A COMPUTERIZED SYSTEM FOR SPOT DETECTION AND ANALYSIS OF TWO-DIMENSIONAL ELECTROPHORESIS IMAGES

R. Appel, M. Funk, C. Pellegrini
Centre Universitaire d'informatique
24, rue du Général-Dufour, CH-1211 Geneva 4, Switzerland

D. Hochstrasser, A.F. Müller
Hôpital Cantonal Universitaire
1211 Geneva 4, Switzerland

ABSTRACT

The 2-dimensional electrophoresis technique is very powerful to determine the composition of a biological sample. The complexity of the image produced by this process needs the use of a computerized system.
The purpose of this paper is to describe a system, called MELANIE, with which the user can process the image of a 2-dimensional electophoresis gel, separate and quantify the different proteins and compare semi-automatically two different pictures. The algorithms involved in the different steps of the process are discussed.

RESUME

L'électrophorèse bi-dimensionnelle est une technique très efficace pour déterminer la composition en protéines d'une substance biologique. La complexité de l'image produite exige le recours à un système informatique.
Le but de cet article est de décrire un système, appelé MELANIE, qui permet à l'utilisateur de traîter l'image d'une électrophorèse bi-dimensionnelle, de séparer et de quantifier les différentes protéines et de comparer semi-automatiquement deux images différentes. Les algorithmes mis en oeuvre dans les différentes étapes du traitement sont présentés.

KEYWORDS: image processing, medical computer graphics, pattern analysis.

I. INTRODUCTION

I.1 Objectives

The 2-dimensional electrophoresis technique is very powerful to separate complex mixtures of proteins. It is estimated that it is possible to separate 19,000 proteins on a single gel (1) (2). Because of this large number, the only practical solution for using 2-dimensional electrophoresis is through a computerized system. The main tasks of the system are :

- image digitization,
- noise reduction or elimination,
- spot detection or separation,
- normalization and standardization.

Programs have been developed at the University of Geneva which constitute the first part of a large project called MELANIE (Medical ELectrophoresis ANalysis Interactive Expert-system). These programs are intended for image processing and medical diagnosis and are based on Artificial Intelligence methods. This research project is multidisciplinary and involves groups of researchers in medicine and in computer science (at the University of Geneva), in biochemistry (at the Federal Polytechnic Institute in Zürich) and in computer science (at NIH-Bethesda, Maryland, USA). The wide range of activities in this domain, in particular the work of Harrington et al. (3) and of Wiederkehr and Vonderschmitt (4), confirms the present interest in adapting the 2-dimensional electrophoresis technique to a clinical environment and of developing a new diagnostic tool.

This paper describes only the image-processing section of MELANIE.

I.2 Main problems in the processing of two-dimensional electrophoresis images

The aim of processing 2-dimensional electrophoresis images is the detection of spots and their characterization by their coordinates and the density of protein they contain.

The biochemical technique of electrophoresis produces images that are not of high quality and which contain a lot of noise. This noise is of three types :

- background noise,
- horizontal and vertical streaks,
- "noise" superimposed on spots.

On the one hand, the image-processing program must compensate for these noises and, on the other hand, must have the following qualities in order to be used in

an operational environment :

    (a) display of all the significant spots,
    (b) fast processing,
    (c) portability.

II. TWO EXISTING SYSTEMS

In the field of computerized spot detection in electrophoresis images, one can cite two important works : TYCHO (5) and ELSIE (6).

II.1 TYCHO

The TYCHO system runs on a PDP-11/60 using an AP-120B array processor. The purpose of the system is :

- background subtraction,
- spot detection,
- Gaussian spot modelling, pattern matching and comparison.

TYCHO is based on a Gaussian model of the spots. This method has many good qualities, but it is not adapted to our needs; also the execution time is too long and the qualitative aspect of the original image is not preserved.

The general approach adapted for eliminating the background noise and the streaks is to find the minimum element in some selected region surrounding each image pixel. The regions are vertical and horizontal line segments, which are used to scan the image in x and y directions. This method allows background noise and streaks to be removed. Unfortunately, despite the use of an array processor, the CPU time necessary to process an image is much too long to be used for an interactive system such as MELANIE. For this reason, only the general method of background and streaks subtraction was retained, but this is implemented in MELANIE in another form.

II.2 ELSIE

The ELSIE system, developed by M. Miller on a VAX-11/750, is a sophisticated and complex set of programs, which perform a large number of operations on 2-dimensional electrophoresis images. Among them we can mention :

- spot detection,
- density measurements,
- image comparison.

The algorithm used to detect the spots is based on the principle that a point p belongs to a spot if the following condition is satisfied :

$$J(P) = \sum_P F(x,y) * G(x,y) < S$$

where : F is a second degree function of two variables;
    G is the discrete function describing the image;
    P is a square region centered on the point p;
    S < 0 is a predefined limit.

