Reprinted from ANALYTICAL AND QUANTITATIVE CYTOLOGY JOURNAL
Vol. 1 No. 1, March-April 1979

Use of the Positive Difference Transform for RBC Elimination in Bone Marrow Smear Images

Peter Lemkin, Ph.D.
Lewis Lipkin, M.D.

The use of an algorithm with a broad range of utility in digital image processing (the positive difference transform) is illustrated. Its effectiveness as a procedure for selective elimination of hemoglobin-containing images is demonstrated. A set of heuristics, employing information concerning nuclear hemoglobin content, is shown to discriminate nucleated erythrocytic cells from those of the leukocyte series.

From the Image Processing Unit, Division of Cancer Biology and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, U.S.A.

Address reprint requests to: Peter Lemkin, Ph.D., Park Bld., Room 417, NIH, Bethesda, Maryland 20014.

An unprocessed digitized image of a biologic object of any but the most minimal complexity must be segmented into its component parts before meaningful measurement can be made. Depending upon the goal of the analysis, one may select or reject classes of subobjects, i.e., begin to segment the image. In the case of images of Romanowsky-stained material which has been scanned at several selected wavelengths, one can initially separate the cell images into classes corresponding to biochemical and histogenetic distinctions. Hemoglobin, with its particular specific absorption peak (between 410 and 420 nm, depending on its combination with oxygen or other components), serves as a marker for the broad class of rubricytic cells, except for the most primitive. Mature red blood cells (RBCs) are ubiquitously present in dense marrow smears. Since the usual intent of bone marrow analysis is to perform differential nucleated cell counts, the complexity of the image is best reduced as early as possible by the removal of RBCs from the image.[5]

An algorithm is presented for eliminating formed elements with hemoglobin content (mature reticulocytes) from Wright's-stained bone marrow smears imaged through a light microscope. The algorithm subtracts a normalized gray scale image acquired using a blue filter from a normalized gray scale image acquired with a green filter using a positive difference function. The result is a gray scale image.

Methods exist for eliminating RBCs from peripheral blood images.[2,12] These methods use the fact that hemoglobin has a maximum light absorption in the region of 420 nm, as noted above. Thus, finding the RBCs necessitates finding the discriminant threshold of a blue scan corresponding to the RBC pixels. Referring to the gray scale histogram, using the valley between two peaks as the optimal threshold for separating the components corresponding to the two peaks was suggested in a paper by Prewitt and Mendelsohn.[10]

Unfortunately, noise due to optics, digitization and illumination and sensor nonuniformities tends to cause problems with this type of threshold selection. In addition, in complex images such as clusters of cells as seen in bone marrow, other effects tend to make the use of a single threshold difficult, as will be detailed below. Thus, applying gray level thresholding of the blue scan to produce a binary image results in a noisy RBC mask because of sensor nonuniformity as well as "salt and pepper" noise. In addition,

0190-0471/79/0101-0067/$02.00/0

image content affects selection of the threshold, making optimal threshold selection difficult.

Materials and Methods

The bone marrow images used in the experiment were obtained using the Real Time Picture Processor hardware, RTPP,[4,7,8] and the Buffer Memory Monitor System image processing system, BMON2.[6,9]

The slide was that of a Wright's-stained aspirated normal marrow smear scanned at 800 × on an Axiomat microscope (Zeiss Optical, New York, New York). A Vidicon camera was used with video data digitized to eight bits (although the video amplifiers produce about a seven-bit dynamic range) and stored in real time in the buffer memory of the RTPP (0 is white, denoted Tmin; 255 is black, denoted Tmax). The light source was a 200-w mercury arc filtered at 420 and 546 nm by 10-nm-wide bandpass interference filters. The area scanned was 256 × 256 pixels at about 0.15μ/pixel, giving about a 130μ square image area.

Six images (numbered 1 through 6) were selected, illustrating complex clusters of leukocytes and reticulocytes in various developmental stages and geometric configurations. The green (GN) and blue (BL) scans are shown in Figure 1.

RBC Elimination

Problems with Thresholding the Blue Image. As noted, the blue image is not always easily segmentable into RBCs and other objects by thresholding because the RBC threshold for complex scenes is so variable. Figure 2 illustrates several images taken at increasing thresholds for three different blue images

