(TH - \|C_x^o(p) - C_x^s(p)\|_{L^*a*b}) \cdot w_{ij} / (S_I \cdot Z)] \quad (9)$$

where  $w_{ij}$  denotes the jointed length between  $R_i$  and  $R_j$ ,  $TH$  is the threshold to judge significant difference, and  $Z$  is a normalization factor. We set the threshold  $TH = 10$  and the normalization factor  $Z = 100$  for no of clusters = 2.

**Table 1. The intra region error of the Fuzzy C-means algorithm and the Relational Fuzzy C-means algorithm on seven data samples.**

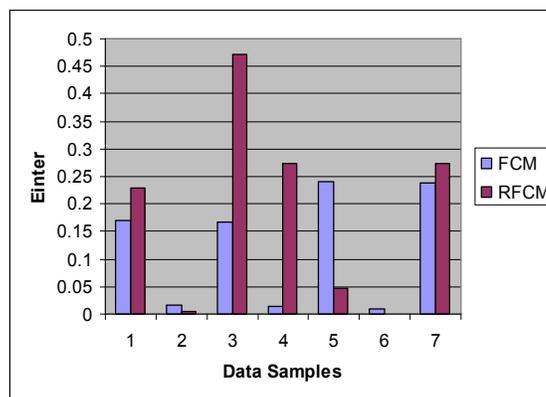
No	Eintra Segmented Image by FCM	Eintra Segmented Image by RFCM
1	0.96712	0.90608
2	0.91501	0.98128
3	0.46035	0.55992
4	0.93101	0.136
5	0.98302	0.86736
6	0.95079	0.99768
7	0.83952	0.26396



**Figure 10.** The intra region error of the Fuzzy c-means algorithm and the Relational Fuzzy c-means algorithm on seven data samples.

**Table 2** The inter region error of the Fuzzy C-means algorithm and the Relational Fuzzy C-means algorithm on seven data samples.

No	Einter Segmented Image by FCM	Einter Segmented Image by RFCM
1	0.17	0.22978
2	0.01604	0.0040217
3	0.16751	0.47053
4	0.015092	0.27255
5	0.24	0.047613
6	0.0091133	0
7	0.23718	0.27355

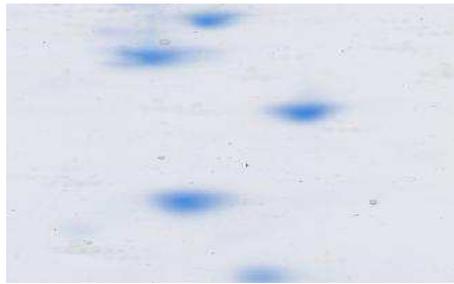


**Figure 11.** The intra region error of the Fuzzy c-means algorithm and the Relational Fuzzy c-means algorithm on seven data samples.

The shaded cells in the two Tables 1 and 2 represent the improvement caused by the new algorithm versus the original one. Notice that for the Eintra evaluation metric, the new RFCM algorithm enhanced the results on 4 data samples over 7 data samples (Percentage 57%) and in case of non-improvement the difference was about 10% in the worst case (data sample 3) which means that the new algorithm was able to detect protein spots more precisely. Moreover, according to the Einter evaluation metric, the new RFCM algorithm – except for data samples 3 and 4- had reduced the over-segmentation or at least didn't increase so much its rate.

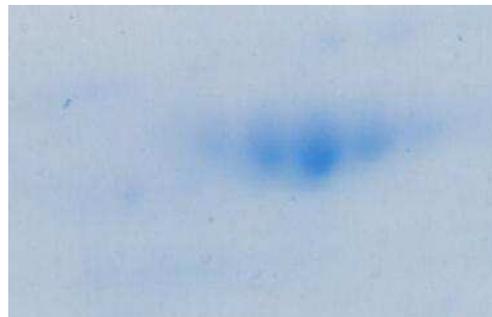
## 6.2 Evaluation by Expert user

Images used in this section were provided by Prof Dr Maha ElDemellawy, Medical Biotechnology Department, Mubarak city for Science and Technology, Borg ElArab, Alexandria, Egypt. We applied the Relational Fuzzy C-means algorithm to compute total number of spot pixels in the image while we knew the number of pixels per spot by choosing, with the help of the expert user, a spot and calculating its size and number of pixels inside it.



