0196-4763/80/0103-0193$0.00/0
CYTOMETRY

Vol. 1, No. 3, pp. 193-199, 1980
Printed in U.S.A.

Estimation of the Performance of an Array-Processor Oriented System for Automatic Pap Smear Analysis[1, 2]

Wolfgang Abmayr, Peter Gais, Karsten Rodenacker and Georg Burger
Institut für Strahlenschutz, Gesellschaft für Strahlen- und Umweltforschung mbH, München, West Germany

Received for publication February 8, 1980; accepted July 17, 1980

The rapid progress in computer technology makes possible the automatic analysis of thousands of cells on a slide in the field of automatic uterine cancer cytology. Our approach starts with the high-resolution scanning of visually selected and classified single cells determining the training set for discriminant analysis. On the basis of more than 15 morphologic and textural features measured, correct classification results of 95% are reached. Our aim is to speed up our image processing system by means of an Array-Processor, the model AP 120 B from Floating Point Systems. In this study a realistic time estimation of the scanning, segmentation, feature extraction and classification of about 40,000 cells on a slide is performed.

Cell image analysis has proven to be a powerful tool for single cell recognition. It implies the scanning of cells with high spatial and photometric resolution, the processing of the digitized images and the cell classification in the feature space.

The application of the method in the automation of diagnostic cytology requires the screening of a statistically significant and cytologically relevant sample of cells in a patient's specimen.

In this procedure image data processing is most time consuming. This is especially true for software oriented sequential procedures implemented on general purpose minicomputers. The total time required for the analysis of a single cell can be as long as a few minutes. The long cell analysis times restrict the use of such systems to research in cancer prescreening projects, as the collection of single cell data bases, the development of feature extraction and selection algorithms and the demonstration of object finding, segmentation and classification routines (2, 4). The systems are normally too slow for operational tests on total cell samples, containing between 10,000 and 50,000 objects per slide, and hence provide no possibilities for the clinical validation of the system's operation characteristics for specimen classification. A system capable of scanning a total cell sample on a slide has to be 100 to 1000 times faster, which seems only achievable by parallel processing strategies (6).

There are several possible approaches along these lines. One example is the Leitz TAS[3] system in which parallel processing algorithms based on the concepts of mathematical morphology (10, 12) are realized in hardware modules. With this system it is possible to detect essentially large and/or dark nuclei in a microscopic field, eliminating artefacts as well as most of the normal cells. This method has been applied to AFS-stained slides (17) and to Papanicolaou (PAP)-stained slides (11). The positive alarm rates can be adjusted corresponding to the accepted false negative alarm rates. It could be demonstrated that false negative slide classifications can be practically avoided with only a few percent false positive alarms. For the latter an improved high resolution image analysis must be performed, or the system operates as an interactive one, where the preselected objects have to be examined visually by a cytopathologist.

Another approach is the home made array-processor system FAZYTAN (15). A cell image is described by several hundred features derived from the primary scanning pattern of the nucleus of PAP-stained cells. The number of features used is reduced by applying principal axis transformation. In this case feature extraction is restricted to image transformations by two-dimensional local operations. Because of the fast feature

[1] Supported in part by the Federal Ministry of Research and Technology (BMFT).
[2] Presented at Automated Cytology VII, Asilomar, California, November 25-30, 1979.

[3] Texture Analysis System of Leitz-Wetzlar, West Germany.

extraction, it seems possible to analyse each object on a slide. No true slide classification results have been reported up to now.

The concept followed by our own laboratory is based upon the use of a commercially available and programmable array-processor, to combine the advantages of a fast hardwired processor with an easily programmed general purpose computer. It is supported mainly by our investigations on the finding and recognition of individual cellular and noncellular objects in monodisperse PAP-stained cervical specimen on glass slides.

Time Constraints in Automated Cytology

The following considerations allow rough time estimates for specimen classification in automated cytology.

A slide with a scanning area of 20 × 20 mm^2 containing ideally 40,000 cellular objects is split into roughly 1,500 non-overlapping scanning fields of 512 × 512 pixels with 1 μm pixel spacing, 6,000 fields with 0.5 μm pixel spacing or 24,000 field with 0.25 μm pixel spacing. In order not to lose the cells touching the boundary the scanning fields must, however, overlap to the extent of the diameter of such cells.

