



# Normalizing the background and removing the trend in one-dimensional DNA fingerprint images

Akbar Akbari\*, Fritz Albrechtsen

*Department of Infomatics, University of Oslo, P.O. Box 1080, Blindern, N-0316 Oslo, Norway*

## Abstract

Maximizing an individual's genetic information from its DNA fingerprint image depends on the number of bands distinguished from the background. To approach this goal, the background should be normalized while the information is preserved. Morphological operators have been used by some authors to normalize the background for two-dimensional gel images. Methods such as mean, median and "maxpolygon" are presented in this work to normalize the background in DNA fingerprint images. Mean and median methods will lead to some deformations. Visual evaluation of the results show that the original shape of the column signals are better preserved by the maxpolygon.

© 2003 Elsevier B.V. All rights reserved.

*Keywords:* Background normalization; Removing trend; Maxpolygon; DNA

## 1. Introduction

Getting the maximum genetic information in a DNA fingerprint image that belongs to an individual, depends on the number of bands detected from the image. Therefore, the detection of all DNA bands having intensities that exceed the background noise will be one of the important steps in DNA fingerprint analysis. The problem in implementing this step directly on DNA fingerprint images is that the background intensity level, upon which the DNA bands appear, is not only non-uniform over an image but also varies between images. Therefore, it is necessary to normalize the background (remove the trend) in the image, so that the DNA bands are better distinguished from the background.

Performing this task requires a background image which can be defined as an image prepared in the

same way as a DNA fingerprint gel image, but without the DNA fingerprint included. Generally, a background image can be estimated by defocussing or removing the objects (DNA bands) from the original image. Subtracting the background image from its corresponding DNA fingerprint image would result in a background normalized image. Since preparing such a background image is not possible, it must be estimated from the DNA fingerprint image itself. In this work, methods such as mean and median filtering, morphological operators [1–3] and maxpolygon are tested on a number of DNA fingerprint images to estimate their corresponding background images. These methods could be applicable to other electrophoretic methods such as sodium dodecylsulfate–polyacrylamide gel electrophoretic (SDS–PAGE) separations of proteins. Then estimated background images will be used to normalize the background in their corresponding original images. A combination of morphological operators have been used by some authors [2,3] to estimate

\*Corresponding author. Tel.: +47-2285-2401.

E-mail address: [akbara@ifi.uio.no](mailto:akbara@ifi.uio.no) (A. Akbari).

background normalized images for two-dimensional gel images. A structuring element that depends on the objects in the image should be used in this method. To apply mean or median filters, a filter size should be defined which depends on the size of the DNA bands in the image. It may be necessary to iterate the filtering to remove or defocus all DNA bands in the image. In that case the number of iterations depends on the size of the filter.

The maxpolygon method is based on the global maximum in the image and the maximum value in a sliding window over the column signals in the DNA fingerprint image. Visual inspection will be used to evaluate the results obtained by these methods.

## 2. Experimental

### 2.1. Materials

The DNA fingerprints which we are going to use in this project, are produced with the 33.6 probe. The

reason is that the 33.6 probe can produce high resolution fingerprints from the root vole *Microtus oeconomus* as well as from several non-rodent mammals [5]. Throughout this article we will refer to images generated on X-ray film by DNA fingerprints in the electrophoresis gels as gel images. We are dealing with one spectral band, i.e., images with gray values between 0 (black) and 255 (white). A rotating drum scanner are used for image digitization in this work. A sampling frequency between 1 and 2 pixels per 200  $\mu\text{m}$  is deemed to be adequate [4]. DNA fingerprint images produced by slab gel electrophoresis are often degraded by different types of noise such as film grain noise and noise from different other sources, for example electrophoresis, membrane, hybridization process in making DNA fingerprints, dirt on the film (due to holding the film with hand and washing the film), etc. To enhance the images we have applied a homomorphic low pass filter [6,7] as a preprocessing operation.

Fig. 1b illustrates the image obtained by applying the homomorphic low pass filter on the image in Fig.

