COMPUTERS AND BIOMEDICAL RESEARCH 14, 272–297 (1981)

GELLAB: A Computer System for 2D Gel Electrophoresis Analysis I. Segmentation of Spots and System Preliminaries

P. F. LEMKIN AND L. E. LIPKIN

Image Processing Section, Division of Cancer Biology and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20205

Received September 8, 1980

GELLAB is a computerized analysis system for two-dimensional gel electrophoresis images. It is particularly concerned with tools for the automation of comparisons among multiple gels rather than simple gel-pair analyses. Interrogation of and experimentation with the spot data base may then be performed in order to extract measurements on particular spots within the set of gels. In the case of gels partitioned into multiple classes, it is possible to search for spots in the spot data base having statistically significant differences between classes. This first paper deals with: (1) some gel properties which bear on image acquisition and image analysis, (2) the broad image acquisition environment, beyond mere scanning, such as auxiliary data, standarization and calibration, and (3) segmentation as a means of spot extraction and characterization. Our segmentation algorithm is based on problem domain heuristics and particularly on properties of the second derivative. It is therefore largely shape and density independent. The segmenter, as the other GELLAB programs, is written in SAIL and runs on either a DECSYSTEM-10 or -20.

1. INTRODUCTION

This is the first of a series of three papers dealing with computer aided and fully automated analyses of two-dimensional electrophoretic patterns. The 2D polyacrylamide gel electrophoresis (PAGE) technique is a rapidly developing biochemical tool, applicable to a wide variety of problems in basic biochemistry, molecular biology, genetics and clinical research (*1*).

The PAGE technique may separate hundreds to a thousand or more components as a matrix of spots because the variables which determine electrophoretic mobility in each of the two dimensions are effectively orthogonal to each other. Extent of movement in the first dimension is determined by isoelectric focusing over a pH gradient while in the second dimension the sodium dodecyl sulfate (SDS) interaction with protein moieties results in a mobility which is a function of molecular weight.

The perhaps two decimal orders of magnitude increase (over 1D techniques) in the number of polypeptide and/or protein fragments ("spots") detectable in a mixture (*2*) is a major cause for efforts at computerized analysis, detection and intergel comparisons (*1, 3–7*). In addition, there has been an increase in complexity of output accompanying this great increase in discriminatory power

0010-4809/81/030272-26$02.00/0

that strains the limits of unaided human analytical ability. As we shall see, the need for analytic assistance is further increased by the non-linear spatial mapping of corresponding moieties in comparable gels.

The successive papers in this sequence will be progressively more concerned with the generation of data structures, strategies and tactics for their employment in the analyses of *sets* of gels; i.e., comparisons, both qualitative and quantitative, among multiple gels which, for example, reflect successive values of a dose or time variable in an experiment or the clinical course of a patient.

Figure 1 illustrates the steps performed in the GELLAB 2D gel analysis procedure. Gel images are first acquired (digitized and stored), then spots are segmented in each gel, spots are paired between gels and a standard gel using a small set of manually defined landmarks, and finally a spot data base is constructed and analyzed. Final output of such a system is in the form of both

