Pakistan Journal of Applied Sciences 1 (3): 256-259, 2001
©Copyright by the Science Publications, 2001

DNA Band Detection Utilizing Neural Networks and Scale Space Analysis

Adnan Khashman
Department of Computer Engineering, Near East University, Lefkosa, Mersin 10, Turkey

Abstract: Low-contrast images, such as DNA autoradiograph images, provide a challenge for edge detection techniques, where the detection of the DNA bands within the images and locating their position is vital. In addition, the speed of recognition, high computational cost, and real-time implementation are also problems that haunt image processing. Thus, new measures are required to solve these problems. This paper reports on a new approach to solving the aforementioned problems. The novel idea is based on combining neural network arbitration and scale space analysis to automatically select one optimum scale for the entire image at which scale space edge detection can be applied. This approach to edge detection is formalized in the Automatic Edge Detection Scheme (AEDS). The AEDS is implemented on a real-life application namely, the detection of bands within low-contrast DNA autoradiograph images. An accurate comparison is drawn between the AEDS and a recently developed edge detection technique namely, the Grammar-Based Multiscale Analysis Technique (GBMAT).

Keywords: AEDS, Neural Networks, Scale Space Analysis, Edge Detection, DNA Bands

Introduction

The work that is presented within this paper is aimed at investigating and demonstrating the suitability of applying the AEDS to complicated images, in terms of the objects contained within them. The image, of a DNA autoradiograph, is particularly difficult to analyze due to its low-contrast nature. The DNA bands within the autoradiographs can be very difficult to identify. This application shows the strength of the AEDS in detecting objects within low-contrast images in comparison to another technique that has been recently developed.

The detection of DNA autoradiograph bands has been successfully addressed by Abdul Aziz (1993). In his approach, Abdul Aziz detected the edges of the DNA autoradiograph bands using a grammar-based multiscale analysis technique (GBMAT). The GBMAT, which was developed in 1993, was shown to be far superior to many conventional edge-based region extraction techniques. Comparisons with conventional edge detection techniques such as the Prewitt operator or Sobel operator, clearly indicated the improvement in detecting low-contrast objects. Edge detection using AEDS provides far better results than when using GBMAT and consequently than when using other techniques, such as the conventional edge-based Sobel or Prewitt operators. However, the main advantage over previous techniques is the versatility and capability of AEDS in providing good edge detection and quick results.

The GBMAT for edge detection will be briefly described, and then applied to images of DNA. The results of the application of this technique will, then, be compared to the results of applying the AEDS.

DNA Autoradiograph Images: DNA autoradiographs are X-ray films showing rectangular bands which represent the A, C, G and T bases within the DNA. These bands vary in their separations, widths and darkness. In the autoradiograph images, the background varies in intensity, i.e. the level of noise. This variation combined with the variety in the bands themselves, leads to it being extremely difficult to analyze the image Khashman (1999). The need for detecting and locating these bands is vital for the determination of the DNA sequence of a particular sample under analysis.

There are 14 DNA autoradiograph images available for the full implementation of the AEDS. These are labeled as: DNA 1, DNA 2, ..., DNA 14. The last two images, namely DNA 13 and DNA 14 will be used for the purpose of comparing AEDS and GBMAT band detection.

Fig. 1 shows the fourteen images in their original grey format. Some of the images shown in this figure appear to be very low contrast and hence the DNA bands can not be clearly seen. The aim of implementing the AEDS is to identify the edges of the bands from the background noise. This kind of problem, where the images are complex and of very low contrast, is a typical example where the AEDS can be efficiently utilized.

Band Detection Using AEDS: The AEDS is based on combining the three fields of scale space analysis, edge detection and neural networks. The result is a system that delivers very quick edge detection of objects within low-contrast images, through the automatic selection of a single optimum scale for applying the scale space edge detection to an entire image.

There are two phases in the implementation of AEDS. The first is implementing scale space analysis of the DNA autoradiograph images, thus providing the ideal detection, represented through the ideal scales, σ_{Ideal}, for the various DNA images. This involves the use of a fast format of the Laplacian of the Gaussian (FLoG) edge detection operator Chen *et al.* (1987); Marr and Hildreth (1980), as shown below in equation (1). The FLoG operator is convoluted with a number of images that represent the training set of images for the neural network, at seven scales in scale space. The standard deviation (σ) of the Gaussian function in the FLoG operator is variable and it dictates the amount of smoothing to be imposed on the image prior to edge detection Sotak and Boyer (1989). The high computations involved in multiscale processing, can be marginally reduced if a suitable scale is found, and then used Lindeberg (1993) and (1996). A criterion for selecting the 'ideal' edge detection at one ideal scale (σ_{Ideal}) is set up, based on the convolution results using 3-dimensional objects Khashman and Curtis (1995).

