PREPROCESSING OF ELECTROPHORESIS SAMPLES FOR SUBSEQUENT CLASSIFICATION

Ondřej Krupka

Master Degree Programme (2), FEEC BUT

E-mail: xkrupk03@stud.feec.vutbr.cz

Supervised by: Martin Vítek

E-mail: vitek@feec.vutbr.cz

Abstract: This paper deals with the preprocessing of the elektroforeogram images that were created with respect to highest quality possible. The main goal is the preprocessing of the images based on the contrast enhancement and the detection of lines and bands.

Keywords: EEICT, electrophoresis, preprocessing, segmentation

1. INTRODUCTION

Gel electrophoresis is a significant separation method. It is using a difference between various particle's mobility in an electric field. It is the most common method used for biopolymer, nucleic acid and protein separation. [2]

This paper deals with an agarose gel electrophoresis. The observed sample of DNA is placed into the gel's sample wells. Then the gel is placed into an electrophoresis tank filled with appropriate buffer which allows the conductive connection. After a DC is applied on the tank, samples begin to move from the cathode to anode. The distance travelled depends on the sample weight. Based on the *electrophoretic mobility*, lighter molecules travel longer distances than the heavier ones. After a desired amount of time, the sample is withdrawn and placed under an UV light and captured by a digital camera. This visualization process is allowed by a UV-sensitive dye, which is a part of the gel. [3]

2. PREPROCESSING OF THE SAMPLE IMAGE

The captured image is usually very far from an ideal one. There are many types of interferences that occurs in the process of electrophoresis, for example the *smile effect*, which is caused by an uneven heating of the gel during the process. Another examples are the *bad resolution of sample fragments*, *band blur* etc. [1], [2]

In the matter of a good classification, it is necessary to get as good image as possible. A good preprocessing algorithm of the image is needed to eliminate the interference. [1]

The used algorithm block diagram is on the **Figure 1**.

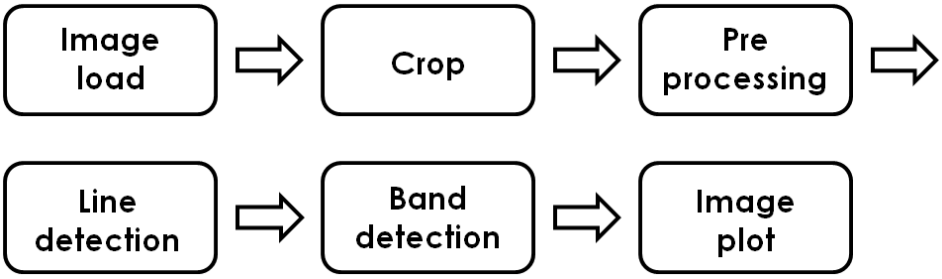

Figure 1: Used algorithm block diagram

2.1. INPUT IMAGES

Sample images were made in the DBME laboratory, the set of eleven images was made with various parameter settings, e.g. voltage, duration and gel density. The image quality was the main goal and after some adjustments in the process, good quality images were achieved – **Figure 3**. Those images are the input for the proposed method.

2.2. CONTRAST ENHANCEMENT

First of all, image contrast is needed to be repaired. The image is automatically cropped by smoothing the image via a median filtering, rounding the values higher than 0.95 to 1 (this is the background of the image) and then detecting the border of the gel, which is not equal to 1, from left and right side. Image is then normalized, converted into grayscale and if the bands are white with black background, the image is inverted. Then the *piecewise linear transformation* is done, this increases the spread of higher used values in the histogram and contrast is enhanced. Next step is to use the *gamma correction*, which is a non-linear contrast transformation. The γ value is in the interval of $<0,6;2,5>$ and $\gamma<1$ is used for overexposed images and $\gamma>1$ is used for underexposed images. Those values are set up manually. The right choice of transformation parameters is very crucial for the method and those values vary from image to image. [1]

2.3. LINE DETECTION

For the line detection, a standard deviation of the image is computed. This 1D signal is smoothed by a median filter and inverted. The local maxima are detected with a windows based on the width of the cropped image in pixels divided by a specified number of lines on the image. Positions of peaks are saved.

2.4. BAND DETECTION

Similar to the line detection, but this time, the average is computed (higher peaks are appearing), smoothed and inverted. Local maxima are detected with a threshold of 15% of the average maximum - **Figure 2**. Once more, detected positions are saved.

