Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Computational Geometry & Artificial Vision, Malta, September 15-17, 2005 (pp112-116)

A simple and effective detection technique of 2D electrophoresis image protein spots

N. Nafati, M. Samson, and B. Rossi
IFR50, Proteom Plate-Form
Faculty of Medicine. University of Sophia Antipolis
Av Valombrose, 06172, Nice
FRANCE
http://www.unice.fr

Abstract: The proteomic analysis is a set of process involving modern analytical techniques: -Realization of 2D Electrophoresis-Gels (2DE-G). - Processing of 2DE-G image and location of new or modified proteins. - Peptidic identification by mass spectrometry. - Location in data bases of proteins whose peptidic card correlates with respect to that found. The 2DE-G image processing is a crucial stage because it allows to localize pertinently the protein spots candidate for a peptidic identification by mass spectrometry. One of the major axes of this stage turns around the protein-spots detection. The reliability of this stage depends among others on the quality of the detection tools. In this paper we present a segmentation technique combining both the Top-Hat transformation and the watershed. The segmentation process is an iterative process leading to a simple and effective detection of protein spots. Before this segmentation process, image must be preprocessed in order to improve the image quality in term of signal to noise ratio. Image preprocessing algorithm uses the sub-band technique combined with Retinex Method. The detection algorithm operates by finding the threshold gray level that minimizes the entropy of the fuzziness measure.

Key-Words: 2D Electrophoresis-Gels (2DE-G). Protein spots. Detection. Top-Hat Transformation. Watershed Transformation. Sub-band technique. Retinex Method. Entropy of the fuzziness measure.

1 Introduction

The final goal of the proteomic analysis is to indentify the peptidic card of the new and\or modified proteins (proteins X) within a series of 2D electrophoresis gels. These are obtained after orthogonal separation following the isoelectric point and the molecular mass on polyacrylamide gel. Often, gels are realized with solubilised proteins resulting from cells or from bodies. These proteins can be used as targets of drug (medicine) or marker pens of disease [10][15].

Here, we present a segmentation technique combining both watershed and top-hat transformation. We will give a comparative results detection made by the proposed method and by the commercial PDQUEST software.

To observe marker pens or a treatment impact, the researchers, the biologists, doctors realize several gels: control gel, treated gels and then accomplish comparative image analysis in order to localize possible proteins X. Specialized 2DE-G image processing mainly leans on two processing tools: the protein-spots detection and the image matching [15]. It is clear that the quality of the 2-nd depends on that of the first. In this article we treat only the first tool: Spots Detection Tool.

Fig.1 below shows the result of the protein-spots recognition made by the specialized BIORAD PDQUEST software. The found spots are marked by symbols (+). Inside the box, it is seen that each protein spot is schematized by several signs +, thus indicating the presence of several spots, whereas actually one should observe only one sign +.

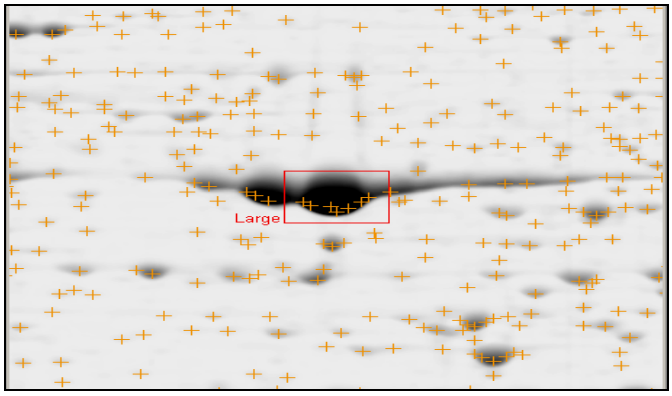

Fig.1: Piece of 2D electrophoresis Gel. Signs + represent the protein spots detected by the commercial PDQUEST software. This last one is dedicated to 2DE-G image processing.

