Proceedings of the 6th Nordic Signal Processing Symposium - NORSIG 2004
June 9 - 11, 2004
Espoo, Finland

EVALUATION OF NOISE IN DNA FINGERPRINT IMAGES PRODUCED BY HYBRIDIZATION TECHNIQUES

Akbari A. and Albregtsen F.

Department of Informatics, University of Oslo,
P.O.Box 1080, N-0316 Oslo, Norway
e-mail: akbara@ifi.uio.no, fritz@ifi.uio.no

ABSTRACT

DNA fingerprint images produced by hybridization techniques are often degraded by different types of noise such as film grain noise and noise from electrophoresis, membrane, hybridization process, etc. The purpose of noise reduction is to obtain an image that differ as little as possible from the noise free image. The basic difficulty of the noise reduction technique is that, if applied indiscriminately, they tend to blur the image. Further, the effectiveness of each noise reduction technique tends to be affected by the type of noise to be filtered and the characteristics of the image to be processed. To perform an effective noise filtering, a noise model which describes the noise type is required. Methods such as SVD (Singular Value Decomposition) with power spectrum and local statistics (such as mean and variance) are applied to estimate the noise characteristics in hybridized DNA fingerprint images. The results obtained indicate that these type of images are affected by both additive and multiplicative noise. Homomorphic low pass filter and iterative *KNN* (K nearest neighbor) filter are applied to reduce the noise.

1. INTRODUCTION

The fundamental units of data in DNA fingerprint[1] studies are the number of bands[2] exhibited in individual lanes on a gel. So the importance of noise filtering in DNA fingerprint images is to preserve the bands and reduce the noise. Throughout this article we will refer to images generated on X-ray film by DNA fingerprints in the electrophoresis gels. We are dealing with one spectral band, i.e., images with gray values between 0 (black) and 255 (white). Both multi- and single-locus DNA fingerprint images are tested in this work.

The images for this work was partly provided by forensic medicine institute Oslo, and partly by biological department University of Oslo.

[1] The characterization of one or more features of an individual's genome by developing a DNA fragment band (allele) pattern.

[2] A band is visual image representation of a particular DNA fragment on an autoradiograph [1].

A basic difficulty with noise removing in a DNA fingerprint image, if applied without care, tends to remove the very weak bands or migrate the very close bands into one. In this paper, first we are going to evaluate the noise in DNA fingerprints produced by hybridization techniques and then applying smoothing algorithms which attempt to resolve the conflict between noise elimination and edge degradation.

2. ESTIMATION OF NOISE CHARACTERISTICS

Noise that degrades the image can be divided into two different categories, signal independent (additive) and signal dependent (multiplicative).

An image which is degraded by the signal independent noise can be expressed by :

$$f(x, y) = g(x, y) + n(x, y) \tag{1}$$

where $f(x, y)$ is the noise corrupted image, $g(x, y)$ is noise free image, and $n(x, y)$ is a sample from a random sequence of white noise at the point (x, y).

If the noise depends on the image signal, then the noise corrupted image will be subjected to multiplicative noise and can be expressed by:

$$f(x, y) = g(x, y)v(x, y) \tag{2}$$

where $v(x, y)$ is a signal dependent noise sequence.
Some images are subject to both additive and multiplicative noise and can be expressed by:

$$f(x, y) = g(x, y)v(x, y) + n(x, y) \tag{3}$$

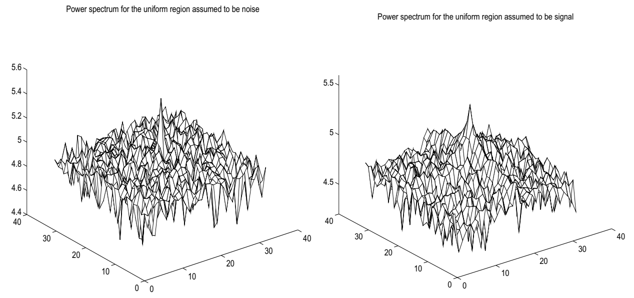
where $g(x, y)$, $v(x, y)$ and $n(x, y)$ are statistically independent.
The film grain noise arises from the statistical nature of the photographic process. In the density domain the film grain noise is additive while it is multiplicative in the intensity domain [2], [3], [4]. The most frequently used method for estimating the noise parameters, is to choose homogeneous and featureless image areas and then use different algorithms

such as SVD and power spectrum [2] or local statistics for these areas.

