Determination of band position for evaluation of DNA/RNA molecular weight from gel electrophoretic image

FRANTIŠEK RACEK
Department of Weapons and Ammunition
University of Defence
Kounicova 65, Brno
CZECH REPUBLIC

TEODOR BALÁŽ
Department of Weapons and Ammunition
University of Defence
Kounicova 65, Brno
CZECH REPUBLIC

MARTIN MACKO
Department of Weapons and Ammunition
University of Defence
Kounicova 65, Brno
CZECH REPUBLIC

JAROSLAV TŘEŠŇÁK
Oprox, a.s.
Vnitřní 10, Brno
CZECH REPUBLIC
http://www.oprox.cz

J. BENEDIK
Institute of Applied Biotechnologies
Služeb 4
108 52 Praha
CZECH REPUBLIC

Abstract : The paper deals with part of image processing of agarose electrophoresis gel image. Determination of band density and position using band modeling is described there.

Key-Words: Image processing, Electrophoresis, band position, band density

1 Introduction

Electrophoresis is the standard method used to separate, identify, and purify DNA/RNA fragments and proteins. The technique is simple, rapid to perform, and capable of resolving fragments of DNA/RNA that cannot be separated adequately by other procedures. Furthermore, the location of DNA/RNA „band“ within the gel can be determined directly by staining with low concentrations of the fluorescent intercalating dye (like ethidium bromide or SYBR green). Bands containing as little as 1-10ng of DNA can be detected by direct examination of the gel in ultraviolet light. If necessary, these bands of DNA can be recovered from the gel and used in further examinations.

Agarose and polyacrylamide gels can be poured in variety of shapes, sizes, and porosities and can be run in a number of different configurations. The choices within these parameters depend primarily on the sizes and number of the fragments being separated. Polyacrylamide gels are most effective for separating small fragments of DNA (5-500bp). Their resolving power is extremely high, and fragments of DNA that differs in size as little as 1bp can be separated from one another.

Agarose gels have a lower resolving power than polyacrylamide gels but have greater range of separation. DNAs fro 100bp to approximately 50kb in length can be separated on agarose gels of various concentrations. Agarose gels are usually run in horizontal configuration in an electric field of constant strength and direction. Larger DNA fragments, up to 10 000kb in length can be separated by pulse-field gel electrophoresis, in which direction of the electric flux is changed periodically.

When one fragment of DNA/RNA is expected i.e. after simple end-point PCR (polymerase chain reaction), the length of expected fragment can be easy estimated by comparison with the marker size – DNA ladder. The more difficult situation is when multiple bands are expected like after multiplex PCR, RFLP (restriction fragment length polymorphism), or pulsed field gel electrophoresis. These methods are often used in molecular diagnostics in medical research i.e. in diagnostics and specifications of human and animal pathogens. In these cases, the computer analysis (image processing of the digitalized image of agarose gel) of multiple fragments is necessary to receive relevant results.

2 Agarose Gel Electrophoresis Image Processing

As said above an image of agarose gel electrophoresis image (EI) is analyzed. To analyze EI we have divided EI image processing to the following steps:

1. basic editing of image,
2. finding of patterns,
3. transformation of patterns to the 2D signal,
4. finding of candidate of bands,
5. modeling of bands for determinations of band density and position,
6. compensation of image deformation (distortion),
7. calculation of DNA/RNA molecular weight.

Several patterns are usually in one EI and two kinds of patterns can be found there. First one is regular pattern representing information about DNA/RNA fragments. Second one is pattern of molecular weight standard representing a scale factor in EI (DNA ladder). Examples of both regular pattern and pattern of standard are in Figure 1. There are eight patterns in this EI. Two of patterns (left one and right one) is DNA ladder and the others are tested samples.

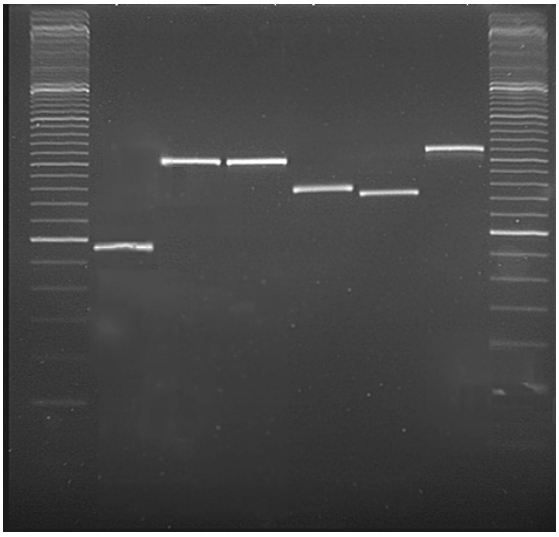

Fig. 1 Gel electrophoresis

White horizontal elements in patterns are bands. Combining information about position and size of bands in one or more patterns allowed us to identify DNA/RNA fragments [1, 2, 3].

