# Improvement of Electrophoretic Gel Image Analysis

Ivan Bajla*, Igor Holländer, Kornel Burg

Austrian Research Centers Seibersdorf
A-2444 Seibersdorf, Austria
ivan.bajla@arcs.ac.at

*on leave from the Institute of Measurement Science,
Slovak Academy of Sciences, Bratislava, Slovak Republic

## Abstract

A novel methodology of DNA gel image analysis has been proposed that is based on 2D image processing methods instead of previously used 1D Gaussian deconvolution. The algorithms specifically tailored to band boundary detection and image intensity homogeneity characterization have been developed. The proposed methodology involves a modified type of user interaction.

**Keywords**: electrophoresis, gel image analysis, 2D edge detectors, geometry-driven diffusion filters

# 1 Introduction

The development of gel electrophoresis as a method of separating and analyzing DNA has significantly influenced the progress achieved in molecular biology in the last 20 years. The electrophoresis is a method that separates molecules on the basis of their size, electric charge, and other physical properties. The *DNA gel electrophoresis* ([1]) refers to the technique in which DNA macromolecules are forced across a span of gel, which is a colloid in solid form, motivated by an electrical current. Two types of gel matrices, *agarose* and *polyacrylamide* gels, are mostly used for DNA gel electrophoresis. In these gels the electrophoretic mobility of macromolecules is thought to be determined primarily by the volume fraction of pores within the gel that the macromolecules can enter. Hence, a mixture of DNA molecules of different sizes will separate into discrete blobs (the so-called *bands*) in the process of electrophoresis.

Samples of DNA, digested by specific restriction enzymes into segments of different sizes, are loaded in the separated wells at the front edge of the gel. After turning on the power supply voltage each sample runs in its own trace (the so-called *lane*). A fluorescent dye is used for marking the positions of the DNA bands in the gel which is then photographed to provide a permanent record of the electrophoresis experiment. In the recent years a number of software systems have been developed for digitization of the photographs of electrophoretic gels, processing and analysis of the digitized images to yield the final information on DNA fragment mobility, i.e., the band positions in each individual lane given in molecular weight (DNA base pairs).

Certification of seed material according to its geographical origin is a relevant problem in present day forestry. The Molecular Biology Group of the Biotechnology Unit of the Austrian Research Centers Seibersdorf developed a chloroplast DNA based marker system ([2]) that is capable to differentiate geographical areas for oak species *Quercus robur*. The system is planned to be put in practice in the near future, however, the available software packages for gel image analysis are not capable to resolve small differences in DNA fragment migration and time consuming manual approach has to be applied in a number of cases. Novel image processing algorithms, which can decrease sensitivity to image artifacts encountered in imperfect gels (frequent cases) and increase the credibility and reliability of semi-automatic gel image analysis, are needed

# 2 Problems with correct band detection and effective user interaction

## 2.1 Ill-posed problem of 1-D signal deconvolution

Most of gel images generated in practice suffer from various types of distortions and degradations. To mention just the most frequent of them: –geometrical distortion of the position of the entire image, –horizontal lane deformation (*smiles*), –spatial brightness nonhomogeneity, etc. Correction of defects of these types does not represent an essential problem and effective tools for these tasks are available in software systems.

However, noise and local artifacts, often present in gel images used in the chloroplast DNA-based marker system, cause essential problems of proper band detection, particularly if low contrast bands occur. The failure of the standard software tools, which work satisfactorily in perfect gel images, is related to properties of the 1D deconvolution operation conventionally used in the software systems. The reasons underlying these problems can be described within the theoretical framework of 1D signal deconvolution. This operation is conventionally defined as a solution of a linear first-kind Fredholm integral equation:

$$\int_0^1 K(s,t)\, f(t)dt = g(s), \qquad 0 \le s \le 1, \tag{1}$$

with a convolution type kernel $K(s,t)$. In contemporary programs the 2D band image structures in a lane are represented by an averaged 1D profile $g(s)$ of the intensities along the lane. The solution $f(t)$ of the equation (1) is searched as a 1D Gaussian deconvolution applied to the profile $g(s)$. According to [3] deconvolution of signals corrupted by noise and other defects encounter the following crucial problems: (i) ill-conditioning of this inverse problem, (ii) the function $K(s,t)$ is unknown and can only be estimated from some convenient functional class. Furthermore, in the case of gel images this function cannot serve as a model for image blurring mechanism, and the real 2D problem of gel image analysis is simplified to a 1D problem. For improvement of band detection operation we propose to regularize the original gel image by some 2D noise filter, and, instead of using 1D deconvolution, to develop specific band boundary (edge) detectors and lane image homogeneity indicators (measures).

