

## XII. COGNITIVE INFORMATION PROCESSING\*

### Academic and Research Staff

Prof. M. Eden	Prof. D. E. Troxel	Dr. J. I. Makhoul
Prof. J. Allen	Prof. I. T. Young	Dr. D. M. Ozonoff
Prof. B. A. Blesser	Dr. R. R. Archer	Dr. O. J. Tretiak
Prof. T. S. Huang	Dr. G. H. Granlund	F. X. Carroll
Prof. F. F. Lee	Dr. E. G. Guttman	Deborah A. Finkel
Prof. S. J. Mason	Dr. K. R. Ingham	E. R. Jensen
Prof. W. F. Schreiber		J. S. Ventura

### Graduate Students

J. Bonn	M. Hubelbank	R. J. Shillman
W. T. Borroz	F. H. Ives	D. G. Sitler
J. E. Bowie	J. W. Klovstad	J. R. Sloan
B. E. Boyle	H. S. Magnuski	A. A. Smith
P. Coueignoux	P. L. Miller	R. D. Solomon
J. R. Ellis, Jr.	B. Minkow	J. H. Stahler
I. S. Englander	O. R. Mitchell, Jr.	W. D. Stratton
A. M. Fakhr	J. A. Myers	H-m. D. Toong
A. E. Filip	D. O'Shaughnessy	K. P. Wacks
S. G. Finn	R. A. Piankian	H. M. Wolfson
D. W. Hartman	R. S. Putnam	K. H. Yung
	R. San Antonio	

#### A. RECOGNITION AND ANALYSIS OF BIOLOGICAL OBJECTS BY COMPUTER

This report describes the latest applications of the modular picture processing package (MP<sup>3</sup>) developed in our group. This package was originally prepared by J. E. Green and O. J. Tretiak, and further developed by J. E. Green. In the past year certain modifications have been made that allow greater flexibility in MP<sup>3</sup>. At present, the package is used for analysis of erythrocytes (red blood cells) both normal and those infected with malarial parasites. Projects are under way to adapt MP<sup>3</sup> to recognize and analyze chromosomes and reticulocytes. The software was written for implementation on a PDP-9 computer with 32K core memory and 10<sup>6</sup> word disk memory.

Functions performed by this package can be divided into three main categories.

##### 1. Receiving and Storing Digitized Pictures

Each picture is sampled on a rectangular array of points by a mechanism that will be described. The brightness corresponding to each point is stored as a 6-bit number, packed 3 per word in the PDP-9 computer. These brightness values are stored as a one-dimensional array.<sup>1</sup> Resolution and size of the picture can be varied as desired. The only factor that limits the resolution is the scanning device and the only factor

---

\*This work was supported by the National Institutes of Health (Grant 5 PO1 GM14940-06).

## (XII. COGNITIVE INFORMATION PROCESSING)

limiting the size is the core memory required to store and analyze the picture. For example, on the PDP-9 computer, a picture no larger than 150 columns  $\times$  190 rows can be processed.

Two input sources are available to the computer (Fig. XII-1).

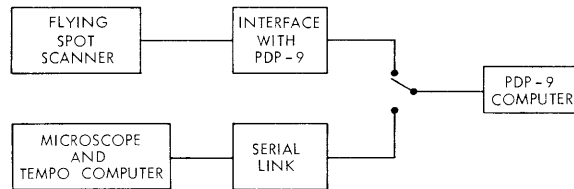


Fig. XII-1. Input sources available to the PDP-9 computer.

### (a) Flying-Spot Transparency Scanner (System I)

System I is a high-resolution flying-spot scanner designed and built in this laboratory.<sup>2</sup> This system is not directly linked to the microscope, and for this reason it is necessary to take a photomicrograph of several microscopic fields and scan these photographs on the transparency scanner. The photographic film is Kodak Panatomic X. Exposures are chosen so that the linear region of the dose-response (H-D) curves is used. After scanning the transparency (negative), the total digitized brightness is sent to the computer through an interface.

### (b) Microscope and Tempo Computer (System II)

This system employs several different devices to scan pictures directly from a microscopic field. A block diagram of the system is shown in Fig. XII-2. The software written for implementation on the Tempo computer is capable of scanning the desired microscopic field with a given resolution at the required magnification. A powerful feature of this system is the capability of the scanner to average each picture column an arbitrary number of times which can be specified by the user. The effect of such averaging is to reduce the noise in the scanning process, if the noise is zero mean and additive.

After each picture is scanned, it can be sent to either of the available PDP-9 computers (BIP or FRM). At this point the digitized picture can either be recorded on a Dec-tape and/or analyzed as it is received.

The advantage of System II over the System I is that scanning is done directly and not through recording microscopic fields on a transparency, and hence the time that would be required for photographing and processing each field is saved. Also, there is no restriction on exposing the film within the linear region, and other sources of

inaccuracy in photography can be removed. A drawback of this system is that there is a loss in reproducibility, since it is rather difficult to return to scan a given field for the second time after that field has been varied.

