GENETIC BASED SPOT DETECTION METHOD IN TWO-DIMENSIONAL ELECTROPHORESIS IMAGES

CarlosFernandez-Lozano,Jose A. Seoane, Daniel Rivero, Julian Dorado

***Abstract**: The image analysis of two-dimensional electrophoresis images is a difficult task were authors were not able to find any other work in the literature handling with evolutionary computation in combination with a second order operator for edge detection. In this work, a novel Genetic Algorithm-based protein detection method from two-dimensional electrophoresis gel images is presented. Such a method makes use of a second order operator for edge detection by means of a Genetic Algorithm-based technique. The proposed method is able to detect proteins in two-dimensional gel images, but a reduction in the False Positive ratio is necessary. A manually selection process should be done by the clinicians to reduce this ratio; this represents a bottleneck due to the number of proteins in each imagein order to discriminate real proteins detected. The goal here was to avoid the loss of time caused by the manual revision of proteins detected by the image analysis software packages. To decrease this ratio, binary and real coded Genetic algorithms were probed and BLX-alpha crossover function was chosen. A comparative test with Z3 and Melanie 3.0, two-dimensional electrophoresis image analysis software packages, is done in order to check the accuracy of the proposed method. All images used for these tests are available on the Internet (http://www.umbc.edu/proteome).*

***Keywords**: Image Analysis, Real Coded Genetic Algorithm, Optimization, Electrophoresis, Edge Detection.*

***ACM Classification Keywords**: F.1.1 Models of Computations – Self-modifying machines. I.5.2 Design Methodology – Classifier design and evaluation. I.5.4 Applications – Computer Vision. I.4.6 Segmentation – Edge and feature detection.*

Introduction

Proteomics is the study of protein properties in a cell or tissue to obtain a global integrated view of disease, physiological and biochemical processes of cells and regulatory networks. One of the most powerful and widely used techniques to analyze complex protein mixtures extracted from cells, tissues, or other biological samples are two-dimensional polyacrylamide gel electrophoresis. In this method, proteins are classified by molecular weight (MWt) and iso-electric (pI) using a controlled laboratory process and digital imaging equipment. Most diseases are manifested at the level of protein activity; hence genetic information itself is not enough to predict protein profiles in a sample. In this field, to discover the genetic component of several diseases, association studies are usually used to establish a statistical relationship between genetic variables of a population and a phenotype[Pendergrass et al 2011]. Proteomics seeks to correlate directly the involvement of specific proteins in a given disease state. Separation of proteins of a sample could be done with several different techniques such as chromatography, mass spectrometry and two-dimensional gel electrophoresis among others. In recent years there has been a transfer of research interests from genomics to proteomics [Tyers et al. 2003].

Two-dimensional gel electrophoresis (2D-electrophoresis) has been, from the beginning of the proteomic research, the main protein separation technique, and even before than proteomics became a reality itself. The main advantages of this approach are its robustness, its parallelism and its unique ability to analyze complete proteins at high resolution, keeping them intact and being able to separate them entirely [Rabilloud et al. 2010].However, it presents also several drawbacks as its very high sensitivity to the dynamic range and quantitative distributions issues.