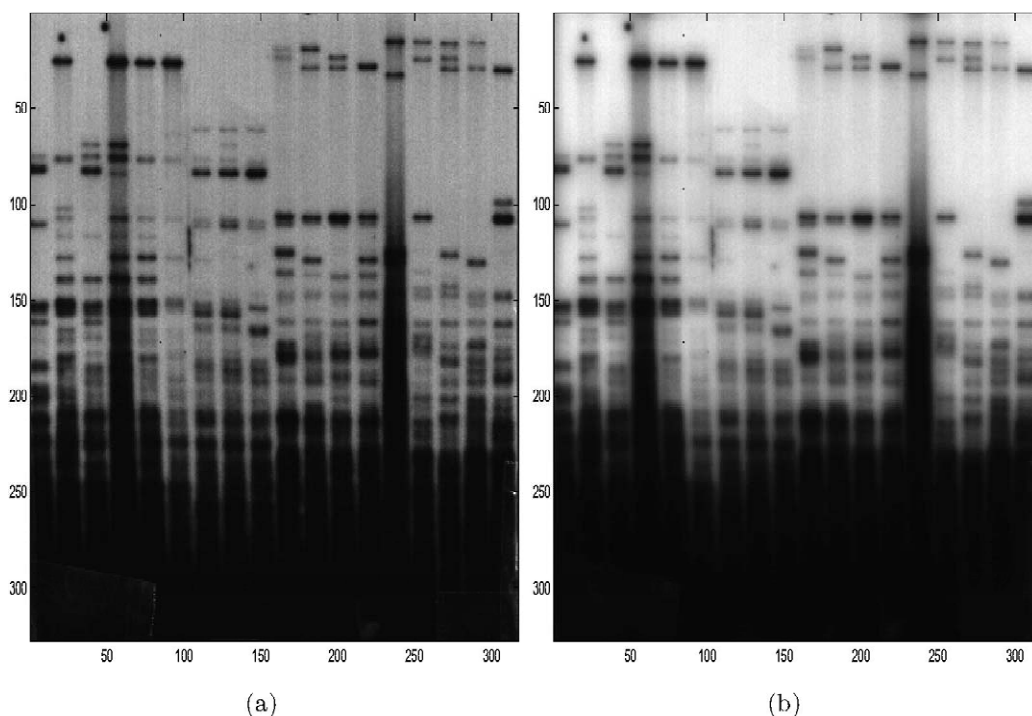


Fig. 1. (a) DNA fingerprint image from *Microtus oeconomus* hybridized to 33.6. (b) Image obtained by applying a homomorphic low pass filter on the image in (a).

1a. Through the rest of this paper, the image in Fig. 1b will be referred as the original image.

## 2.2. Mean and median filtering

A background image is estimated by using a mean or median filter to remove or defuse the DNA bands in the original image. Using mean or median filters to estimate a background image will be in the sense of defocussing the DNA bands on the original image. These methods depend on the size of the filter which again depends on the size of the DNA bands in the image. The first step in preparing an estimated background image is defocussing or removing all DNA bands from the original DNA fingerprint image. Using a mean filter with a size larger than the largest DNA band on the original image will defocus all the DNA bands in the image. The image produced after this operation reflects mainly the variations in the background and will be called the background estimated image. An image that is free from the background variations and contains only

DNA fingerprint bands from the original image, can be estimated by subtracting (absolute difference) the background estimated image from its corresponding original DNA fingerprint image. The image obtained by this process is called for the background normalized image and is negated to provide a better visualization.

The background image in Fig. 2a is estimated by applying a mean filter with the size of  $lane-width \times 50$  (with a scanning resolution of  $200 \mu m$  the lane-width is approximately 15 pixels and therefore a filter with the size of  $15 \times 50$  could be applicable) on the image in Fig. 1b. The background normalized image in Fig. 2b is obtained by absolute differencing of the image in Fig. 1b and its corresponding estimated background image in Fig. 2a, followed by negation. Fig. 3 illustrates the signal from column 22 in both the original image in Fig. 1b and the background normalized image in Fig. 2b.

Visual inspection of the column signals (intensity profiles) in Fig. 3 indicates the deformation and loss of information (indicated by arrows) on the back-