The number of rectangular fields with sidelengths d to be scanned in case of p micron overlap relative to that without overlap is simply given by

$$u = \frac{d^2}{(d-p)^2}$$

u is called the overlapfactor

Let us assume the upper diameters for suspicious cells to be 64 μm. In case of a 0.25 μm pixeldiameter and a field size of 128 × 128 μm^2 the overlap factor is 4. The total number of fields scanned amounts then to 96,000.

A time of about 9 msec would be required for the complete analysis of one field to obtain a complete slide classification in about 15 min. In case of 0.5 μm pixel distance the time per field is reduced to about 88 msec. This enormous processing speed has not been achieved in the previously mentioned

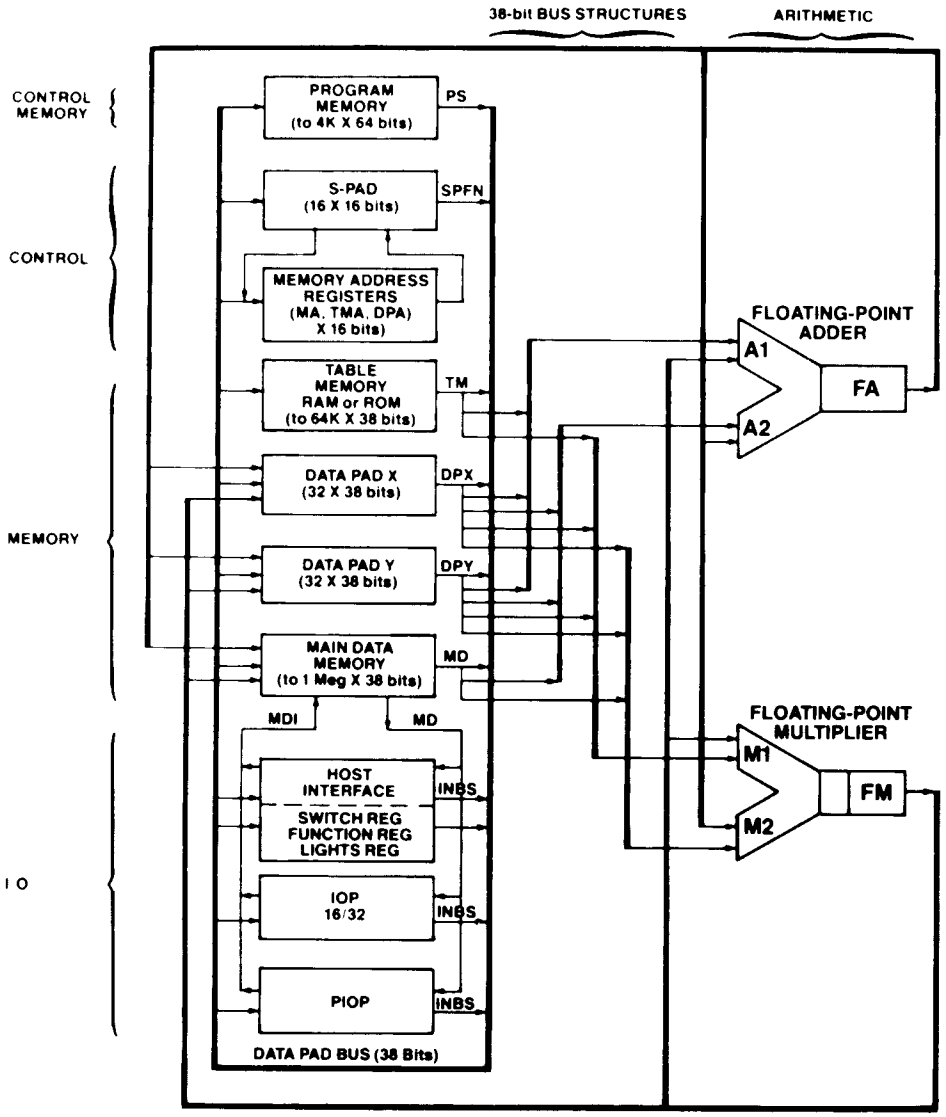

FIG. 1. Architecture of the AP-120 B from Floating Point Systems (3).

