Automatic Lane Extraction in Hemoglobin and Serum Protein Electrophoresis Using Image Processing

Samira Khodabakhshi✉, Hamid Hassanpour

Department of computer engineering, Shahrood university of technology, shahrood, Iran

s.khodabakhshi.s@gmail.com; h.hassanpour@shahrood.ac.ir

Abstract

Image analysis is an image processing technique that aims to extract features or information from images. Image analysis in medicine has a special place because is a basis for disease diagnosis for physicians. Electrophoresis is a laboratory separating technique. Electrophoresis images are created during the electrophoresis process. Serum protein and hemoglobin electrophoresis test are the most common test in Laboratory. Each lane represents a patient's blood sample. Lane extraction is one of the most important steps in analyzing of electrophoresis images. This paper, proposed automatic lane extraction method based on Otsu thresholding and Savitzky-Golay smoothing. The results show the proposed method has 98% accurate.

Keywords: Electrophoresis, Lane Extraction, Otsu, Savitzky-Golay

1. Introduction

Electrophoresis is one of the most well-known separation techniques in laboratories. It was discovered in 1807 by Reuss. He observed that the clay particles are dispersed in the water affected by uniform electric field [1]. Electrophoresis is motion of dispersed particles within the fluid under the influence of a uniform electric field [2]. In electrophoresis process, hemoglobin or serum proteins move under the influence of an electric field across a cellular network such as cellulose acetate gel [3].Cellulose acetate electrophoresis images have eight blood samples from different patients. Each of samples is called lane and each of part in lane is called band. Serum protein and hemoglobin electrophoresis images are shown in Fig1. The first step in analysis electrophoresis image is lane extraction. In this phase, the use of image processing technique is very effective.
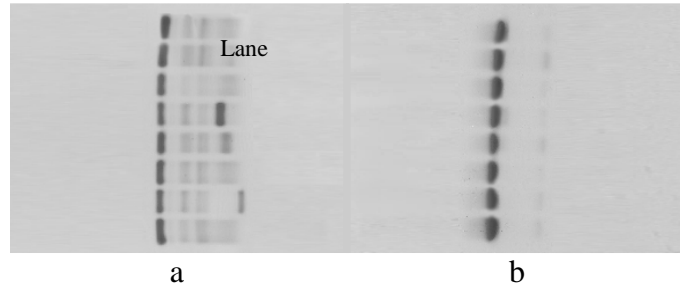
Figure1. (a)Serum protein electrophoresis image, (b) hemoglobin electrophoresis image

Lin et al [4] represented a computational method for comparing lanes in DNA electrophoresis images. Kaabouch et al [5] introduced an automatic procedure for the analysis of DNA and protein that thresholding and Improvement faint images are outstanding features of this method. Caridade et al [6] introduced an automatic digital image processing method for DNA electrophoresis, this method automatically identify number and location of lane and position of the bands in each lane. Mohdnoor et al [7] introduced a method for extracting bands from background in the electrophoresis gel images using Otsu multilevel thresholding based on PSO. In [8], was introduced way to identifying and tracking the center of lane as the first step in DNA image analysis.

This paper presented the automatic lane extraction method in hemoglobin and serum protein electrophoresis. This method includes pre-processing and lane extraction steps. In Pre-processing stage to calculate the rate of changes in the vertical profile of image and Savitzky-Golay smoothing are identified the area of lanes. In lane extraction stage the appropriate threshold determined using Otsu thresholding technique. In this paper first introduces Otsu thresholding technique and Savitzky-Golay smoothing technique. Then, the proposed algorithm is discussed. Afterwards, the result will be expressed.

1.1. Otsu thresholding

Thresholding is one of the most common techniques used for image segmentation. The main idea of thresholding is select optimal threshold for separate objects from the background. Otsu method was introduced as a way for two-level thresholding in 1979 [9]. In this method assumes that there are two distributions (background and foreground). Hence threshold is chosen such that to minimize the inter-class variance of two distributions [10]. This method chooses an optimal threshold based on the analysis of the gray level histogram.