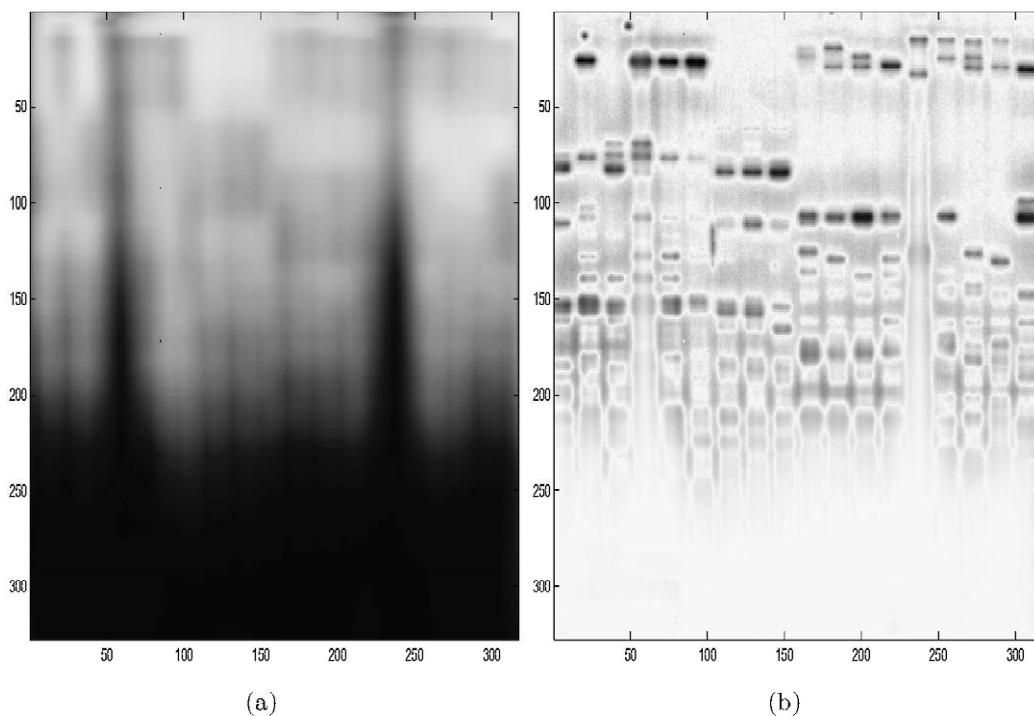


Fig. 2. (a) Estimated background image obtained by using the mean filter with a window size  $15 \times 50$  on the image in Fig. 1b. (b) Normalized background obtained by subtracting (absolute difference) the image in (a) from the image in Fig. 1b followed by negation.

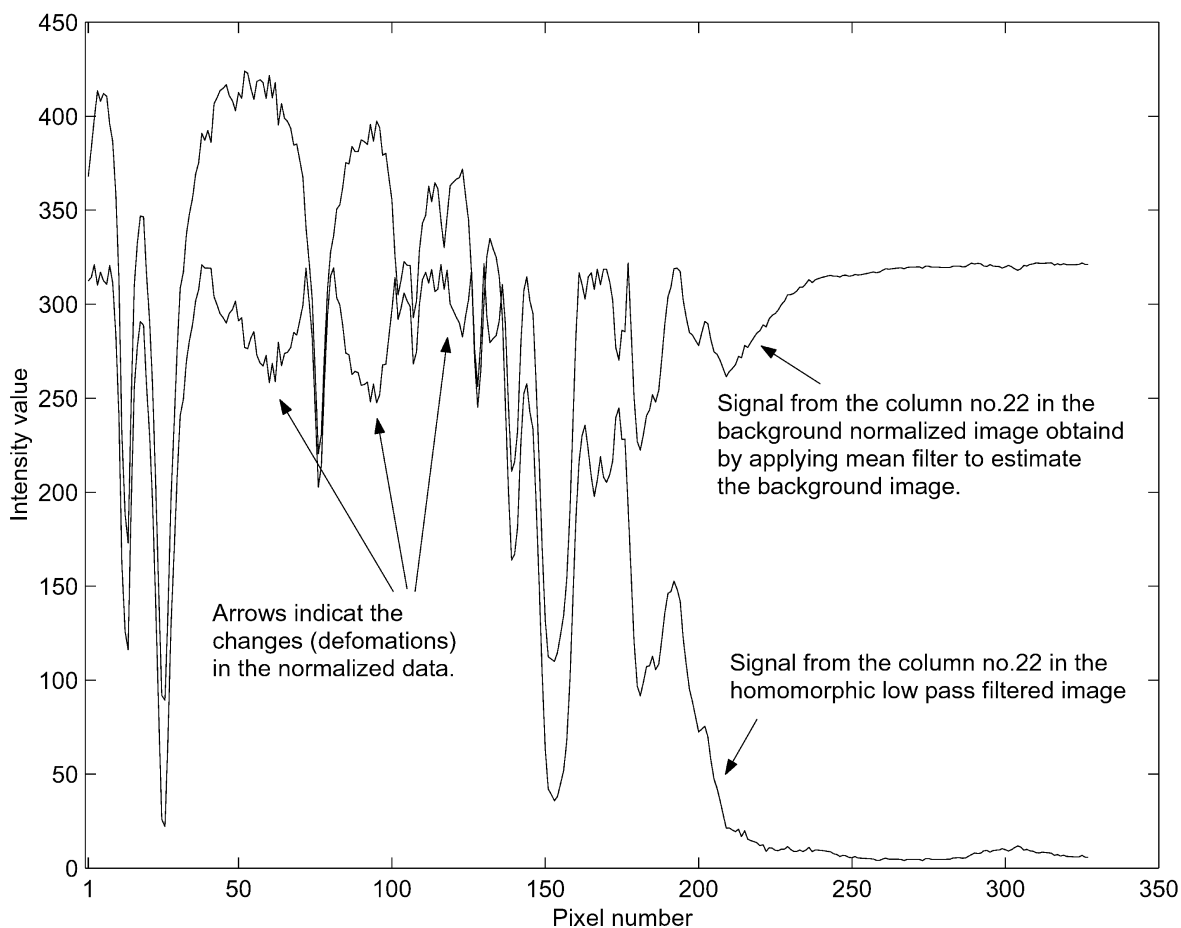


Fig. 3. The signal from column 22 in both images Figs. 1b and 2b for comparison.