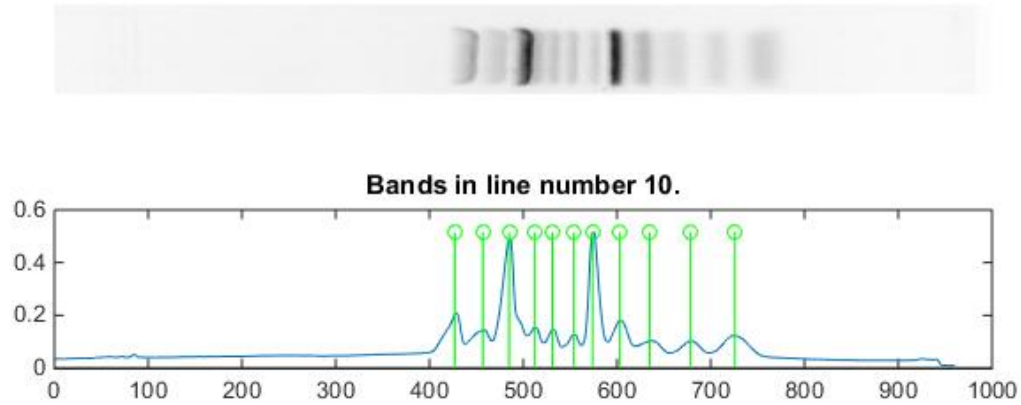

Figure 2: Band detection of Figure 3. a), with a threshold of 15% signal´s maximum

2.5. OUT COMING IMAGES

Based on the saved positions, the image is segmented – **Figure 3**. Almost all bands are detected, there are some false positive and false negative detections, but those depend more likely on the quality of the image – there are detections of sample wells that were not removed by the preprocessing - Table 1.

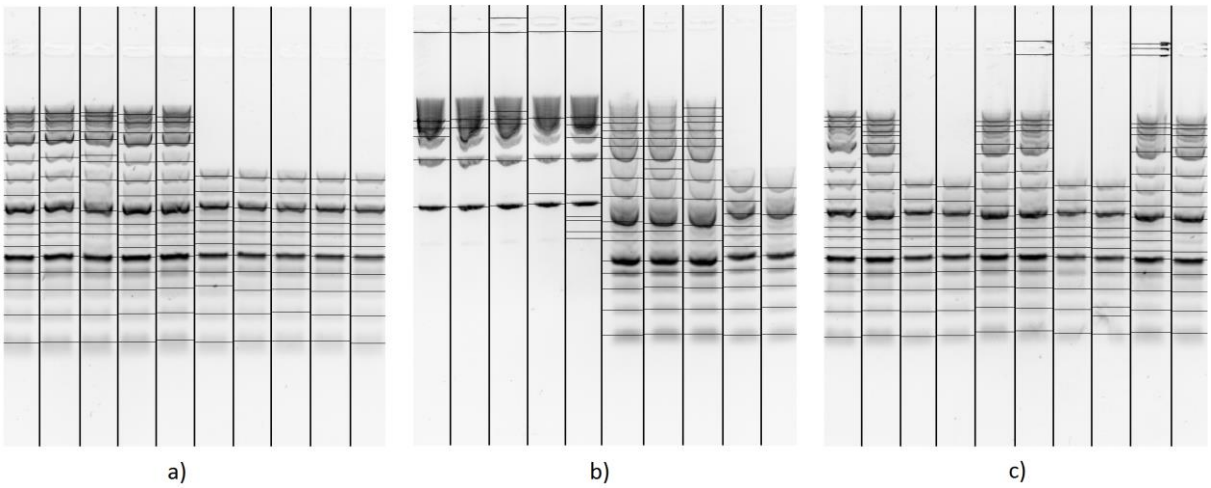

Figure 3: Output images a) One of the first images - blurred bands, curved lines and good contrast – 93.24% successful detection b) Fourth attempt – some blurred bands, some not, mostly straight lines and good contrast – 83.65% successful detection c) Later attempt – mostly clear bands, straight lines and good contrast. 91.45% successful detection

	Band	No band
Detection	89.45%	5.61%
No detection	10.55%	-

Table 1: Results of the method

3. CONCLUSION

From given images, it can be said, that the method is successful – almost 90% of bands were detected. Sensitivity of the method is 0.89 and positive prediction value is 0.94. Images were made almost ideal and with right contrast enhancement and the preprocessing in general, the segmenta-

tion itself is not difficult. Lines are detected without any errors, bands are overwhelmingly detected right on all of the images. Parameter adjustments on the *piecewise linear transformation* had to be made for some images with worse contrast caused by dull Sigma and Biolabs 100bp ladders.

A minority of detections are false positive, or false negative, but this is often caused by an error of the gel, for example a sample well puncture or by a damaged gel. Even with blurred bands, the detection results are very good.

Images, that were preprocessed and segmented by this method, can further be used for electrophoreogram classification, using the cluster analysis.

REFERENCES

[1] SKUTKOVA, Helena, Martin VITEK, Sona KRIZKOVA, Rene KIZEK and Ivo PROVAZNIK. Preprocessing and Classification of Electrophoresis Gel Images Using Dynamic Time Warping. 2013, vol. 8, pp. 1609–1622.

[2] AUSUBEL, Frederick M. Current protocols in molecular biology. Media, Pa.: J. Wiley, order fulfilment, c1987-, 2 v. (loose-leaf). ISBN 97804715033782-.

[3] CHEN, Peter. COLLEGE OF DUPAGE. Electrophoresis [online]. [cit. 2014-11-16]. Available at: http://bio1151.nicerweb.com/Locked/media/ch20/electrophoresis.html