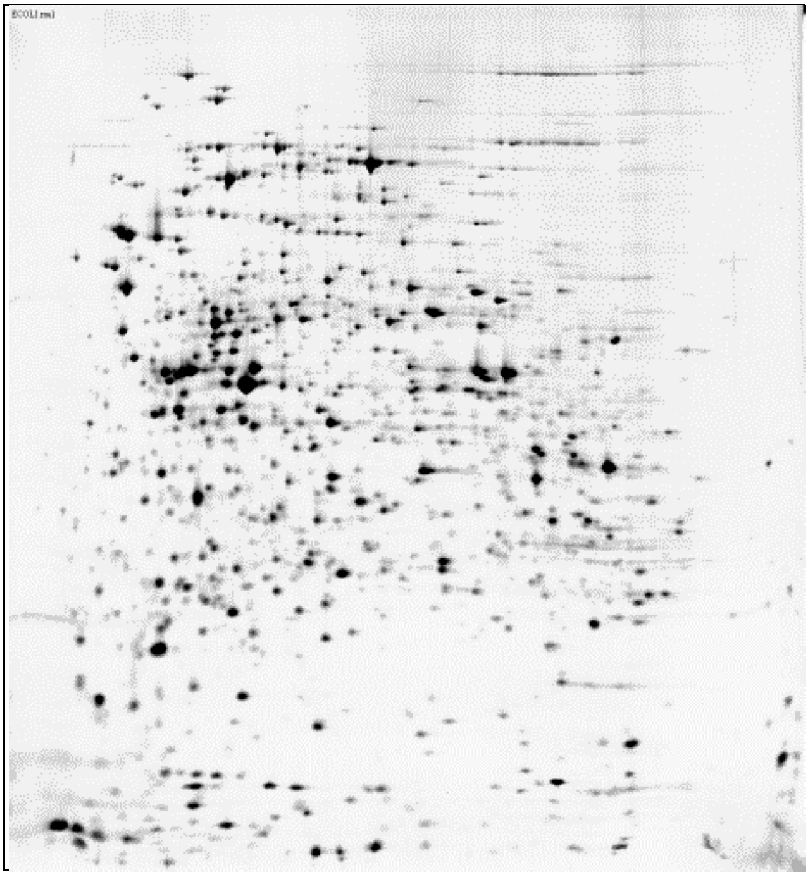

Figure 1. Standard two dimensional gel image. Downloaded from http://www.umbc.edu/proteome/images01

To deal with this kind of images, Figure 1, is a difficult task, because even if there is a partial list of online gel databases where its possible to download 2D gel images (http://www-lecb.ncifcrf.gov/EP/table2Ddatabases.html), there is not a commonly accepted ground truth, but existent as proposed in [Lemkin; Marten]. Another aspect that makes the work difficult from a computer vision point of view is that both proteins and background noise seem to follow a Gaussian distribution [Tsakanikas 2009]. The inter- and intra-operator variability in manual analysis of these images is also a big drawback [Millioni et al. 2010]. The main problems in analyzing these images are due to the fact that the proteins will be expressed in different ways in different images, being common that several groups of proteins appear only in one of the images of the pair. Furthermore, there are several problems related to the experimental conditions and acquisition of the images, which make the correlation a very complex task. Thus, it is expected a high variability in the images. Hence, 2D-electrophoresis denoising is of great importance [Xin et al. 2011], and several approaches have been proven (spatial filtering, wavelets, median-modified wiener filter, total variation, among others) but work is on going.

The typical workflow in a 2D-electrophoresis analysis starts with a protein detection process, which appear in the images as dark spots, and where the size of the spot depends on the amount of protein in the sample, continued with a matching process between images and ending with a statistical analysis of the expression differences among matched proteins.

This work is focused in the first step of the workflow, protein detection. Most of the methods proposed in the literature, present very simples algorithms; their simplicity causes a high false positive rate (FPR) [Rogers et al. 2005]. A manually selection process should be done by the clinicians to reduce this FPR, this represents a bottleneck due to the number of proteins in each image [Voss et al. 2000]. Several software packages seek to solve the protein analysis workflow [Dowsey et al. 2010]. Focusing on the protein detection, these software packages implements different segmentation methods ranging from smoothing or morphological operators [Raman et al. 2002], watershed-based [Conradsen et al. 1992] to geometric algorithms for edge detection [Efrat 2002]. There is also a detection method that is based on a surface criterion, not on an intensity criterion in the 2D-electrophoresis images surface, proposed by [Langella et al. 2008]. The method proposed in this work intends to be as much conservative as possible due to the time cost that implies for a clinician the analysis of an image with hundred of false proteins detected in order to discard them.

In order to evaluate the accuracy of this work, [Marten] dataset is used.It is composed with two real 2D-electrophoresis images, with proteins manually selected, and eleven artificial images created with Matlab for spot quantification tasks. The images are in tiff format, 8-bit greyscale and the artificial images with 300 dpi. This dataset were used in a quantitative comparison and evaluation of two software packages [Raman et al. 2002], so we used this work to compare the accuracy of our method.

The aim of this paper is to develop an Evolutionary Computation-based method for the optimization of a second order operator for protein detection in 2D-electrophoresis images.