ground normalized image. Estimating a background image by using a median filter can be performed in the same way as described for the mean filter. Results obtained (not shown here) by using a median filter is mostly the same as mean. The processing time for a median filter depends on the implementation of the algorithm [8,9].

### 2.3. Morphological operators

A background image can also be estimated by using morphological operations on the original image. In Refs. [2,3] it is stated that a background image can be estimated by using gray-level morphological opening which is a gray-level morphological erosion followed by a gray-level mor-

phological dilation [10–12]. A structuring element is required to perform the opening operation. Such a structuring element can be obtained from the original image by choosing one of the DNA bands in the image. The same structuring element will be used for both the erosion and dilation operations. We choose the structuring element to be one of the large DNA bands in the image. Since the images which we work on are positive images, the structuring element will be negated and the morphological closing operation (morphological dilation followed by morphological erosion) will be applied to estimate the background image. Fig. 4b indicates the estimated background image obtained by using the structuring element in Fig. 4a for morphological closing on the image in Fig. 1b. The background normalized image in Fig. 4c

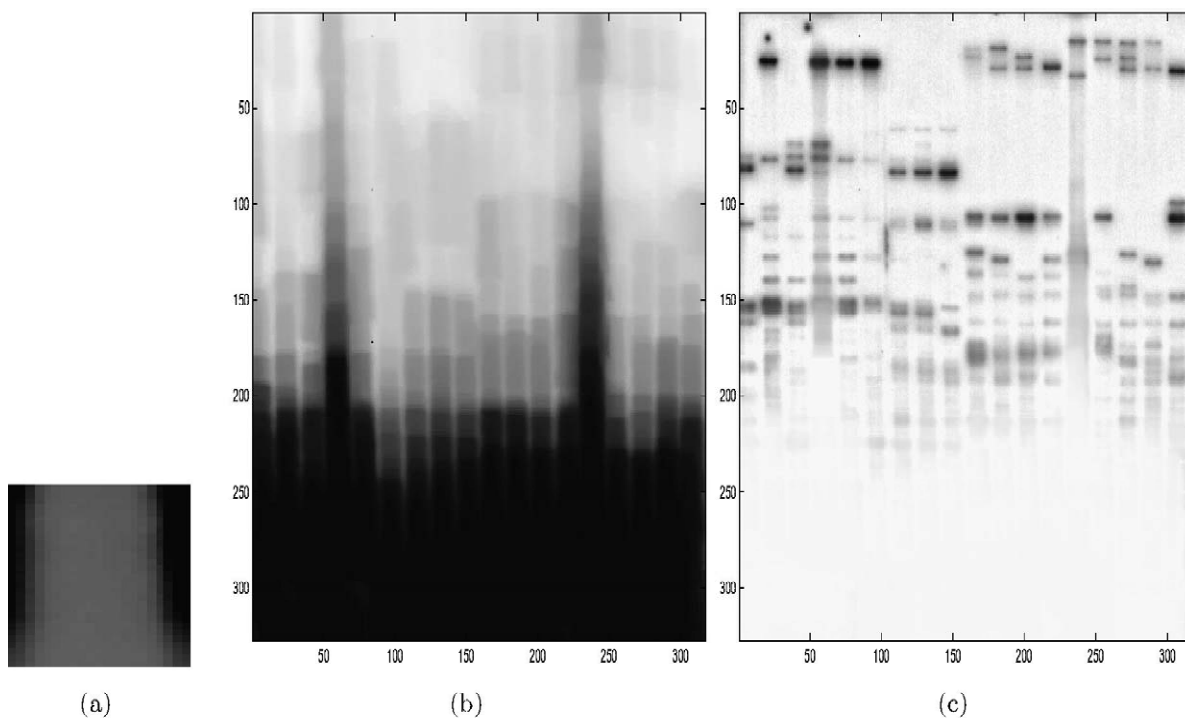


Fig. 4. (a) The structuring element chosen from the original image and negated for using in the morphological closing operation. (b) Estimated background image obtained by using the structuring element in (a) for morphological closing on the image in Fig. 1b. (c) The background normalized image obtained by subtracting (absolute difference) the image in Fig. 1b and the image in (b), followed by negation.

is obtained by absolute differencing of the image in Fig. 1b and its corresponding estimated background image in Fig. 4b, followed by negation process. Fig. 5 illustrates the signal from column 22 in both “original” image (Fig. 1b) and the background normalized image in Fig. 4b.