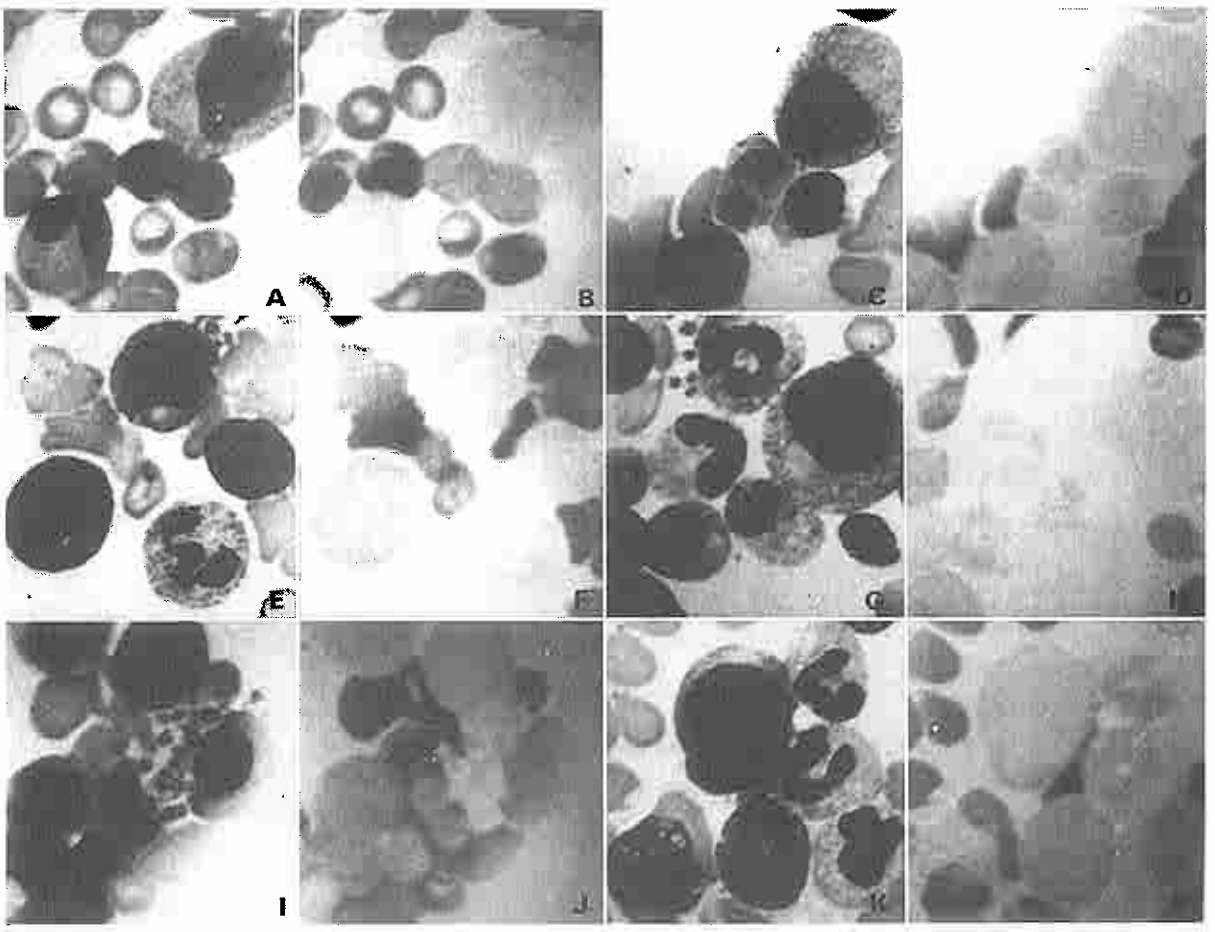

Figure 1
Six original green and blue scanned images. (A) 1-green. (B) 1-blue. (C) 2-green. (D) 2-blue. (E) 3-green. (F) 3-blue. (G) 4-green. (H) 4-blue. (I) 5-green. (J) 5-blue. (K) 6-green. (L) 6-blue.

starting at the approximate RBC threshold Trbc (110) and at two values above it (120 and 130). These values were obtained by examining the blue image histograms shown in Figure 3. The blue histograms show additional peaks for some normoblasts in Figure 3B while the normoblast peak in Figure 3J is not apparent.

The two normoblast nuclei (which contain some hemoglobin) in Figures 2A-2D are not as dense as the RBCs, yet they will effect a shift in the RBC threshold. If the threshold is raised slightly to remove them, the RBCs start to fragment. In Figures 2E-2H, those RBCs overlying one another or squeezed in between other cells as well as the crushed RBC in the center left of the image are denser than isolated RBCs in other parts of the images. In Figures 2I-2L, there is the case of a RBC overlying a white cell, with the RBC's density in between that of the isolated RBCs and those RBCs squeezed or overlying one another.

Thus, there is no unique threshold for eliminating all of the RBC content from the image due to local variation in RBC density depending on how it presents itself in conjunction with other objects in the image.

Another technique, developed by Bacus,[1] called the "whitening transform," maps two colored images into a "colored" image C and a "density" image D. This is done by computing the principal eigenvectors and eigenvalues for each image to maximize separation. Bacus makes two assumptions: that the principal eigenvector lies along a 45° line in the original bivariate space and that the variances in either direc-

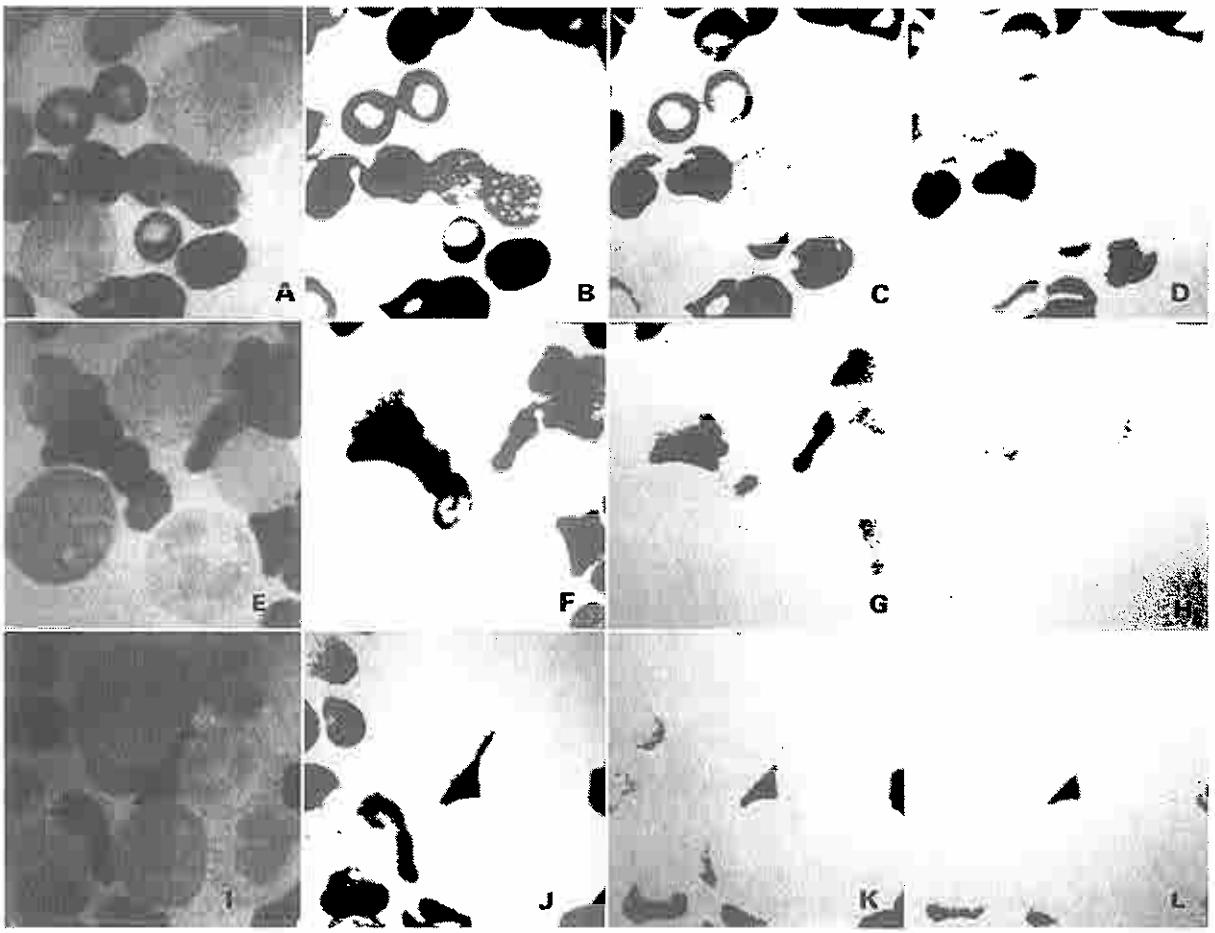

Figure 2
Three blue (BL) images sliced at about the RBC threshold, Trbc and above (110, 120 and 130). (A) 1-BL. (B) 1-BL sliced 110. (C) 1-BL sliced 120. (D) 1-BL sliced 130. (E) 3-BL. (F) 3-BL sliced 110. (G) 3-BL sliced 120. (H) 3-BL sliced 130. (I) 6-BL. (J) 6-BL sliced 110. (K) 6-BL sliced 120. (L) 6-BL sliced 130.