high-resolution systems, using either general purpose computers and array-processors or even hardwired systems. It may be realized by special array-processor systems (7, 8), which are currently in the planning stage. It should, however, be mentioned that in practice the time constraints will be less stringent. First, it seems proven by the existing investigations on cell populations that the number of atypical cells is high enough to allow in most cases only a small part of the slide area to be scanned for a reliable sample diagnosis. Second, it should be possible, by improving the segmentation algorithm, to increase the ideal cellular object density up to 15,000 objects per cm^2. Finally, if it proves true that cell nuclei analysis only yields sufficient information the density of cells can be increased by at least a factor of 10 and hence the total area, the necessary overlap distance and the data processing speed can be lowered considerably.

Architecture of AP 120 B Array-Processor and System Layout

The greatest advantage of dedicated minicomputers (compared to large scale computers) is their low cost. They have a lower computational speed (about 0.5 mips or less) than the large scale computers, but this can be overcome by array processors which are designed as attached machines. Our particular machine is a Floating Point System AP 120 B Array-Processor (AP) (Fig. 1). The AP features include (3) the following. (*a*) Independent storage for programs, data and constants; (*b*) independent Floating Point multiplier and adder units; (*c*) two blocks of 32 Floating Point registers; (*d*) independent address indexing and counting by integer arithmetic unit with integer registers; (*e*) precision enhancement by 38 bit internal Floating Point format with convergent hardware rounding.

The large instruction length and multiple data paths permit parallel operation of all of these elements so that, for example, a floating point adding, floating point multiplying, branch on condition, index increment, memory storage and a load to and from floating point and integer registers, can be simultaneously executed in a single 167 nsec cycle. The result of this architecture is a processor with a maximum computation rate of around 12 mips, which is approximately the raw speed of a CDC 7600 computer but in the price range of a minicomputer.

The AP 120 B is internally synchronous; thus there is no need for internal handshaking between arithmetic units, memories and processors. Data and results are available at precisely determined times. This synchronous approach allows a step by step program debugging.

Some array processor systems, such as the AP-400 from Analogic, are designed as asynchronous multiprocessor machines. Asynchronous design necessitates an interaction protocol (handshaking) to synchronize communication between each processor. This leads to a more complex machine which is difficult to program.

The AP 120 B is designed in a typical ring structure; the user has no contact with the operating system. Access to the AP 120 B is accomplished through the host's Fortran program by calling mathematical routines supplied by Floating Point Systems. A Fortran compiler is available which translates Fortran-code into Fortran-callable routines for the AP processor.

In Table 1 the processing time of typical image processing operations on both binary and gray-level images are listed and compared to operations performed on the TAS-system. Parallel binary operations in a field of 3 × 3 pixels are executed more rapidly on the AP 120 B than in the TAS-system, whereas the TAS can perform a threshold operation much

Table 1
Processing Time for Typical Operations for Array Processor compared to TAS-Operations

Type of operation	Time Required			
	AP 120 B		TAS	
	Pixel (μsec)	Field 256 × 256 (msec)	Pixel (μsec)	Field 256 × 256 (msec)
Binary-image				
AND	0.04	2.7	0.1	20
OR	0.04	2.7	0.1	20
EXOR	0.04	2.7	0.1	20
SHIFT	0.06	3.9	0.1	20
THRESH	1.50	98.3	0.1	20
MAGR	2.00	131.1	0.1	20
SHRINK (3 × 3)	0.16	10.0	0.1	20
BLOW (3 × 3)	0.16	10.0	0.1	20
BITE	0.08	5.3	0.1	20
SHRINK (8 × 8)	1.20	80.0	0.1	20
Gray-level				
ADD	0.8	52.4		
SUB	0.8	52.4		
MULT	0.8	52.4		
DIV	1.7	111.4		
CONVOLUTION (3 × 3)	4.5	204.9		
GRADIENT	1.4	111.4		
LAPLACIAN	2.1	137.6		
MEDIAN	18.7	1223.0		
HISTOGRAM	1.3	85.2	100 × 0.1	2 sec

faster than the AP 120 B. The TAS system also executes SHRINK and BLOW operations in fields of 8 × 8 pixels more rapidly. Gray-level operations and linear filtering can be processed rapidly in the AP 120 B, whereas a nonlinear Median-smoothing operation needs more than 1 sec for a field of 256 × 256 pixels.