Our paper presentation is as following:

1. General algorithm

2. Preprocessing process
3. Fuzzy Minimum error criterion
4. Results
5. Conclusion

2 General algorithm

The global algorithm is done according the following processes:

- Sub-band and Retinex preprocessing

- Protein spots detection process
-- Top Hat Transformation
-- Optimal threshold computation
-- Whatershed Transformation

- Spot domains computation

3 Sub-band and Retinex preprocessing

In proteomic image analysis context, the notified noise must be minimized in order to avoid parasite spots detection and consequently to optimize the proteomic image analysis cost.

Generally, a good quality of image is translated by a good Signal to Noise power Ratio (SNR), or by good Maximum Signal Power to Noise power (PSNR). In other words, less is the background and noise, better is the contrast. So, the finality is to reduce the noise and to improve the contrast of the image. To carry out such an objective, one broke up the image (sub-band decompostion) into two categories of frequencies: low and high frequencies [10,12,13,14], then one applied the method retinex [3,4] only to the low frequencies. As the Fig.2.2 shows it, one notes an improvement of the image quality.

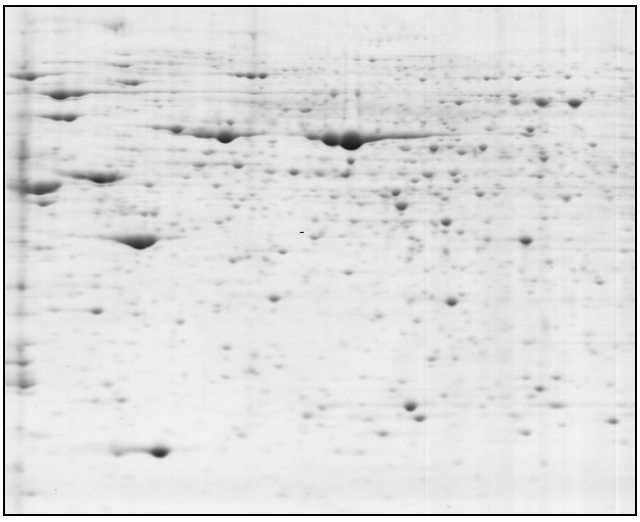

Fig.2.1: origin image

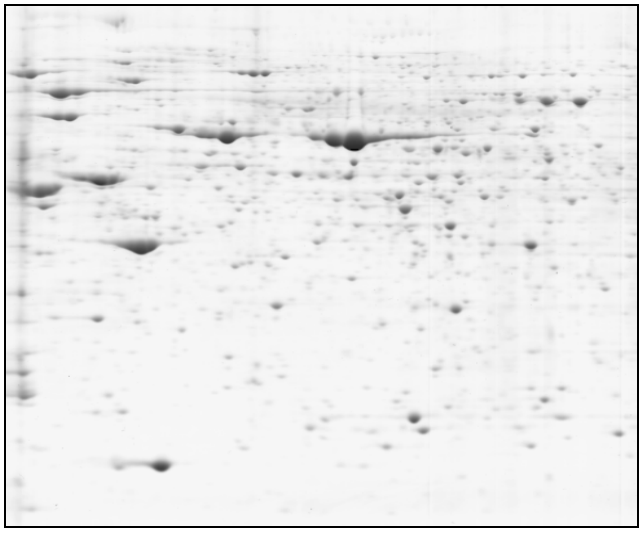

Fig.2.2: Preprocessed image by sub-band and retinex combination techniques.

4 Protein Spots Detection process

In the literature, numerous applied techniques can lead to more or less satisfactory results. One can find those that are direct and that base themselves on the image sweeping with a geometrical element and an exam of maxima inside these regions, one can meet derivative methods which base themselves on the analytical properties of the continuous surfaces, the analytical methods of the convexity which base themselves on the correlation of the image with a mask (template matching) [1][5][17][19]. These methods depend on the curve model (geometrical shape). Methods used in our approach got from the mathematical morphology and do not depend on the shape. These methods are [9] [17,18,19]: - Top Hat Transformation (THT) and Watershed Transformation (WT).