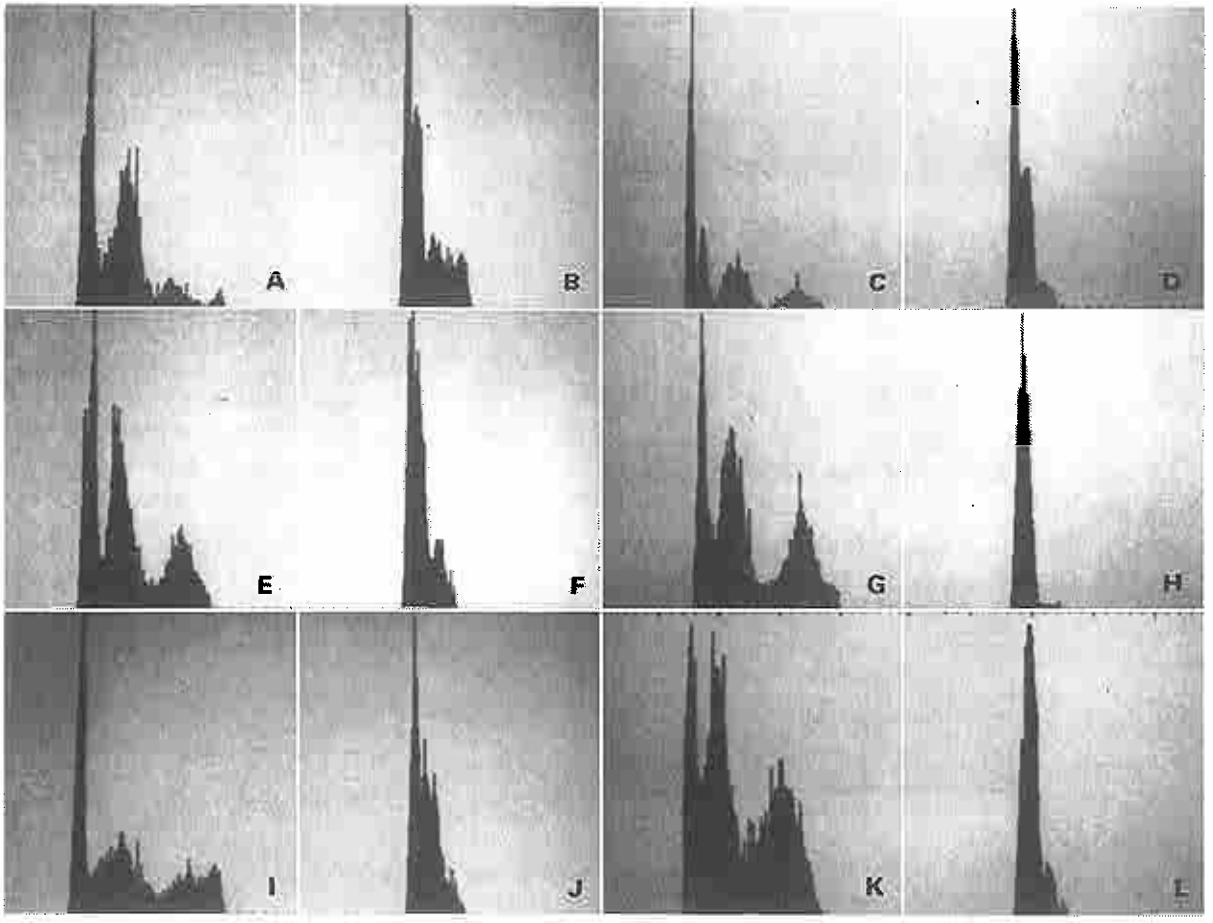

Figure 3
Six gray level frequency distribution histograms of the original green (GN) and blue (BL) scanned images, with grid marks every 10 gray levels in the range of 0:255. (A) 1-GN. (B) 1-BL. (C) 2-GN. (D) 2-BL. (E) 3-GN. (F) 3-BL. (G) 4-GN. (H) 4-BL. (I) 5-GN. (J) 5-BL. (K) 6-GN. (L) 6-BL.

tion are relatively constant. For the blue and green images used in our experiments, this transform is given by the equations

$$C = K(BL-GN) \quad \text{and} \quad D = K(GN+BL).$$

Figure 4 shows data for evaluating the whitening transform for images 4, 5 and 6. Note that it will not work very well for image 4 but may work (to some extent) for images 5 and 6. The lack of discrimination seems to be due to overlapping RBCs which spread out the blue (hemoglobin) contributions in a continuum depending on the extent of overlap. This overlap problem is especially common with marrow smears but does not occur with the spinner-produced smears of peripheral blood used by Bacus.

Because of the problem of RBC overlap, a nonlinear technique was developed in order to eliminate as much as possible of the different RBC populations (normal RBCs, rouleaux, RBCs overlapping white blood cells).

Image Normalization. In order to eliminate local density effects, the two images are normalized by their density extrema. The extrema are computed during histogram computation for the two images. Figure 3 shows the histograms for the green and blue images. The normalized blue and green image transforms are defined by four equations:

$$\begin{aligned} \text{gmin} &= \text{MIN}(g(x,y)|g(x,y) \text{ in GN}),\\ \text{gmax} &= \text{MAX}(g(x,y)|g(x,y) \text{ in GN}),\\ \text{bmin} &= \text{MIN}(g(x,y)|g(x,y) \text{ in BL}),\\ \text{bmax} &= \text{MAX}(g(x,y)|g(x,y) \text{ in BL}). \end{aligned}$$