ELSIE offers the user a very powerful tool for detection and comparison of spots. However, it has some inconvenient properties :

- the total number of operations is too large,
- small spots are not always detected,
- the use of a third-degree function to approximate the background noise introduces some imperfections (bad limit conditions, lack of precision to approximate small differences, etc.).

III. MELANIX - THE IMAGE PROCESSING PART OF MELANIE

III.1 Introduction

Collaborating with Dr. Miller's group allowed us to benefit from previous work done in the processing of 2-dimensional electrophoresis images. We have based our developments on the existing ELSIE system, but we have implemented new algorithms for noise subtraction and for spot detection. Moreover, we have augmented the range of image display options by adding to the common 2D pictures in grey levels or in false colors, a pseudo-3D representation, for which the user can interactively define or modify the parameters (see Fig. 3 and following figures).

III.2 MELANIX

MELANIX is the image-processing subsystem of MELANIE, and it executes the following sequence of operations :

- background and streak subtraction,
- noise on spot reduction,
- detection and separation of the spots,

- computation of the surface of spots and of protein density.

In addition to this basic set of operations, MELANIX offers the user various display options : grey levels, pseudo-colors, 2D representations and pseudo-3D representations (see illustrations). To produce the pseudo-3D representation, and in order to emphasize the details without loosing the global view of the image, we consider the 2-dimensional gel as a projection of a three-dimensional picture with the values of the pixels representing the third dimension.

An algorithm to remove the hidden lines gives the picture a better quality and makes the small spots easier to visualize and to detect.

Below are the detailed descriptions of the four phases of image-processing.

III.2.1 Background and streak subtraction

The basic idea used in the TYCHO system for background and streak subtraction is improved in MELANIX to give better results and run faster.

The method consists of the following steps:

1) Determining a horizontal segment of n pixels length centered on a given pixel p(i,j).

2) The minimum value over the segment is placed in a matrix at the position (i,j).

3) Steps 1 and 2 are repeated for every pixel of the gel.

At the end of the process we have a matrix of the same resolution as the image. In order to avoid fuzziness at the edges of the streaks, we determine the maximum values over the matrix using the same method as explained above. Finally, the matrix represents the value of the background noise and the horizontal streaks. We subtract this matrix from the image.

The same process is performed vertically. Figures 3 to 8 illustrate the different steps of this process.

In order to reduce the CPU time necessary to execute the algorithms of background and streak subtraction, the points of the overlapping segment are organized in a

binary-tree. This structure allows a very fast determination of the minimum (or maximum) value. When the segment is shifted to the adjacent pixel (vertically or horizontally), the extraction of the outgoing pixel and the insertion of the incoming one is easily achieved. A great amount of time can be saved using this method.

An option in the system allows the user to eliminate the background noise without suppressing the streaks. This is done by considering a segment which is diagonal to the picture. This is an interesting possibility, because the streaks are produced by proteins which may be significant in certain cases and we want to keep them visible.

III.2.2 Noise on spot elimination

This step consists mainly in the elimination of the noise local to a spot. This is done by the following smoothing algorithm.

Given a function F(X,Y), this function is "smooth" if its Laplacian equals zero or :

$$\nabla_F = - \frac{\partial^2 F(X,Y)}{\partial x^2} + \frac{\partial^2 F(X,Y)}{\partial Y^2} = 0$$

Thus we compute the Laplacian of the discrete function G(X,Y), describing the gel, at each point (X_i,Y_j) and by using the Gauss-Seidel method we compute its solutions.

$$\nabla G(X_i,Y_j) = 2G(X_i,Y_j) - G(X_{i+1},Y_j) - G(X_{i-1},Y_j)$$
$$+ 2G(X_i,Y_j) - G(X_i,Y_{j+1}) - G(X_i,Y_{j-1})$$

$$\nabla G(X_i,Y_j) = 4G(X_i,Y_j) - G(X_{i+1},Y_j) - G(X_{i-1},Y_j)$$
$$- G(X_i,Y_{j+1}) - G(X_i,Y_{j-1}) = 0$$

By factorisation

$$G(X_i,Y_j) = (G(X_{i+1},Y_j) - G(X_{i-1},Y_j)$$
$$- G(X_i,Y_{j+1}) - G(X_i,Y_{j+1}))/4$$

$G(X_i,Y_j)$ is then taken as the new value of the point (X_i,Y_j). This is in fact the mean of the four neighboring points of the point (X_i,Y_j). This is done for all the points of the original image and it produces the smoothed image.