Results of this phase represent the training data for the neural network arbitration in the second phase.

$$\nabla^2 G(x,y) = h(x)g(y) + g(x)h(y) \tag{1}$$

whereby

$$h(\xi) = \sqrt{A}\left(1 - \frac{\xi^2}{\sigma^2}\right)e^{-\frac{\xi^2}{2\sigma^2}}$$

and

$$g(\xi) = \sqrt{A}e^{-\frac{\xi^2}{2\sigma^2}}$$

The second phase in the implementation of the AEDS is training a neural network to recognize σ_{Ideal} for an input image. This is based on the hypothesis that σ_{Ideal} is a function of noise Khashman and Curtis (1997) and Khashman 2000, as in equation (2).

$$\sigma \quad \infty \quad \frac{1}{N} \tag{2}$$

where, σ is the ideal scale (σ_{Ideal}) and N is the amount of noise present within the image. The trend of this non-linear relationship is what the neural network will be trained to recognize. For the purpose of training the neural network, seven out of the fourteen images available are used. These images are: DNA 1, DNA 2, ..., DNA 7. The remaining seven images, DNA 8, DNA 9, ..., DNA 14, will be used to test the generalization properties of the neural network.

The multilayer perceptron neural network, which has been developed for the AEDS, is based on the back propagation learning algorithm, with the total number of three layers, comprising, input layer, hidden layer and output layer. The Acquisition of the training data and presenting it to the neural network is very important, and care should be taken when selecting the training data. Manipulating the large amount of data available, when dealing with images, can be very computationally expensive and hence can take a long time. However, the use of a Sun-Sparc 10 running the UNIX operating system, together with C-language source code, provided a quick and powerful tool to optimize training time.

The neural network converged and learnt in 4 hours, whereas the running time for the generalized neural network was 0.63 seconds. Table 1 lists the final parameters of the successfully trained neural network. Table 2 shows the total running time for the AEDS at the various scales. The robustness, flexibility and speed of the AEDS has been demonstrated through this application. The recalling of the seven training images was 100% successful, where all training images were allocated their correct ideal scales. The generalization of the neural network, has also earned a similar success rate of 100%. The neural network recognized the correct ideal scales 12 for the remaining seven images, which it had not been trained on before. Two of the generalization images, DNA 13 and DNA 14 can be seen in Fig. 2a.

Band Detection Using GBMAT: This implementation is based on band extraction by grammar specification. There are two criteria on which this implementation is based. The first is the maximum scale space lifetime criteria. This is based on the assumption that DNA bands have a maximum scale space life time. In other words, the edges of the bands undergo all scale space events without annihilation, whereas weaker background noise edges do not survive the many events occurring on them at higher scales and thus disappear. The second criteria is based on the assumption that unwanted noise in the background have tree automata with spans equal to or greater than unity. The edges within the images are detected using the LoG operator over 6 space scales.

Edge curve classification: All edge curves within the image, including bands and background, are classified and labelled according to the criteria applied. This is either the maximum scale space life time or the unity tree span.

Scale space events string representation: These are descriptions of all scale space events occurring on each labelled edge curve.

Construction of a tree automaton to extract the DNA autoradiograph bands according to the applied criteria.

Parsing: This is based on parsing all scale space events strings describing every single edge curve within the image using both tree automata. This will result in two types of detection: the first is based on the first criteria of maximum scale space lifetime and the second is based on the tree unity span.

Combine both criteria, by parsing the scale space events strings using both tree automata according to the two criteria, and then ANDing the results. This will result in a third and final edge detection of the bands.

Fig. 2b shows the edge detection of DNA bands within DNA 13 and DNA 14, as obtained using GBMAT.

Comparison: It can be noticed from Fig. 2b that the detection of all the bands has not been achieved using GBMAT. Many weaker DNA autoradiograph bands have not been detected due to the application of the tree unity span criteria. On the other hand, stronger background noise and unwanted features have still survived both criteria. The criteria used for the detection is too general to produce exclusive detection of the DNA bands. In addition, the final detection, which is performed by combining (i.e. ANDing) both criteria, is too extreme in the sense that it eliminates the sought after bands as well as some of the unwanted background noise. Moreover, and most importantly, the time consumed in detecting these bands is quite high and the implementation procedure is very computationally expensive.