Then, the normalized images GN′ and BL′ are com-

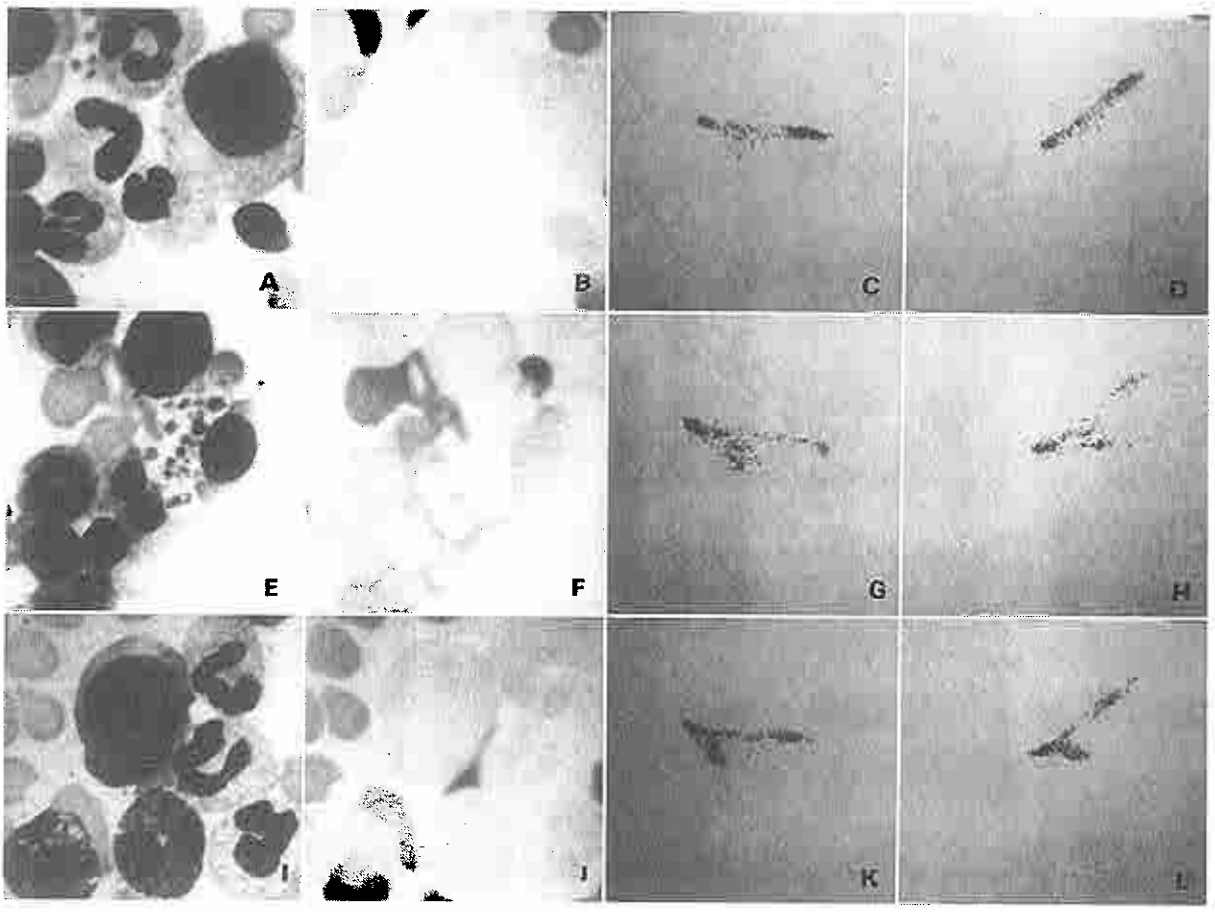

Figure 4
Comparisons of joint gray scale distributions of GN versus BL and C versus D for three images. The x-axis (left = 0) of the joint distribution is GN (C) while the y-axis (top = 0) is BL (D). (A) 4-GN. (B) 4-BL. (C) 4-GN vs. BL. (D) 4-C vs. D. (E) 5-GN. (F) 5-BL. (G) 5-GN vs. BL. (H) 5-C vs. D. (I) 6-GN. (J) 6-BL. (K) 6-GN vs. BL. (L) 6-C vs. D.

puted using linear interpolation to fill an eight-bit dynamic range according to two equations:

$$GN'(x,y) = \frac{(gmax - gmin)}{Tmax} \cdot (GN(x,y) - gmin)$$

$$BL'(x,y) = \frac{(bmax - bmin)}{Tmax} \cdot (BL(x,y) - bmin).$$

Although in principle it is possible for the normalizations to be biased by a few noise pixels at either extreme of the gray scale range, this has turned out not to be a problem in the experiments performed. If it were a problem, it could be corrected by ignoring very low percentages of outlying pixels. Figure 5 shows the GN′ and BL′ normalized images, which have better contrast than the original images shown in Figure 1 because of the contrast stretching effect of the normalization.

Positive Difference Transform. The positive difference transform,[11] PDT, is an asymmetric arithmetic operation which forces negative gray values to zero. This has the effect of losing some of the information. Figure 6 illustrates the positive differences GB = PDT(GN′,BL′) and BG = PDT(BL′,GN′) computed using

$$PDT(g1,g2) = \begin{cases} g1\text{-}g2, & (g1\text{-}g2 > 0) \\ 0, & \text{otherwise.} \end{cases}$$

over all pixels in the image. As can be seen, the second difference (BG) is not very useful for white cell extraction, whereas the first difference (GB) contains the desired subtraction of hemoglobin-formed objects. The BG difference seems to contain hemoglobin-formed objects as well as background.

Results of RBC Extraction. Figure 7 shows the histograms and the smoothed histograms of the GB images. The GB gray level distribution histograms