—

III.2.3 Spot detection and separation

Once the image is processed for noise subtraction, one has to detect and separate the spots in order to give the user a synthetic view of the gel and to be able to perform the computation of the protein density for each spot.

The algorithms implemented in MELANIX for spot detection and separation are completely different from those used in TYCHO or ELSIE. They are based on the following concepts :

(1) the Laplacian of that function is positive at it's peak;

(2) the minimum value of the second derivative with respect to x and y is negative on the boundary of the edges of a spot or in the region separating two very close spots.

A point p belongs to a spot if the following condition is satisfied :

$$\min \left\{ -\nabla G(p) - L,\ \frac{\partial^2 G(p)}{\partial y^2} - R,\ \frac{\partial^2 G(p)}{\partial y^2} - C \right\} > 0$$

where L, R and C are positive constants.

III.2.4 Computation of the surface of the spots and of the protein density

The method used in ELSIE was implemented for MELANIX. This method consists of computing the integral for every spot. The density of protein is then determined in relation to certain reference spots whose concentration is precisely known.

At the present stage of development, MELANIX performs the complete processing of an image of medium complexity (512 x 512 pixels and approximatively 500 spots) in around 10 minutes of VAX-11/780 CPU time (running UNIX BSD 4.2).

REFERENCES

(1) O'FARREL, P.H.
High resolution two-dimensional electrophoresis of proteins.
J. Biol. Chem. 250, 4007-4021 (1975).

(2) ANDERSON, N.L. et N.G.
High resolution two-dimensional separation of human plasma proteins.
Proc. Natl. Acad. Sci. USA, 1977, 5421-5425.

(3) HARRINGTON, M.G.; MERRIL C.R.; GOLDMAN, G.; XU, X.H.; McFARTIN, D.E.
Two-dimensional electrophoresis of cerebrospinal fluid proteins in multiple sclerosis and various neurological deseases.
In electrophoresis, 1984, vol.5, p. 236-245.

(4) WIEDERKEHR, F.; VONDERSCHMITT, D.J.
Zweidimensionale Elektrophorese des liquor cerebrosbinalis verschiedener neurologischer Patienten; (submitted 1985).

(5) COULTEC, B.P.; ANDERSON, N.G.
The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns.
In Clin. Chem., vol. 27,11, p. 1807-1820 (1981).

(6) MILLER, M.J et al.
Computer analysis of two-dimensional gels : semi-automatic matching.
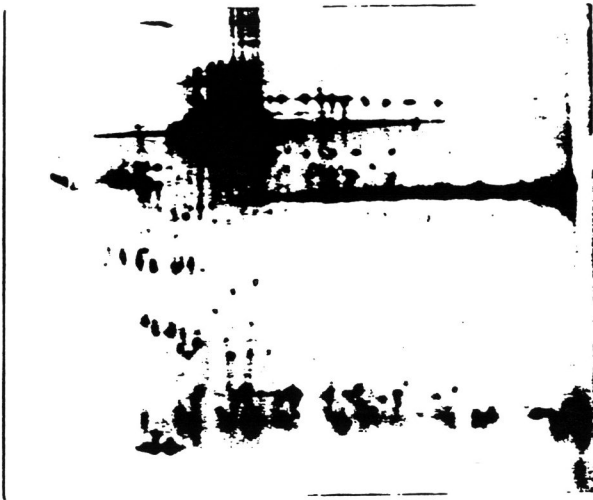Clin. Chem., vol.28,4, p. 867 (1982).

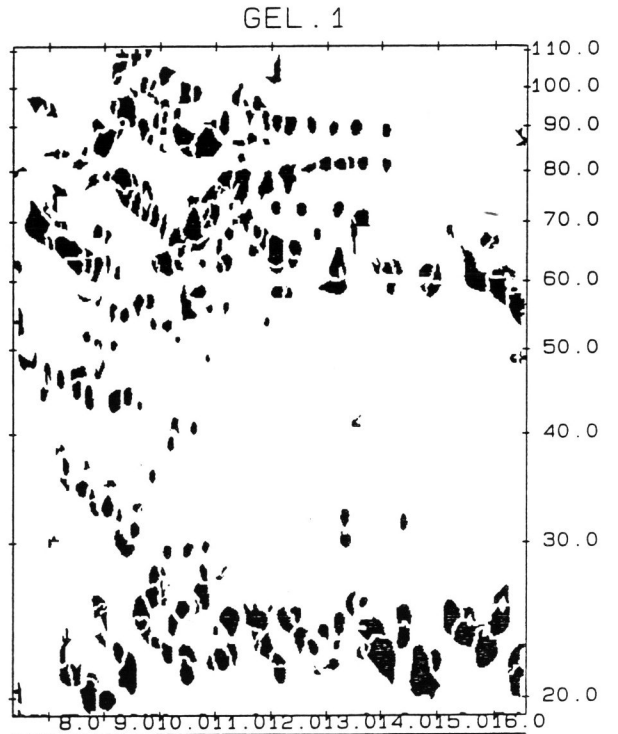Figure 1. An electrophoresis image.

