COMPUTERS AND BIOMEDICAL RESEARCH 12, 279–289 (1979)

Preprocessing of Electron Micrographs of Nucleic Acid Molecules for Automatic Analysis by Computer

L. LIPKIN,* P. LEMKIN,* B. SHAPIRO,* AND J. SKLANSKY†

*Image Processing Unit, Division of Cancer Biology and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20014; and †School of Engineering, University of California, Irvine, Irvine, California 92717

Received June 21, 1978

A technique is proposed for computer preprocessing of digitized electron micrographs of adenovirus type 2 RNA strands to facilitate their automatic segmentation and subsequent analysis. This technique, the notch filter, is able to remove image shading errors due partly to sample preparation, electron microscope optics, photographic processing, and electronic acquisition of the digital image.

1. INTRODUCTION

This paper presents a methodology for computer preprocessing of electron micrographic (EM) images of nucleic acid molecules. In particular the technique is illustrated by use of spread images of adenovirus Type 2 molecules. The preprocessing methodology was developed as part of a research program for finding methods for automatic segmentation, and morphologic analysis of such images, now underway in the Image Processing Unit of the National Cancer Institute.

It is only comparatively recently (*1*, *2*) that nucleic acid biochemists have begun to understand the morphologic significance of such electron micrographs. The observations of various features, e.g., "hair pins," "loops," "trees," etc., which are regularly recurring in relatively constant position, are indicative of, for example, the occurrence of reversed complementary sequences of bases of various lengths. A sequence and its corresponding reversed complement occurring close together (such as 3 to 20 intervening bases) would appear, under appropriate preparative conditions, as a hair pin while a long interval sequence would be more likely to result in a loop.

This kind of morphologic data becomes more and more significant when available on large numbers of molecules. Ultimately, as precise base sequencing of specific nucleotides becomes available for the same molecules on which we have precise

0010-4809/79/030279-11$02.00/0

measures (*3*, *4*), we anticipate the construction of a good model of biochemical–morphologic correlation.

As is the case with almost every natural image of biological significance, shading constitutes a partial obstruction to direct image analysis. By shading error, we mean an effective nonuniformity of the background density of the final image. One source of error is the sample itself. When the sample is prepared, it is almost impossible to get a completely uniform substrate layer. This in turn creates a variable density of background in the photographic negative. Local and global heterogeneities in the illuminating electron beam (such as off center alignment, etc.) will also result in significant local variance in background brightness and thus contribute to the shading problem. The camera–film processing sequences may also contribute to shading effects, although film grain, etc., is apt to be of higher spatial frequency than those effects noted above. Finally, the electronic image acquirer and digitizer will also introduce shading effects since the detector, in this case a plumbicon, has an intrinsically irreducible nonuniformity in light sensitivity across its face.

An overall discussion of the shading problem and a possible solution in a context involving instruments such as a plumbicon is given in (*5*). We present a less complex, computationally less expensive method, which is effective in the context of the RNA images. We have chosen to treat the shading error in such a system as if it were produced by a single source rather than trying to correct for each source of error separately.

As noted above the computer analysis of such EM micrograph images is hindered by the large scale shading. Any system which attempts to extract the RNA strands for further analysis must deal with this problem one way or another. Although photographic laboratory techniques such as "dodging" and/or "burning in" can correct for some of the shading errors, it is difficult for a global process to be applied to locally variant effects so as to correct all of the shading errors. Other photographic enhancement techniques such as using high-contrast copy film do not help since such nonlinear point transformations are applied homogeneously over the image. Furthermore, photographic enhancement would be difficult to implement as a reliable procedure in a production setting. This is especially true in those cases where the variation in substrate thickness causes multiple dark bands of varying width and orientation which differ in each sample specimen.

That the images consist of a relatively small number of very dark strands on a relatively light background is seen in Fig. 1. This fact is used in designing a digital filter to shift individual picture elements belonging to the background to a uniform level of background. Similarly, pixels belong to the RNA strands (and dark artifacts and RNA fragments) are shifted to a uniform darker level.

Ito and Sato, in analyzing DNA electron micrographs (*6*), used manual threshold slicing of line fragment template matching filters (suggested by (*7*)) to isolate the strands. Depending on the amount of shading present, this technique may work quite well or may have difficulty.