Our system uses an Aximomat microscope for TV-image pick-up with automatic focussing and a fast scanning-stage of 10 kcycle/sec. The TV-plumbicon tube is interfaced via a fast AD-converter to the AP 120 B Array-processor memory, which has a capacity of 384 kbytes. Images from 64 × 64 pixels up to 512 × 512 pixels are transmitted in one scan into the AP-memory.

The TV-image can be displayed on a TV-monitor by refreshing from the AP 120 B main data memory. The whole system is controlled by the host-computer, a Siemens 330 (Fig. 2).

The programs for the Array-processor are written in host-computer Fortran. The main-data-memory is divided into 2 pages. In page 0 the entire TV-scanned gray-level image of 512 × 512 pixels is stored in a packed mode and in page 1 eight Bit planes of 256 × 256 pixels are located, as well as a workspace of 16 K 38 Bit-words (Fig. 3). The workspace is necessary, because before processing the packed image must be unpacked.

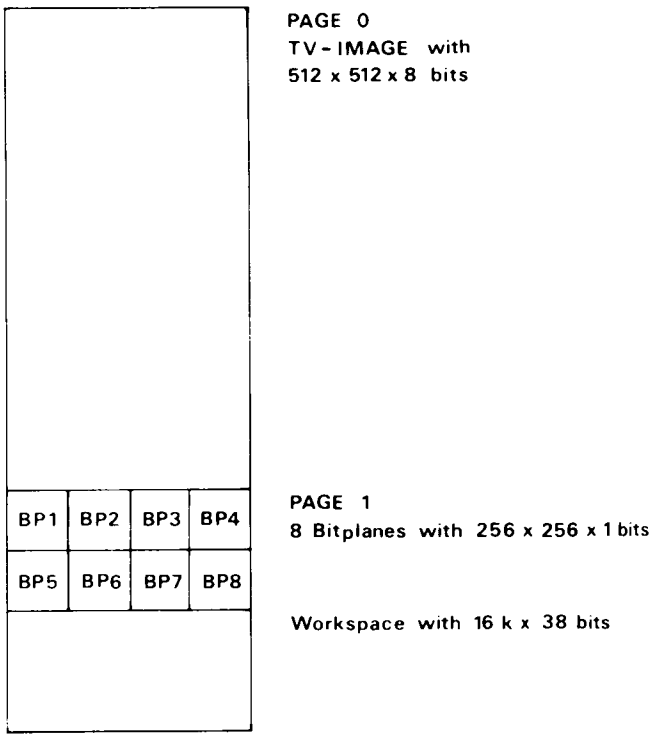

FIG. 3. Organization of the Main-Data-Memory of the AP 120 B.

Design of an AP-Program for Fast Cell Image Analysis

A program for fast cell-image analysis and sample classification using the AP 120 B will be discussed. The program has been simulated on the Siemens 330 minicomputer and part of it is implemented on the AP 120 B. In Figure 4 a flow chart of this program is shown (2).

The program is split into seven steps: (*a*) scanning of a field with shading correction; (*b*) rough segmentation of the field into single objects; (*c*) sequential extraction of single objects; (*d*) fine segmentation of single objects into nucleus and cytoplasm; (*e*) feature extraction from objects; (*f*) classification of single objects; (*g*) specimen classification.

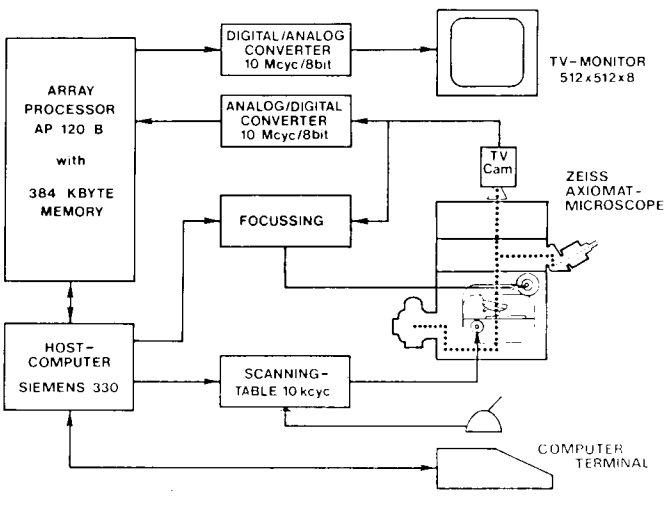

FIG. 2. System layout with the Axiomat microscope, the TV-camera and the Array-processor.