If different spatial positions in a film are used to measure optical densities, the statistical correlation between samples are zero when the samples are spaced farther apart than the grain size of the film. Experimental analysis indicates (see references in [3]) that these sampling conditions are satisfied and therefore film-grain noise may be considered as a random white noise process.

3. ESTIMATING NOISE CHARACTERISTICS WITH SVD AND POWER SPECTRUM.

For evaluating the noise model by the use of SVD (Singular Value Decomposition) [2], we extract a subsection of the image which is assumed to be noise (no data) and call it the noise section, and another subsection from a part of the image which is assumed to be data (a DNA band) and call it the data section. The size of the selected uniform sections are 32x32. The selected data section is from one of the largest DNA bands in the image.



(a) (b)

Fig. 1. *(a) Power spectrum for a relatively uniform section (32x32) assumed to be noise in a DNA fingerprint image. (b) Power spectrum for a data section (32x32) from the same image as in figure (a)*

First the SVD for both (noise and data) sections are computed and then the corresponding eigenvalues for SVD in both sections are computed. The SVD-eigenvalue plots (figure 2) for the assumed uniform noise and signal sections indicate a relatively wide band level. This can be confirmed by analyzing the power spectra for these sections. The power spectrum for each section is obtained from the Fourier analysis of the corresponding subsection and are illustrated in Fig. 1 (a) and Fig. 1 (b) respectively.

From SVD and power spectrum for the noise and signal sections, can be concluded that the process of noise changes as a function of the underlying object brightness [2]. This indicates that the noise may be multiplicative or at least that it is signal dependent. It may be multiplicative because the noise variance is greater at higher film density [2]. To confirm the conclusion about multiplicative noise, different uniform regions in the image should be tested. SVD is a supervised method and relatively time consuming becaue it should be tested with different homogenous windows.

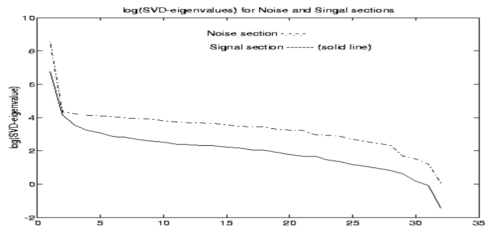


Fig. 2. *SVD-eigenvalues for the noise section indicated by "-.-.-" and for the data section indicated by "solid line".*

4. ESTIMATION OF NOISE CHARACTERISTICS BY LOCAL STATISTICS

Using homogeneous (featureless) image blocks to estimate the noise characteristics with SVD and power spectrum is a supervised method, and the image might not have homogeneous areas which are large enough, and in order to have a good estimation, the block size and the number of blocks should be large. To avoid these problems there is an unsupervised method in which the original image is divided into small image blocks (Fig. 3) to estimate the noise characteristics.

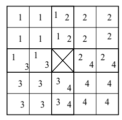


Fig. 3. *A 5x5 window is divided into 4 subwindows of size 3x3. The center pixel should be included in all subwindows.*

The noise variance of a local area can be estimated by the local variance of a flat area. Based on this idea, noise variance can be estimated adaptively. Noise variance can also be estimated by using a running window on the image and calculating the windows statistics such as mean and variance. This will not exclude the edges from the analysis because the window may not be homogeneous due to the presence of edges. To exclude the edge effects from the analysis the main window is divided into sub-windows (we have used four sub-windows Fig. 3) in which the center pixel from the main window will be included in all sub-windows. Then the mean and variance for all sub-windows are computed and the sub-window with the lowest variance is selected. Therefor the selected sub-window will have the most homogeneous area around the center pixel in the main window. In this way the effect of edges are greatly reduced. This technique is also called *maximum homogeneity* [6, 7, 8]. A scatter plot for local variance (σ^2) versus local mean squared (μ^2) for sub-windows will be used to estimate the noise characteristics. Since the outliers representing the edges are to some extent removed, a least square algorithm can be used to fit a straight line to the cluster on the scatter plot. In the case of additive noise, a large number of samples will cluster along a line of constant variance. If the noise is multiplicative, they will cluster along a sloped line. An example of the analysis used in this work, is the image i Fig. 4 which is a small part of a single locus DNA fingerprint image. We first compute the mean (μ) and variance (σ^2) for the main window to show the edge effects in the scatter plot (Fig. 5(a)) and then selecting the sub-window with the minimum variance in the main window to reduce the edge effects (Fig. 5(b)). Based on equation 3, the interception of the regression line in Fig. 5(b)(with reduced edge effects) with vertical axis (variance) gives an indication to the presence of both additive and multiplicative noise in the image. Negative slope of the regression line is due to the presence of few objects with low intensity and large bright area in the image.

