MEDINFO-83, van Bemmel/Ball/Wigertz, Editors

ANALYSIS OF ELECTROPHORESIS GELS BY AN IMAGE PROCESSING SYSTEM

U. Engelmann and H.P. Meinzer

German Cancer Research Center,
Inst. of Documentation, Information & Statistics, Im Neuenheimer Feld 280, D-6900 Heidelberg, F.R.G.

Considerable information relating to the molecular weights and purity of samples can be obtained in medical and biological research using either polyacrylamide PAGE or agarose gel electrophoresis AGE. Gel images can be analysed by the computer using image processing methods. PIC, such an image processing system, developed at the German Cancer Research Center, was used to tackle this problem. The result is a high modular, flexible, easy-to-use system.

1. INTRODUCTION

Autoradiograms and photographs are used as a permanent record of PAGE or AGE in medical or biological laboratories. These autoradiographs can be used to estimate the molecular weights of polypeptides or DNA fragments and thereby to order the fragments. Presentally such estimations are performed manually which is both time consuming and inexact.

In cooperation with the Institute of Virus Research we at the Institute of Documentation, Information and Statistics (both at the DKFZ) are trying to adapt the computer for the evaluation of pictorial information.

We have first developed a general software tool for image processing problems (3). Using this as a basis we have examined specific problems relating to the interpretation of results visualised as either autoradiographs or photographs (4).

2. MATERIAL

One of the main areas of research at the Institute of Virus Research is virus DNA. In this context the production of many gels is necessary. With the help of these gels the molecular weights of DNA fragments generated by various restriction endonucleases can be estimated and their relative order within a genome determined.

Such gels conform in their construction and method of evaluation. They all consist of several tracks. Tracks run from the top to the bottom, are non everlapping and contain bands which extend from the left to the right edge of the track. Bands indicate the presence of DNA material at the position where they are observed. The relative mobility of a DNA fragment provides information as to the molecular weight of the fragment. Comparison of several tracks yields further informations,e.g. the DNA sequence or the location of a DNA fragment within the genome.

An image processing system for the analysis of gel images should be able to localise tracks and bands of a gel image. Localisation having been achieved, further evaluations should be possible (determination of the sequence or molecular weights).

3. METHODS

3.1 A BASIC SOFTWARE TOOL: THE INTERPRETER PIC

In image processing there is a wide field of basic operations which are quite general (2,5,10,11). Thus, in our institute we developed the computer language PIC for basic image processing operations (3,8,9). The advantage of an interactive language as opposed to other solutions, such as subroutine packages or menu-driven program systems has been discussed earlier (4,7,8,9).

The language, PIC, consists of approximately 90 commands subdivided into nine groups: dataset management functions, display functions, simple image manipulations, histogram modifications, gradient operators, smoothing algorithms, Fourier operations, arithmetic and Boolean operations and auxiliary functions. Useful special features include: abbreviations of most of the commands and the processing modes batch and conversationally; through the use of a special input processor, which was developed in our institute, the building of common or private abbreviations, synonyms and procedures consisting of PIC commands is supported (12); the most important "special" feature is a user interface via command

CALL Subroutine.

With this latter command user-written FORTRAN subroutines can be dynamically loaded at runtime and executed. Therefore PIC can be enlarged by the user himself for his own specific problems. It is also useful for testing new routines of PIC without changing the grammar of the language.

The analysis of gel images is based on this software tool. Thus, we could discard basic operations such as LOAD, STORE or DISPLAY and image preprocessing operations like SCALE, WINDOW or FILTERING. We only had to create solutions for the specific problem.

3.2 SPECIAL TOOLS FOR THE ANALYSIS OF GEL IMAGES

The main image processing problem is the segmentation of the image into tracks and bands. The software for solving these problems was developed and tested via the command CALL Subroutine. Later new commands were created and integrated into PIC. Thus, the interpreter has a new domain for the analysis of gel images. The new commands will be introduced and the semantic to which the commands refer will be described in more detail.

For reading the commands it must be explained that all expressions in brackets can be excluded. The sign "|" is the logical OR function. All commands can be abbreviated. For clarification the long form has been used.