The aim is to find the position of bands in regular patterns and to compute DNA/RNA molecular weight of the bands with regard to patterns of standard.

3 Problem Solution

The status of EI image processing is: patterns are found, and the modified pattern 2D signals are computed in all the patterns. The candidates of patterns are also found in all patterns.

It is not sufficient to consider the position of the band vertex as the position of the band because of:

- a band is not axially symmetrical. The shape of the band is characterized by its rapid entering edge, the rounded peak and gentle descending edge. In many cases the shape of the band can be considered as axially symmetrical. This is caused due to narrower width of the band or the band is displayed over small number of pixels.
- It is not possible to find the vertex of the band because it is not displayed due to saturation of the image. Example of this case you can see in Figure 2, where the right band is saturated. Saturation of the image caused that the peak of the band is cut and its upside is flat.

Next possibilities attached with finding of the band position are:

- Band, in the extreme, can be displayed over three pixels only.
- Band is lost in the noise signal.

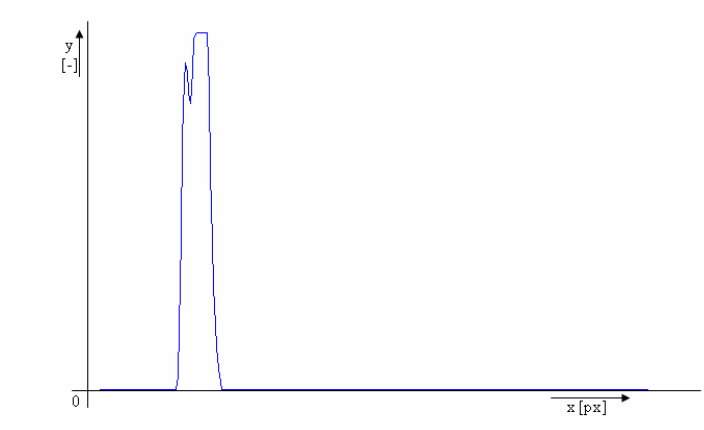

Fig. 2 Pattern 2D signal

Not only position as an input parameter for molar weight computing has to be found. Density as a function of band size has to be evaluated, too. Density D [-] of the band we define as:

$$D = W \cdot I \qquad (1)$$

Where W is weighting function [px^{-2}], I is a surface of the band [px^2].

3.1 Band Approximation

Band approximation is affected due to following difficulties.

- Indetermination of band beginning and its end because these are lost in noise signal.
- Indetermination of band beginning and its end because of using baseline.
- Two bands can be connected. It means, the beginning or end of the band can be situated in the body of the other band. In this case it is not possible to use method Valley to Valley for beginning or end finding, respectively. The case of connected bands you can see in Figure 2.

Said problems lead to the use of band modeling. This means that sufficient mathematical function was searched to fit the nature of the band and to be suitable for band approximation. We have decided to use function defined as:

$$f(x) = A \cdot e^{-\left(\frac{\ln(x)-B}{C}\right)^2} . \qquad (2)$$

Where A is a parameter characterizing high of the band [-], B is a parameter characterizing position of the band [-] and C is a parameter characterizing width of the band [-].

Parameters A, B, C of the model were estimated using nonlinear least squares regression [5]. Before starting regression, a group of sampled variables $y(x_i)$ corresponding with current band has to be separated. Therefore Valley to Valley method was used for separation. It is valid for the segment of variables $y(x_i)$ which is used for approximation: monotonically decreasing segments of variables $y(x_i)$ to the left/right from the vertex of the band are considered as a band. Thus, for the left limit $y(x_l)$ is valid: $y(x_l) \leq y(x_l\text{-}1)$ and for right limit $y(x_p) \leq y(x_p+1)$.

More complicated situation causes in the case of approximation of saturated band. Using all the variables $y(x_i)$ between left and right limit should cause bad approximation. Therefore the entering edge and descending edge have to be only used for approximation. Result of approximation using entering and descending edge is seen in Figure 3. Variables $y(x_i)$ of entering or descending edge used for approximation are bordered with red stars in Figure 3. Model of that band is than drawn by black curve.