4.1. Top Hat Transformation (THT)

There are two types of THT: white top hat and black top hat transformation [1][6].

The method used here is the "black top hat" transformation. It allows to discover the protein spots. The THT corresponds to the thresholding difference (threshold value T) of the closed image and the initial image. The closed image corresponding to the " background ". The thresholding of the image consists here in finding the maximal value of intensity from which it is possible to get protein patterns in image. At first time and before the detection process, protein spots must be separated from other phenomenon in 2DE-G image. This separation from background is done by applying Top-Hat transformation [1]. So, the original gray scale image I(x,y) is first opened by a cylinder

structuring element B with a radius of 10 pixels. The resultant image is then subtracted from the original, producing a different image which retains spots information. The process can be expressed as follow.

$$I_d = I - (I \circ B)$$

The output image $\mathbf{I_d}$ is then thresholding by T value.

4.2 Optimal threshold computation

Before applying the WT to the $\mathbf{I_d}$ image, this one must be converted into binary image (black and white). To be done, it is necessary to find an automatic and optimal threshold of the segmentation process, in other words a value of below which we wish that all the pixels take the value 0 (black) and above which the pixels take value 1 (white). Our automatic computation process is based on the minimization of the entropy of the fuzziness measure[2][7,8,9]. The first step is to define the membership function. One possible condition for the membership function is that, the smaller the difference between the gray level of any pixel and the mean of its class, the greater will be the value of the membership function. A membership function that fulfills this condition and suitable for the 2DE-G image is :

$$\mu(x) = \begin{cases} \dfrac{1}{1+|x-\mu_0|/c} & if \quad x \le t \\ \dfrac{1}{1+|x-\mu_1|/c} & if \quad x > t \end{cases}$$

With

$$\mu_0(t) = \frac{\sum_{i=0}^{t} i.h(i)}{\sum_{i=0}^{t} h(i)} \qquad and \qquad \mu_1(t) = \frac{\sum_{i=t+1}^{255} i.h(i)}{\sum_{i=t+1}^{255} h(i)}$$

In the equation above, t signifies a given threshold gray level, C is constant that represents the difference between the maximum and minimum gray levels present in the grayscale image, μ_0 is the mean value of background pixel class, and μ_1 is the mean value of specimen pixel class. The upper equation of the membership function applies to the background pixels and the lower equation applies to specimen pixels. In either case the membership function assigns a numerical probability between 0.5 and 1 to the degree that pixels belongs in one of the two classes. The second step is to determine a measure of the segmentation fuzziness for a given threshold **t.** One method for measuring fuzziness is based on the idea of the fuzzy set entropy, which is calculated using Shannon's function, or:

$$H_f(x) = -x\ln(x) - (1-x)\ln(1-x)$$

The entropy of the entire image is then given by:

$$E(t) = \frac{1}{N}.\sum_{i=0}^{255} H_f(\mu(i)).h(i)$$

In the equation above, the summation is taken over all of the possible gray levels, and N is the total number of pixels in the image. The process of the optimal threshold computation is done by finding the threshold gray level that minimizes the entropy of the fuzziness measure.

4.3 Watershed transformation

WT is a mathematical morphology method which constitutes the main tool of images segmentation [1][5][9][18]. This technique uses the description of images in geographic terms. Indeed, an image can be perceived as a relief if one associates the gray level of every point to a height. It is then possible to define the watershed as being the crest forming limit between two ponds hillsides. To obtain it, it is necessary to imagine the dumping of a relief in some water, by clarifying that, the water can penetrate into the valleys only by its minima. The watershed is represented with the points where two separate lakes join during the dumping (see Fig.3).