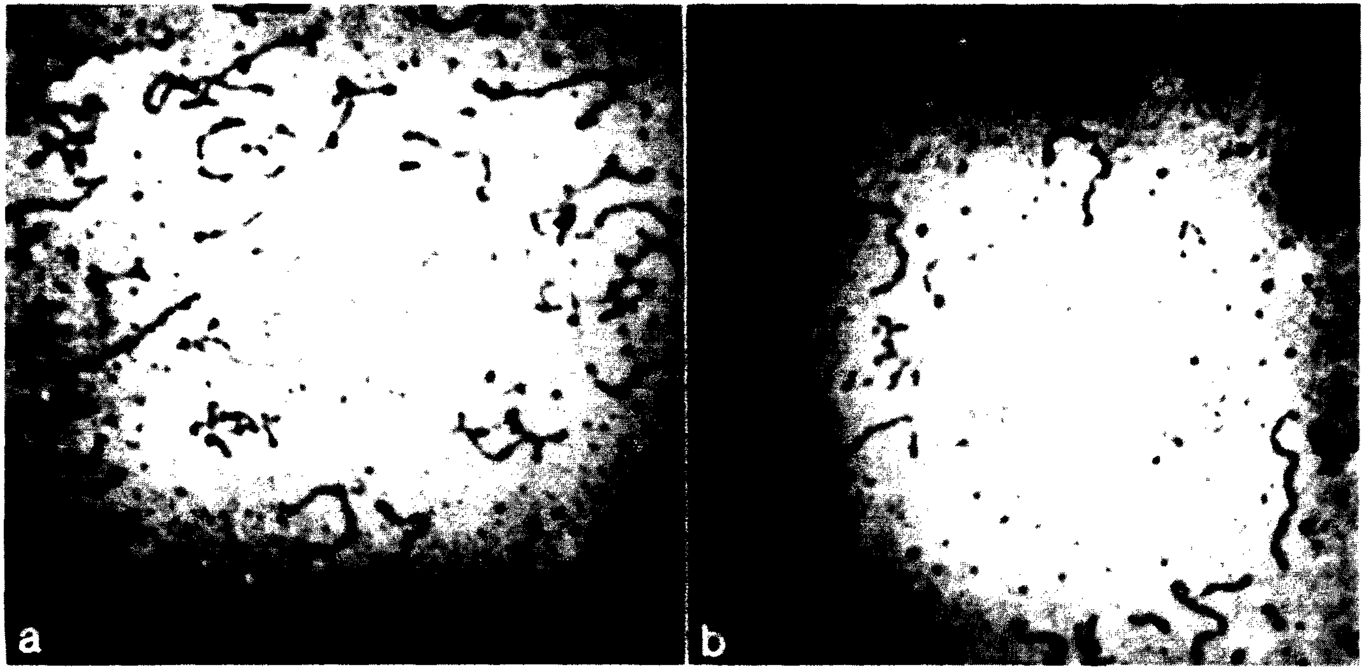

FIG. 1. Adenovirus type 2 (a) ribosomal RNA image from EM micrograph negative No. 006323, (b) m-RNA image from EM micrograph negative No. 006361.

2. Materials and Methods

2.1. Specimen preparation

Several EM micrographs of ribosomal and messenger RNA from adenovirus type 2 were obtained from J. Maizel, J. Meyer, M. Sullivan, and H. Westphal (Laboratory of Molecular Genetics, NICHLMG, N.I.H.).

The EM micrograph negatives were obtained using a Philips 300 electron microscope at 10 000×, with Kodak 4489 film, developed with D76 for 3 min. Adenovirus type 2 *m*-RNA was prepared as previously described (*8*). Ribosomal RNA from rabbit reticulocytes was prepared as previously described for HELA ribosomal RNA (*9*). The RNA samples were prepared with 70% formamide solution, 0.01-*M* TRIS (hydroxymethlaminomethane), and 0.001-*M* EDTA onto distilled water. Two such EM micrographs are shown in Fig. 1a and b. The ribosomal RNA image in Fig. 1a has a lower overall shading error than the *m*-RNA image in Fig. 1b. In this instance, much of the shading error is attributable to the digitizing process.