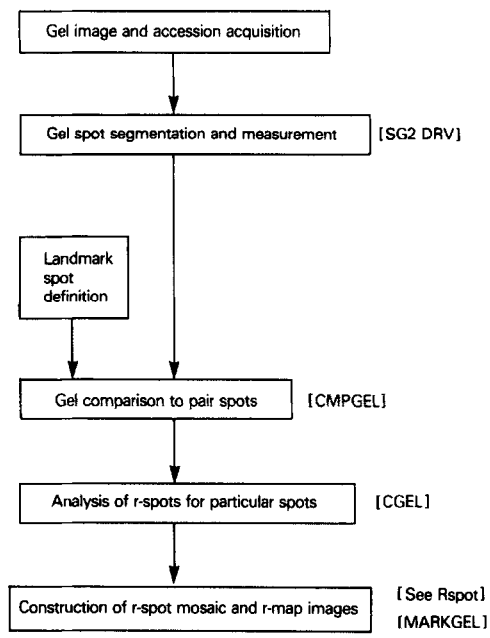

FIG. 1. Block diagram of the 2D-gel analysis GELLAB system. Programs associated with major steps of GELLAB are indicated in "[. . .]". Gel images are acquired by scanning with a vidicon TV camera interfaced to a picture memory and saved on the computer. Accession information about the set of gels is also used to update an accession file. The gel images are then segmented and measurements made of the spots which are found. Landmark spots are then manually selected which are either known proteins or well-defined spots spaced fairly evenly throughout the gel. Using gel image flicker alignment, the landmark spots are aligned for all of the gels with a representative gel (R-gel). This information and the raw segmentation data are then used to pair spots in the remaining gels with the R-gel. The set of gel pairings with the same R-gel may be merged together to form a set of sets of equivalent R-spots called the composite gel data base (CGL). Thus a R-spot (supposedly) contains the same spot from all the gels in which it occurs.

labeled gel image maps where statistically interesting spots have been marked as well as numeric spot data to support these findings.

This first paper in the series is concerned with some initial aspects of the analytic process; it briefly touches on image acquisition and digitization and concentrates on image segmentation and spot extraction. The segmentation algorithm is discussed in detail so as to facilitate implementation on other hardware systems and in differing software environments. Discussion of the nature of the gels themselves is limited to those aspects of their generation, staining (and other modes of spot detection) which bear directly on the computerized procedures for their analysis. Some remarks on the software environment in which the overall system is implemented are appended. The second paper covers spot pairing using landmarks to locally align subregions (*8*). The third paper discusses construction of multiple gel data bases given sets of paired spots and the subsequent types of searches and analyses which can then be performed (*9*).

Characteristics of a Gel That Affect Processing It as an Image

For the most part, the resultant geometric position of a spot bears no relation to function or the origin of the protein it represents. Closely related proteins or polypeptides may be separated by considerable distances while functionally unrelated materials may be distributed in close proximity. In contrast to more conventional images such as X-rays or microscopic fields, the image "structure" (i.e., local adjacencies, inclusions, etc.) in gels provides little information to facilitate the analysis. Individual spots, unless overlapped by other spots or contaminated by artifacts, are much simpler images than, say, the image of a cell in a blood smear. Thus once a spot has been isolated, its analysis and characterization as an *individual* at least in a single gel, is relatively simple.

A more serious difficulty is the lack of point to point reproducibility of gels (i.e., noncongruence), even of those derived from the same sample, and even of those from a single run on the same apparatus. This is due to a large number of preparative factors, including local heterogeneities in polyacrylamide texture and/or local concentration, local temperature variations, heterogeneities of ampholine concentration, etc. All of these variables and perhaps others less well-recognized act to reduce the responsibility of mobility of fragments in one or another dimension. This results in sets of gels which are *not congruent* but which are affine in the sense of some transformation. In other words, comparable spots within a set of gels will have corresponding neighbors but will not be located at necessarily exactly the same distances from these neighbors in every or any instance. They will show a local superimpossibility, which is maintained for surrounds of varying extent. It is this absence of simple direct correspondence coupled with the large numbers of spots that makes some automated assistance a necessity.

Hardware

This has been described in (*3, 10–14*). Briefly the system consists of a DECSYSTEM-2020 controlling two Real Time Picture Processors (RTPPs). Each RTPP consists of a PDP8e acting as a display processor controller for the sixteen 8-bit gray value (256 gray values) 256 × 256 picture element (pixel) frame buffers. One of the systems has a TV camera and frame buffer hardware which enables a TV frame to be acquired in 1/10 second. The time shared 2020 processor has 384K words of memory and two large disk pack drives as well as two magtape drives.

Image Acquisition

Data acquisition is accomplished by scanning backlighted gels or gel autoradiographs with a Vidicon camera interfaced to our RTPP picture processing computer (described in (*3*) with respect to 2D gel flicker analysis). The Vidicon camera has a Nikon-N auto 1:2 28-mm objective lens routinely set to f8 with the autoradiograph film mounted 69 cm away and backlighted on an Aristo type T-12 uniformly illuminated light box (Port Washington, NY). A type 1009 NBS Neutral Density (ND) wedge is mounted at the bottom of the illuminated area. The TV camera system video amplifier is adjusted to maximize contrast over the dark end of the range of the ND wedge (with the preservation of information at the white end of the gray scale range in the image). The effective resolution of the image is about 250 μm/pixel (picture element). For use with the 120 size film, a 55-mm micro-Nikkor lens is fixed at *f*8 and set for a distance of 55.5 cm. An alternative scanning protocol positions the gel image 42 cm from the TV camera, which yields about 170 μm/pixel resolution.

