IDENTIFICATION OF RELATIVE PROTEIN BANDS IN POLYACRYLAMIDE GEL ELECTROPHORESIS (PAGE) USING MULTIRESOLUTION SNAKE ALGORITHM

Metin Nafi Gürcan[†], *Mehmet Koyutürk*[†], *H. Serkan Yıldız*[†], *Rengül Çetin-Atalay*[‡], *A. Enis Çetin*[†]

[†] Dept. of Electrical and Electronics Engineering
[‡] Dept. of Molecular Biology and Genetics
Bilkent University
Bilkent, Ankara TR-06533, Turkey
E-mail: gurcan@ee.bilkent.edu.tr
Phone: (90) 312-266 4307 Fax: (90) 312-266 4126

ABSTRACT

Polyacrylamide Gel Electrophoresis (PAGE) is one of the most widely used techniques in protein research. In the protein purification process, it is important to determine the efficiency of each purification step in terms of percentage of protein of interest found in the protein mixture. This study provides a rapid and reliable way to determine this percentage. The region of interest containing the protein is detected using the snake algorithm. The iterative snake algorithm is implemented in a multiresolutional framework. The snake is initialized on a low resolution image. Then, the final position of the snake at low resolution is used as the initial position in the higher resolution image. Finally, the area of the protein is estimated as the area enclosed by the final position of the snake.

1. INTRODUCTION

Polyacrylamide Gel Electrophoresis (PAGE) is one of the most widely used techniques in protein research [1]. Denaturing Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) is a significant method used for the separation of proteins based on the migration of negatively charged proteins depending on their molecular weight in an electrical field. This procedure is often used during protein purification process. Its advantage is that proteins can be visualized as well as separated. SDS-PAGE technique provides information about protein concentrations of the protein mixture applied on a certain lane of the PAGE and the degree of purity of a particular protein in this mixture during protein purification process. Each band in the SDS-PAGE image represents a different protein band.

In the protein purification process, it is important to determine the efficiency of each purification step in terms of percentage of protein of interest found in the protein mixture. This study provides a rapid and reliable way to determine the percentage. Currently, the percentage is determined by eye and the results are quite subjective. In the currently available gel analysis systems, the analysis is usually based on the one dimensional profile of manually extracted lanes. The positions of the proteins of interest are determined from the peak points in the horizontal profile and the areas of the rectangular regions are measured. Finally, these measurements are used for efficiency calculation. However, rectangular approximations are not accurate representations of the protein areas.

In our scheme, the region of interest containing the significant protein is detected using the snake algorithm [4]. The iterative snake algorithm is implemented in a multiresolutional framework. The snake is initialized on a low resolution image. Then, the final position of the snake at the low resolution is used as the initial position in the higher resolution image. The area of the protein is estimated as the area enclosed by the final position of the snake.

In Section 2, the characteristics of the SDS-PAGE images are described. In Section 3, the snake algorithm is reviewed. In Section 4, the multiresolution implementation of the snake algorithm is described. In Section 4.1, the computational complexity of the proposed scheme is analyzed and simulation results are presented.

This work was supported by the Technology Development Foundation of Turkey under TTGV-199.

0-7803-5073-1/98/$10.00 ©1998 IEEE

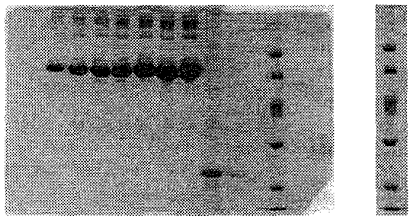

Figure 1: (a) Denaturing PAGE image, (b) a lane of the denaturing PAGE image

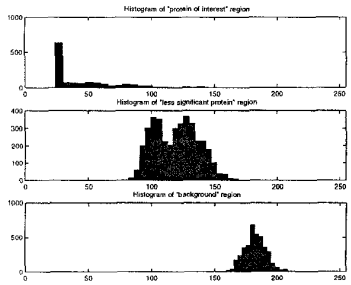

Figure 2: Histograms of three different regions of a PAGE image.

2. SDS-PAGE IMAGES

Figure 1(a) shows a typical SDS-PAGE gel image. The vertical stripes containing several protein bands in this gel image are called lanes and one of them is shown in Figure 1(b). The SDS-PAGE image is composed of three parts. The background does not carry any useful information. The second part corresponding to the images of protein bands have darker appearance. The proteins of interest appear as dark stripes and constitute the third part. The ratio of total area of the dark regions to the total area of other proteins is the measure of the efficiency of each purification step. The histograms of the three regions corresponding to the significant protein, other proteins and the background on a typical lane are given in Figure 2.

The regions containing protein of interest are nearly elliptic regions which have the smallest gray level on the image. Thus, the boundary of such a region is an elliptic closed contour having small total intensity and high gradient.