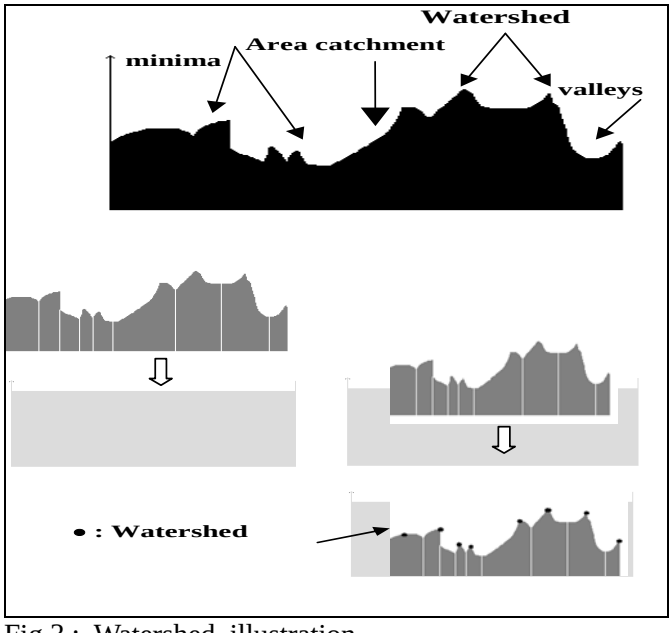

Fig.3 : Watershed illustration

4 Algorithm Results

As mentioned above, the protein spots detection is started by applying the THT process in which the radius cylinder structuring element B is 10 pixels. The output of this process is given in Fig.4. Then and in order to obtain the binary image, the computation process of the optimal threshold is executed. The obtained value that minimize the fuzziness measure criterion is equal to 87. Fig.5 is the output of the segmentation process. After, the algorithm operates with watershed transformation that gives the result showed in Fig.6. At the end we launch the process of the protein domains computation. The output of this last is given in Fig.7.