Theoretical Background

The method proposed in this work intends to assist in 2D-electrophoresis image analysis by studying a Laplacianbased edge detector for protein detector by means of a Genetic Algorithm-based technique.

Related Works

The authors were not able to find any other work in the literature handling with evolutionary computation in combination with a second order operator for edge detection with 2D-electrophoresis images, however [Iakovidis et al. 2006] proposed an unsupervised protein spot detection in 2D-electrophoresis images based on a genetic algorithm. This approach involves three steps: they starts performing a wavelet-based noise reduction step, following with a segmentation process into regions around the local maxima of the image intensities and finishing with a detection and model-based quantification of the spots within each region using a GA. This algorithm searches the parameters of multiple diffusion models that fit the characteristics of possible spots. In their work, each chromosome is a fixed length string of real values representing the parameters of the corresponding diffusion model. As this real-coded genetic algorithm (RCGA) uses a uniform crossover function, a high mutation probability of 0.8 was chosen.

Genetic Algorithm

Briefly, Genetic Algorithms (GAs) are a search technique for solving optimization problems based on the principles of genetics and natural selection. It was developed by Holland in the 1970s [Holland 1975] inspired by Darwin's theory of evolution, the survival of the fittest. An initial population of individuals is evolved by means of genetic operators: selection, crossover and mutation. The selection operator ensures the survival of the fittest,

while the crossover represents the mating between individuals, and the mutation operator introduces random modifications. GAs possesses effective and efficient exploration and exploitation to simultaneously explore and exploit a large number of possible solutions [Goldberg 1989]. Using the crossover operator, GA combines the information for parents to produce new solutions, in order to better exploit the good solutions found in the search space. By making use of the mutation operator, new information is introduced in the population in order to explore new and promising areas of the search space. At the end of the process, GA makes the population of solutions converge to the global optimum. Hence, beginning from an initial population, and evaluating it according to a quality criterion (fitness function), new populations are created using crossover and mutation operators, and individuals are selected according to their fitness value to become a part of this new generation.

A chromosome in the population, which represents candidate solutions to the concrete problem, is commonly a string of binary numbers with the same length of the vector [Goldberg 1990], which is solution to the problem. Not all the problems are in accordance with a binary coded solution. For this reason, non-binary representations such as floating point number codification of the chromosome are used. This codification appears in the late 80's [Lucasius et al. 1989] and is mainly use since then for numerical optimization on continuous domains.

The crossover operator combines the features of two parents to form two offspring; it means that this operator is a method for sharing information. In binary-coded GA the classical crossover operator is the simple crossover, but there are also more powerful operators such as the multi-point and uniform crossover. In the case of RCGA, each crossover operator generates a different number of new solutions, in the way that is sometimes necessary that this operator is accompanied by some selection criterion. Numerous crossover operators have been developed for RCGAs, starting in the latest 80's with the simple crossover developed by Goldberg or the flat crossover operator proposed by Radcliffe. These initial crossover operators implements a depth search methodology that can lead to a premature convergence of the population, because the new offspring is generated only in the region enclosed by the parents. In order to avoid this problem, new crossover operator as blend crossover (BLX-alpha)[Eshelman et al. 1993], flat crossover (is equivalent with BLX-0.0) or simulated binary crossover (SBX) [Agrawal et al. 1994] among others.

Difference of Gaussians

Difference of Gaussians (DoG) is a computer vision technique for edge (notable grey level variations between sides) detection using a second order Laplacian of a Gaussian based filter (LoG). Second order derivative (SOD) methods finds edges by looking for zero-crossing points after filtering the image. These methods are very sensitive to noise, so a filtering process is required in order to decrease the high frequency components of the noise that quadratically increases in the frequency domain. Hence, SOD usually does not be used for edge detection without a smoothing filter. The first step intends to smooth the image with a Gaussian filter, finishing with an edge enhancement process convolving a Laplacian operator through the image.