GEL.1

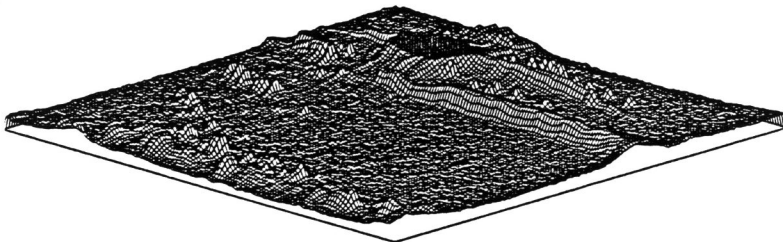Figure 2. The image of fig. 1 after the detection and separation of spots.

Figure 3. A pseudo-3D representation of the digitized original image.
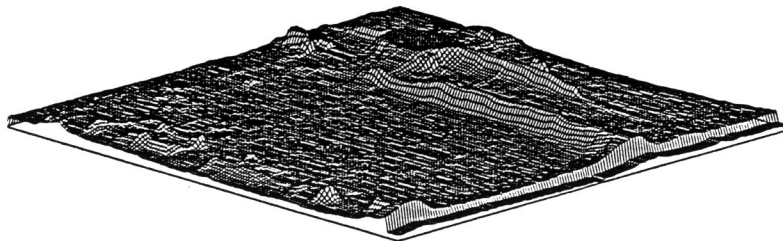

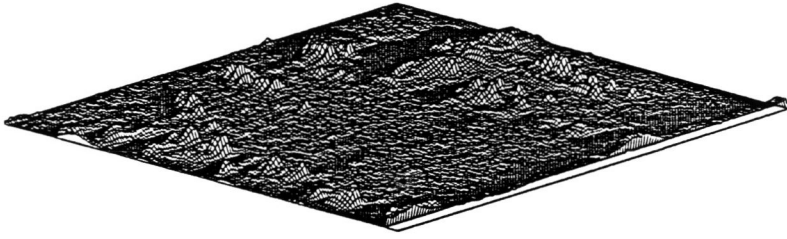Figure 4. The value of the background noise and the streaks in the x direction.

Figure 5. The image after the subtraction of the background noise and of the streaks in the x direction.
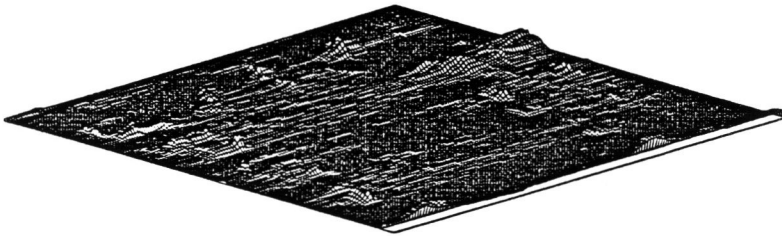

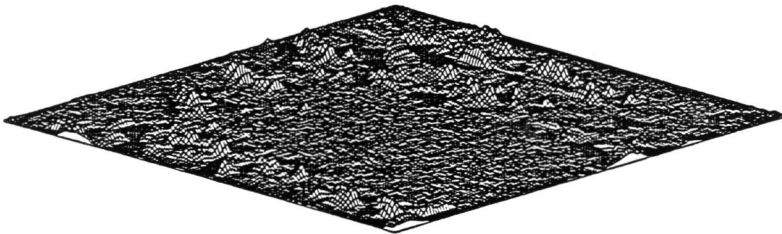Figure 6. The value of the streaks in the y direction.

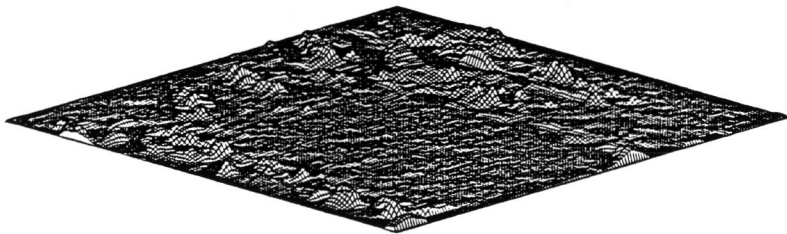Figure 7. The image after the subtraction of the streaks in the y direction.

Figure 8. The final image after noise and streaks subtractions and after smoothing (this is a pseudo-3D representation of the image of fig. 2)