After preprocessing an image with PIC it must be segmented into tracks. This can be achieved by the command

```
FIND TRACK (MANUAL)
```

In the first mode (when MANUAL is excluded) the operation is performed fully automatically. Tracks are calculated with the help of the standard deviation of the gray levels in the columns of the image. Columns, which contain bands have a higher standard deviation as compared to those without bands. Thus, the column standard deviation histogram can be calculated. On this a variable thresholding operation is performed (ie. the threshold is newly determined for each picture). For every possible threshold from the minimum to the mean value of standard deviation, the number of resulting tracks is calculated. The most frequent resulting number of tracks indicates the point of optimal thresholding.

Using this method normally results in the correct segmentation of the image. However occationally incorrect segmentation occurs and it is then necessary to use the manual mode. For this a graphic terminal is required. After the column projection of the image is displayed on the screen by a line-drawing algorithm, the user can enter the regions of tracks into the line-drawing with the help of a cursor.

Segmentation of the tracks yields a track array where boundaries of tracks are stored.

The results of track segmentation can be presented with the commands

```
DISPLAY TRACKS

PRINT TRACKS (ON (WHITE) PAPER)
```

With the first command two diagrams will be plotted on the screen. The first diagram shows the column projection and the second the column standard deviation histogram. In both diagrams the track boundaries are drawn in. Additionally, the optimal threshold is shown in the standard deviation histogram, so that the segmentation can be verified visually. The second command lists the track boundaries (from the track array) on the screen or prints them on paper (on white paper, if so desired).

With the command

```
ELIMINATE BACKGROUND
```

all columns outside the tracks can be set to a negative gray level. If the image is displayed with one of the display routines of PIC, only track columns will be visible.

The next step is the localisation of bands within the tracks. For this the algorithms take advantage of some a priori knowledge concerning the structure of the images:

- Bands extend from one side of the track to the other.
- The vertical maximums of a band build a connected path from the left to the right hand boundary of the track.

Using this knowledge we have the command

```
FIND PEAKS VERTICAL
```

which calculates the relative maximums in the columns.

From the set of maximum pixels, paths from the left to the right hand boundary of the tracks can be determined with the command

```
FIND PATH
```

This calls a backtracking algorithm (2,11), which examines every maximum pixel on the left boundary position in order to establish whether there is a path to the other boundary. Only the maximum pixels to the immediate right, upper right or lower right of the point concerned are considered in order to establish this path. During the search of the path the actual path is

stored in a stack where all possible alternatives to a path pixel are also registered. If the algorithm runs into a deadlock, a pop-operation is performed on the stack and an other path is examined. If no further pop-operation is possible (stack empty), no path exists from the actual starting point, whereas in the other case a path is found which marks the maximum line of the band.

This command (ie. the hidden algorithm respectively) yields satisfactory results in the form of an object array which contains track and row positions.

Results of operations up to this step can be presented in various ways. For visual presentation and verification, common display routines of PIC can be used where only band pixels (the path) are visible. All other pixels are set to a negative gray level. A list of all band positions and the track, to which they belong, can be displayed on the screen or printed out with the command

PRINT BAND (ON (WHITE) PAPER).

As tracks are often shifted relative to each other we introduced a command

CORRECT TRACKS.

Thus it is possible to align tracks. This is very useful for further evaluations.

For the estimation of molecular weights of fragments it is necessary to enter first known molecular weights (electrophoresed in a reference track) by a command

ATTACH MOLECULAR WEIGHTS.

Unknown molecular weights can then be interpolated by

INTERPOLATE WEIGHTS.

Results can be printed out with

PRINT MOLECULAR WEIGHTS (ON (WHITE) PAPER).

One byproduct is the command

PLOT TRACK DENSITIES (MEDIAN|MEAN|MIDDLE).

This plots a diagram of the optical densities of the tracks as produced by densitometers. If the keyword MEAN or MEDIAN is entered the mean or median value of the track row is used. Alternatively the pixel values of the middle of the track are used.

For routine application it is of course not necessary to type in all commands listed above. The high modularity is very useful for the training of users of the system and for maintenance. In routine application many commands are included in one powerful procedure.