**Figure 12. 2D Gel Electrophoresis image of human myocardial membrane proteins**

Figure 12 shows 2D Gel Electrophoresis image of human myocardial membrane proteins. In a 128x128 2DGE image, Pixels per Spot (average spot size in pixels) = 176 pixels.  
No. of spot pixels according to the RFCM algorithm = 1850 pixels  
No. of Protein spots (automatic quantification using equation 7) =  $1850/176 = 10.5111$   
Protein Spots (i.e about 10 spots)  
No. of Protein spots (detected by Expert user quantification) = 10-15 Protein Spots



**Figure 13. 2D Gel Electrophoresis image of human myocardial membrane protein 50-54kDa fraction**

Figure 13 shows 2D Gel Electrophoresis image of human myocardial membrane protein 50-54kDa fraction. In a 128x128 2DGE image, Pixels per Spot (average spot size in pixels) = 200 pixels.

No. of spot pixels according to the RFCM algorithm = 4075 pixels

No. of Protein spots (automatic quantification using equation 7) =  $4075/200 = 20.375$  Protein Spots (i.e about 20 spots)

No. of Protein spots (detected by Expert user quantification) = 15-18 Protein Spots

## 7. Discussion

In this work, we presented a new algorithm based on the notion of fuzzy relations to segment and detect protein spots in 2-D gel electrophoresis images and used it for automatic quantification of protein spots.

For future work, we suggest the development of fuzzy relations to obtain better results. The parameters representing the degree of closeness must be defined for enhancement and improvement of the RFCM algorithm. The use of intuitionistic fuzzy relations to identify the degree of non-closeness between pixels and the hesitation margin can be investigated also. The automatic quantification of Protein Spots in 2D gel electrophoresis images may open a new direction of research in the proteomics field. Knowing that there are diseases, like cancer[25,26], that affect the quantity of Protein spots in 2DGE images, we can suggest a machine learning system that uses the quantity of protein spots in a 2DGE image as a factor that classifies the cells into normal cells and abnormal cells and this can be considered as a step towards diagnosis of a disease using 2DGE images analysis.

Now a question is raised: Is the quantity of protein spots in a 2D gel images the only factor affected by Cancer disease or the color intensity or the shape too. The existence of more than one factor means that the suggested machine learning system can take an input vector of multi-dimensions which yield to more accurate classification of normal and abnormal cells.

## Acknowledgment

Special thanks for Prof.Dr. AbdelFattah AbdelMongy Sadakah, president of University of Tanta and Prof Dr. Yasser Refat, president of Mubarak City for science and technology.

## References

- [1]N.G. Anderson, A. Matheson, N.L. Anderson, "Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology", *Proteomics* 1:3, 2001.
- [2] V.C. Wasinger, S.J. Cordwell, A. Cerpa-Poljak, J.X. Yan, A.A. Gooley, M.R. Wilkins, M.W. Duncan, R. Harris, K.L. Williams, I. Humphrey-Smith, "Large-scale amino-acid analysis for proteome studies", *Electrophoresis*, 16, 1995.
- [3] W.P. Blackstock, M.P. Weir, "Proteomics: quantitative and physical mapping of cellular proteins", *Trends Biotechnol.* 17 (1999).
- [4]O'Farrell PH, "High resolution two-dimensional electrophoresis of proteins", *Journal of Biological Chemistry*, 250:4007-4021, 1975.
- [5] Klose J, "Protein mapping by combined isoelectric focusing and electrophoresis in mouse tissues: A novel approach to testing for induced point mutations in mammals", *Human genetk* 26:231-243, 1975
- [6]Klose J, Kobalz U., "Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome", *Electrophoresis* 16:1034-1059, 1995.