In this case represents a two-dimensional image with l gray levels, which the range of each gray-level is$[0,1,2,…,(L-1)]$. The number of pixels with gray level i is given byn_i. So the probability of gray level i in an image is shown as follows:

$$p_i = \frac{n_i}{N} \tag{1}$$

Suppose that the image with a threshold t is divided into two classes C1 and C2 (background and foreground). Class C1 shows pixels with gray levels $[0,1,2,…,(t-1$ and C2 shows pixels with gray levels$t,…,L-1$. The mean of two classes C1 and C2 is calculated in equation (2) and (3):

$$\mu_1 = \sum_{i=0}^{t-1} \frac{ip_i}{q_1} \tag{2}$$

$$\mu_2 = \sum_{i=t}^{l-1} \frac{ip_i}{q_2} \qquad (3)$$

Such that q_1 and q_2 obtained with (4) and (5):

$$q_1 = \sum_{i=0}^{t-1} p_i \qquad (4)$$

$$q_2 = \sum_{i=t}^{L-1} p_i \qquad (5)$$

Class variance is defined by (6) and (7):

$$\sigma_1^2 = \sum_{i=0}^{t-1} \frac{(i-\mu_1)p_i}{q_1} \qquad (6)$$

$$\sigma_2^2 = \sum_{i=t}^{L-1} \frac{(i-\mu_2)p_i}{q_2} \qquad (7)$$

Otsu introduces inter-class variance by (8):

$$\sigma_w^2(t) = q_1\sigma_1^2 + q_2\sigma_2^2 \qquad (8)$$

Otsu method ensures that t is the optimal threshold that minimizes inter-class variance.

1.2. Savitzky-Golay Smoothing

Savitzky and Golay introduced method for data smoothing based on a least-squares polynomial approximations in 1964 [11]. They showed that fitting a polynomial to set of input data and then evaluate the resulting polynomial at a point within the approximate area equivalent to discrete convolution with the constant impulse response. The purpose of the Savitzky-Golay was smoothing the noisy data obtained from chemical spectral analyzer. They showed that the least squares smoothing reduce noise while maintain shape and height of peaks [12].

Savitzky method is generalized moving average method. This method is based on the values of the convolution and convolution function. The convolution concept can be extended by the moving average. In general, C shows convolution values. In (9) is shown the convolution process [11]:

$$Y_i = \frac{\sum_{i=-m}^{i=m} C_i y_{j+i}}{N} \qquad (9)$$

Where y_i, Nm and Y_i respectively denote the value width, normalization factor, the number of left and right neighborhood data and the new value. The Idea of Savitzky method is to find the C_i coefficients. Indeed, the basic idea of the Savitzky method is to estimate a high degree polynomial within the moving window. For each point y_j a polynomial fit to $2N + 1$ point in the moving window, also the new position point i (center of window) obtained with calculate Y_i. Then the window is shifted one forward and again a least squares polynomial is calculated for the new position the center of the moving window. In this case, for each window is obtained the polynomial coefficients, which doing it would be difficult and time consuming [13]. Hence, Savitzky and Golay in their paper got tables that showed C_i coefficients for the window size of 5 to 25 points, polynomial of degree 2 to 5 and Derivatives up to order 5 with normalization factor [12]. So when the input signal convolved to the C_i coefficients, least squares polynomial fitting done be automatic. This is inherent feature of the Savitzky method [14].

Input parameters for the Savitzky method is polynomial of degree and number of smoothing points. Select the number of points is very important. Because, if the number is too low, a new error in the data is created and if there is too much information on the spectral data will be lost. The Choice Degree of polynomial is a very important. Because, higher degrees preserved width and height attributes of signal, while done less smoothing in the broader peaks. Thus, the smoothing parameters are different in various applications [15].