The scanned images are acquired using the PIXMTA subsystem of BMON2 (*11*) and saved on magnetic tape for later analysis. The RTPP digitizes the video output to 8-bits with 256 gray values ranging from 0 (white) to 255 (black). An image consists of a 512 × 512 pixel array. The (0,0) x,y position is the upper left-hand corner and (511,511) represents the lower right-hand corner. The actual dynamic range of the gray scale data is slightly greater than 7-bits but of this probably only 6-bits of gray scale resolution is actually valid. The TV camera video amplifier is adjusted to yield the largest dynamic range such that it does not saturate "white" video input. Since the spot information will be in the darker linear range there is no problem of distortion for photographically nonsaturating spots. Fifty images are easily scanned in about one hour or less.

Although the transfer function of a Vidicon TV camera is nonlinear and has a maximum dynamic range from 0 to about 2.00 D, it may still be used to perform densitometry under some conditions. These include: (1) the majority of the material to be measured *is not* near a saturating end of the camera's dynamic range, (2) the nonlinear gray scale to OD transfer function be computed and well behaved over the range of data to be measured, and (3) all calculations

involving total OD/spot be done in the density domain. Even then, appreciable errors on very dark spots may occur due to saturation at the dark end of the dynamic range and to various optical problems having to do with glare and under-representation of dark spots.

Calibration

Using the TOTDENSITY program on the RTPP the user interactively defines the computing window (the region in the gel image where the spots are located) taking care to omit gross artifact and the ND wedge on the gel. Using this program, the ND wedge is also calibrated by computing a gray scale histogram of a 20 pixel wide computing window positioned by the user across the wedge. Peaks in the smoothed histogram are matched automatically with the actual OD values of the wedge. The wedge gray value peak information and gel computing window position are automatically updated into the accession file GEL.ID which is then transferred to the DECSYSTEM-20. A piecewise linear function can be generated for the ND wedge as a function of gray value so that all scanner output values are mapped to densities (OD). Table I shows part of a typical accession file. Figure 2 illustrates a gel with the ND wedge 2a, its computing window 2b, the ND wedge sample window 2c, and its ND wedge histogram calibration curve 2d. The corresponding piecewise linear OD calibration function is drawn (black) over the histogram.

Segmentation: Nature of the Image

Image acquisition is only the first stage in making spot information available for automated processing. In any image some provision for separating out "pictorial information of interest" (in this case the spots) from "background" and "noise" is necessary before spot positions and spot properties can be compared. Consequently, a spot extraction algorithm must be capable of (1) detecting, (2) defining the extent of, and (3) measuring the density of a spot under a wide variety of actual gel image conditions.

In the general field of image processing the segmentation problem is one of the most important and ubiquitous. Almost all real images resist simple thresholding as a solution to pictorial partitioning or segmentation and despite the simplicity of spot structure, 2D gels are no exception. The thresholding operation may be thought of as redrawing an image such that all values higher than a certain gray value are retained while all others are set to white.

Morphology

The vagueness of spot morphology and inhomogeneity of background complicate gel images. Spots frequently touch each other, resulting in overlapping spots or spots that extend for a considerable distance into an overlapping region, so their tails overlap. Spots have no distinct boundary, but occur most often as an effectively continuous Gaussian-like distribution. Lutin