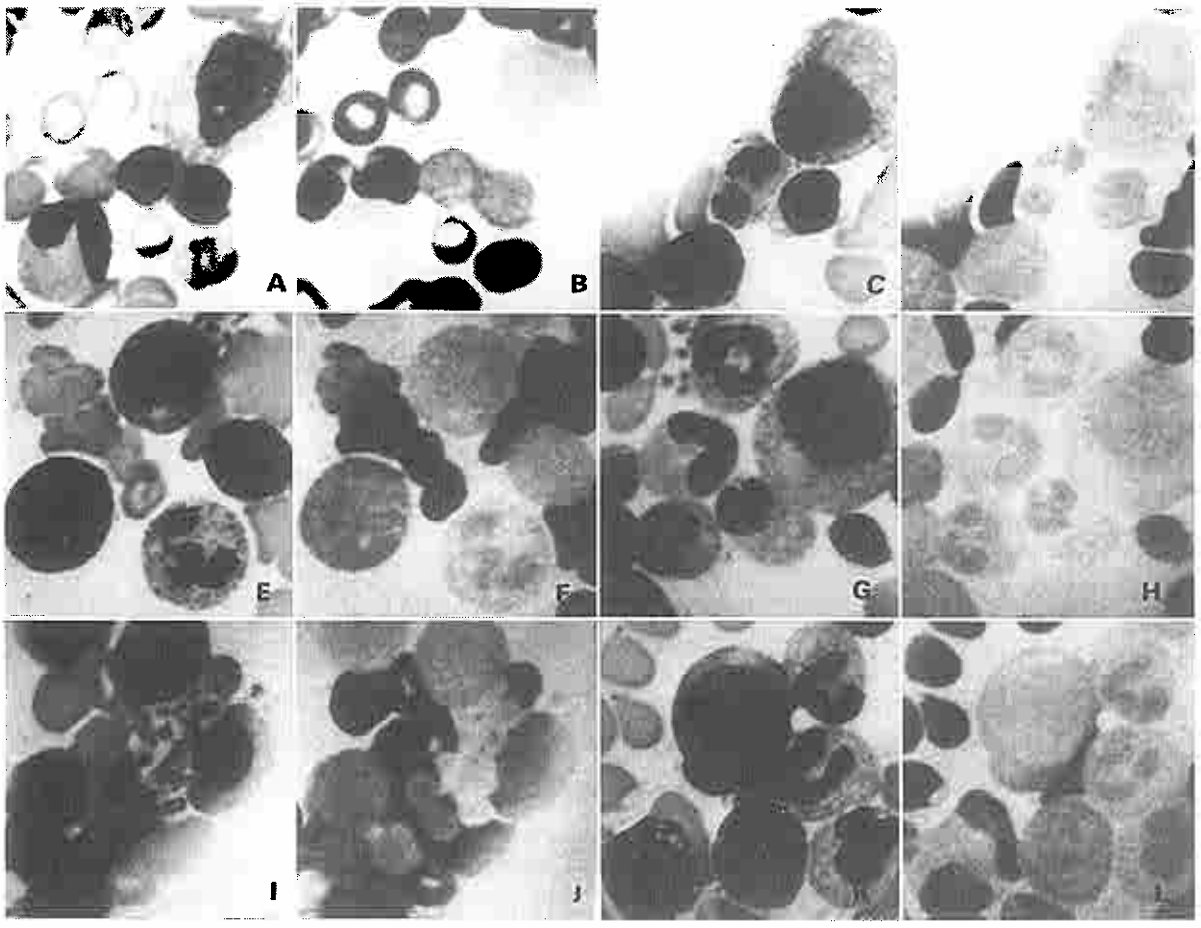

Figure 5
Six normalized green and blue images. (A) 1-GN'. (B) 1-BL'. (C) 2-GN'. (D) 2-BL'. (E) 3-GN'. (F) 3-BL'. (G) 4-GN'. (H) 4-BL'. (I) 5-GN'. (J) 5-BL'. (K) 6-GN'. (L) 6-BL'.

can be interpreted as follows: Tmin (lowest value of the histogram) to Tcyto (first minima in the histogram) is the range for plasma proteins and background, Tcyto to Tnuc (second minima in the histogram) is the range for cytoplasm and parts of platelets, and Tnuc to Tmax is the range for nuclei.

If, as it turns out, Tcyto and Tnuc are fairly constant over a wide range of image content and complexity, then no histogram analysis of the GB histograms is necessary. This is helpful since these histograms are noisy because of normalization. In looking at the GB histograms one notes that Tnuc centers around 60, fairly independently of the content of the image, and that Tcyto centers around 10, also fairly independently of the image content. Because the cytoplasm gray scale distribution intersects that of the background, Tcyto is dropped to five to keep down fragmentation. Figure 8 shows the six images sliced at [Tnuc:Tmax] with very good extraction of the nuclei. The failures in Figure 8F were caused by the morphology of the basket cell, which was heavily vacuolated and had overlapping RBCs, and the heavy granulation of the mature eosinophil. Both presented problems with the definition of the nuclear border.

In summary, a nonlinear gray scale transform for multispectral bone marrow images can be used successfully to eliminate formed elements containing hemoglobin. This transform also has the property that nuclear and background thresholds are fairly constant and independent of image content. This algorithm is thus a useful preprocessing step in segmenting images of bone marrow smears.

Erythrocyte Extraction. Erythrocytes, because of their hemoglobin content and its characteristic absorption at 419 nm, are segmentable as a group using a threshold discriminant on the blue scan. This dis-