The image processing operations used are applied to gray level as well as to binary images. The principal algorithms are based on local neighbourhood operations in binary images as well as digital filtering.

A field is scanned by a TV-camera and software shading correction is performed in the AP. Rough thresholding of cytoplasm is performed by defining a threshold, depending on the standard deviation of the background. This results in a binary image, which is preprocessed by BLOW and SHRINK operations. Depending on their extent it is possible to eliminate very small and very large objects, and also, objects touching the image frame. In the next step single objects are extracted with a grassfire procedure. In this procedure, one object after another is selected and a frame around the object is determined. This is followed by the precise segmentation of cytoplasm and nucleus (5), which is performed by thresholding, and binary image operations for mask generation. Three types of features are extracted: morphologic features from the masks, photometric (cytochemical) features from the original histograms and textural features from the histograms of the filtered images. About 40 features are generated (1).

Single object classification is performed with hierarchical classifiers, based on linear discriminant analysis at each node. The number and type of features used is different for each node. The features are ranged for discrimination according to their F value. The numbers of features finally used at each node were chosen by trial and fail. It ranges from one feature to a maximum of 15 features. For single cell multiclass discriminant analysis, about 15 features have proven to be highly

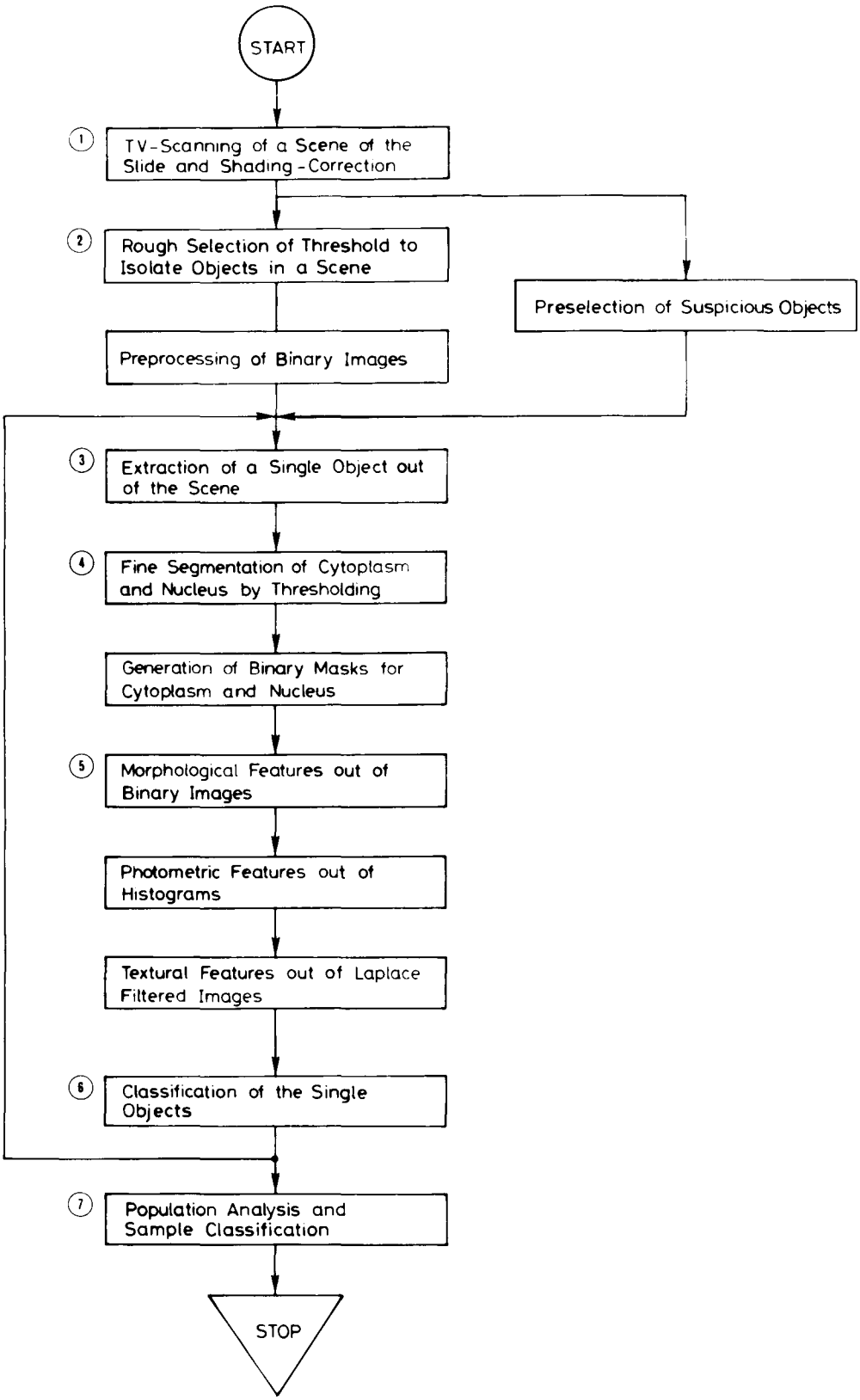

FIG. 4. Flow-chart of an AP-program for Fast Cell-Image Analysis.

specific. For this study the data base contains about 3000 cells of 12 different classes. Figure 5 shows an example of field segmentation into single cells by thresholding method and binary image processing procedures.