LoG is an isotropic operator, hence applies equally in all directions, and gives both positive and negative values. It finds regions within an image with outstanding changes in the intensity values, being L(x,y) the Laplacian of an image, and I(x,y) the pixel's intensity value, the Laplacian is defined in (1) as

$$\mathrm{L}(\mathrm{x},\mathrm{y}) = \frac{\partial^2 \mathrm{I}}{\partial \mathrm{x}^2} + \frac{\partial^2 \mathrm{I}}{\partial \mathrm{y}^2} \tag{1}$$

The calculation of the Laplacian begins with the convolution of a 2-D Gaussian kernel for smoothing the image. A two dimensional Gaussian kernel with equal variance value, is defined in (2) as

$$\mathrm{kernel} = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\left(\frac{\mathrm{x}^2+\mathrm{y}^2}{2\sigma^2}\right)} \tag{2}$$

With the smoothed image, derivatives are calculated in (3) in the way that

$$\mathrm{LoG} = -\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

This process is a very expensive computational task, so LoG is approximated with de DoG (4)operator that is calculated as the difference of two Gaussians with different standard deviation values

$$\mathrm{DoG} = \frac{1}{\sqrt{2\pi}}\left[\frac{1}{\sigma_1} e^{-\frac{(x^2+y^2)}{2\sigma_1^2}} - \frac{1}{\sigma_2} e^{-\frac{(x^2+y^2)}{2\sigma_2^2}}\right] \quad (4)$$

Edges in the nature are multiscale structures. Marr and Hildreth [Marr et al. 1980] firstly proposed an edge detection method that exploits this, with a multiscale edge detection filter. Hence, that physical phenomenon may be detected over different spatial sizes or scales.

Scale Space

Witkin introduce the concept of scale-space [Witkin 1983] and a method for extract the information within images across a continuum of scales. The scale-space of an image f(x,y) is defined as all zero-crossings of its LoG in the way that

$$\{\psi(x, y; \sigma)\} = \{(x, y; \sigma)\} | \zeta(x, y; \sigma) = 0, \quad (5)$$

$$(\partial\zeta/\partial x)\text{^}2 + (\partial\zeta/\partial y)\text{^}2 \neq 0, \sigma > 0, \quad (6)$$

$$\zeta(x, y; \sigma) = \{\nabla^2 g(x, y; \sigma) * f(x, y)\}. \quad (7)$$

Sigma ranges from zero to infinite, so continuous surfaces are formed in the scale-space. There is no loss of information using a Gaussian filter for smoothing an image, because this filter determines signals unequivocally associating a spatial constant with it [Wu et al. 1990].

Proposed Method

As regards with DoG, it is necessary to provide values to the following parameters:

- Scales number.
- Gaussian's sigma value for the first DoG.
- Scaling factor for the successive DoG's sigma values. It must be applied with the previous sigma value and the result must be higher than it. Convolving a kernel with a low sigma value leads to the detection of small size edges; hence it is more noise sensitive. On the contrary, convolving a kernel with a high sigma value in the scale space leads to smooth the signal, and only the most significant edges are detected.
- For each scale, kernel's size is modified. Chebyshev's Inequality states that for any normal distribution, 99,7 % of results will fall between +3 and -3 standard deviations. Taking this into account and avoiding the odd results, kernel's size s is calculated as

$$s = 3\sigma.2 + 1 \tag{8}$$

- A threshold in percentage from the maximum value of the detected pixels' grey level is calculated in order to consider remaining pixels as belonging to a protein. With these pixels, a region of interest is generated. The centroid of this region is considered as the center of the protein.

A fixed length RCGA is proposed in combination with a DoG for protein detection in 2D-electrophoresis images. Each individual in the population is in real-coded codification with a fixed length of 4 gens.

- Scales number. Ranging between 2 and 5.
- Gaussian`s sigma value for the first DoG. Ranging between 1 and 20.
- Scaling factor for the successive DoG's sigma values. Ranging between 0,5 and 2.
- A threshold. Ranging between 0 and 1.

Parameters are normalized ranging between 0 and 1, such that

$$v' = \frac{v - \min(r)}{\max(r) - \min(r)} * (\max(r') - \min(r')) + \min(r') \tag{9}$$

Where v' is the new value, v is the value to be normalized and r and r' are the initial and the final range of values.

Fitness function

The fitness function in this work is defined as the sum of two factors, one related to the distance measured between the proteins detected and the ground truth, one related with the number of proteins detected and another one related with how the protein detected fits to a Gaussian peak.