Visual inspection of the column signals (intensity profiles) in Fig. 5 indicates that the background trend is removed. But, there are some deformations (indicated by arrows) on the background normalized signal. The process of morphological operation is very much dependent on the selected structuring element.

#### 2.4. Maxpolygon

Maxpolygon is an adaptive algorithm to estimate the background signal for each column in the image. Each column in the image is processed separately as a one-dimensional signal. A one-dimensional sliding window will run over each column in the image. The

size of the window depends on the scanning resolution and the size of the DNA bands in the image. With a scanning resolution of 200  $\mu\text{m}$  we found by experiment that a window with the size of  $1 \times 50$  could be adequate. At each point on the column signal, the global maximum in the sliding window is selected to be the value for the corresponding point in the estimated background signal for that column. The results obtained by this process on each column in the image will be the background signal for the corresponding column in which all valleys are removed. The word “valley” is used here because the DNA bands in the images appear toward black and the background is toward bright. Therefore DNA bands appear as valleys on the column signals in the image. Fig. 6 shows the estimated column signal obtained by using this algorithm on column 22 in Fig. 1b.

An estimated background image will be obtained when all columns signals in the image are processed by this algorithm. The result obtained by using this

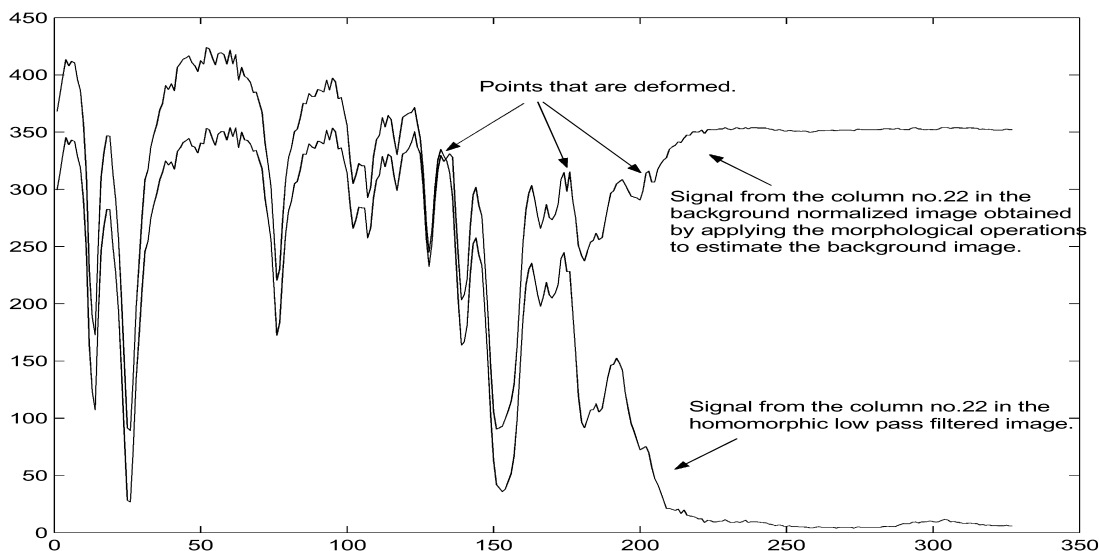


Fig. 5. The signal from column 22 in both images Figs. 1b and 4c for comparison.

algorithm on the image in Fig. 1b is illustrated in Fig. 7a.

The estimated background image which is obtained with this algorithm, can be combined with its corresponding original DNA fingerprint image to make the image in which its background is normalized (background trend is removed). This is based on

the difference between each point on the estimated background image and the global value in the original image. The following algorithm is used to make the background normalized image.