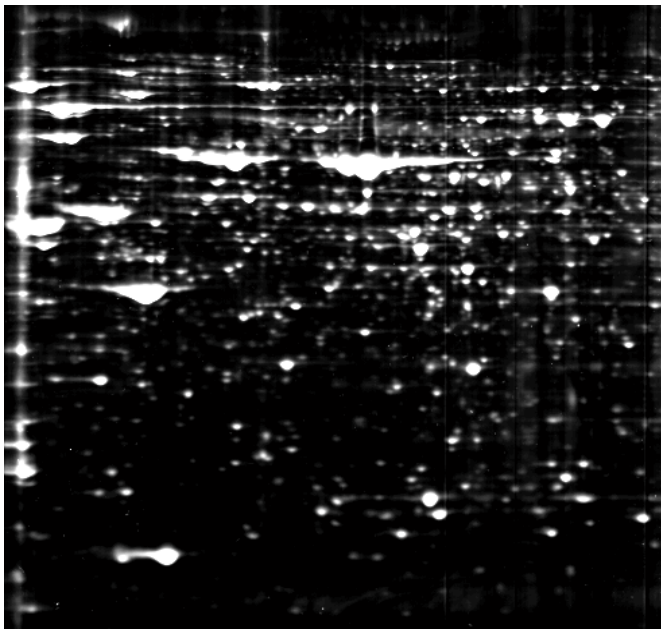

Fig.4: Result of the THT process

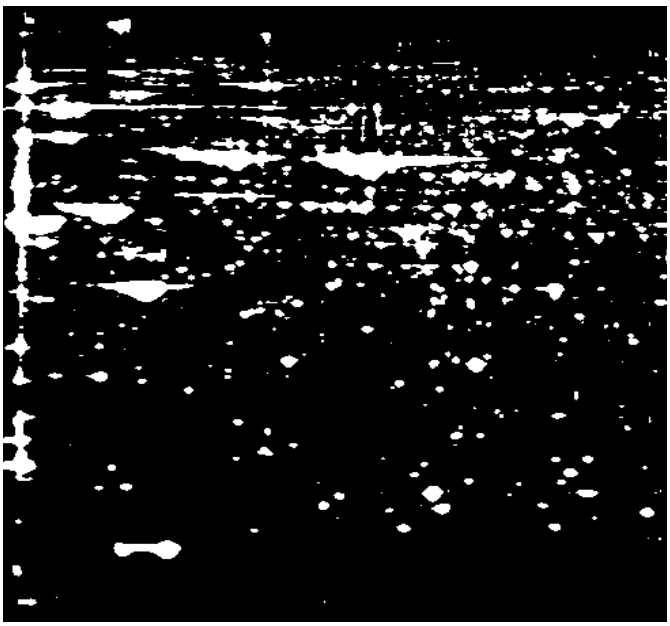

Fig.5: Binary image result

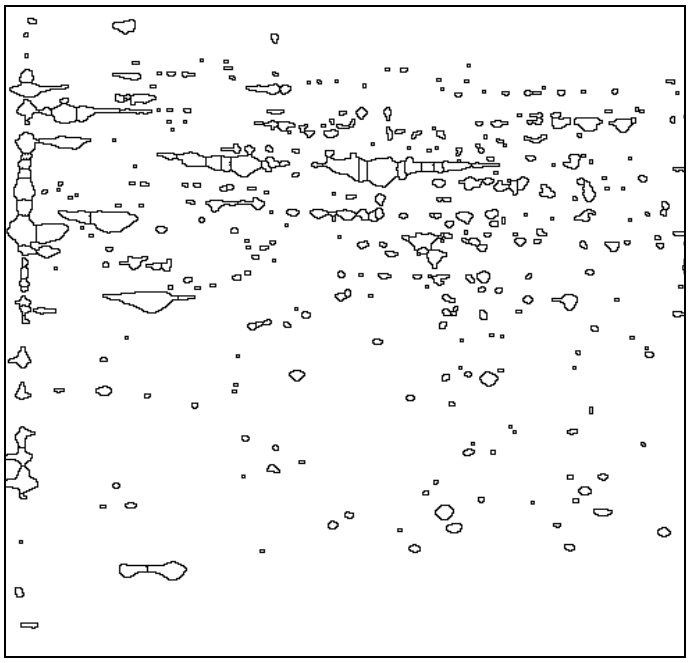

Fig.6: Result of the WT process.

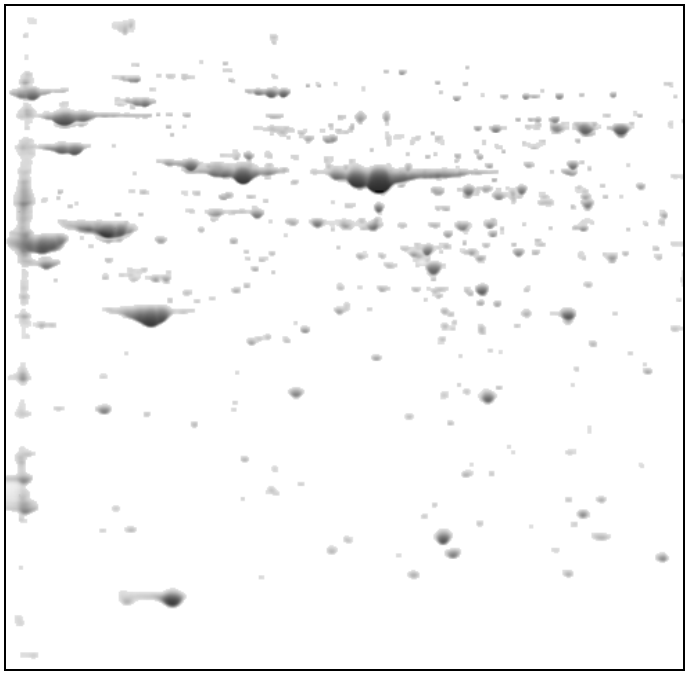

Fig.6: Detected protein spots

3 Conclusion

The given results show the effectiveness of the watershed transformation coupled with the "Top Hat" transformation. The suggested method generates less parasitic spots (not like PDQuest software for example). In this context, the associated WT and CHF technique was thus considered to be relevant for the protein spots detection. We note out that the market processing software majority require subjective criteria, involving a cost in analysis time and risks of error. Our method is

objective since it generates an optimal threshold according to the fuzziness measure theory and does not require knowledge in image processing.