Distance

Distance is measured between the proteins detected and the ones belonging to the ground truth using the Minkowski distance,

$$L_q(\vec{x}, \vec{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^q \right)^{\frac{1}{q}} \tag{10}$$

more concretely when q=1, known as the Manhattan distance.

$$L_1(\vec{x}, \vec{y}) = \sum_{i=1}^{n} |x_i - y_i| \tag{11}$$

In order to compare this method with Marten's ground truth, as it is not a list of centers coordinates in the dataset, we use the method proposed in [Langella et al. 2008] to calculate the centers of the proteins, avoiding proteins detected not originally in the Marten's dataset. It is important to note that centers in [Langella et al. 2008] are calculated as the point in the image where most of the closest surroundings pixels migrate, according with the maximum slope direction. In this paper, centers are calculated using the centroid of the region detected, so a slight difference is anticipated, ranging between 0 and 6 pixels, when detecting the same protein in the image.

Number of proteins

Pondering the number of proteins detected with the number of proteins detected in the ground truth is done with the following factor

$$\text{factor} = \left| 1 - \frac{\text{size}(d)}{\text{size}(r)} \right| \tag{12}$$

Experimental Results

Experiments were performed to evaluate the performance of the proposed method on a set of real 2D-electrophoresis images digitized at 727x666-pixels and 998x1087 at 8-bit grey level depth. Each gel contains approximately 1000 proteins. For detailed methods and guidelines used for counting the protein spots manually, please refer to http://www.umbc.edu/proteome. Also a set of artificial 2D-electrophoresis images generated with Matlab and digitized at 400x400-pixels at 8-bit grey level depth, and 300 dpi was used. With the set of artificial gels a 100% of true positive (TP) protein detected was achieved, with no false positives (FP) for the whole set of images. These results as expected are due to the fact that this set is not very complicated. There is not noise in the image, and proteins are perfectly isolated, with a perfect Gaussian shape, and the point with a 255 grey level value is in the middle of the spot, and it is the center of it.

A population of 500 individuals was used, and two-point, uniform, SBX and BLX-alpha crossover functions were tested. In each generation 5 individuals of the population were maintained. A high mutation probability is necessary with the two-point and uniform crossover functions. This is due to the real-coded codification of the chromosome and that these crossover functions have better results with BCGA, and it is necessary to add new genetic information to the offspring according with[Janikow et al. 1991], RCGA need a high mutation rate to provide with enough diversity. For this fitness functions, this value is evaluated with values ranging from 0.7 to 0.9. The crossover probability is ranging from to 0.8 to 0.95. In the case of the BLX-alpha, the alpha value is ranging from 0.1 to 0.5. Better results where obtained with alpha=0.5.

The mean and the median of the errors were calculated and it is observed that with the median, the FPR decrease significantly. Table 1 shows the results for each crossover function with the median. In [Raman et al. 2002] a quantitative test was devised to determine how well Melanie 3.0 and Z3 (image analysis software packages) detect proteins spots. In this comparision they obtained true, false and extraneous spots. They compared the manual spot detection on a spot-by-spot basis to identify the spots missed or no detected, and extraneous spots was then calculated using formula (13).

$$\text{extraneous spot count} = software\ count - manual\ spot\ count + total\ missed\ spots \quad (13)$$

Table 1.Spot detection analysis with gel (a) and gel (b) and different crossover functions. Total spots detected expressed as percentage of the manual count.

GEL	FUNCTION	TRUE SPOTS	FALSE SPOTS	EXTRA. SPOTS
A	Uniform	39	61	280
	Two-point	57	43	63
	SBX	52	48	85
	BLX-0.5	81	19	17
B	Uniform	41	59	217
	Two-point	50	50	43
	SBX	46	54	51
	BLX-0.5	76	24	27

Obtained results in Table 1, are graphically represented in Figure 2. These results indicate that the crossover function with less extraneous spots detected is the BLX-0.5 crossover function for both gel images.