The negatives were scanned using the Image Processing Unit's BMON2 system (*10*, *11*) running on the real-time picture processor (RTPP) (*12*, *13*, *14*). A plumbicon TV scanner is used with a 720-line resolution coupled to a Quantimet film scanner. This film scanner has multiple sources of shading error due to the illumination system and optics. The digitized images contain 256 × 256 picture elements with 256 possible gray levels (although far fewer are seen due to the low contrast of the negatives). The dynamic range of the image processing system is white (for which the gray level is digitized at the value 0) to black (255). The effective magnification of the image is 53 Å/pixel (film scanner/negative: 1 cm/186 pixels) × (1/10 000 magnification).

The scanned images are photographically negative (because we scan the EM negatives). We therefore, complement the digitized image with respect to gray value so that the strands are darker than the background. The entire gray level range of the image is biased toward one end of the dynamic gray scale range of the image processing system since it consists mostly of background. The images are shifted toward the middle of this range (for aesthetic purposes without the loss of any gray scale information) by subtracting a uniform gray value across the images. This causes the gray level distributions of the complemented images to start at approximately gray level 50 which facilitates subsequent photographing of the images from the RTPP TV display.

2.2. The notch filter

The digital notch filter, discussed in (*15*), is a transformation used to remove low spatial frequencies from a signal. An $n \times n$ pixel sampling window is moved through the image and its average is subtracted from the center point of the window for each point in the image. For RNA EM micrographs where strands appear to be several pixels wide, a value for n of 32 is used. When applied to digitized RNA images, it is used to remove the low spatial frequencies corresponding to image shading.

A dc offset value, B, is added to the value computed by the filter at each point. This requires two passes of the algorithm (the first being used to compute B). The filter is computed in [1].

$$G_j(x,y) = Gi(x,y) - \mathrm{AVG}n \times n(x,y) + B. \qquad [1]$$

Let

$$n1 = (n/2)-1. \qquad [2]$$

Then,

$$\mathrm{AVG}n \times n(c,r) = \frac{1}{n{*}n} \sum_{\substack{x=c-n1:\\ c+n/2}} \sum_{\substack{y=r-n1:\\ r+n/2}} Gi'(x,y). \qquad [3]$$

To handle the edge conditions, the image is reflected about the boundaries using [4]

$$Gi'(x,y) = Gi(R(x),R(y)). \qquad [4]$$

$R(\cdot)$ is the reflection function and computes the reflected coordinate value on the edges at 0 and 255, under the following conditions in [5].

$$\begin{aligned} R(p) &= -p, && p < 0,\\ &= 255-(p-255), && p > 255,\\ &= p, && \text{otherwise.} \end{aligned} \qquad [5]$$

During the first pass B is computed in [6]. The final output image is computed during the second pass

$$B = -\min_{\text{All } x,y} (Gi(x,y) - \mathrm{AVG}n \times n,i(x,y)). \qquad [6]$$

To increase the efficiency, AVG$n \times n,i$ is computed iteratively in [7.1] and [7.2]

$$\mathrm{AVG}n \times n(0,r) = \frac{1}{n{*}n} \sum_{y=r-n1:r+n/2} \mathrm{COLSUM}(x,r) \qquad [7.1]$$

and,

$$\begin{aligned} \mathrm{AVG}n \times n(c,r) = {}& \mathrm{AVG}n \times n(c-1,r) - \mathrm{COLSUM}(c-n/2,r)\\ &+ \mathrm{COLSUM}(c+n/2,r), \end{aligned} \qquad [7.2]$$

$$\mathrm{COLSUM}(c,r) = \sum_{y=r-n1:r+n/2} G'(c,y). \qquad [8]$$

Although this transformation is specified here for even n, it may easily be extended to odd n taking the symmetry of the sampling window into account.

2.3. *Experiment*

The two digitized RNA micrograph images are notch filtered and the strands in the resulting images segmented. This is illustrated here in several steps to show the improvement of the images after filtering. The gray scale frequency distribution histograms of the two original RNA images are shown in Figs. 2a and b. As can be seen, the RNA gray values are distributed over a range of gray values which overlaps that of the background. The minimum histogram value in Fig. 2b (used as a