Time estimates for the analysis of microscopical fields (scenes) have been performed by simulating the AP system on the Siemens PR 330. The time references in reference 3 could be established because the AP is a completely synchronous machine. Results are listed in Table 2. The most time consuming step in the whole procedure is the extraction of single objects from a scene. The time required for this step can be reduced by using more-efficient methods than the grassfire procedure, or to build special hardware for this purpose. For blood cell counters special hardware is developed, which process a scene in less than 20 msec (9). In our AP-system this step takes about 77% of the computing time. The precise segmentation is also rather time consuming, especially in PAP-stained slides. The total time for processing one field containing about 20 cells takes 37 sec.

The most important step is the one before the extraction of a single object out of the scene is performed. Appropriate fast algorithms, performing field transformations, may be applied to eliminate not only very small and very large objects, such as image background, isolated leukocytes and cell agglomerates, respectively, but also other definitely unsuspicious objects. Such a strategy is proposed in reference 10 and tested on clinical material (11, 17). Similar strategies are described by Poulsen (13), and Read *et al.* (14). At the moment it is possible to reduce the number of remaining suspicious objects to less than 1%.

A further time reduction can be achieved by nucleus stain-

Table 2

Time Estimates for the Evaluation of Pap-Stained Slides with Scanning Fields of 512 × 512 Pixels and a Pixel Distance of 1 μm[a]

	Time/field in msec	%
TV-scanning and shading correction (512 × 512 pixels)	844	2.2
Rough Thresholding	327	0.8
Preprocessing of binary image (256 × 256 pixels)	614	1.3
Extraction of a single object	1,437 × 20	77.3
Fine segmentation of cytoplasm and nucleus	169 × 20	9.1
Extraction of morphologic features	2 × 20	0.0
Extraction of photometric features	74 × 20	4.0
Extraction of texture features	89 × 20	4.8
Totals	37,185	100.0

[a] to analyze all 40,000 cells on a slide 2,000 fields have to be processed (20 cells/field). The whole time, required, is about 20 hr.