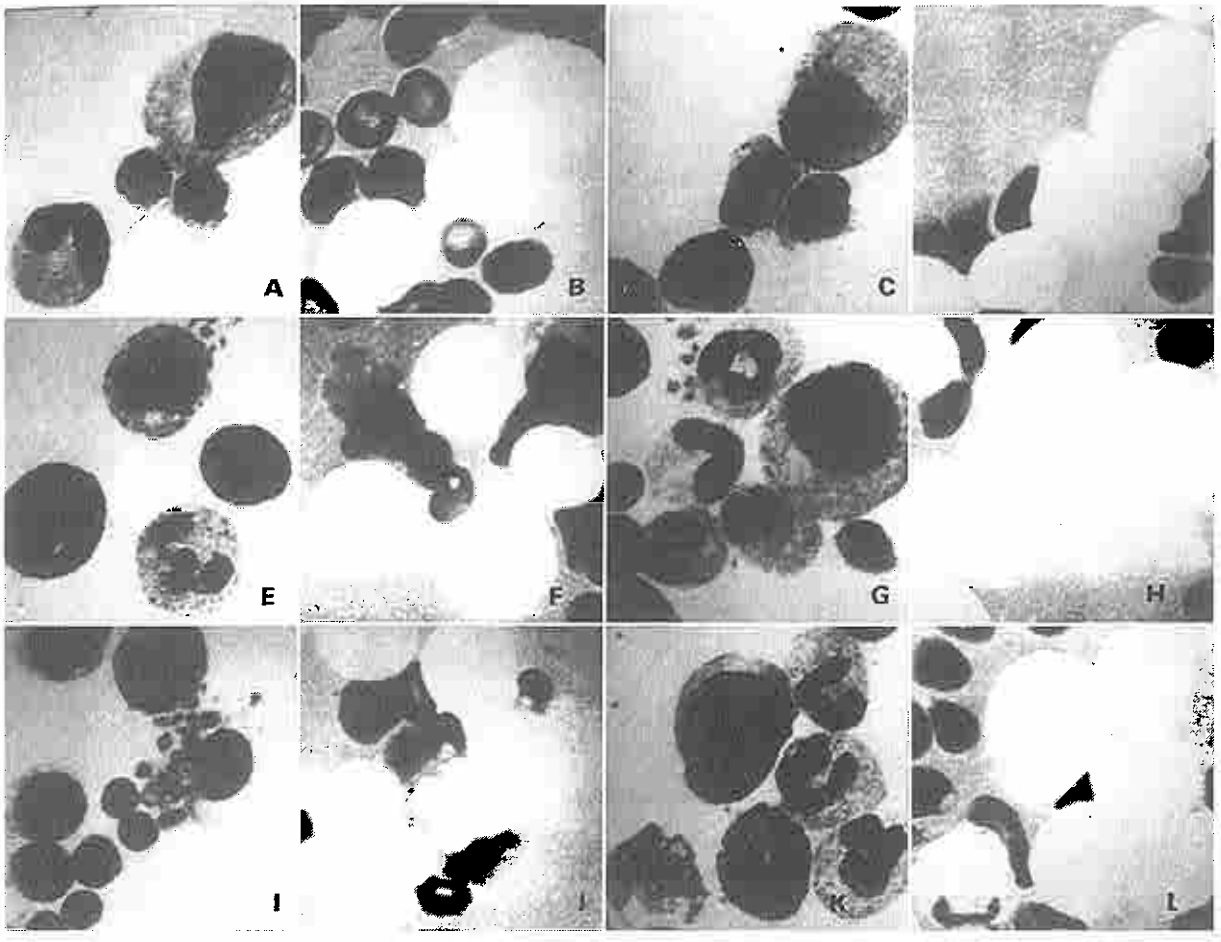

Figure 6
Positive difference transform for six images. (A) 1-GB. (B) 1-BG. (C) 2-GB. (D) 2-BG. (E) 3-GB. (F) 3-BG. (G) 4-GB. (H) 4- BG. (I) 5-GB. (J) 5-BGF. (K) 6-GB. (L) 6-BG.

criminant threshold is defined as the valley of the blue density histogram. The discriminant divides the image into two regions: leukocytes, platelets and background versus RBCs and nucleated erythrocytes. Figure 9 shows the blue histograms of the six images in the data base with the RBC threshold valley obvious in only some of the cases. As Bacus points out,[1] scarcity of RBCs in the image might cause problems in finding the valley.

Occasionally there are extra valleys in the histogram due to overlap of RBCs with other RBCs in rouleaux or with leukocytes as well as to the occurrence of smashed RBCs, which tend to have a lower hemoglobin content. The overlapping causes additional valleys at the higher density histogram than would be due to RBCs occurring alone; the smashed RBCs cause additional valleys at the lower density histogram.

Figure 10 shows the blue images sliced at Trbcs determined manually for each of the blue image histograms.

Nucleated Erythrocyte Extraction

RBC precursor cells, the most common of which are normoblasts, are nucleated cells which appear late enough in the red cell line to contain some hemoglobin.[3] In blue light (419 nm), the hemoglobin in such cells appears less dense than in RBCs but denser than in leukocytes. It often appears as a coarse texture rather than a continuous dense region.

The following algorithm determines which nucleated cells are possibly nucleated RBCs. It uses the RBC hemoglobin threshold Trbc with the blue image and the nucleus-connected component image CCnuc, segmented from the GB image. This is done by first defining a mask image, MASKrbc, of all of the red

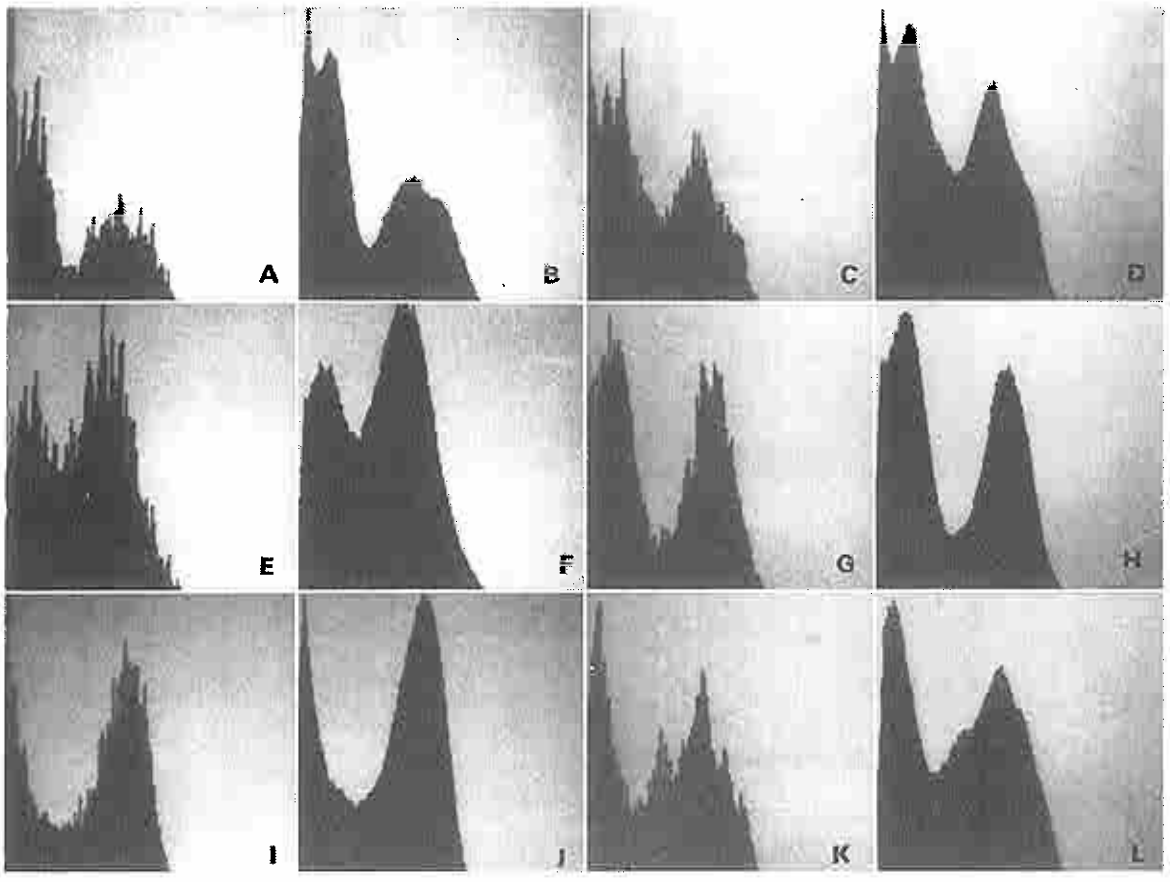

Figure 7
Histograms of positive differences for six images, with grid marks every ten gray levels over the range of 0:255. (A) 1-H(GB). (B) 1-SMOOTH(GB). (C) 2-H(GB). (D) 2-SMOOTH(GB). (E) 3-H(GB). (F) 3-SMOOTH(GB). (G) 4-H(GB). (H) 4-SMOOTH(GB). (I) 5-H(GB). (J) 5-SMOOTH(GB). (K) 6-H(GB). (L) 6-SMOOTH(GB).