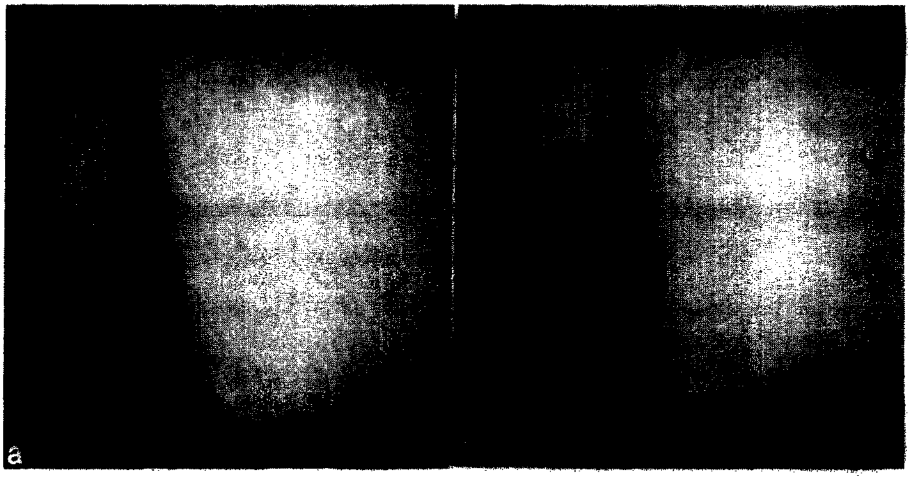

FIG. 2. Gray level frequency distribution histograms of (a) ribosomal RNA image, (b) m-RNA image. The range is from 0 (white) to 255 (black) with the histogram being scaled to the maximum value.

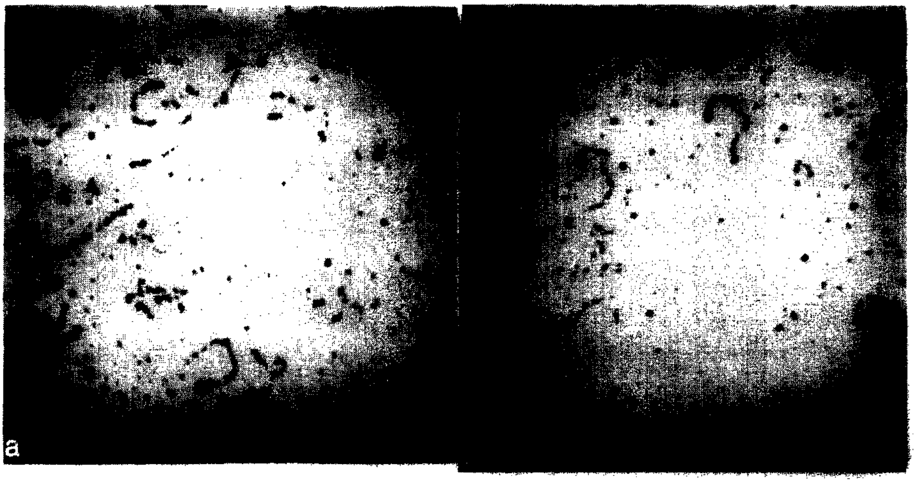

FIG. 3. Gray level threshold sliced original images of (a) ribosomal RNA image, (b) m-RNA image. The threshold range of (a) is [115:255], and of (b) is [115:255].