2. The proposed algorithm

The aim of this paper is extracts lanes of electrophoresis images. In Fig 2 Stages of the algorithm presented is shown.
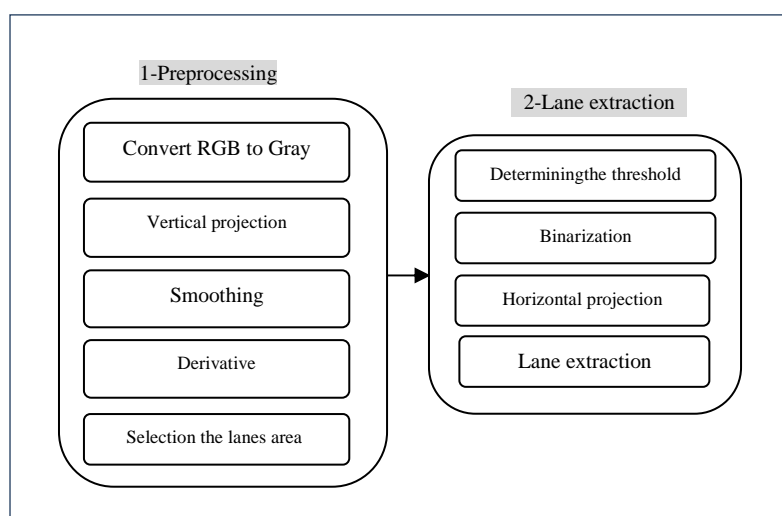

Figure2. The proposed algorithm

2.1 Preprocessing

First the RGB electrophoresis images are converted into gray. The lane places in the middle of a gel also should be extract area of lanes. Hence, we obtained vertical projection of the gel. Vertical projection is similar to a signal that it can be seen that the roughness. So we used the Savitzky filter to smooth vertical projection of the image. We choose 4 and 41 for degree of polynomial and size of window. Fig3 is showing the vertical projection and smoothing of serum protein electrophoresis image.
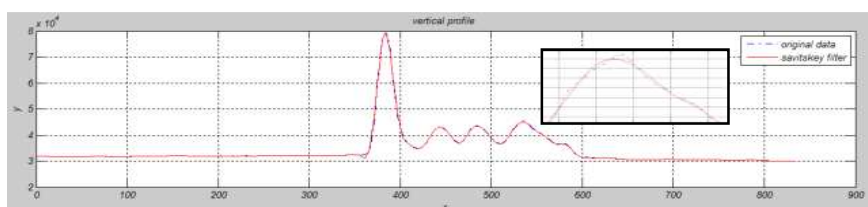

Figure3. The vertical projection and smoothing gel Figure 1

The numbers of vertical axis are very large numbers, which are not suitable for select area of lane. So, the changes in the curve obtained with the first order derivative. The part of the image contain lanes the curve changes is very large and the curve changes is

very small and close to zero in where there aren't any lanes. Hence, these changes in the curve can be used for Identification the area of lanes. In Fig 4 are shown the derivative and area lanes extracted of image.
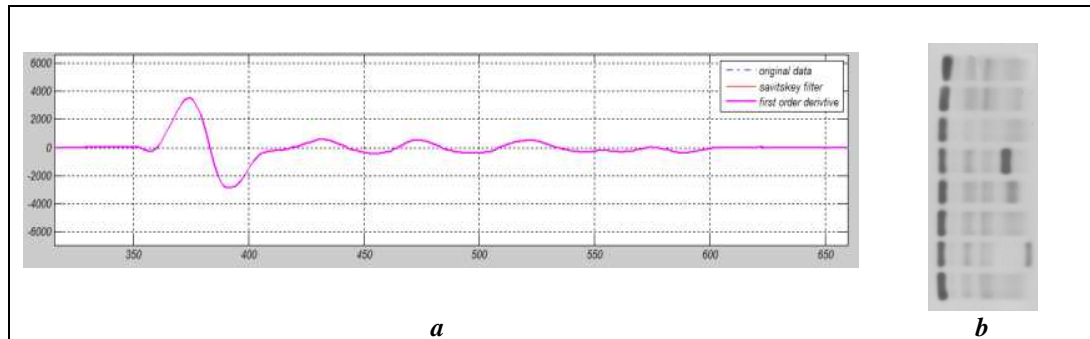

Figure 4.(a) Derivative vertical projection of Serum protein electrophoresis, (b) the area Lanes

The Steps of extraction area lanes in hemoglobin electrophoresis images is exactly the same serum protein electrophoresis.