## 2.2 Need of interaction paradigm change

Due to the high variability in quality of gels, it is not possible to design a universal band detection algorithm capable to resolve all practical situations in a fully automatic mode without any user intervention. Therefore, the existing software systems incorporate specific tools for interactive settings of the parameters of band detection algorithms. These tools are based on the multiparametric interaction paradigm: (i) the user is prompted to vary some adjustable parameters to influence the Gaussian deconvolution operation, (ii) a complex result of the repeated attempts of band detection (a re-generated lane profile and the image with the graphical markers indicating band positions) is displayed. This paradigm is illustrated in Fig1 in which windows used for band detection in the system *GelComparII*™ (Applied Maths, Belgium) are displayed. The new result may be either completely correct or it can contain errors of both types (undetected true bands and falsely detected bands). The user has no possibility to influence the band detection operation in dependence on one of two crucial types of errors separately. If he does not wish to continue a troublesome process of searching for the optimum combination of the numerical values, an alternative to the *assisted complete detection* is provided. Namely, the user can reject or add necessary detections directly by positioning the cursor at the proper location. However, this process is again laborious, and moreover it is not precise.

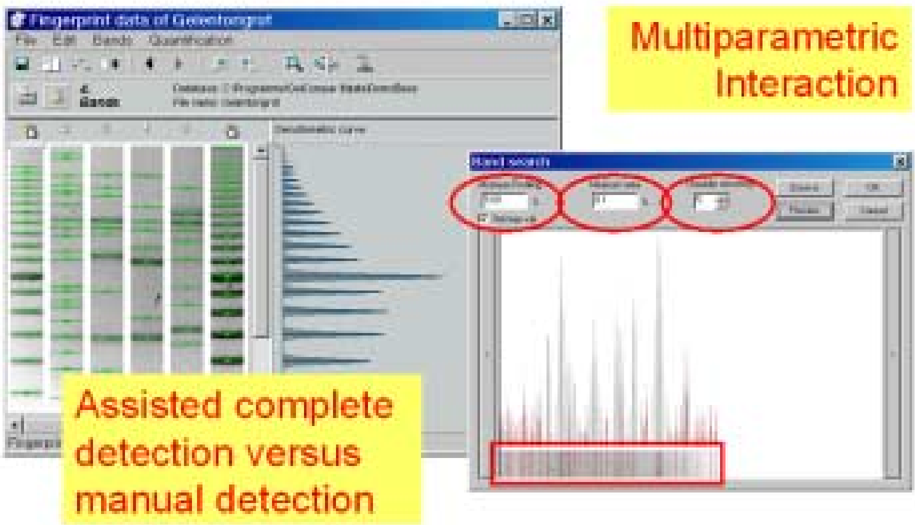


Figure 1: Two basic windows dedicated to band detection in the system *GelComparII*™. Three parameters (*minimum profiling*, *minimum area*, and *shoulder sensitivity*) are required for 1D deconvolution of the density profile.

# 3 A novel approach to band detection and user interaction

Based on the investigation of critical cases of real gel image analysis we proposed the following novel methodology of gel image processing:

1. to regularize the gel image with an appropriate nonlinear 2D-filtering algorithm,
2. to apply a first stage linear operator *Detband(i)* for detection of horizontal band boundaries as a function of lane rows *i,*
3. to smooth the values of *Detband(i)* with an appropriate window filter (to suppress maxima within the area of the lane background),
4. to couple the band boundaries for a final *bandbox* generation,
5. to apply a second stage operator for eliminating false band detections caused by image artifacts or redundant *Detband(i)* local maxima in the lane background,
6. to calculate mobility distances of the bands embedded in the bandboxes.