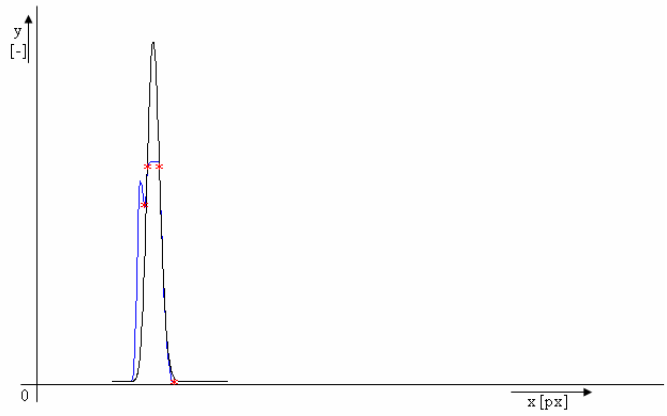

Fig. 3 Approximation of satured band

3.2 Model Verification

Presented here information originate from verification of EI image processing of 144 real EI received from Lab.of Molecular Diagnostics. They were bmp or jpg black and white images. They were real images used in general medical practice and no selection was done. It means, no requirements were set on image resolution, image quality, number of patterns in one EI, number of bands in one pattern, etc.

Error of estimation of A, B, C parameters was computed as a residual R [4]:

$$R = \sum_i \left[y(x_i) - f(x_i, A, B, C) \right]^2 . \qquad (3)$$

Firstly we looked for dependence of residual R on number of pixels N used for band approximation. It can be said that 6657 bands in 144 images was processed. Number of pixels used for approximation ranges from 2 up to 21. In Figure 4 you can see frequency n_N of separate number of pixels N which were used for approximation. As it is seen, most frequency number of pixels was 4.

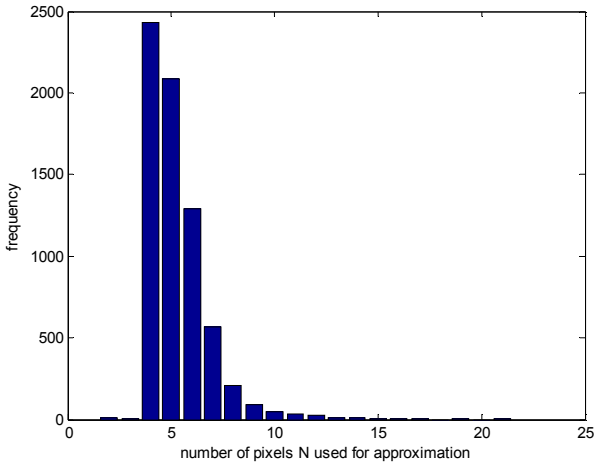

Fig. 4 Frequency of number of pixels used for approximation

In Figures 5a – 5d you can see residuals R computed during A, B, C parameters estimation. Presented there examples are valid for number of pixels $N \in \langle 4,7 \rangle$. Final result of this validation is summarized in Figure 6. There are presented average residuals $\overline{R}$ computed for concrete number of pixels N. It is typical, the greater frequency n_N the lower average residual $\overline{R}$.

The significant advantage of band approximation is attainments of subpixel resolution of determination of band position. Result of said is presented in Figure 7, where average displacement of band position is displayed as a function of concrete number of pixels N. Average displacement is defined as:

$$\overline{\Delta} = \frac{1}{n_N} \sum_{i=1}^{n_N} \Delta_i \quad . \tag{4}$$

Where Δ_i is displacement of i^{th} band position, computed:

$$\Delta_i = P_{Ti} - P_{Ai} \quad . \tag{5}$$

Where P_{Ti} is i^{th} band position determined as a top of i^{th} band peak, P_{Ai} is i^{th} band position determined from parameter B of i^{th} band model.

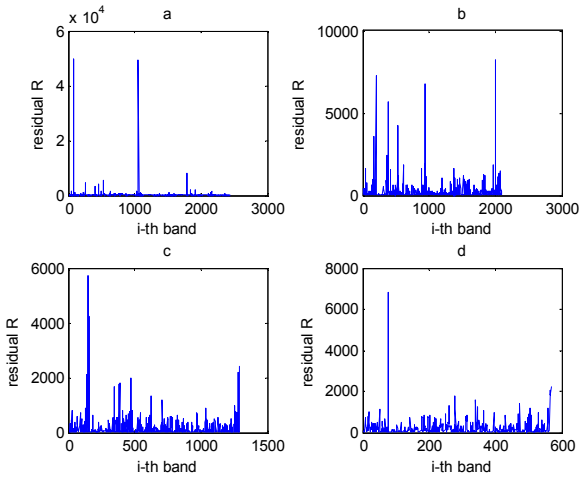

Fig. 5 Residuals computed during band approximation, a – bands with 4 variables $y(x_i)$ used for approximated, b) bands with 5 variables $y(x_i)$ used for approximated, c) bands with 6 variables $y(x_i)$ used for approximated, d) bands with 7 variables $y(x_i)$ used for approximated