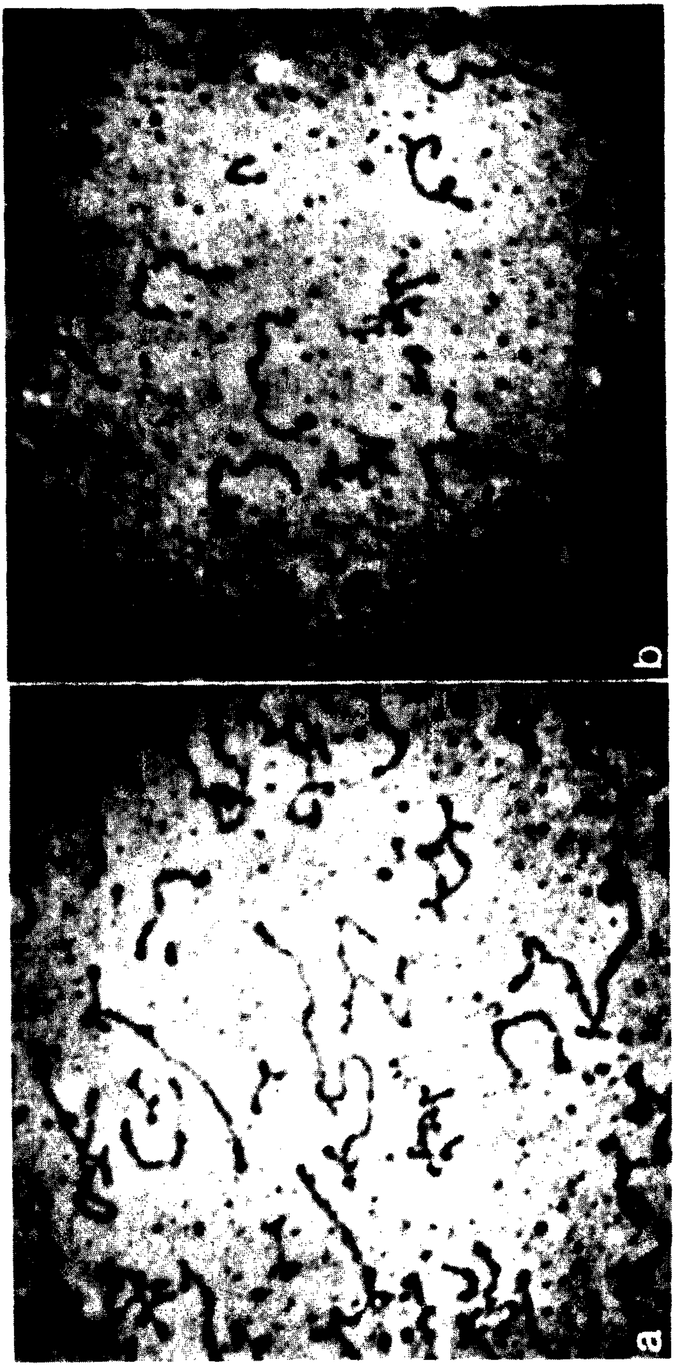

FIG. 4. Notch filtered images with an averaging window of size 32 × 32 for (a) ribosomal RNA image, (b) m-RNA image.

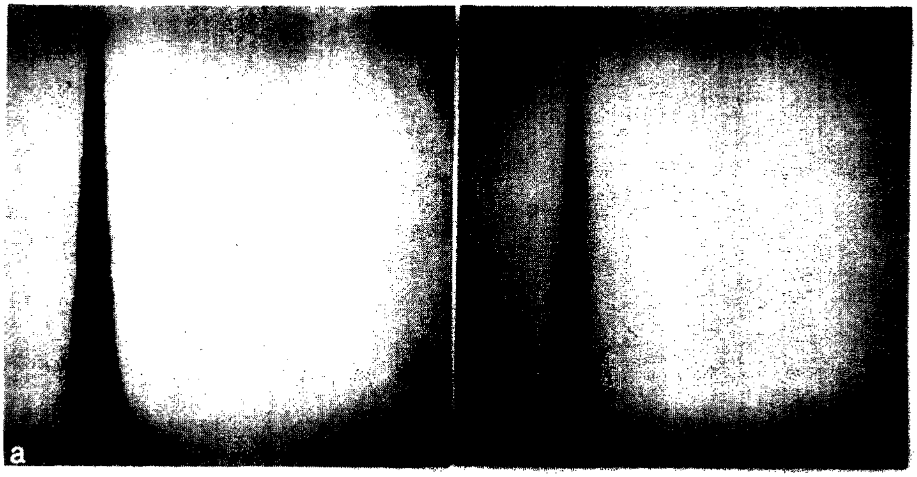

FIG. 5. Gray level frequency distribution histograms after applying the notch filter of (a) ribosomal RNA image, (b) m-RNA image. The range is from 0 (white) to 255 (black) with the histogram being scaled to the maximum value.

threshold) does not separate the RNA strands from the background but rather the shaded region from the nonshaded region. Figures 3a and b show the original images' gray value threshold sliced at values producing good boundaries for a majority of the RNA strands. However, it is still impossible to select an optimal threshold slice range to pick out all of the boundaries.