In [4] we explored the nonlinear image filtering method that is based on the principle of geometry-driven diffusion (GDD), given by this equation:

$$\partial I / \partial t = div \left[ c\left( \left| \nabla I(x,y,t) \right| \right) \nabla I(x,y,t) \right], \qquad (2)$$

where $\nabla I(x,y,t)$ is the gradient of an image intensity function I(x,y,t) , the conductance c(.) is a function of spatial coordinates x, y and t is time which corresponds to the index of the iteration step in discrete implementation. Tests with this GDD-filter applied to individual lanes showed its good regularization properties. The GDD-prefiltering of the lane image enables to use a simple linear detector of horizontal band boundaries.

The linear band boundary detector is defined for all rows $i$ as the *cumulative row difference* operator $D_i = \sum_j Dif_i^{\,j}$, where *j* are the indices of the lane image columns and $Dif_i^{\,j} = \left| I_{i+1,j} - I_{ij} \right|$

for the image intensities $I_{ij}$ in the given lane. Thanks to the image regularization via the GDD filtering good estimates of band boundaries are obtained as local maxima $\{d_k\}$ of the cumulative row difference operator. To simplify the further description we display in Fig.2 a vertical lane in horizontal position (bands are oriented vertically now).

For elimination of false detections (red crosses in Fig.2) which occur in some places of rather inhomogeneous background a suitable mechanism of coupling the local maxima $\{d_k\}$ into maxima doublets is needed. Such doublets would represent the bands themselves and remaining detections could then be interpreted as false. For this purpose we proposed a stepwise constant function $S_j$ that is defined in the following way. For each couple of neighboring band boundary detections two symmetric rectangular regions $\Omega_1$ and $\Omega_2$ are introduced. For these regions medians of intensities are calculated.

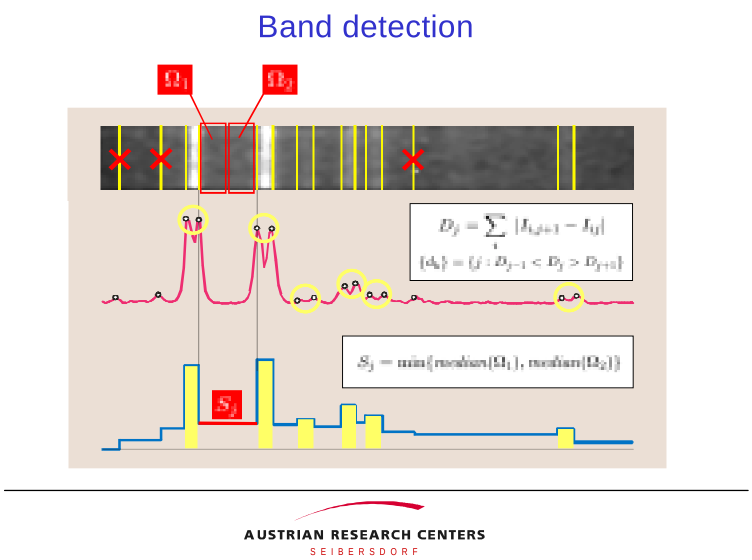


Figure 2: Illustration of the construction of the stepwise constant function $S_j$ and maxima doublets for discrimination of the band detections from the non-band detections (this is an animated Powerpoint2000 presentation).

Finally, the corresponding constant value of the function $S_j$ is given as the minimum of two medians. Thus, the true band boundary detections can be separated from the false detections by the criterion of maxima doublets which correspond to the positive "teeth" of the function $S_j$ (see Fig.2). The advantage of this approach is that the remaining uncoupled detections can be deleted automatically.

The second stage operator is needed to improve the band boundary indication. Several measures of local intensity nonhomogeneity have been tested, e.g., a measure proposed in [4]. We propose a specific indicator of band boundaries. It serves for final checking whether the paired detections do not comprise a false element, and also for more precise calculation of the mobility distances. The design of the improved band boundary indicator is illustrated in Fig.3.