$$I_{BN}(x, y) = [\max(\text{original image}) - I_B(x, y)] + I(x, y) \quad (1)$$

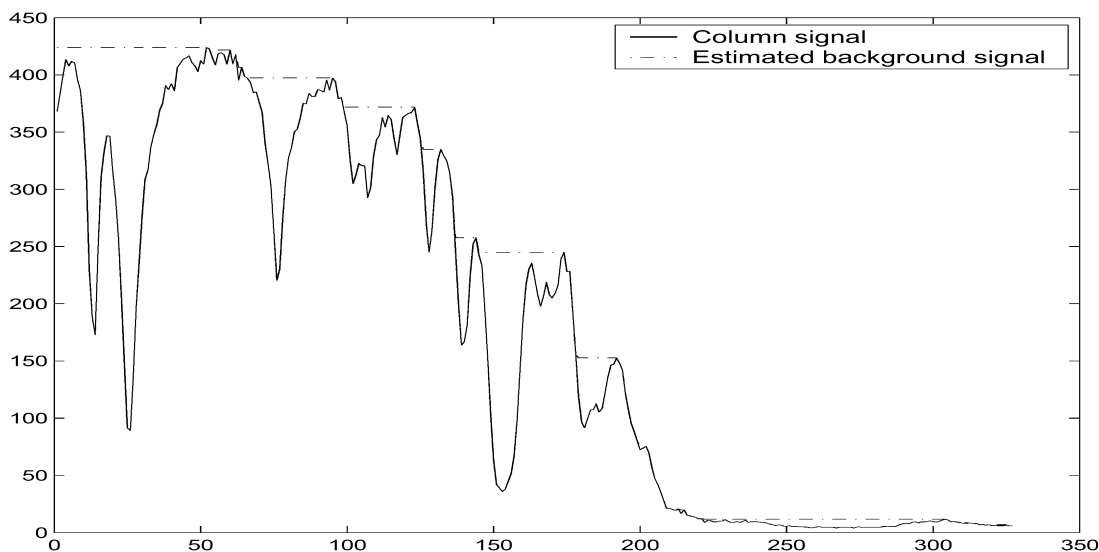


Fig. 6. The signal from column 22 in both images Figs. 1b and 7a. The signal presented by dashed line is obtained by using the algorithm described in Section 2.4 with a sliding window of the size  $1 \times 50$  on the image in Fig. 1b.

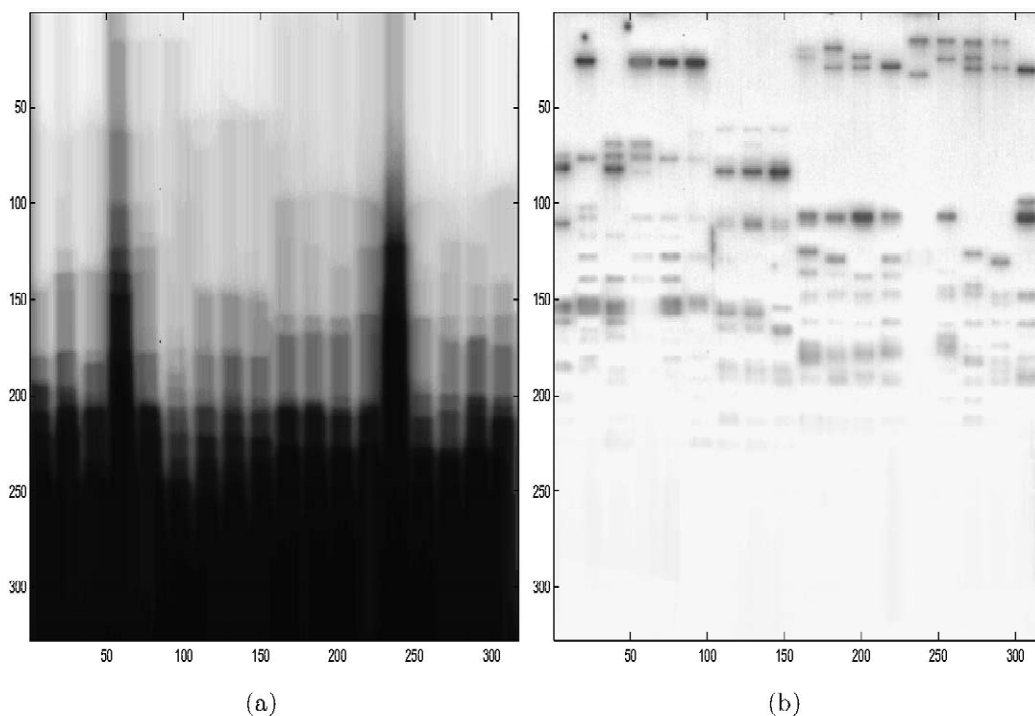


Fig. 7. (a) The background estimated image obtained by using the maxpolygon algorithm on the image in Fig. 1b. (b) The background normalized image obtained by the algorithm described in Eq. (1) in Section 2.4.

where  $I_{BN}(x, y)$  is intensity value in the background normalized image;  $I_B(x, y)$  is intensity value in the estimated background image;  $I(x, y)$  is intensity value in the original DNA fingerprint image;  $\max(\text{original image})$  is global maximum in the original DNA fingerprint image. The background normalized image obtained by using Eq. (1) on the images in Figs. 1b and 7a is illustrated in Fig. 7b.