3. THE SNAKE ALGORITHM

Snakes are deformable contour models increasingly used in image segmentation [4, 5]. They are planar curves. The shape and location of the curve is determined according to an objective function. Snakes are influenced by image structure and pulled toward the edges and lines on the image. The energy of a snake is defined in a way that its total energy becomes minimum when the snake lies around the region of interest. The total energy of the snake that defines a closed contour v(s) is defined as:

$$E(v(s)) = \oint_v (E_{int}(s) + E_{ext}(s))ds \qquad (1)$$

where E_{int} and E_{out} represent the internal and external energies of the snake, respectively. The internal energy, E_{int} is determined according to the length and bending of the snake and defined as

$$E_{int}(s) = \alpha_s \frac{dv}{ds} + \beta_s \frac{d^2v}{ds^2} \qquad (2)$$

where the first term denotes the length energy and the second term denotes the bending energy. The external energy, E_{ext}, represents the energy created by image structure depending on the location of the snake on the image. Two main components of external energy are gradient and intensity and these affect the line and edge energies of the snake. The external energy of the snake is defined as

$$E_{ext}(s) = w_{line} I(x(s), y(s)) - w_{edge} |\nabla I(x(s), y(s))| \qquad (3)$$

where the first term denotes the line energy, the second term denotes the edge energy, and $I(x(s), y(s))$ denotes the intensity of the image at location (x, y) along the snake contour parameterized by s. The parameters $\alpha, \beta, w_{line}, w_{wedge}$ are set in such a way that the energy should be minimum on the region of interest.

At each iteration of the snake algorithm, the snake moves to a location having less energy. In our application, a fast greedy algorithm described in [6] is used. The algorithm is adapted to this problem in such a way that the resultant snake has an ellipse like shape. This is provided by putting constraints on the internal angles of the snake. In our algorithm, the important parameters that affect the convergence of the algorithm are the energy coefficients, the search radius and the number of snaxels which are the control points of the snake curve.

4. MULTIRESOLUTIONAL FRAMEWORK

The snake algorithm is implemented in a multiresolution framework. The original SDS-PAGE image is first processed by the filter banks described in [2, 3] and low resolution quarter size subimages, $I_{ll}, I_{lh}, I_{hl}, I_{hh}$ are obtained. The snakes are initialized on the low-low subimage. The initial form of the snakes are circles centered at the peak points of the 1-D profile of

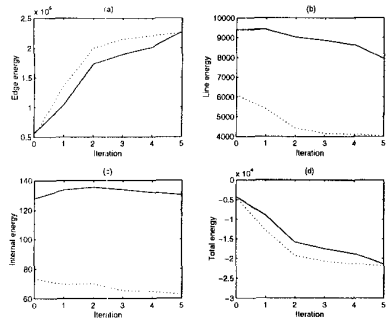

Figure 3: Change of different energy components. Solid lines indicate the energy changes in the high-resolution image and dashed lines show the energy changes in the lower-resolution image

the lane image. Another initialization is based on the I_{lh}, I_{hl}, I_{hh} subimages which contain the edge information about the original SDS-PAGE image. Therefore, the edges of the protein regions can be detected from these subimages and used as the initial positions of the snakes.

After the initialization on the low resolution image, the snake adapts itself to the contour around the protein region in an iterative manner. The final snake obtained on the low resolution image is used as the initial estimate on a finer resolution SDS-PAGE image. This process is repeated until the snake has closely covered the protein regions in the original full-resolution SDS-PAGE image. The advantage of this multiresolutional approach is three-fold: i) The snake converges better, ii) it brings robustness to the process, i.e., it is not easily affected by the scanning artifacts in the original SDS-PAGE image, and iii) computational cost is less as smaller images are processed in lower resolutions.

The behavior of the different energy components and the total energy of a snake at several iterations at different resolutions are given in Figure 3 for the SDS-PAGE image shown in Figure 2. At the beginning of the convergence process, the slope of the total energy of the snake in the low resolution image is higher than that of the snake in the high resolution image as shown in Figure 3(d). In other words,

$$\left| \frac{E_{low}[k+1]}{E_{low}[k]} \right| > \left| \frac{E_{high}[k+1]}{E_{high}[k]} \right| \quad (4)$$

This means that the snake converges faster on the low resolution image. This observation verifies the validity of the multiresolutional approach. In our simulation studies we use only one level of subband decomposition but this methodology can be considered in multiple resolution levels as well.

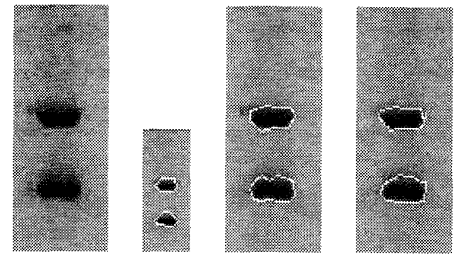

Figure 4: Protein of interest is enclosed within the snake : (a) Original lane image (b) Snake obtained using only the low-resolution image (c) Snake obtained using only the high-resolution image (d) Snake obtained using the multiresolutional approach

The energy parameters defined in Equations 2 and 3 $(\alpha, \beta, w_{line}, w_{edge})$ must be adapted to the energy level changes in different resolutions. In the multiresolutional scheme, at the final iteration of the low resolution image,α, β and w_{line} values are halved and the iterations are continued in the higher resolution image. At the low resolution these values are $\alpha = 4, \beta = 4, w_{line} = 0.2, w_{edge} = -0.1$.