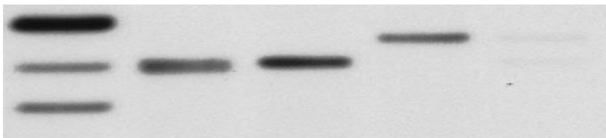
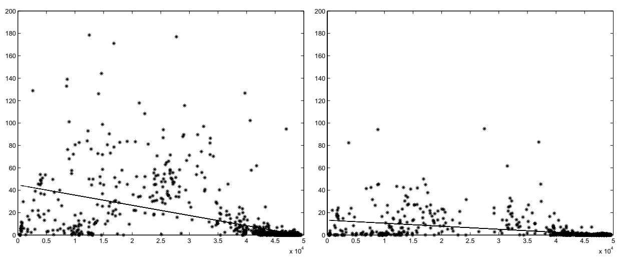


Fig. 4. *DNA fingerprint image used for the analysis of noise characteristics.*



(a) (b)

Fig. 5. *(a) Local variances versus local mean squared in blocks of size 5x5 on the image in Fig. 4. (b) Local variances versus local mean squared in the sub-windows (3x3) with minimum variance in each block of size 5x5 on the image in Fig. 4. For a better visualization of the points in the scatter plot we have divided the image into non-overlapping blocks of size 5x5 instead of using the running window on each pixel.*

5. HOMOMORPHIC FILTER

The basic nature of the image $F(x, y)$ may be characterized by two components, the amount of source light incident on the scene being viewed and the amount of light reflected by the objects in the scene. These are called the illumination ($i(x, y)$) and reflectance ($r(x, y)$) components. An image can be expressed as the product of illumination and reflectance components[9].

$F(x, y) = i(x, y)r(x, y)$

In this part the image is first transformed into the frequency domain where a linear convolutional relationship between the original scene and the degraded image can be established. The linear filter which is applied will suppress the noise and enhance the image. At last the image is transformed back into the space domain. This process is known as homomorphic filtering (Fig. 6). The control of this system over the illumination and reflectance requires a filter function that affects the low and high frequency components operating on the image in the frequency domain. Since the ideal filters suffer from blurring and ringing, a smoother transition in the frequency domain filter such as Butterworth filter can give a better results applied on the DNA fingerprint images. That is due to the migration of very closed bands into one band.

$$H(u, v) = \frac{1}{1 + [\frac{D(u,v)}{D_0}]^{2n}} \tag{4}$$

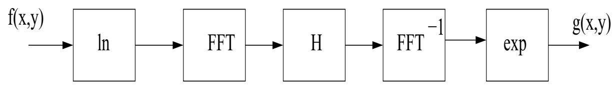


Fig. 6. *The process of homomorphic filtering*

We applied the homomorphic filtering with Butterworth low pass filter with cutoff frequency ($D_0 = 40$ in equation 4 and dynamic range ($\lambda_L = 0.99$ and $\lambda_H = 10$ referring to Fig.4.32 on page 191 in [9]) on the image in Fig. 4. The resulting image is illustrated in Fig. 7.

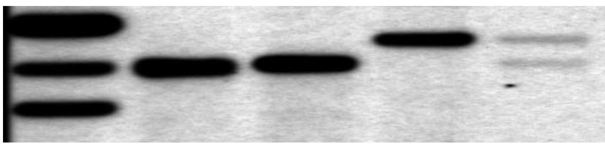


Fig. 7. *The result of applying homomorphic filtering described in text on the image in Fig. 4*

As it can be observed from Fig. 7, the two bands on the last lane which could be hardly detected in the original image in Fig. 4 can be easily detected after applying the homomorphic filter on the original image. The values for dynamic range and cutoff frequency may vary from image to image. We can conclude that the homomorphic filtering process could be useful to strengthen the very weak DNA bands.