References:

[1] S. Beucher. Segmentation d'images et morphologie mathématique, *Doctorate Thesis.* Paris School of Mines. France. June 1990.

[2] E.. David Jansing, Thomas A . Albert, Darrel. L. Chenoweth. Two-dimensional entropic segmentation, *Pattern Recognition Letters 20.* pp. 329-336. 1999.

[3] B. Funt, Florian Ciurea, and JohnMcCann. Retinex in MaLab, *Color Science. Systems and Applications.* pp. 112-121. 2000.

[4] B. Funt. Kobus Barnard. Michael Brockington. Vlad Cardei. Luminance-Based Multi-Scale Retinex. Simon Fraser University, *Proceedings AIC Colour 97 Kyoto 8th Congress of the International Colour Association.* May 1997.

[5] L. A. Gee and M. A. Abidi, Segmentation of Range Images Using Morphological Operations: Review and Examples, *Proc. of SPIE Conf. on Intelligent Robots and Computer Vision XIV*, Vol. 2588. pp. 734-746. Philadelphia. October 1995.

[6] R. C. Gonzalez and R. E. Woods, *International Edition. Digital Image Processing. Second Edition.* Prentice-Hall. 2002.

[7] H. Haussecker, H. R. Tizhoosh, Fuzzy Image Processing, I. Handbook of Computer Vision Application, Edited by B. Jagne, H. Haussecker, and P; Geisster, Academic Press 1999.

[8] E. D. Jansing, T. A. Albert, D. L. Chenoweth, Two Dimensional Entropy Segmentation. Pattern Recognition Letters 20, Letters 20, 1999, pp. 329-336.

[9] O. Lezoray. Supervised Automatic Histogram Clustering and Watershed segmentation. Application to Microscopic *Medical Color Image. Image Anal Stereol.* Pp. 113-120. 2003

[10] P. Lio, Wavelets in bioinformatics and computational biology : state of and perspectives, Bioinformatics, Vol. 19, 2003, pp 2-9

[11] F. Meyer and S. Beucher. Morphological Segmentation. *Journal of Visual Communication and Image representation.* Vol. 1, n° 1, pp. 21-49. Sept. 1990. Academic Press.

[12] N. Nafati. Contribution à l'étude du filtrage, de la décomposition en sous-bandes et de la reconstruction appliqués au traitement des images. *Thèse de doctorat*. CNAM de Paris 1995.

[13] N. Nafati.. An iterative technique to design 2D QMF banks under constraints by the method of Lagrange multipliers, *IEEE-SP Time Frequency and Time-Scale Analysis.* 1996.

[14] N. Nafati.. Synthèse itérative et simultanée de banc de filtres biorthogonaux de reconstruction parfaite avec des critères adaptés aux applications de codage de la parole et de l'image, *GRETSI.* pp. 1085-1088. Grenoble France. Septembre 1997.

[15] P. Nugues. Interprétation de gels d'électrophorèses 2D, *Thèse de Doctorat. Unversité de Nancy*. 1989.

[16] T. D. Pham, A New Approach for Calculating Implications of Fuzzy Rules, IEEE International Conference on Artificial Intelligence Systems, pp. 71. 2002.

[17] J. Serra. Image Analysis and Mathematical Morphology. Academic Press. New York, pp. 436–437. 1982.

[18] L. Vincent, and P. Soille. Watersheds in digital spaces and efficient algorithm based on immersion simulations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. pp. 583-593. 1991.

[19] S. Yi-Chen Shao. L. C. Chen, Object Segmentation in Elevation Space Using Mathematic Morphology, *Proc. ACRS 2001 - 22nd Asian Conference on Remote Sensin,*. Vol. 1. pp. 227-232. Singapore. Novembe 2001.