cells in the blue image greater than gray value Trbc. Then each nucleus in the image is tested to see whether it also overlies an RBC. To avoid problems with image noise causing a wrong decision, a minimum area is used to validate the decision. The algorithm is given by equations:

"Extract hemoglobin containing formed object"
MASKrbc(x,y) = 255, (Trbc Leq BL(x,y) Leq Tmax)
0, otherwise.

and

"Test nuclei for hemoglobin content"
Nucleated RBCs = {i|(i in CCnuc) and
(Area(CCnuc(i) & MASKrbc) <t)}.

Figure 11 shows BL&CCnuc and BL&CCnuc& MASKrbc for images 4 and 5 (the latter image contains several normoblasts while the former does not). Table I shows the nuclear area and the area corresponding to hemoglobin content (above threshold Trbc) for each nucleus for cells in these images (in square microns). For threshold t set to 10 square microns, the normoblasts in image 5 are correctly detected while the nuclei in image 4 are correctly ignored.

Having identified the nucleated RBCs, a modified RBC mask image MASKrbc' may be created which has the nucleated RBC nuclei removed. It is computed over the range of (x,y) using

MASKrbc'(x,y) = 0, (i is nucleated RBC)
and (MASKrbc(CCnuc(i) >0)
MASKrbc, otherwise.

Discussion

A technique has been proposed which offers the ability to remove hemoglobin-containing formed elements relatively independently of the quantity of hemoglobin. This is especially important in marrow, where such formed elements are common and where a unimodal distribution model of hemoglobin cell content is not always valid because of cell overlap,

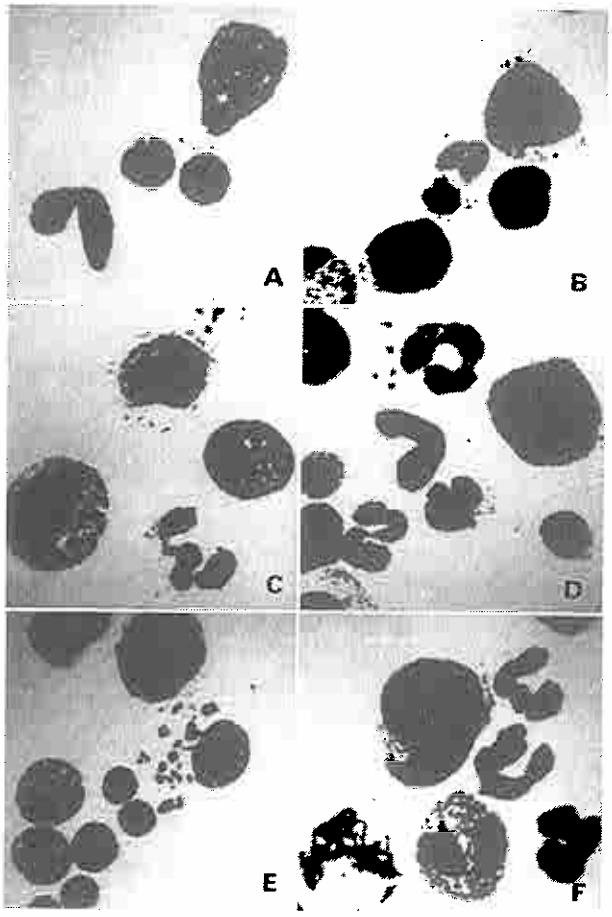

Figure 8
Six GB images sliced at 60:255. (A) 3-GB. (B) 4-GB. (C) 5-GB. (D) 6-GB. (E) 7-GB. (F) 8-GB.

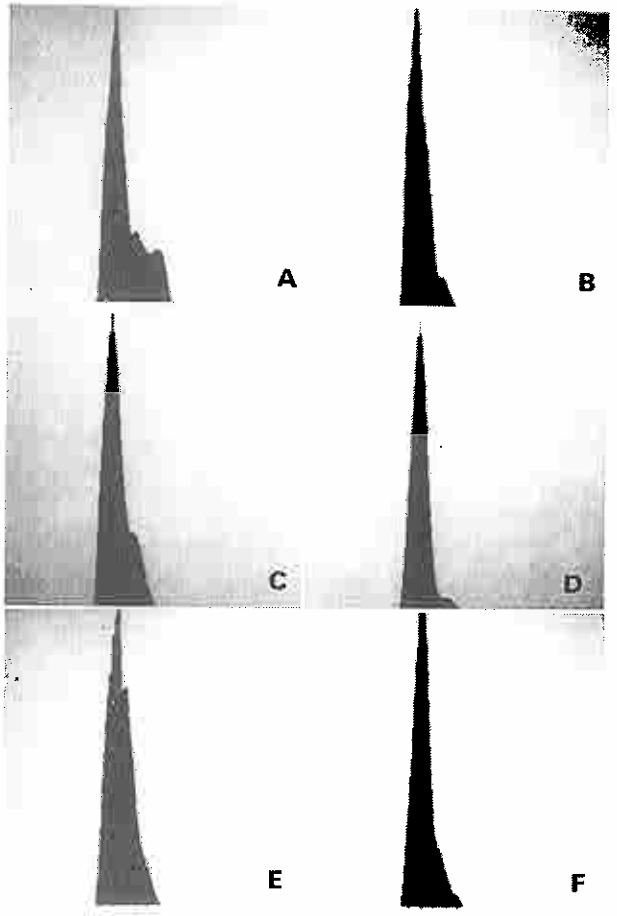

Figure 9
Smoothed gray value distribution histograms of the smoothed blue images, with grid marks every ten gray levels over the range 0:255. (A) 1-BL. (B) 2-BL. (C) 3-BL. (D) 4-BL. (E) 5-BL. (F) 6-BL.

