AUTOMATIC LANE DETECTION AND SEPARATION IN ONE DIMENSIONAL DNA GEL IMAGES

Akbari A.*, Algregtsen A.

Department of Informatics, University of Oslo, P.O.Box 1080, N-0316 Oslo, Norway

* Corresponding author: e-mail: akbara@ifi.uio.no; fritz@ifi.uio.no

Keywords: lane detection, lane separation, DNA

Summary

Motivation: In this paper we describe the automatic detection and separation of lanes in DNA gel images, applying an iterative low-pass filter and equivalent width algorithm. DNA gel images often contain more than one lane in which, each lane contains either DNA for an individual or standard markers. Each time DNA for an individual in one lane is going to be analyzed or compared with DNA from other individuals within or between gel images, the location of each lane within the gel images should be recomputed. Isolating the lanes in a DNA gel image and making a specific image file for each lane makes the process of analyzing and the comparison of individuals more feasible. In addition the storage space will also be reduced.

Results: A simple iterative low-pass filter followed by equivalent width algorithm is applied to localize and separate the lanes in DNA gel images. Results obtained by this method are reliable and will reduce further computations when the DNA for an individual is going to be compared with other individuals within or between gel images.

Availability: available on request from the authors.

Introduction

DNA gel images often contain more than one lane, where each lane contains DNA for one or more (mixed samples) individuals. The efficiency of both within and between gel comparisons will be increased by having DNA fingerprints for each individual on a separate image (file). That is because the data for each lane (individual) would be available directly from its image file instead of searching through the whole gel image every time it is needed.

The focus of this work will be applying an iterative low-pass filter and equivalent width algorithm to locate and separate lanes in one dimensional gel images. A number of pre-processing steps should be applied on the image prior to the lane detection and separation. First a low-pass homomorphic filter is applied to enhance the image followed by an edge preserved noise filtering algorithm. Then a background normalization operation is applied to the resulting image. Gel electrophoresis and different probes for hybridization are used to produce the one dimensional DNA fingerprints and the images are generated on X-ray film (Kirby, 1990; Burke, 1989; Debenham, 1992). Ideally, the lanes on the electrophoresis should be equidistant. But, DNA are often degraded by different types of noise, e.g. film grain noise and noise from other sources, such as electrophoresis, membrane, the hybridization process, etc. Due to these degrading facts, lanes on the gel images are not equidistant (Fig. 1), and therefore the process of separating lanes from gel images will be complicated. The gel images we are working with have no non-migrated bands that represent the electrophoresis wells which could be used as a reference point for each lane. On the contrary we have no a-priori information for the start and the end of each lane. Sometimes, handling the gel into the membrane will also cause localized stretching on the gel. If such kinds of distortions are too large, the image should be adjusted using a suitable geometrical transformation algorithm. Otherwise, dropping a few pixels (depending on the scanning resolution) from each side of each lane will help to separate the lanes from the gel image. The gel image in Fig. 1a indicates that there are no specified edges in the vertical direction for lanes in the image.
The width (X-direction) of the DNA bands on each lane varies due to their radioactive strength. The spaces between lanes are almost indistinguishable on the lower part of the image. Since the lanes on the DNA fingerprint gel images have no specified edges, artificial edges will be made for each lane on the image. To do so, a 2-D. mean filter of the size (3, ImageHeight/3) is used to smooth the DNA bands along each lane (Fig. 1b).

**Fig. 1.** (a) A DNA fingerprint image contains lanes which are not equidistant due to noise and other effect in the image; (b) the image obtained by using a two dimensional mean filter of the size (3, ImageHeight/3) on image (a).

**Methods and Algorithms**

**Iterative Low-pass Filter.** An approach for separating the lanes from a gel image is based on the one dimensional signal obtained by averaging the intensity for each column in the gel image into the horizontal axis. Figure 2, illustrates the average signal computed from the gel image in Fig. 1a.