2.2 Lane extraction

At this stage, the goal is to extract lane from the picture. The first we should convert image to binary. So we need to select a suitable threshold for the image. If the threshold is not the correct, the bands that are colorless eliminate in the binary process. Therefore these lanes are not extracted. So to solve this problem, Otsu method used to select the threshold. As a result the threshold is set to 0.67. So by choosing this threshold all bands will keep in the binary process. Fig 5 shows the result of thresholding in electrophoresis image with Otsu method and without Otsu method.
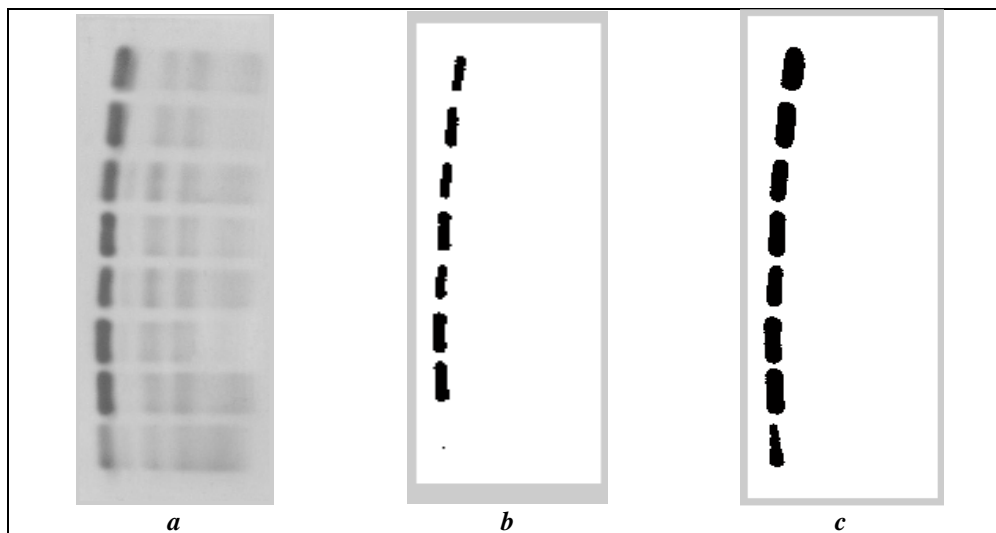

Figure 5. (a) Electrophoresis image, (b) Unsuitable threshold, (c) suitable threshold

In some electrophoresis images, the lanes may well not be separated from each other. In this case, the selections of appropriate threshold with Otsu method plays important role in separating the lanes. After the thresholding image, we obtain the horizontal

projection. Since the values of binary array are zero and one, so position of bands will be easily identified in horizontal projection. Fig 6 shows the horizontal projection of Fig 5(c).
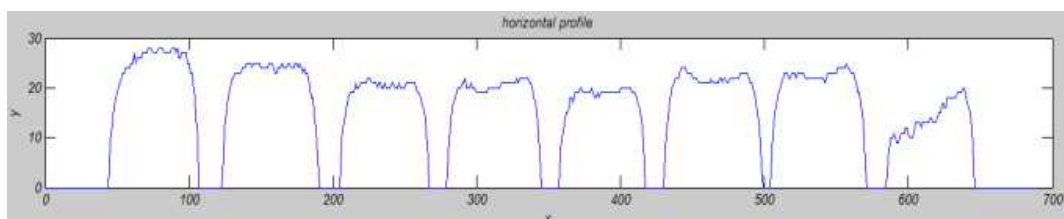

Figure 6.Horizontal projection of Fig 5(c)

As can be seen in Fig 6, the background is indicated with zero and the other parts are the lanes in the image. So we can extract the lanes from image. Fig 7 shows the lanes extracted from the Serum protein electrophoresis (Fig 5(c)).
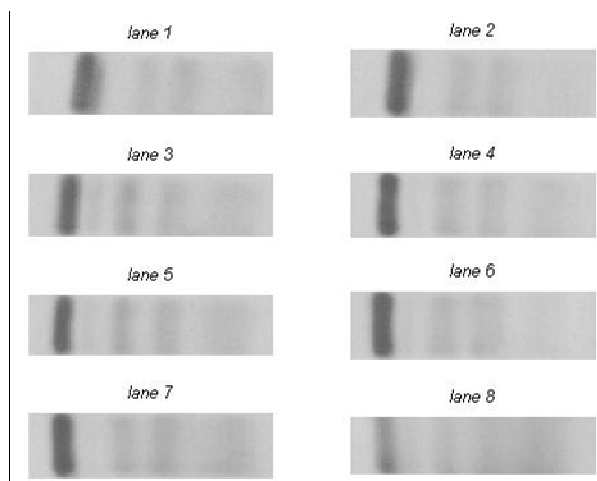

Figure7.Result of lane extraction

As shown in Fig7, all lanes extracted from electrophoresis image. Lane extraction stages for Hemoglobin electrophoresis image are the same as Lane extraction stages in serum protein electrophoresis images.