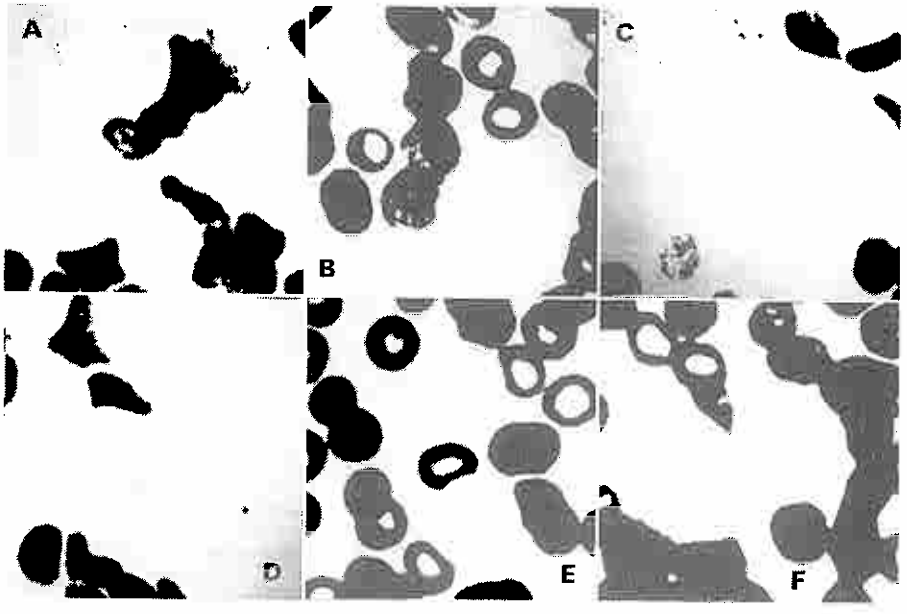

Figure 10
The blue images threshold sliced at Trbc: 225. (A) 1-Trbc:110. (B) 2-Trbc:110. (C) 3-Trbc:110. (D) 4-Trbc:108. (E) 5-Trbc: 110. (F) 6-Trbc:110.

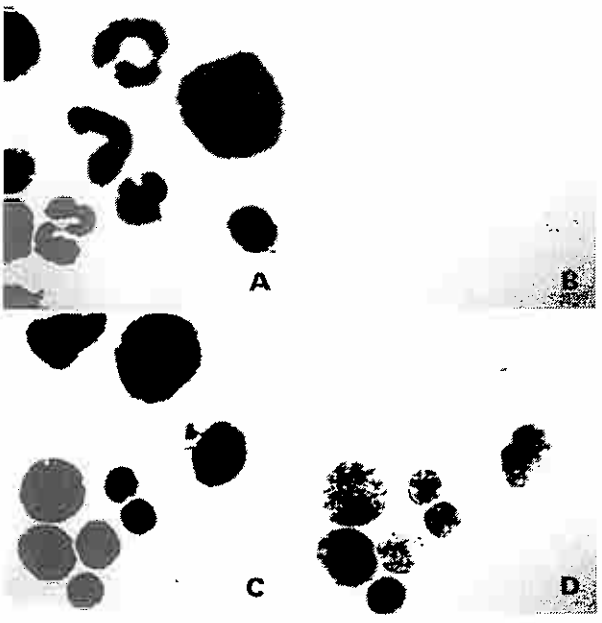

Figure 11
Blue image pixels defined by CCnuc&BL and BL&CCnuc& MASKrbc. (A) 4-CCnuc&BL. (B) 4-CCnuc&MASKrbc. (C) 5-CCnuc&BL. (D) 5-CCnuc&MASKrbc.

Table I *Hemoglobin Content of Nuclei*

Image #	CC# i	Total nuclear area (CCnuc)	Hemoglobin content of nucleus area (CCnuc(i)& MASKrbc)
4	2	48.2	0.02
4	3	63.8	0.00
4	4	160.4	0.10
4	5	54.3	0.00
4	6	26.2	0.75
4	7	40.1	0.00
4	8	47.2	0.15
4	9	39.5	0.02
4	10	32.8	2.37
4	11	20.1	0.15
5	2	59.3	0.22
5	3	112.5	6.22
5	4	55.3	31.18
5	5	62.0	45.16
5	6	17.3	13.66
5	7	18.2	14.40
5	8	30.8	12.93
5	9	47.1	43.39
5	10	18.9	18.90

The content is measured by the area of the blue image (for each nucleus CCnuc(i) above Trbc. In order to contrast the nuclei, the areas of the nuclei are given as well. All areas are given in square microns.

debris and the occurrence of RBC precursor cells. The technique has been successfully demonstrated in the cases presented in this paper. In addition, a technique has been proposed for using the nuclear hemoglobin content of nucleated cells to discriminate nucleated erythrocytes from leukocytes, which it did successfully in the test cases.

Acknowledgments

The authors would like to thank George Carman of Corvallis, Oregon, and Morton Schultz and Bruce Shapiro of the Image Processing Unit for their contributions to the design and construction of the RTPP system, on which this algorithm was developed. Parts of this paper were taken from a chapter in the Ph.D. dissertation of one of the authors.[5]

References

1. Bacus JW: A whitening transform for two-color blood cell images. Pattern Rec 8:53-60, 1976
2. Bacus JW, Aggarwal RK, Belanger MG: Computer recognition of microscopic images. EASCON 73.A-73.F, 1975
3. Bessis M: Living Blood Cells and Their Ultrastructure. New York, Springer-Verlag, 1973
4. Carman G, Lemkin P, Lipkin L, Shapiro B, Schultz M, Kaiser P: A real time picture processor for use in biological cell identification: II. Hardware implementation. J Histochem Cytochem 22:732-740, 1974
5. Lemkin P: Bone Marrow Smear Image Analysis. Ph.D. Dissertation, Computer Science Department, University of Maryland, College Park, 1978
6. Lemkin P: Buffer memory monitor system for interactive image processing. NCI/IP Technical Report 21b, April 1978
7. Lemkin P, Carman G, Lipkin L, Shapiro B, Schultz M: Real time picture processor—description and specification. NCI/IP Technical Report 7a, June 1977
8. Lemkin P, Carman G, Lipkin L, Shapiro B, Schultz M, Kaiser P: A real time picture processor for use in biological cell identification: I. System design. J Histochem Cytochem 22:725-731, 1974
9. Lemkin P, Lipkin L: BMON2—buffer memory monitor system for interactive image processing of biological images. Comput Programs Biomed (submitted for publication)
10. Prewitt JMS, Mendelsohn MI: The analysis of cell images. Ann NY Acad Sci 128:1035-1053, 1966
11. Rosenfeld A: Picture Processing by Computer. New York, Academic Press, 1969
12. Young IT, Paskowitz IP: Localization of cellular structures. IEEE Trans Biomed Eng 22:35-40, 1975