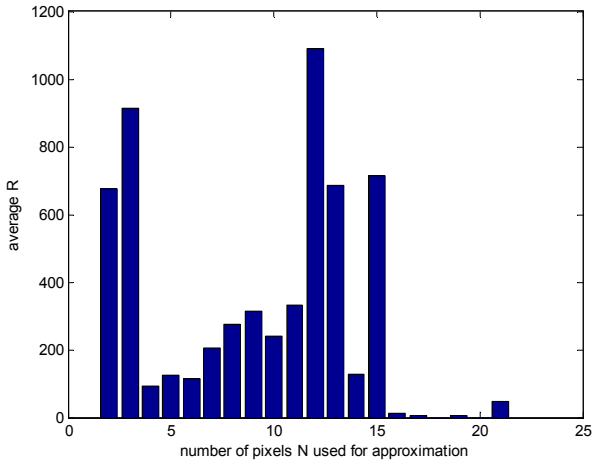

Fig. 6 Average residual as a function of number of pixels used for approximation

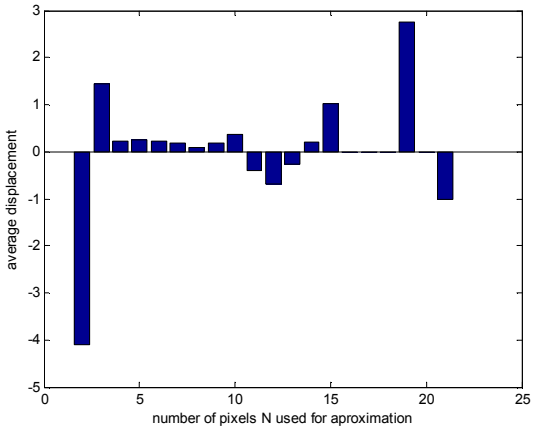

Fig. 7 Average displacements as a function of number of pixels used for approximation

In Figures 8a – 8d are presented computed displacement Δ_i for number of pixels $N \in \langle 4,7 \rangle$.

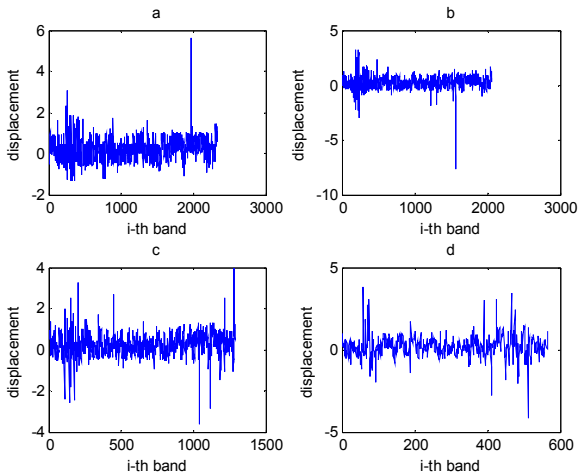

Fig. 10 Displacements of band positions a) 4 variables $y(x_i)$ used for approximation, b) 5 variables $y(x_i)$ used for approximation, c) 6 variables $y(x_i)$ used for approximation, d) 7 variables $y(x_i)$ used for approximation

4 Conclusion

Electrophoresis is the standard method used to separate, identify, and purify DNA/RNA fragments and proteins. Agarose gel electrophoresis image can be successfully processed. Finding position and size of the bands is one of the important steps during agarose gel electrophoresis image processing. By the use of presented model it is possible to reach sub-pixel resolution of band position and compute density of band. Although the presented displacements seem to be small, these displacements are really significant when the DNA/RNA molecular weight is computed from band position.

Applicability of processes presented in this article was confirmed through using that in general medicine practice and creating commercial program of the EI image processing based on presented processes.

References:

[1] Sambrook J, Fritsch EF, Maniatis T.:Molecular Cloning. Cold Spring Harbor Lab. Press, 1989

[2] Horváth R, Dendis M, Schlegelová J, Růžička F, Benedík J. : A combined AFLP-multiplex PCR assay for molecular typing of Escherichia coli strains using variable bacterial interspersed mosaic elements. Epidemiol Infect. 2004 132(1):61-65.

[3] Reiner Westermeier. Electrophoresis in Practice. A Guide to Methods and Applications of DNA and Protein Separations Fourth, revised and enlarged Edition, 2005. ISBN 3527311815

[4] Weisstein, Eric W. "*Residual*." From MathWorld—A Wolfram Web Resource. 2007 http://mathworld.wolfram.com/Residual.html

[5] Weisstein, Eric W. "*Nonlinear Least Squares Fitting.*" From MathWorld--A Wolfram Web Resource. 2007 http://mathworld.wolfram.com/NonlinearLeastSquaresFitting.html