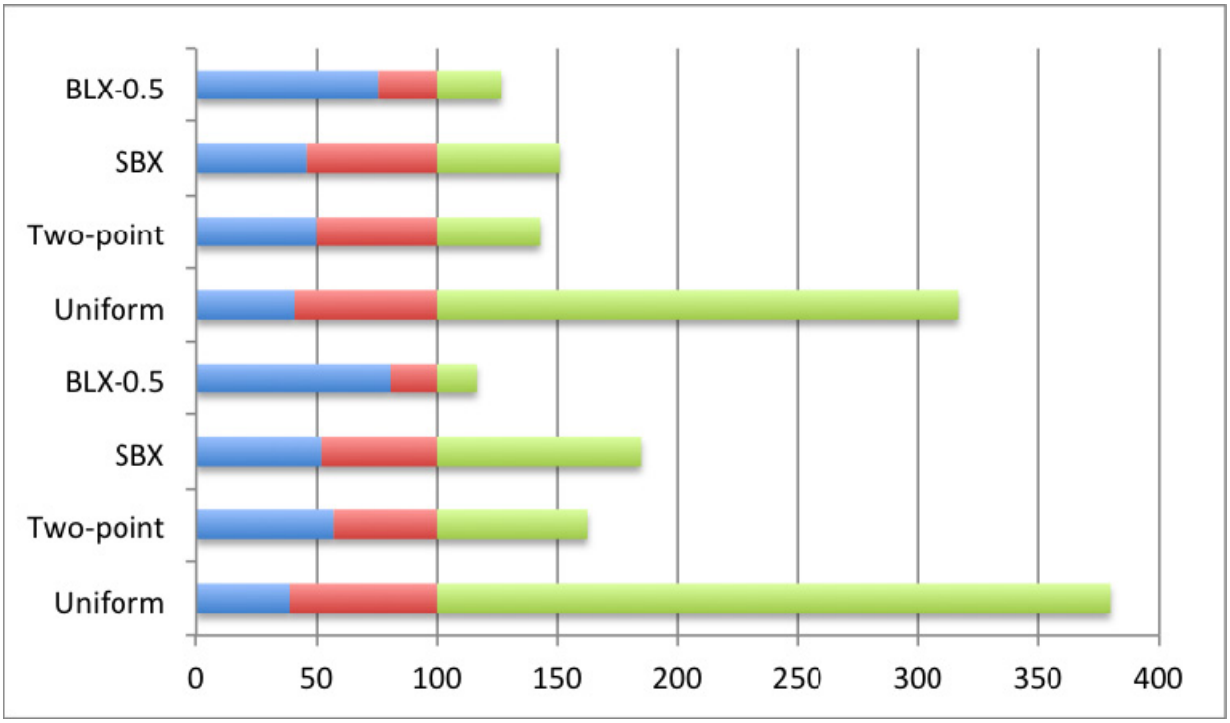

Figure 2. Spot detection analysis with gel (a) and gel (b) using uniform, two-point, SBX and BLX-0.5 crossover functions.

Z3, Melanie 3.0 quantitative tests results [Raman et al. 2002] and BLX-0.5 shown in Table 2.For gel (a) and gel (b), each row of the tables shows the number of true, false and extraneous spots detected for each method, or different settings of the same method.

Table 2. Spot detection analysis comparative with Z3 and Melanie 3.0 using gel (a) and gel (b). Total spots detected expressed as percentage of the manual count

GEL	METHOD	TRUE SPOTS	FALSE SPOTS	EXTRA. SPOTS
A	Z3(default)	89	11	14
	Melanie(default)	93	7	350
	Melanie(adjusted)	87	13	138
	BLX-0.5	81	19	17
B	Z3(default)	89	11	6
	Melanie(default)	94	6	165
	Melanie(adjusted)	89	11	31
	BLX-0.5	76	24	27

Graphically Z3, Melanie 3.0 quantitative tests results and BLX-0.5 shown in Figure 3. Proposed method obtained better results in gel (a) than Melanie in both, adjusted and default settings experiments for the number of extraneous spots, and slightly worst than Z3 with default settings. In gel (b), proposed method is better than Melanie with both adjusted and default settings.