6. K-NEAREST NEIGHBOR FILTER

Based on our experiments and an evaluation of some non-linear adaptive noise smoothing filters for both additive and multiplicative noise by Mastin [10], an iterative *KNN* filter was applied to remove the noise from DNA fingerprint images while preserving edges. By iterating, window sizes incrases with iteration, while k as a percentage of neighborhood pixels decreases. This will substantially smooth the noise while preserving edges. Increasing the window size while reducing the effect of isolated or small clusters of noise points by reducing the relative value of k on each iteration is due to moving up the image pyramid [11]. The result of applying the *KNN* filter with two iteration ($3x3$, $k = 6$) followed by ($5x5$, $k = 12$) on the image in Fig. 8(a) are illustrated in Fig. 8(b) and Fig. 8(c) respectively. The intensity profile in Fig. 8(c) indicates that the noise are significantly removed while the edges are preserved. As seen in Fig. 8(d), a conventional low pass filter would remove some of the weak/close - but significant bands (see arrows).

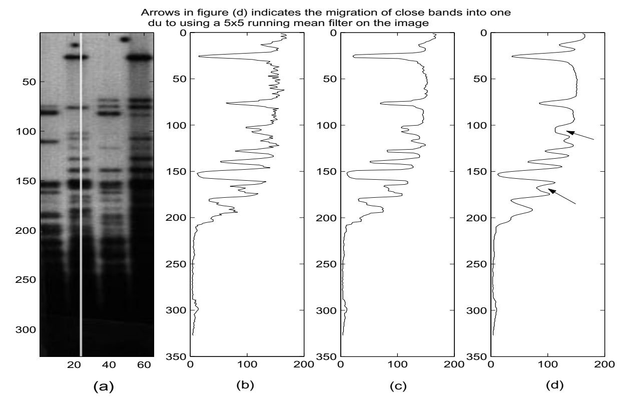


Fig. 8. *(a) The line along which the intensity profile has been drawn, is shown on the noise corrupted image. (b) Profile for a column in the noise corrupted image. (c) Profile in the filtered image after two iteration (w1=3x3 with k1=6, w2=5x5 with k2=12.). (d) Profile in the filtered image using a running 5x5 mean filter.*

7. CONCLUSIONS

The results of noise estimation in sections 3 and 4 indicate that the DNA fingerprint images produced by hybridization techniques are affected by both additive and multiplicative noise. Application of homomorphic low pass filter with a suitable dynamic range and cutoff frequency on the DNA fingerprint images which contain very weak DNA bands can strengthen the weak bands. Iterative KNN filter can be applied to reduce the noise and preserve the edges in DNA fingerprint images. This could be useful where the very close DNA bands should be distinguished and not migrated into one band.

8. REFERENCES

[1] Committee on DNA Forensic Science: An Update, National Research Council, *The Evaluation of Forensic DNA Evidence*National Academy Press, 2nd ed. 1998.

[2] H. C. Andrews, B.R Hunt *"Digital Image Restoration."* PRENTICE-HALL, INC. 1977.

[3] W.K. Pratt *"Digital image processing".* third ed. 2001, pp. 342-343.

[4] D.G. Falconer *"Image Enhancement and Film Grain Noise."* Optica. acta Vol.17 693-705, 1970.

[5] Jong-Sen Lee *"Speckle Analysis and Smoothing of Synthetic Aperture Radar Images."* Computer Graphics and Image Processing Vol.17, 1981, 24-32.

[6] F. Tomita and S. Tsuji, *"Extraction of multiple regions by smoothing in selected neighborhoods,"* IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-7, pp. 107-109, Feb. 1977.

[7] M. Nagao and T. Matsuyama, *"Edge preserving smoothing,"* Computer Graphics and Image Processing, vol. 9, pp. 394-407, Apr. 1979.

[8] W. Higgins, *"A flexible implementation of maximum-homogeneity filtering for 2-D and 3-D images,"* IEEE Trans. Signal Processing, vol. 39, pp. 2325-2331, Oct. 1991.

[9] Rafael C. Gonzalez and Richard E. Woods *"Digital Image Processing"* 2nd ed. 2002 Prentice Hall.

[10] Gary A. Mastin *"Adaptive Filters for Digital Image Noise Smoothing: An Evaluation"* Computer Vision, Graphics and Image Processing Vol.31, 1985, 103-121.

[11] T. Pavlidis. *"Structural Pattern Recognition."* Springer-Verlag, Berlin, 1977. page 302.