The automatic edge detection scheme AEDS, provides the solutions to the above problems and furthermore provides better edge detection for the low contrast objects. Visual inspection of the band detection using AEDS in Fig. 2 in comparison to the band detection using GBMAT, shows that the first has provided much better results, where the problematic bands can be identified clearly. Weaker DNA bands are also detected in the AEDS results and the background noise is kept to a minimum

Edge detection using AEDS provides far better results than when using GBMAT and consequently than when using any other techniques, such as the conventional

Table 1: Final neural network parameters for detecting DNA bands

Hidden Nodes (H)	Learning Rate (η)	Momentum Rate (α)	Initial Weights Range (R)	Error level (ε)	Iteration (I)	Train Time (tt)	Run Time (rt)
70	0.0075	0.21	[-0.273-+0.273]	0.009	4205	4 hrs	0.63 s

Table 2: AEDS total running time at the various scales

Ideal Scale (σ_{Ideal})	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6
Scale Recognition Time (seconds)	0.63	0.63	0.63	0.63	0.63	0.63	0.63
Edge Detection Time (seconds)	1.00	2.66	6.03	12.91	26.29	53.92	107.58
Total AEDS Time (seconds)	1.63	3.29	6.66	13.54	26.92	54.55	108.21

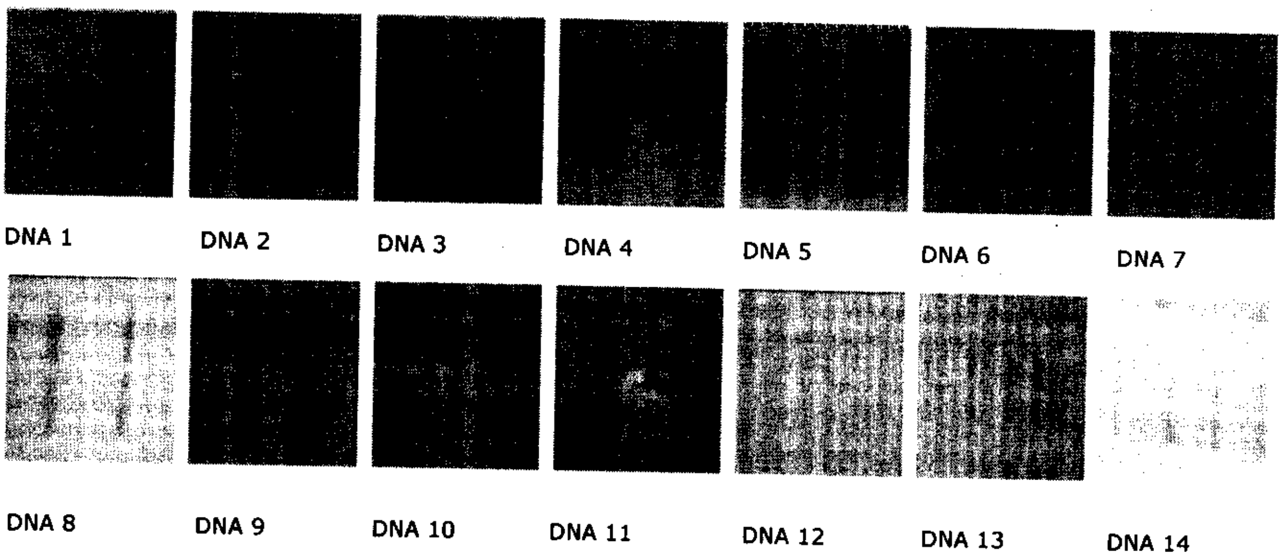

Fig. 1: Images of DNA autoradiograph (1-14)