4. IMPLEMENTATION

This software tool runs on an IBM 3032 with operating system TSS (time sharing system). For the graphic output and display of images a TEKTRONIX 4014 or 4112 is needed. The software is written in FORTRAN IV and needs approximately 10 Mbyte main storage. This allows to hold three integer images of a size up to 900 per 900 pixels and four real images up to the size of 256 per 256 pixels simultaneously in the main storage. The analysis of input string (commands) is done by a parser which was generated by the parser and lexical analyser generating system PAULA, developed in our institute (6). The building of synonyms and procedures is supported by a self-developed input processor (12). For the graphic output we use the TECTRONIX packages TCS (terminal control system) and AG II (advanced graphic two) (6).

5. CONCLUSION

Experience with the command language PIC shows that the user can understand and work with it very easily. The commands for the analysis of gel images are very simple. As the analysis operations are modular they can easily be maintenanced. Collections of several commands in a powerful procedure has improved the acceptance considerably. The system can be used by non-computer-oriented personnel. Image analysis by computer can partly yield better results than manual evaluation and relieve the researcher of this time consuming work.

6. ACKNOWLEDGEMENTS

We would like to thank Gerhard Zinser for digitalisation of images and the colleagues in our institute who helped us using the parser and lexical analyser generators and the input processor. Special thanks go to Chris Gray from the Institute of Virus Research for advising us in the biological-chemical coherencies and for his cooperation.

7. REFERENCES

(1) BECKER, N., OSTERBURG, G., SCHADEWALDT, K., PAULA - Generator fuer LL(1)-Parser und lexikalische Analyseprogramme, DKFZ, Heidelberg, Technical Report No. 10 (1977).

(2) CASTLEMAN, K.R., Digital Image Processing. (Prentice Hall, Englewood Cliffs, 1979).

(3) ENGELMANN, U., MEINZER, H.P., PIC - Ein Interpreter zur Bildbearbeitung, DKFZ, Heidelberg, Technical Report No. 25 (1982).

(4) ENGELMANN, U., Entwicklung eines Interpreters zur Analyse von Gelbildern mit Methoden der Bildverarbeitung, Diplomarbeit Fachbereich Medizinische Informatik, Universität Heidelberg/Fachhochschule Heilbronn (1983).

(5) GONZALEZ, R.C., WINTZ, P., Digital Image Processing (Addison-Wesley-Publishing Comp., Massachusetts, 1977).

(6) HAHNE, H., TEKTRONIX: TCS und AG-II, DKFZ, Heidelberg, Technical Report No. 8 (1976).

(7) MEINZER, H.P., Command Languages in Application Programming, in Lindberg, D.A.B., Kaihara, S. (eds.), MEDINFO 80, (North-Holland Publ. Co., Amsterdam, New York, Oxford, 1980).

(8) MEINZER, H.P., An Interpreter for Matrix Graphics, in Moore, R.R., Barber, B., Reichertz, P.L., Roger, F. (eds.), Medical Informatics Europe 82 (Springer, Berlin, Heidelberg, New York, 1982).

(9) MEINZER, H.P., ENGELMANN, U., A Language for the Interactive Manipulation of Digitized Images, in Lang, M. (ed.), 6th International Conference on Pattern Recognition, Munich, 1982 (IEEE, Computer Society Press, Silver Spring, MD, 1982).

(10) PRATT, W.K., Digital Image Processing (John Wiley Sons, New York, Chichester, Brisbane, Toronto, 1978).

(11) ROSENFELD, A., KAK, A.C., Digital Picture Processing (Academic Press, New York, 1976).

(12) SCHADEWALD, K., MERX, R., KYNAST, W., Der Input-Processor, DKFZ, Heidelberg, Technical Report No. 18 (1980).

MEDINFO 83

Proceedings of the Fourth World Conference
on Medical Informatics
Amsterdam, August 22-27, 1983

Edited by

JAN H. VAN BEMMEL

Department of Medical Informatics
Faculty of Medicine
Free University
Amsterdam

MARION J. BALL

Computer Systems and Management Group
Temple University
Philadelphia
Pennsylvania

and

OVE WIGERTZ

Department of Medical Informatics
Linkøping University
Linkøping
Sweden

PARTICIPANTS EDITION

Part 1

1983

NORTH-HOLLAND - AMSTERDAM · NEW YORK · OXFORD