The notch filter was applied to the ribosomal RNA and a *m*-RNA image with the resultant filtered images shown in Figs. 4a and b, respectively. Notice that the major

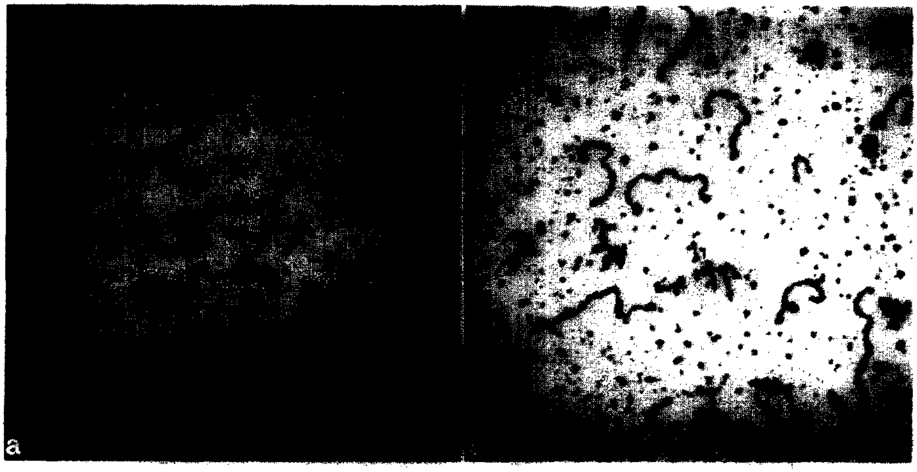

FIG. 6. Threshold sliced notch filtered images of (a) ribosomal RNA image, (b) m-RNA image. The threshold range of (a) is [65 : 255], and of (b) is [80 : 255].

shading inconsistency has been corrected. Figures 5a and b show the gray scale histograms for the notch filtered images. Figures 6a and b show the filtered images' threshold sliced at thresholds which now include all of the RNA.

Given the uniform background/foreground images produced by notch filtering, it is now possible to begin processing the RNA strands. A simple boundary follower segmentation algorithm (discussed in (*10*, *11*)) which eliminates strands less than 100 pixels in perimeter was applied to the filtered images. It successfully segmented many of the strands—although some were fragmented because of gaps of low density in the RNA. Some fragmented strands were removed because each of the pieces had perimeters less than 100 pixels. The gaps correspond to possible artificats in the sample/photograph/digitizing process. In addition to acually segmenting the boundaries, the use of a lower bound on perimeter sizing in the segmentation process effectively removes the small compact artifacts from the image. Figures 7a and b show the boundaries of the segmented images.

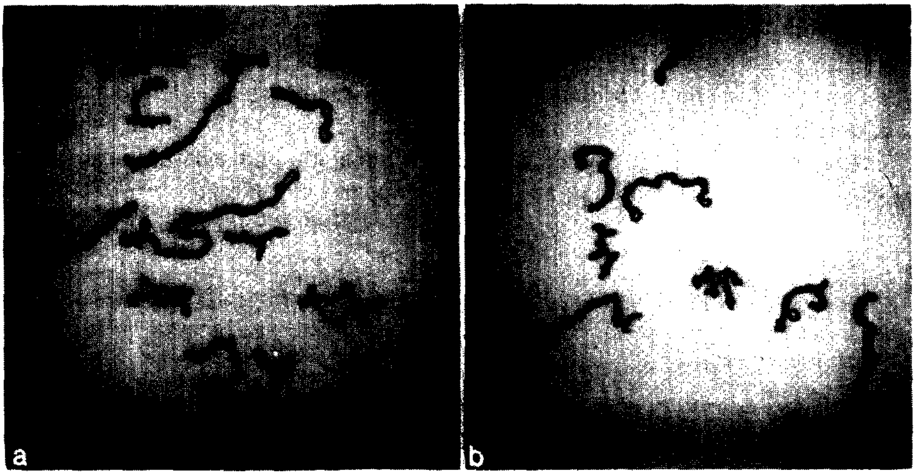

FIG. 7. Boundary follower segmentation of notch filtered images of (a) ribosomal RNA image, (b) m-RNA image. The boundary follower saved objects with boundaries greater than 100 pixels from the threshold sliced images in Fig. 5.

3. RESULTS

As can be seen from Figs. 4 and 6, the notch filter effectively removes the shading which adds complexity to analyzing the RNA EM micrographs. Having solved the overwhelming shading problem, two other problems now become apparent in attempting to automatically segment the molecules. The first is the occurrence of many small, relatively compact blobs. The second is the filling of gaps along the molecules which would cause them to be fragmented.

Both of these problems are under study and will be reported on subsequently.