Figure 4 shows an original lane of the image as well as the final positions of different snakes obtained using different methodologies. The result in Figure 4(c) is obtained after six iterations of the snake algorithm on the original image. Figure 4(d) shows the result of the multiresolutional approach. The snake obtained at the low resolution after two iterations (shown in Figure 4(b)) is used as the initial position of the snake in the high resolution image. Another three iterations in the high resolution image result in Figure 4(d). The comparison of Figures 4(c) and (d) indicates that the final snake converges better to the boundary of the protein area if a multiresolutional approach is used. Additionally, the number of computations required in the multiresolutional approach is smaller.

Figure 5 shows the results of the initialization procedure based on the edge information extracted from I_{lh}, I_{hl}, I_{hh} subimages. Figure 5(a) depicts the initial position of the snake at the subimage I_{ll}. The six snaxels are located on the extracted edges of the protein region. This snake takes the form shown in Figure 5(b) after only one iteration. This form is very close to the final position of the snake on the low-resolution image obtained with the other initialization procedure based on the 1-D profile of the lane image. Figure 5(c) shows the final position of the snake on the high resolution image after two iterations. There is no significant difference between this result and the result shown in Figure 4(d) obtained with the other initialization proce-

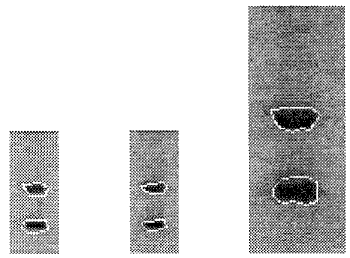

Figure 5: Protein of interest is enclosed within the snake : (a) Initial position of the snake located on the extracted edge (b) The snake in (a) after one iteration on low-resolution image (c) Snake obtained using the multiresolutional approach with the initial snake in (a)

dure. Edge information based initialization procedure produce similar results with less number of iterations. However, this initialization procedure requires the extraction of edges from the subimages.

4.1. Computational Complexity

For each iteration of energy calculations, the number of computations depends on the length of the snake and the radius of the search window. With search radius of n pixels and snake length of N pixels, the number of computations for a single iteration of energy calculations is $0(Nn^2)$ [6]. In this study, the number of snaxels and the radius of the search window are halved in the lower resolution images. Therefore, the ratio, R, of the total computational complexity in the low resolution image to that in the high resolution image is

$$R = \frac{(n+1)^2}{2 \times (2n+1)^2} \tag{5}$$

where $(n+1)^2$ and $(2n+1)^2$ are areas of search windows at the high and low resolution images, respectively.

Table I shows the number of times each energy component is computed on high and low resolution images at the last iteration. The total number of energy component computations in the low resolution image is reduced to approximately $\frac{1}{4}$th of that in the high resolution image. This figure is almost equivalent to the value we get if we take $n = 2$ in Equation 5.

Table 1: **Number of computations for high and low resolution**

Resolution	Edge	Line	Length	Bending
High	3517	3517	648	324
Low	756	756	132	66

5. CONCLUSIONS AND FUTURE WORK

In this work, we present a multiresolution snake algorithm for the identification of relative protein bands in PAGE images. The algorithm is more accurate compared to usual eye-exams or some commercially available programs that utilize rectangular approximations. Furthermore, the algorithm is fast as a multiresolutional framework is used.

The multiresolutional snake framework can be used in other applications such as the detection of mass and stellate lesions in mammograms using the snake algorithm [7].

6. REFERENCES

[1] R. C. Allen, C. A. Savaris and H. R. Marier, *Gel Electrophoresis and Isoelectric Focusing of Proteins : Selected Techniques*, Walter de Gryten: Berlin, 1984.

[2] C. W. Kim, R. Ansari and A. E. Cetin, 'A class of linear-phase regular biorthogonal wavelets,' *IEEE International Conference on Acoust., Speech, and Signal Processing (ICASSP'92),* San Francisco, USA, March 1992.

[3] Ömer Nezih Gerek, A. Enis Çetin, "Linear/Nonlinear Adaptive Polyphase Subband Decomposition Structures for Image Compression," *IEEE Int'l. Conf. on ASSP., ICASSP'98,* Seattle, WA, USA, May 12-15, 1998.

[4] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Computer Vision*, vol. 1, no. 4, pp. 321-331, 1988.

[5] F. Leymarie and M. D. Levine, "Tracking Deformable Objects in the Plane Using an Active Contour Model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 617-633, 1993.

[6] D. J. Williams and M. Shah. "A fast algorithm for active contours," *IEEE Trans. on PAMI*, vol. 4, pp. 592-595, 1990.

[7] Mehmet Koyuturk, Metin Nafi Gurcan, A. Enis Cetin, "Mammogramlarda Kutle Lezyonlarinin Yilan Algoritmasiyla Otomatik Olarak Sezinlenmesi," *6. Sinyal Isleme ve Uygulamalari Konferansi Kitapcigi*, 28-30 Mayis, 1998, Kizilcahamam, Ankara. (Detection of Mass Lesions in Mammograms Using Snake Algorithm) (in Turkish).