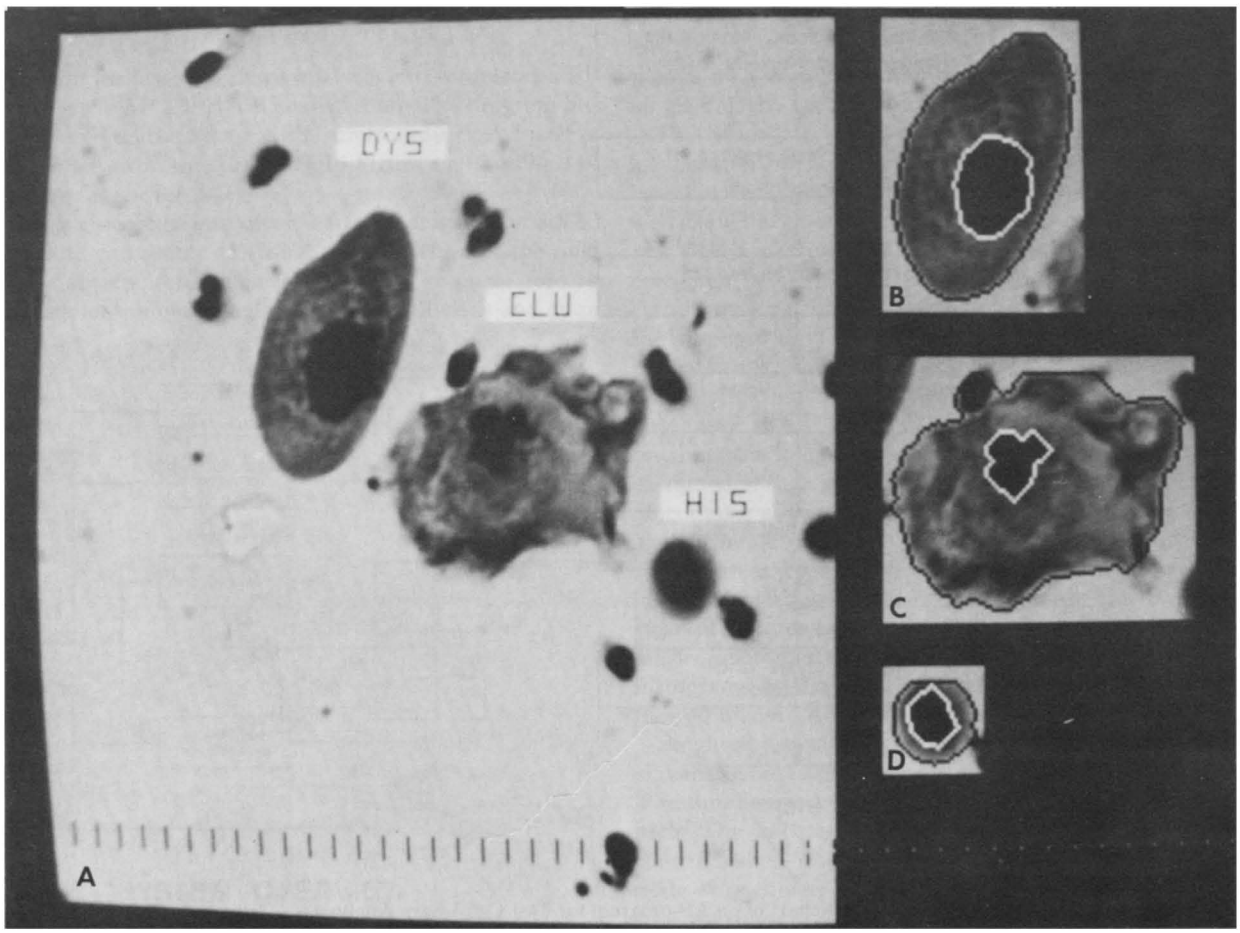

FIG. 5. Segmentation of a TV-scanned field of cells by means of thresholding and binary image processing procedures. *A*, complete scene; *B*, dysplasia cell; *C*, cluster; *D*, histiocyte.

ing techniques proposed by Tanke *et al.* (16). Using this technique, only ¼ of the processing time is required. The extraction of single objects and the segmentation is also much easier. In this case about 500 msec are required to evaluate one cell.

Conclusion

The AP 120 B is a commercial Array-processor which is designed for signal processing operations such as the fast fourier transform, correlation, etc. The architecture is not optimally designed for image processing in the space domain because the access to single pixels is realized by an unpack and pack procedure. After the unpacking of a pixel in one word of the 38 bit Floating Point representation, a pixel operation can be performed. But the AP allows easy program developing and testing by using a simulator and debugger. A Fortran compiler is available.