Visual inspection of the column signals (intensity profiles) in Fig. 8 indicates that the background trend on the column signal is removed and the original shape of the signal is better preserved than in the previous methods.

### 3. Discussion

(1) Estimating a background normalized image using a mean or median filter depends on the size of the filter and the size of the objects (DNA bands) in

the image. Visual inspection of some column signals in the background normalized image obtained by these methods indicates the changes between the original and the estimated background images. Fig. 3 shows the intensity profiles for column 22 in both the original image (Fig. 1b) and its corresponding background normalized image (Fig. 2b). Some points which are highly deformed are indicated by arrows in Fig. 3.

(2) Using morphological operations to estimate a background image depends on the selected structuring element. The background trend in the estimated background image is removed. Visual inspection of column signals in the background normalized image obtained by this method indicates small changes. Fig. 5 shows the intensity profiles for column 22 in both the original image (Fig. 1b) and its corresponding background normalized image (Fig. 4c). Some points which are slightly changed are indicated by arrows. The process of morphological closing with a struc-

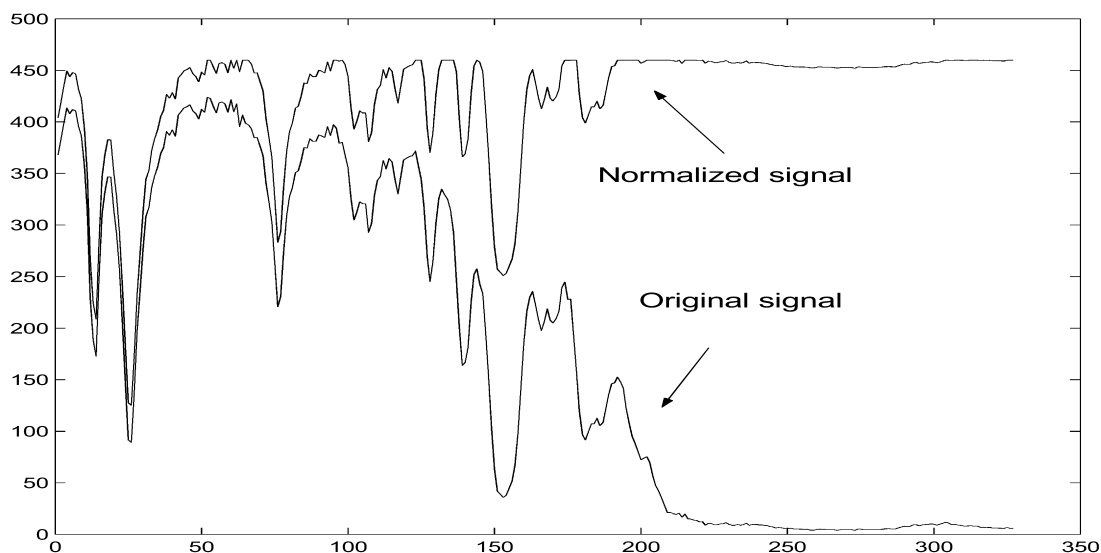


Fig. 8. The signal from column 22 in both images Figs. 1c and 7b for comparison.

turing element of the size  $m \times n$  will require more mathematical operations than a mean or median filter of the same size.

(3) The maxpolygon method is based on the

global maximum in the original image (assuming the image is positive, background is bright) and the size of the sliding window. The size of the window depends on the size of the DNA bands in the image.