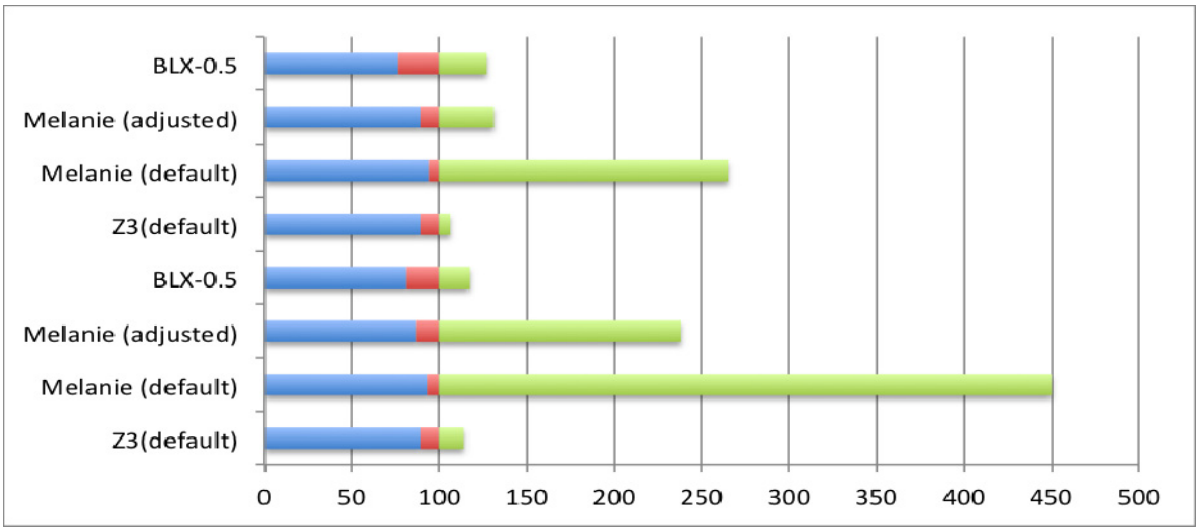

Figure 3. Spot detection analysis comparative with Z3 and Melanie 3.0.

Conclusion

To the best of our knowledge, this is the first work in which the protein identification in two-dimensional electrophoresis gel images is tackled using Evolutionary Computation and a second order operator for edge detection.

According with the FPR our results improved the obtained in [Raman et al. 2002] for Melanie 3.0 image analysis software and is close to Z3, but the obtained results seem to indicate that the GA-based method is conservative and needs more information to discriminate noise from real proteins and increase the number of true positive protein detected. A new Gaussian fitting value factor could be included in the fitness function to evaluate the goodness of fit of a detected protein with a normal distribution to improve the proposed method and increase the number of true positives spots detected.

Acknowledgments

This work is supported by the General Directorate of Culture, Education and University Management of the Xunta de Galicia (Ref. 10SIN105004PR).

Bibliography

[Agrawal et al. 1994] R. B. Agrawal, K. Deb, K. Deb and R. B. Agrawal, Simulated Binary Crossover for Continuous Search Space, 1994.

[Conradsen et al. 1992] K. Conradsen and J. Pedersen, Biometrics, 1992, 48, 1273-1287.

[Dowsey et al. 2010] A. W. Dowsey, J. A. English, F. Lisacek, J. S. Morris, G. Yang and M. J. Dunn, Proteomics, 2010, 10, 4226-4257.

[Efrat 2002] A. Efrat, F. Hoffmann, K. Kriegel, C. Schultz and C. Wenk, Journal of Computational Biology, 2002, 9, 299-315.

[Eshelman et al. 1993] L. J. Eshelman and J. D. Schaffer, Foundation of Genetic Algorithms 2, 1993.

[Goldberg 1989] D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Professional, 1989.

[Goldberg 1990] D. E. Goldberg, Complex Systems, 1990, 5, 139-167.

[Holland 1975] J. H. Holland, Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, University of Michigan Press, 1975.

[Iakovidis et al. 2006] D. K. Iakovidis, D. Maroulis, E. Zacharia and S. Kossida, Proceedings of International Conference on Information Technology in Biomedicine (ITAB), Ioannina, Greece, 2006.

[Janikow et al. 1991] C. Janikow and Z. Michalewicz, Proc. of the 4th International Conference on Genetic Algorithms, 1991.