The time estimates we have made show that it is possible to process one cell in about 1.75 sec, provided all cells are processed sequentially and PAP-staining is used. If only the nucleus is stained and measured, the time is reduced to about 500 msec for 1 cell. These processing times lead to a machine that classifies slides with sample sizes of 40,000 cells in 2000 fields, scanned at 1 μm resolution, in about one day or in case of nuclear image analysis in 6 hr.

The registration of each object in the sample seems necessary in the training phase to develop suitable transformation strategies for cell preselection. The realization of such strategies lead to a machine where only 200 to 1000 cells out of 40,000 have to be processed, depending on the elimination threshold. This kind of system can perform a slide classification in 1.0–4.0 hr.

Under these conditions it seems possible to perform a clinical trial for validation of the whole cell analysis system using the array processor. The next generation of array processors is expected to have higher processing speeds by a factor of 5 to 10. This would then permit the construction of devices for economic and versatile routine application in automatic cytology.

Acknowledgment

The authors wish to express their thanks to Dr. E. Maier from Floating Point Systems GmbH, Munich for his support in Array-processor programming.

Literature Cited

1. Abmayr W, Borst H, Schwarzkopf G, Gais P, Rodenacker K: A feature-selection system for conventional Pap-stained cell images. The automation of cancer cytology and cell image analysis. In: Tutorials of Cytology, Pressman NJ, Wied GL (eds.). Chicago, 1979, pp. 259–268
2. Abmayr W, Burger G, Soost HJ: Progress report of the TUDAB project for automated cancer cell detection. J Histochem Cytochem 27: 604, 1979
3. AP Math Library Manual, vol 2 860-7288-004 and Processor Handbook 860-7259-002 of Floating Point Systems, Inc
4. Bartels PH, Wied GL: Computer analysis and biomedical interpretation of microscopic images: current problems and future directions. Proceeding of the IEEE, 65: 252, 1977
5. Borst H, Abmayr W, Gais P: A thresholding method for automatic cell image segmentation. J Histochem Cytochem 27: 180, 1979
6. Cordella L, Duff MJB, Levialdi S: Comparing sequential and parallel processing of pictures. In: Proceedings of the Third International Joint Conference on Pattern Recognition, 1976, pp. 728–733
7. Danielson PE, Kruse B: PICAP II—a second generation picture processing system. V. International Congress for Stereology, Salzburg, 1979
8. Duff MJB: CLIP 4: a large scale integrated circuit array parallel processor. Proceedings of the Third International Joint Conference on Pattern Recognition, 1976, pp. 728–733
9. Green JE: Rapid analysis of hematology image data, the ADC 500 preprocessor. J Histochem Cytochem 27: 174, 1979
10. Meyer F: Iterative image transformation for an automatic screening of cervical smears. J Histochem Cytochem 27: 126, 1979
11. Meyer F, Van Driel-Kulker AMJ: Automatic screening of Papanicolaou stained cervical smears using the TAS. Acta Microscopica (Suppl 4): 82, 1980
12. Nawrath R, Serra J: Quantitative image analysis: theory and instrumentation. Microscopica Acta 82: 101, 1979
13. Poulsen RS: Automated Prescreening of Cervical Cytology Specimens. Ph.D. Thesis, Mc Gill University, Montreal 1973
14. Read JS, Borovec RT, Bartels PH, Bibbo M, Puls JH, Reale FR, Taylor J, Wied GL: A fast image processor for locating cell nuclei in uterine specimens. The automation of cancer cytology and cell image analysis. In: Tutorials of Cytology, Pressman NJ, Wied GL (eds.). Chicago, 1979, pp. 143–155
15. Reinhardt ER, Erhardt R, Schwarzmann D, Bloss WH, Ott R: Structure analysis and classification of cervical cells using a processing system based on TV. Anal Quant Cytol 2: 143, 1979
16. Tanke HJ, Van Ingen EM, Ploem JS: Acriflavine-Feulgen stilbene staining: a procedure for automated cervical cytology with a television based system (LEYTAS). J Histochem Cytochem 27: 84, 1979
17. Van Driel-Kulker AMI, Meyer F, Ploem JS: Automated analysis of cervical specimens using the TAS. Acta Microscopica (Suppl 4): 73, 1980