3. Conclusion

One of the most important steps in electrophoresis images analysis is lane extraction. In some cases, the lanes may well not be separated due to inaccuracy in doing electrophoresis test. Therefore, the user must specify the range of each lane manually. The algorithm presented in this paper used for automatic extraction lanes in electrophoresis images. Otsu thresholding and Savitzky-Golay smoothing are important image processing techniques that are used in this article. The algorithm is proposed includes preprocessing and lane extraction stages. The goal of the preprocessing stage is to select area of lanes from the entire image. In lane extraction stage, lane extraction is

done with using Otsu method .The results of applying the algorithm presented in this paper shows 98% accuracy on images of database.

4. Acknowledgements

5. References

[1] S.C. Anderson and S. Cockayne, "Clinical Chemistry Concepts and Applications", *Waveland Press, Inc., Long Grove, Illinois,* 2003.

[2] K. M.Ward, C. A. Lehmann and A. M. Leiken, "Clinical Laboratory Instrumentation and Automation Principles Applications and Selection", *Saunders, Philadelphia,*1994.

[3] H. Sagha and M. Sorushniya ,"Comprehensive reference laboratory equipment and diagnostic products", *Tehran, ketabmir,*2003.

[4] C.Y. Lin, Y.T. Ching and Y. L. Yang, "Automatic Method to Compare the Lanes in Gel Electrophoresis Images", *IEEE Transactions on Information Technology in Biomedicine,* 2007.

[5] N. Kaabouch, R. R. Schultz, B. Milavetz and L. Balakrishnan, "An Analysis System for DNA Gel Electrophoresis Images Based on Automatic Thresholding an Enhancement", *IEEE Electro/Information Technology*, 2007.

[6] C.M.R. Caridade, A.R.S Marcal, T. Mendonca, A.M. Pessoa and S. Pereira, "An automatic Method to identify and extract information of DNA bands in Gel Electrophoresis Images*", IEEEEngineering in Medicine and Biology Society*, 2009.

[7] M. H. Mohd Noor, Z. Hussain, K. A. Ahmad and A. R. Ainihayati, "Gel Electrophoresis Image Segmentation with Otsu Method Based on Particle Swarm Optimization", *Signal Processing and its Applications*, 2011.

[8] S.C. Park, I. S. Na, S. H. Kim, G. S. Lee, K. H. Oh, J. H. Kim and T.H. Han, "Lanes Detection in PCR Gel Electrophoresis Images", *11th IEEE International Conference on Computer and Information Technology,*2011.

[9] N.Otsu, "A threshold selection method from graylevel histograms", *IEEE Transactions on Systems, Man and Cybernetics*, 1979.

[10] X. Ye, M. Cheriet and C.Y. Suen, "Stroke-model-based character extraction from gray-level document images", *IEEE Transactions on Image Processing*, 2001.

[11] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures", *Anal Chem*, 1964.

[12] R.W. Schafer," On the Frequency-Domain Properties Of Savitzky-Golay Filters", *IEEE Digital Signal Processing and Signal Processing Education*, 2011.

[13] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, "Numerical Recipes in Fortran 77 The Art of Scientific Computing", *University of Cambridge*, 1992.

[14] J. Luo, K. Ying, P. He and J. Bai, "Properties of Savitzky Golay digital differentiators", *Digital Signal Processing Elsevier,* 2005.

[15] J. Chen, L. Wu, T. Pan, J. Xie and H. Chen, "Quantification Method of Glucose in Aqueous Solution by FTIR/ATR Spectroscopy", *IEEEFuzzy Systems and Knowledge Discovery*, 2010.