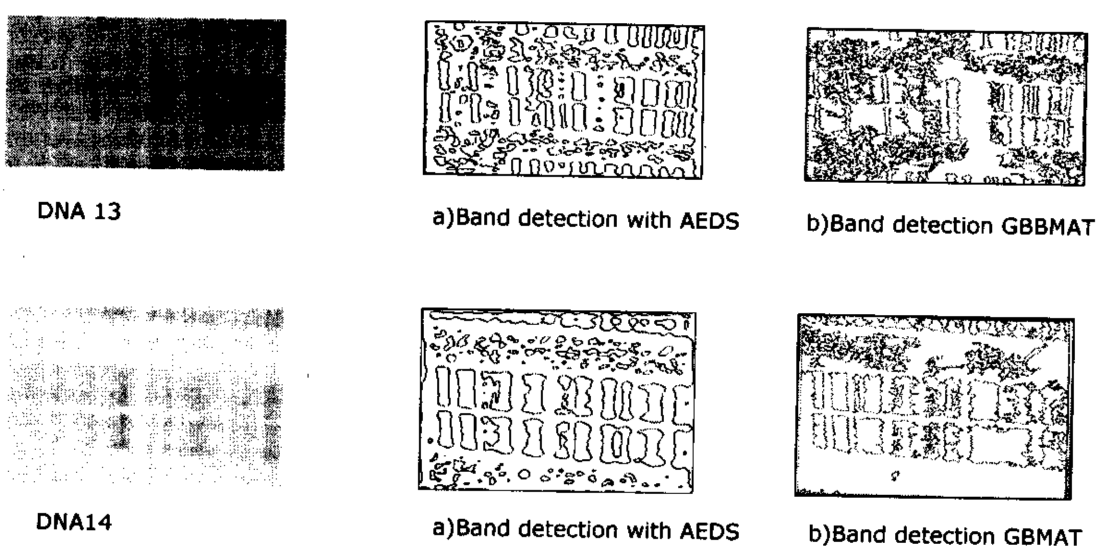

Fig. 2: DNA 13 and DNA 14 Band edge detection: a) using AEDS b) using GBMAT

edge-based Sobel or Prewitt operators. However, the main advantage over previous techniques is the versatility and capability of AEDS in providing good edge detection and quick results. In addition, the user friendly system of AEDS is not very computationally expensive and can be adapted to a wide range of applications.

Conclusions

The automatic edge detection scheme AEDS has been successfully applied to the difficult problem of band detection in DNA autoradiograph images. A remarkable implementation speed of 0.63 seconds is obtained when the neural network is used to generalize, as part of the complete automatic edge detection scheme (AEDS). The total edge detection time including the automatic scale recognition time is in the range of (1.63 - 108.21) seconds. The development of a novel technique in the discipline of image processing has been completed.
The utilization of edge behavior in scale space and the implementation of neural networks to make the decision upon what the ideal scale should be has been researched and successfully implemented. Thus, providing a powerful technique that appears to be superior to any other edge detection technique for similar applications. AEDS is an efficient technique that performs rapid automatic edge detection and bypasses the tedious costly process of going through all the scales in a more conventional scale space analysis technique. The AEDS has also been compared to and shown to be far superior to the previously developed region extraction technique, GBMAT. The advantages and disadvantages of the GBMAT have been described and the ability of the AEDS to overcome the disadvantages, while improving upon the advantages, has also been demonstrated.

References

Abdul Aziz A.Z. 1993. "Region Extraction Using a Grammar-Based Multiscale Analysis Technique", PhD Thesis, Univ. Nottingham, UK.

Chen J.S. , A. Huertas and G. Medioni, 1987. "Fast Convolution with Laplacian-of-Gaussian Masks", IEEE Pat. Analys. and Mach. Intell.,. 9: 584-590.

Khashman A. 1999. "Automatic Edge Detection of DNA Bands in Autoradiograph Images", Proceedings of the IEEE International Symposium on Industrial Electronics , Bled, Slovenia.

Khashman A. and K. M. Curtis 1995. "Scale Space Analysis Applied to 3-Dimensional Object Recognition", IEEE Int. Conf. on Digital Signal Processing , Cyprus, pp. 584-589.

Khashman A. and K. M. Curtis 1997. "A Novel Image Recognition Technique For 3-Dimensional Objects", IEEE Int. Conf. on Digital Signal Processing , Greece.

Khashman A. 2000. "Neural system for monitoring Non-Linear Foetal measurements", Proceedings of the IEEE Nordic Signal Processing Symposium , Sweden.

Lindeberg T. 1993. "Detecting Salient Blob-Like Image Stuctures and Their Scales with a Scale-Space Primal Sketch: A Method for Focus-of-Attention", Int. J. Computer Vision, 11: 283-318.

Lindeberg T. 1996. "Edge Detection and Ridge Detection within Automatic Scale Selection", IEEE Conf. Computer Vision and Pattern Recognition, pp. 465-470.

Marr D. and E. C. Hildreth , 1980. "Theory of Edge Detection", Royal Soc. London, pp. 187-217.

Sotak G.E. and K. L. Boyer. 1989. "The Laplacian-of-Gaussian Kernel: A Formal Analysis and Design Procedure for Fast, Accurate Convolution and Full-Frame Output", Comp. Vision, Graphics and Image Processing, 48:. 147-189.