- [7] Voigt B, Schweder T, Sibbald MJ, Albrecht D, Ehrenreich A, Bernhardt J, Feesche J, Maurer KH, Gottschalk G, van Dijk JM, Hecker M, "The extracellular proteome of *Bacillus licheniformis* grown in different media and under different nutrient starvation conditions", *Proteomics* 6(1):268–281, 2006.
- [8] Höper D, Bernhardt J, Hecker M, "Salt stress adaptation of *Bacillus subtilis*: a physiological proteomics approach", *Proteomics* 6(5):1550–1562, 2006.
- [9] Youngho Kim, JungJa Kim, Yonggwon Won and Yongho In, "Segmentation of protein spots in 2d gel electrophoresis images with watersheds using hierarchical threshold", *LNCS - Computer and Information Sciences - ISCIS 2003*, 2869:389-396, 2003.
- [10] Y. Sugahara, Y. Hayashizaki, I. Tanihata, "An automatic image analysis system for RLGS\_LMS", *Mammalian Genome*, 9:643-651, 1998.
- [11] Takahashi, K., Nakazawa, M., and Watanabe, Y., "DNAinsight: An image processing system for 2-d gel electrophoresis of genomic DNA", *Genome Informatics*, 8:135-146, 1997.
- [12] Jeffrey S. Morris, Brittan N. Clark and Howard B. Gutstein "Pinnacle: a fast, automatic and accurate method for detecting and quantifying protein spots in 2-dimensional gel electrophoresis data", *Bioinformatics*, 24(4):529-536, 2008.
- [13] Umer Zeeshan Ijaz, Safee Ullah Chaudhary, Moon Sang Don, Kyung Youn Kim, "Computational Strategies for Protein Quantitation in 2D Electrophoresis Gel Image Processor for Matlab", *Proceedings of the 2007 Frontiers in the Convergence of Bioscience and Information Technologies*, 129-134, 2007.
- [14] Christopher S. Hoeflich and Jason J. Corso, "Segmentation of 2D Gel Electrophoresis Spots Using a Markov Random Field", In *Proceedings of SPIE Conference on Medical Imaging*, 2009
- [15] D.K. Iakovidis, D. Maroulis, E. Zacharia and S. Kossida, "A Genetic Approach to Spot Detection in two-Dimensional Gel Electrophoresis images", in *Proceedings of International Conference on Information Technology in Biomedicine (ITAB)*, Ioannina, Greece, Oct 2006
- [16] Panagiotis Tsakanikas and Elias S. Manolakas, "Active Contours Based Segmentation of 2DGE Proteomics Images", in *Proceedings of Eusipco 2008*.
- [17] Ji Won Yoon, Simon J. Godsill, ChulHun Kang and Tae-Seong Kim Bayesian, "Inference for 2D Gel Electrophoresis Image Analysis", *Bioinformatics Research and Development* p 343-356, 2007.
- [18] Dunn JC, "A fuzzy relative of the ISODATA process and its use in detecting compact, well separated clusters", *Journal of Cybernetics*, 3(3):32-57, 1974.
- [19] Matthias Berth, Frank Michael Moser, Markus Kolbe, Jörg Bernhardt, "The state of the art in the analysis of two-dimensional gel electrophoresis images", *Applied Microbiology Biotechnology* 76:1223–1243, 2007.
- [20] [www.ccrn.ncifcrf.gov/2DgelDataSets](http://www.ccrn.ncifcrf.gov/2DgelDataSets)
- [21] [bioinformatics.org/lecb2dgeldb](http://bioinformatics.org/lecb2dgeldb)
- [22] H.-C. Chen, S.-J. Wang, "The use of visible color difference in the quantitative evaluation of color image segmentation," in *Proc. ICASSP*, 2004.
- [23] Shaheera Rashwan, Talaat Faheem, Amany Sarhan, B.A. Youssef, "A relational fuzzy c-means algorithm for detecting protein spots in two-dimensional gel images" accepted to be published in *Advances in Computational Biology*, Springer 2010
- [24] Matthias Berth, Frank Michael Moser, Markus Kolbe, Jörg Bernhardt "The state of the art in the analysis of two-dimensional gel electrophoresis images" *Applied Microbiology Biotechnology* (2007) 76:1223–1243
- [25] Ge Zhou, Hongmei Li, Dianne DeCamp, She Chen, Hongjun Shu, Yi Gong, Michael Flaig, John W. Gillespie, Nan Hu, Philip R. Taylor, Michael R. Emmert-Buck, Lance A. Liotta, Emanuel F. Petricoin, and Yingming Zhao "2D Differential In-gel Electrophoresis for the Identification of Esophageal Scans Cell Cancer-specific Protein Markers" *Molecular & Cellular Proteomics* 1:117-123, 2002.
- [26] Gene Kim and MyungHo Kim, "2D Electrophoresis Gel Image and Diagnosis of a Disease" *CoRR cs.CC/0305048*: (2003)
- [27] Steven P. Gygi and Ruedi Aebersold, "Absolute Quantitation of 2-D Protein Spots" *Methods in Molecular Biology*, Vol.112, pages 417-421, February 2008
- [28] Ijaz, U.Z.; Chaudhary, S.U.; Moon Sang Don; Kyung Youn Kim, "Computational Strategies for Protein Quantitation in 2D Electrophoresis Gel Image Processor for Matlab", *Frontiers in the Convergence of Bioscience and Information Technologies*, 2007. FBIT 2007

## Author



### **Heshmat Rashwan**

PhD student in computer science and automatic control department, Faculty of Engineering, Tanta University. Currently, Eng. Rashwan is a research assistant in Informatics Research Institute, Mubarak City for science and Technology, Burg El Arab, Alexandria, Egypt