[Langella et al. 2008] O. Langella and M. Zivy, Proteomics, 2008, 8, 4914-4918.

[Lemkin] P. F. Lemkin, Center for Cancer Research Nanobiology Program (CCRNP). The LECB 2D page gel image data set. http://www.ccrnp.ncifcrf.gov/users/lemkin. Last Access June 2012.

[Lucasius et al. 1989] C. B. Lucasius and G. Kateman, in Proceedings of the 3rd International Conference on Genetic Algorithms, Morgan Kaufmann Publishers Inc., 1989, pp. 170-176.

[Marr et al. 1980] D. Marr and E. Hildreth, Proc R Soc Lond B Biol Sci, 1980, 207, 187-217.

[Marten] R. Marten. Marten Lab Proteomics Page. http://www.umbc.edu/proteome/image_analysis.html. Last Access June 2012.

[Millioni et al. 2010] R. Millioni, S. Sbrignadello, A. Tura, E. Iori, E. Murphy and P. Tessari, Electrophoresis, 2010, 31, 1739-1742.

[Pendergrass et al. 2011] S. A. Pendergrass, K. Brown-Gentry, S. M. Dudek, E. S. Torstenson, J. L. Ambite, C. L. Avery, S. Buyske, C. Cai, M. D. Fesinmeyer, C. Haiman, G. Heiss, L. A. Hindorff, C. N. Hsu, R. D. Jackson, C. Kooperberg, L. Le Marchand, Y. Lin, T. C. Matise, L. Moreland, K. Monroe, A. P. Reiner, R. Wallace, L. R. Wilkens, D. C. Crawford and M. D. Ritchie, Genetic Epidemiology, 2011, 35, 410-422.

[Rabilloud et al. 2010] T. Rabilloud, M. Chevallet, S. Luche and C. Lelong, Journal of Proteomics, 2010, 73, 2064-2077.

[Raman et al. 2002] B. Raman, A. Cheung and M. R. Marten, Electrophoresis, 2002, 23, 2194-2202.

[Rogers et al. 2005] M. Rogers, J. Graham and R. P. Tonge, 2005.

[Tsakanikas 2009] P. Tsakanikas and E. S. Manolakos, Proteomics, 2009, 9, 3877-3888.

[Tyers et al. 2003] M. Tyers and M. Mann, Nature, 2003, 422, 193-197.

[Voss et al. 2000] T. Voss and P. Haberl, Electrophoresis, 2000, 21, 3345-3350.

[Witkin 1983] A. P. Witkin, in Proceedings of the Eighth international joint conference on Artificial intelligence - Volume 2, Morgan Kaufmann Publishers Inc., Karlsruhe, West Germany, 1983, pp. 1019-1022.

[Wu et al. 1990] L. Wu and Z. Xie, IEEE Trans. Pattern Anal. Mach. Intell., 1990, 12, 46-54.

[Xin et al. 2011] H. Xin and F. Zhao, 2011.

Authors' Information

Carlos Fernandez-Lozano *– PhD Student. Information and Communications Technologies Department. University of A Coruña. Campus Elviña S/N. 15071 A Coruña (España); e-mail: carlos.fernandez@udc.es*

Major Fields of Scientific Research: Genetic Algorithms, Support Vector Machines, Information Systems, Biomedical Imaging.

Jose A. Seoane *– Researcher. Information and Communications Technologies Department. University of A Coruña. Campus Elviña S/N. 15071 A Coruña (España); e-mail: jseoane@udc.es*

Major Fields of Scientific Research: Artificial Neural Networks, Data Mining, Bioinformatics, Information Systems, Data Integration.

Daniel Rivero*– Associate Professor. Information and Communications Technologies Department. University of A Coruña. Campus Elviña S/N. 15071 A Coruña (España); e-mail: drivero@udc.es*

Major Fields of Scientific Research: Artificial Neural Networks, Genetic Algorithms, Genetic Programming, Adaptive Systems.

Julian Dorado *– Senior Lecturer. Information and Communications Technologies Department. University of A Coruña. Campus Elviña S/N. 15071 A Coruña (España); e-mail: julian@udc.es*

Major Fields of Scientific Research: Artificial Neural Networks, Genetic Algorithms, Genetic Programming