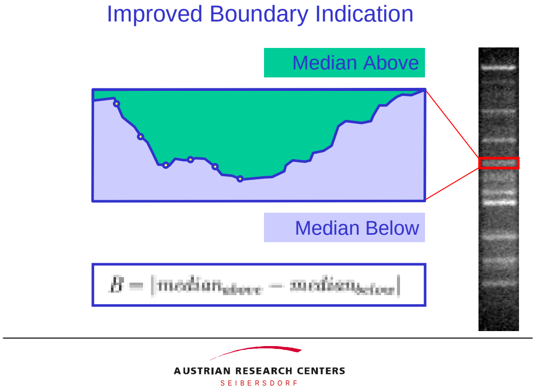


Figure 3: Design of the improved band boundary indicator for each detection. It is given via the absolute difference of the intensity medians in two disjunctive regions (median above, median below). This is an animated Powerpoint2000 presentation.

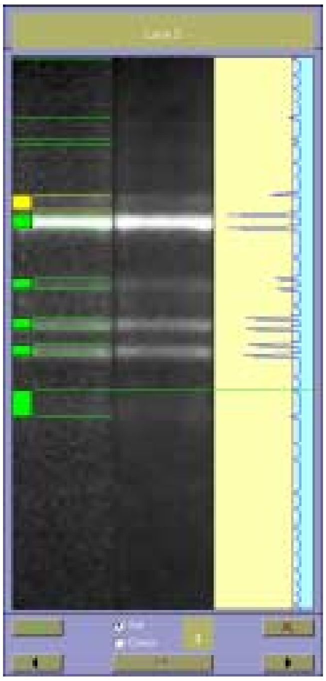

Figure 4: The layout of the band detection window in GAS1: left window shows the original lane, the middle window shows a GDD-filtered version of the lane, the right window serves for displaying a plot of the first-stage operator.

For each individual band boundary from a set of all paired detections a rectangular pixel neighbourhood is defined. An intensity gradient array is calculated for this neighbourhood. For each column of the neighbourhood the position of the maximum gradient value represents a reasonable estimate of a band boundary point. A set of these points constitutes an estimate of the whole band boundary which separates the neighbourhood into two disjunctive regions. The statistical characteristics of these regions may serve for measuring the difference between the background and band intensity distributions. Thus, we propose to compute the *median above* and *median below* and to define the final band boundary indicator *B* simply as their absolute difference. After automatic coupling of the corresponding detections using the stepwise constant function $S_j$ and checking the final bandboxes for false detections, the integral band intensity (density) is calculated. Finally, the inertia axes of the bands are found for which more precise values of mobility distances can be calculated.

The essentially modified interaction is related to the band boundary detection. In Fig.4 the layout of the band detection window is displayed. The whole process can be described as follows.

1. In the first stage (the“set” radio-button), after selecting a lane to be analysed (left window in Fig.4) from the original gel image, the user controls the threshold selected from the set of the *Detband* maxima (right window) and compares detected bandboxes in the left window to band intensity structures in the middle window (the GDD-filtered version of the lane).
2. If some (very low contrast) bands are not detected by the first stage detector, the user can add missing bandboxes manually. Once all the true band boundaries are detected, the user proceeds to the second stage (the “check” radio button).
3. Immediately after pushing the button all uncoupled detections are rejected as false. A final check procedure is then available in the second stage in which the detected bandboxes (green rectangles) are sequentially displayed to the user in the increasing order of the values of the improved band boundary indicator. The user can interrupt the checking procedure in any moment by pushing the “OK” button.
4. The graphical representations of the final mobility distances are then overlaid on the original lane.

# 4 Conclusions

A novel methodology of the gel image processing and analysis has been proposed. It involves a modified scheme of user interaction. Image processing algorithms have been developed and tested which are tailored to specific band boundary detection and characterization of image intensity homogeneity. The pilot implementation of the program system GAS1 is now being tested.

References

[1] D.Tietz (ed): Nucleic Acid Electrophoresis. Springer-Verlag Berlin Heidelberg New York, 1998.

[2] U.Csaikl, K.Hohl, K.Burg: Assessing genetic diversity in Austrian oak populations: cpDNA variation. In: Proc. of the Conference on Diversity and Adaptation in Oak Species, 1998, pp 98-101.

[3] P.Ch.Hansen: Numerical Aspects of Deconvolution. Lecture Note, Department of Mathematical Modelling, Technical University of Denmark, http://www.imm.dtu.dk/~pch.

[4] I.Bajla, I.Holländer: Nonlinear filtering of magnetic resonance tomograms by geometry-driven diffusion. *Machine Vision and Applications*. 1998, Vol.10, p.243-255.