4. Discussion

The shading effect encountered in these examples of adenovirus RNA electronmicrographs is yet another example of the ubiquitous class of difficulties in automatically analyzing biological images by computer. This effect is seen even in the most carefully stained and digitized biological images of excellent contrast. The apparently trivial problem of setting a suitable gray level threshold to accomplish initial scene segmentation rapidly becomes a significant barrier to further progress in automatic image analysis. The true magnitude of the problem is at first obscurred by the eyes' relative insensitivity to changes of low spatial frequency, and perhaps as much by our inherent optimism. In any case it becomes rapidly evident with almost every class of natural digitized images that a single threshold globally applied does not produce morphologically acceptable segmentation. "Natural" boundaries are violated and/or new invalid divisions (fragmentation of RNA molecules) are produced.

Gray level thresholding as an algorithm to produce initial segmentation is ineffective in large measure due to shading, which as we have seen is actually the result of several factors in the image acquisition process. In any case, even a partial suppression of shading facilitates effective scene segmentation. If shading can be suppressed over a broad domain of images and at comparatively small expenditure of resources the prospect of automatic characterization and measurement of such biologically significant materials as viral RNA appears considerably brighter.

ACKNOWLEDGMENT

The authors are much indebted to Dr. J. Maizel of NICHLMG, N.I.H. He not only provided the electron micrographs of adenoviruses on which this work is based, but also many hours of patient explanation and discussion.

REFERENCES

1. Wellauer, P. K. and David, I. B. Secondary structure maps of ribosomal RNA and DNA. I Processing of *Xenopus laevis* Ribosomal RNA and structure of single-stranded ribosomal DNA. *J Mol. Biol.* **89,** 379 (1974).
2. Wellauer, P. K., David, I. B., Kelley, D. E. and Perry, R. P. Secondary structure maps of ribosomal RNA and DNA. II. Processing of *L*-cell ribosomal RNA and Variations in the processing pathway. *J. Mol. Biol.* **89,** 397 (1974).
3. Shen, C-K. J. and Hearst, J. E. Mapping of sequences of 2-fold symmetry on the simian virus 40 genome: A photochemical crosslinking approach. *Proc. Nat. Acad. Sci. USA* **74,** 1363 (1977).
4. Hsu, A. and Jelinek, W. R. Mapping of inverted repeated DNA sequences within the genome of simian virus 40. *Proc. Nat. Acad. Sci. USA* **74,** 1631 (1977).
5. Schultz, M., Lipkin, L. E., Wade, M. J., Lemkin, P. F., and Carman, G. High resolution shading correction. *J. Histochem. Cytochem.* **22,** 751 (1974).
6. Ito, T. and Sato, K. Computer processing of electron micrographs of DNA. In "Digital Processing of Biomedical Images" (K. Preston and M. Onoe, Eds.), pp. 89–100, Plenum, New York, 1976.

7. Holdermann, F. and Kazierczak, H. Processing of gray-scale pictures. *Comput. Graphics Image Processing* **1,** 66 (1972).

8. Eron, L. and Westphal, H. Cell-free translation of highly purified adenovirus messenger RNA. *Proc. Nat. Acad. Sci. USA* **71,** 3385 (1974).

9. Wellauer, P. K. and David, I. B. Secondary structure maps of RNA: Processing of HeLa ribosomal RNA. *Proc. Nat. Acad. Sci. USA* **70,** 2827 (1974).

10. Lemkin, P. "Buffer Memory Monitor System for Interactive Image Processing." NCI/IP Technical Report No. 21, Nat. Tech. Info. Serv. PB261536/AS, Dec. 1976.

11. Lemkin, P. and Lipkin, L. BMON2—Buffer memory monitor system for interactive image processing of biological images. *Comput. Progr. Biomed.*, submitted.

12. Carman, G., Lemkin, P., Lipkin, L., Shapiro, B., Schultz, M., and Kaiser, P. A real time picture processor for use in biological cell identification. II. Hardware implementation. *J. Histochem. Cytochem.* **22,** 732 (1974).

13. Lemkin, P., Carman, G., Lipkin, L., Shapiro, B., Schultz, M., and Kaiser, P. A real time picture processor for use in biological cell identification. I. System design. *J. Histochem. Cytochem.* **22,** 725 (1974).

14. Lemkin, P., Carman, G., Lipkin, L., Shapiro, B., and Schultz, M. "Real time picture processor—Description and Specification." NCI/IP Technical Report No. 7a, Nat. Tech. Info. Serv. PB269600/AS, June 1977.

15. Schwartz, A. A. and Soha, J. M. Variable threshold zonal filtering. *Appl. Opt.* **16,** 1779 (1977).