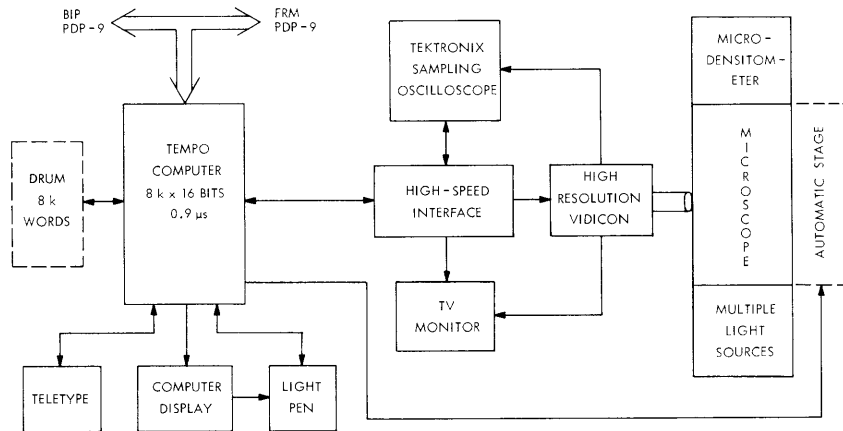


Fig. XII-2. Television-microscope scanning system.

## 2. Locating Objects in the Picture

Since objects are scanned in a rather uniform background, the brightness of a background point is markedly different from that of an object point. Algorithms have been developed to locate the objects in the picture. First, a brightness histogram of all points in the picture is determined, then by appropriate methods<sup>3</sup> thresholds are determined to separate the objects from the background. All cells touching the border are removed from the picture. A systematic search for a point whose corresponding brightness differs from that of the background brightness then begins. Upon encountering the first such point, a contour of the object to which that point belongs is formed. In this way we separate all of the objects from the background.

## 3. Analysis of the Objects

Since different objects, such as red blood cells, chromosomes, and so forth vary in shape and size, MP<sup>3</sup> must be adapted to each class of objects separately. We have used MP<sup>3</sup> to analyze erythrocytes more than for any other kind of cell or object. For this reason, in the rest of this report we shall explain the functions of MP<sup>3</sup> in the analysis of red blood cells.

Initially, after locating a cell, a refinement of the cell is made by finding the brightness histogram of the background in the immediate vicinity of the cell and redefining the threshold values. This process provides improved object definition in local regions of the image.<sup>3</sup>

## (XII. COGNITIVE INFORMATION PROCESSING)

Now the cell is ready for analysis and we are able to extract some of the characteristics of the cell (called features) by automatic processing. The most important features that we can now obtain are the following.

a. Mass [summed optical density (SOD)] obtained by summing the optical density at each point over all of the points in the picture. Optical density corresponding to the  $i^{\text{th}}$  point is defined as

$$(\text{O.D.})_i = \ln \frac{I_o}{I_t},$$

where  $I_t$  is the intensity corresponding to point  $i$ , and  $I_o$  is the background intensity obtained when the local neighborhood of the cell is histogrammed. Furthermore,

$$\text{S. O. D.} = \sum_{\text{all } i} (\text{O. D.})_i.$$

b. Area of the cell is obtained by counting the total number of picture points on and inside the contour specifying the cell.

c. Perimeter of the cell is calculated by counting the total number of points on the contour of the cell.

d. Mass/area is the ratio of the quantity in (a) divided by the quantity in (c). It is a measure of the mass concentration.

e. Perimeter is a measure of the circularity of the cell.

$$2\sqrt{\text{Area}/\pi}$$

For a circle it is equal to  $\pi$  and for any other geometrical shape it is greater than  $\pi$ .

f. Area of the central pallor (if any) is a feature representing a measure of the size of the central pallor. It is obtained by defining a threshold for the brightness of the central pallor and dividing it from the rest of the cell for analysis.

g. Area of inclusions measures such features as Howell-Jolly bodies, Cabot rings, malarial parasites, and nucleus residue.<sup>4</sup>

Feature (a) contains information about the hemoglobin content of red blood cells. In fact, it is possible to show theoretically that the amount of light absorption by the red cells is linearly proportional to the mean corpuscular hemoglobin (MCH) content of the cells.<sup>5</sup> We have conducted a series of experiments using the flying-spot scanner to verify such a relationship. We are now performing the same set of experiments using System II.

A. M. Fakhr

### References

1. J. E. Green and O. J. Tretiak, "Modular Picture Processing Package (MP<sup>3</sup>)," Quarterly Progress Report No. 94, Research Laboratory of Electronics, M. I. T., July 15, 1969, pp. 261-279.

(XII. COGNITIVE INFORMATION PROCESSING)

2. W. F. Schreiber, "The New Scanner" (unpublished Manual, Cognitive Information Processing Group, Research Laboratory of Electronics, M. I. T., Cambridge, Mass., 1968).
3. J. E. Green, "Computer Erythrocyte Analysis," Proc. IEEE Conference on Feature Extraction and Selection in Pattern Recognition, Argonne, Illinois, 1970.
4. B. S. Leavell and O. A. Thorup, Jr., Fundamentals of Clinical Hematology (W. B. Saunders Company, Philadelphia, Penn., 2d ed., 1966).
5. E. G. Guttmann, "Mechanization of the Interpretation of Vaginal Smear," Ph.D. Thesis, Department of Electrical Engineering, M. I. T., October 1970.