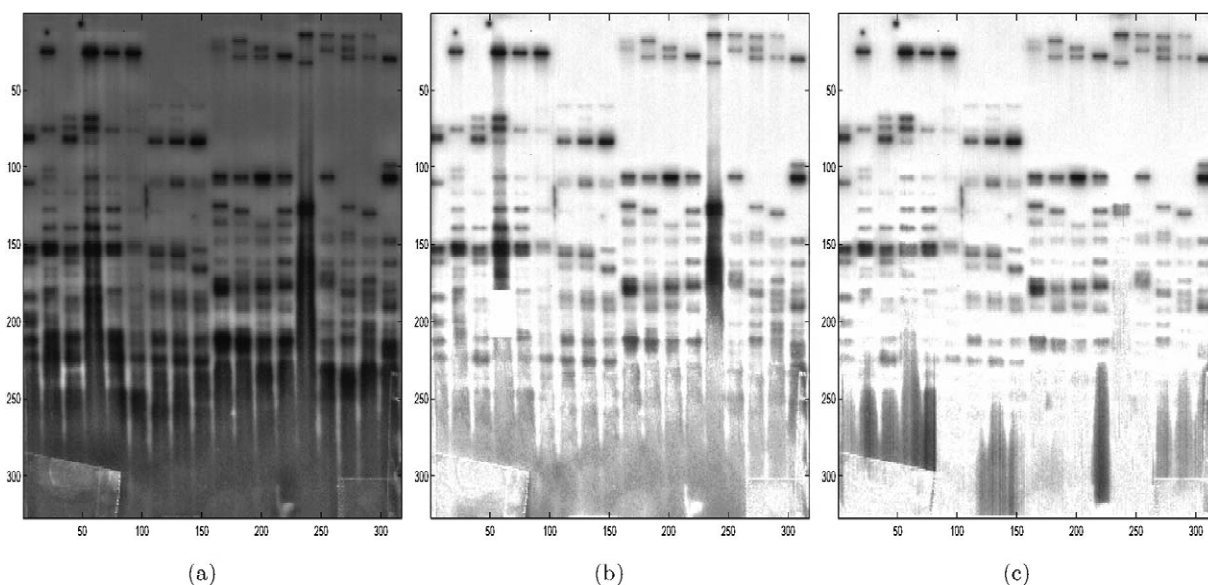


Fig. 9. (a) Background normalized estimated image obtained by using the point-to-point division of the image in Fig. 1b by the image in Fig. 2a. (b) Background normalized estimated image obtained by using the point-to-point division of the image in Fig. 1b by the image in Fig. 4b. (c) Background normalized estimated image obtained by using the point-to-point division of the image in Fig. 1b by the image in Fig. 7a.



The results obtained by this method indicates that the background trend in the image is removed while the information is preserved. Results are shown in Fig. 8.

(4) The point-to-point division of the original image by its corresponding estimated background image obtained by the methods described above, is also tested to estimate a background normalized image. The results are shown in Fig. 9 for comparison.

Visual comparison of images in Fig. 9a,b and c to their corresponding images (in Figs. 2b, 4c and 7b, respectively) obtained by using *absolute difference* indicates that the background is not properly normalized. Therefore, using *absolute difference* is preferred to *point-to-point division*.

#### 4. Conclusion

In this work, four methods are tested on different one-dimensional DNA fingerprint images to normalize their background. Methods such as mean and median filtering depend on the filter size which again depends on the size of the DNA bands in the image. In addition to this, the background normalized images obtained by these methods have some deformation (visually inspected) and loss of information. Using morphological operators to normalize the background for DNA fingerprint images, depend on the structuring element which again depends on the size of the DNA bands in the image. This method has small deformations. The maxpolygon algorithm

depends on the size of the one-dimensional sliding window which can be fixed experimentally. It also depends on the global maximum in the original image. The results obtained by the maxpolygon method illustrate that the information and the shape of the intensity profiles for the column signals in the image are better preserved than the results obtained by other methods described in this work.

#### References

- [1] N.L. Anderson, J. Taylor, A.E. Scandora, B.P. Coulter, N.G. Anderson, *Clin. Chem.* 27 (1981) 1807–1820.
- [2] S. Sternberg, *IEEE Computer* January (1983) 22–34.
- [3] M. Skolnick, S. Sternberg, J.V. Neel, *Clin. Chem.* 28 (1982) 969–978.
- [4] E.S. Gelsema, L.N. Kanal (Eds.), *Automated Analysis of Digital Images Generated by Two-dimensional Gel Electrophoresis*, Elsevier, Amsterdam, 1986, pp. 301–311.
- [5] J.E. Stacy, R.A. Ims, N.C. Stenseth, K.S. Jakobsen, *Nucleic Acids Res.* 19 (1991) 4004.
- [6] T. Stokham, *Proc. IEEE.* 7 (1972) 828–842.
- [7] R.C. Gonzalez, P. Wintz, in: *Digital Image Processing*, Addison-Wesley, Reading, MA, 1987.
- [8] P.E. Danielsson, *Computer Graphics Image Proc.* 4 (1981) 71–78.
- [9] T.S. Huang, G.J. Yang, G.Y. Tang, *IEEE Trans. Acoustics Speech Signal Proc.* 27 (1979) 13–18.
- [10] R.M. Haralick, R. Stanley, S. Sternberg, X. Zhuang, *IEEE Trans. Pattern Anal. Machine Intell.* 9 (4) (1987) 532–550.
- [11] A.C.P. Loui, A.N. Venetsanopoulos, K.C. Smith, *SPIE, Nonlinear Image Processing* 1247, 145–156.
- [12] J.P. Serra, in: *Mathematical Morphology and Image Analysis*, Academic Press, London, 1981.