**Fig. 2.** Signal obtained by averaging the intensity for each column in the image in Fig. 1b.

To be able to use the average signal for computing the positions of the lanes in the gel image, the average signal should be smoothed such that the number of minimum points on the signal being equal to the number of lanes in the image. At the same time the minimum points in the smoothed signal should be approximately equidistant. The number of the lanes in the image is given as input parameter to the program. The approximate distance between two lanes in an image with a sampling frequency of one pixel per 200 µm will be 7–8 pixels. This option can also be changed by the user. Total space between lanes in the image will be approximated by multiplying the (number of (lanes – 1)) with the approximate distance between lanes. The lane width which is...
used to compute the window size for the first iteration of low-pass filter is approximated with dividing the effective width by the number of lanes.

The effective width is: \( \text{(image width)} - \text{(total distance between lanes)} \).

To smooth the average signal a low-pass filter with a window size \((4/5 \times \text{Lane width})\) is used. The window size is chosen to be smaller than the approximate lane width to preserve the information for the lanes in the gel image. We use this argument to generalize this filter applicable on different DNA fingerprint images. Otherwise different window sizes should be chosen for different images. The low-pass filter is used because the low frequency part of the average signal assumed to represent the lane frequencies in the image and the higher frequencies in the signal are assumed to noise. The smoothed signal will be analyzed, if the number of minimum points in the resulting signal is more than the number of lanes on the gel image, the smoothing operation will be performed iteratively until the correct result is obtained. If iterative smoothing is acquired, then the size of the window for next iteration will be adjusted with the factor of \((1/3 \times \text{previous window size})\). The reason of choosing window sizes in this way is to preserve the information about the lanes in the image.

This could be written as follow:

\[
W(0) = \frac{4}{5} \times \text{Lane Width}
\]

\[
W(n) = \frac{1}{3} \times W(n-1)
\]

where, \( n = 1, ..., \) ....

The procedure for computing the approximate lane width the following algorithm is used.

Total distance between lanes = (Number of lanes - 1) * (Approximate distance between lanes).

Approximate effective width = (Image width) - (Total distance between lanes).

Approximate lane width = (Approximate effective width) / (Number of lanes).

As it is outlined above, the resulting lane width will be used to compute the size of the window for the low-pass filter in the first pass. The signal obtained by this filtering process will be analyzed to find out if more filtering iterations should be performed. Each local minimum of the smoothed signal in Fig. 3, represents a point within the corresponding lane in the image in Fig. 1a. To visualize this point, it is illustrated by “white line” along each lane in the image in Fig. 3b.

Information, such as coordinates and intensity levels for each local minimum and its two neighboring maximum points are used to compute the edge coordinates of the corresponding lane in the image. Computing of edge coordinates for each lane is performed by using the equivalent width algorithm (Gray, 1976; Oppenheim, 1983; Marple, 1987; Pristley, 1989). The results of applying the iterative low-pass filtering followed by the equivalent width algorithm on three different DNA fingerprint images are illustrated in Fig. 4.

![Image](image-url)

**Fig. 3.** (a) Signal obtained by using the low-pass filter with window size \( W(0)=9 \) (computed by the procedure described in the text) on the signal in Fig. 2 in only one iteration; (b) “White line” along the lanes in the image represent the corresponding local minima of the smoothed signal in (a).
Discussion

In this work an iterative low-pass followed by equivalent width algorithm is applied to detect and separate lanes in gel images. Results obtained by this method are reliable, but it has some disadvantages. One disadvantage to this method is the approximate distance between two lanes which is the user choice and makes effect on the separated number of lanes. If the number of lanes resulted from the program compared to the number of lanes (in the image) input to the program are different, then the separated lanes should be analyzed visually about the lane/lanes which are not detected.

References


Fig. 4. Images on the second row are obtained by applying the iterative low-pass filtering followed by the equivalent width algorithm on the images on the first row correspondingly.