A SYSTEM FOR ACQUIRING, EDITING AND ANALYZING OPTICAL MICROSCOPE DATA

F.G. Peet and T.S. Sahota
Pacific Forest Research Centre, 506 West Burnside Road, Victoria, British Columbia, Canada V8Z 1M5

ABSTRACT

A system for acquiring, editing, and analyzing optical microscope cell data has been developed. The system has been applied to both insect and plant cell material. The biological results which have been obtained have been significant both in a statistical sense and a biological sense. Below the system is described in terms of its data acquisition, data editing, and data analysis components.

RESUME

On a developé un système pour l'acquisition, la rédaction et l'analyse des données d'un microscope optique. On a appliqué le système aux cellules des insectes et des plantes. Les résultats biologiques, qu'on a obtenu, sont significatives. Le système est décrit dans le texte.

KEYWORDS: digital microscopy, biological image processing, microphotometry

DATA ACQUISITION

The hardware for data acquisition consists of a Zeiss SMPO5 Scanning Microphotometer, a Digital Equipment Corporation(DEC) PDP 11/34 minicomputer with 128K(K=1024) words of memory, a magnetic tape drive, two 88-megabyte disc drives, two 2.5-megabyte disc drives, seven video terminals and two hard-copy terminals. Biological material is placed on a stage. The stage has two scanning motors each with one-half micron step size. These motors are pulsed to generate the scan in the x and y directions. At each step light passes through a small area (pixel) of the specimen to a photo-detector where an analogue output signal is generated proportional to the incident light intensity. This signal is sampled and digitized twenty-nine times and then averaged. The stage is then stepped and the procedure repeats until the whole specimen has been scanned. Each day a check is made of the system's stability and calibration. The computer program which actually acquires the data comprises four sections:

1) a section to measure incident light
2) a section to position the specimen
3) a section to incorporate annotation data
4) a section to acquire the data

DATA EDITING

We are interested only in cell nuclei. Although there are algorithms for finding the nucleus within the cell they are somewhat unreliable (9). We have developed a man-computer interactive procedure for doing this. The hardware consists of a television type display device connected to the computer described earlier. The cell data which has been acquired is displayed on the television monitor. A cursor is moved around the nucleus and a line is traced. When the line is closed the computer program sets all pixels outside of the nucleus to -1 and leaves the nucleus intact. The edited data is then written out to disc as a new file and just before doing so the pixel intensities in the nucleus are converted to absorbances.

DATA ANALYSIS

For each cell sixty-three features are determined. They are:

1) total absorbance
2) average absorbance
3) standard deviation
4) moment of inertia
5) area
6) entropy
7) polar averages (eleven features)
8) histogram of absorbances (twenty features)
9) and transition probabilities (twenty-four features), and
10) condensed nuclear material ratios (two features).

These terms have been defined elsewhere (7).

The best twenty-five of these sixty-three features are then selected by the following procedures. For each feature the values of the ambiguity function, A, and the receiver operating characteristic d', are computed(1). An intermediate value is calculated for each feature which is an average of A and d'. The features are sorted on the basis of their intermediate value and the twenty-five features with the largest intermediate value are selected as being the best. The correlation matrix for these twenty-five is computed. The average correlation, $\bar{r}$, of a feature with the remaining twenty-four is evaluated. Another value is then computed which is a figure-of-merit and is an average of all three variables (A, d', and $\bar{r}$). The best twenty-five features are then sorted on their figure-of-merit values to give an ordered list of the best features.

Once the best features for discrimination are selected, the question to be asked is whether two cell populations are different. We answer this question by first of all applying a normality test(5) and Box's M test(3) to check for equality of the covariance matrices. If the data are normally distributed, and the covariance matrices are equal, Wilks' Lambda test(3) and Hotelling's T^2 test (3) can be applied to determine if the populations are significantly different. If Box's M test fails then the one-sample Hotelling's T^2 test (6) may be applied. It requires only that the data be normally distributed. If the data are not normally distributed and/or the covariance matrices are not equal, one may apply a combination of the Fisher discriminant function and the non-parametric Kruskal-Wallis test(1,4,8).

If it is found that the populations are different, then one may try to determine how well the individual cells in the two populations may be classified. Four classifiers have been implemented. There is one supervised, parametric classifier based on the maximum-likelihood procedure(2). There are two supervised non-parametric classifiers. One is the DSELECT algorithm of Bartels(2), and the other is the Fisher discriminant procedure(4). One unsupervised classifier is available and it is the Basic Isodata algorithm(4).

The system has been applied to different biological systems and differences are being found between populations which were surmised but could not be proven with other techniques.

ACKNOWLEDGEMENTS

The authors wish to thank Prof. P.H. Bartels for his generous assistance with the analysis methods and programs.

Mr. J. Reischmiller designed and built the interface between the microscope and the computer.

Dr. J. Pannekoek, Mr. J. Partridge, Mr. D. Say, and Mr. M. Kendrick assisted with computational aspects of the project. Mr. A. Ibaraki and Mr. S. Farris prepared the biological material.

REFERENCES

1. Bartels PH: Numerical Evaluation of Cytologic Data, III. Selection of Features for Discrimination. Analy Quant Cytol 1(3):153-159, 1979

2. Bartels PH, Olson GB: Computer Analysis of Lymphocyte Images. In Catsimpoolas N: Methods of Cell Separation, Volume 3, New York, Plenum Press, 1980, pp1-67.

3. Bartels PH: Numerical Evaluation of Cytologic Data, VII. Multivariate Significance Tests, Analy Quant Cytol 3(1):1-8, 1981

4. Duda RO, Hart PE: Pattern Classification and Scene Analysis. New York, John Wiley and Sons, 1973, pp114-228

5. Mardia KV, Kent JT, Bibby JM: Multivariate Analysis, London, Academic Press, 1979, pp 20-21, 148-149

6. Morrison DF: Multivariate Statistical Methods. New York, McGraw-Hill, 1967, pp 120

7. Peet FG, Sahota TS: A Computer Assisted Cell Identification System. Manuscript in Preparation.

8. Peet FG, Sahota TS: A Non-Parametric Approach to the Discrimination of Two Cell Populations. Manuscript in Preparation.

9. Taylor J, Puls J, Sychra JJ, Bartels PH, Bibbo M, Wied GL: A System for Scanning Biological Cells in Three Colors, Acta Cytologica 22(1):29-35, 1978