TABLE I

EXAMPLE OF GEL.ID GEL ACCESSION DESCRIPTOR FILE

```
ACCESS. #/PATIENT/BIRTHDATE/RACESEX/EXP DATE/EXP #/CULTURE REAG/AMPH,GEL/
INTRVL BEFR LBLNG/LPLNG ISOTOPE/DURTN LABEL/DURTN OF EXPSR/STUDY/
FILE #/TAPE #/OPT. BACKUP TAPE #/ CAMERA,LENS,DISTANCE/EXPRMNTR*
ND:.05,.20,.35,.50,.66,.80,.95,1.10,1.25,1.41,1.56,1.72,1.87,2.02,2.17
ASBSPT.DA = ASBESTOS E-SPOT FILE
        .
        .
        .
0094.1/P388D1/24 HRS TOTAL/-/12-4-78/#A40/FISCHER'S/3:10, 10%/
0 HRS/C14/24 HRS/1 MONTH/ASBESTOS MACROPHAGE CONTROL/
L00041/R602/--NONE--/VIDICON-MAN,28MM F8,69CM/LIPKIN*
 38 65 96 120 139 156 173 186 197 206 222 0 0 0 0 0 47 441 100 374
0095.1/P388D1/24 HRS TOTAL/-/12-4-78/#A40/FISCHER'S/3:10, 10%/
0 HRS/C14/24 HRS/1 MONTH/ASBESTOS MACROPHAGE UICC AMOSITE/
L00045/R602/--NONE--/VIDICON-MAN,28MM F8,69CM/LIPKIN*
 38 63 95 120 139 156 172 185 197 206 218 0 0 0 0 0 0 77 432 106 327
0096.1/P388D1/48 HRS TOTAL/-/12-4-78/#A40/FISCHER'S/3:10, 10%/
24 HRS/C14/24 HRS/1 MONTH/ASBESTOS MACROPHAGE CONTROL/
L00049/R602/--NONE--/VIDICON-MAN,28MM F8,69CM/LIPKIN*
 39 66 95 121 140 158 173 187 198 207 222 0 0 0 0 0 0 70 449 129 386
0097.1/P388D1/48 HRS TOTAL/-/12-4-78/#A40/FISCHER'S/3:10, 10%/
24 HRS/C14/24 HRS/1 MONTH/ASBESTOS MACROPHAGE UICC AMOSITE/
L00053/R602/--NONE--/VIDICON-MAN,28MM F8,69CM/LIPKIN*
 39 67 97 121 140 159 174 187 198 207 222 0 0 0 0 0 0 53 400 146 394
0098.1/P388D1/72 HRS TOTAL/-/12-4-78/#A40/FISCHER'S/3:10, 10%/
48 HRS/C14/24 HRS/1 MONTH/ASBESTOS MACROPHAGE CONTROL/
L00057/R602/--NONE--/VIDICON-MAN,28MM F8,69CM/LIPKIN*
 39 65 95 120 139 157 174 186 197 207 219 0 0 0 0 0 0 53 373 105 379
0099.1/P388D1/72 HRS TOTAL/-/12-4-78/#A40/FISCHER'S/3:10, 10%/
48 HRS/C14/24 HRS/1 MONTH/ASBESTOS MACROPHAGE UICC AMOSITE/
L00061/R602/--NONE--/VIDICON-MAN,28MM F8,69CM/LIPKIN*
 39 63 95 120 139 157 173 186 198 206 222 0 0 0 0 0 0 71 422 82 413
        .
        .
        .
```

An example of part of a typical gel accession descriptor file. Each data record is four lines. The first four lines of the file define the record field descriptors which are separated by "/" and terminated with a "*". The fourth line of a record is the set of gray value peaks corresponding to the ND wedge calibration. The last four numbers of that line are the computing window for that gel [x1 : x2, y1 : y2].

asserts that this distribution tends to be symmetric in isoelectronic point (pIe) but is skewed in molecular weight (MW) (*5*), although this holds only for ideal nonconglomerate spots. In practice, spots may appear round, oblong, or take on various continuous shapes, particularly when there is excessive loading of material in the gel. In all cases (extreme overloading excluded), however, the center of a spot is its darkest part. Spots may be obscured because certain regions of the gel are susceptible to streaking in both MW and isoelectric axes. Spots may occur within these streaks.

Several factors also contribute to increased background variance. Streaks will often increase the mean background density surrounding spots as well. Regions containing clusters of spots have a considerably higher mean background density then relatively empty ones. Hence, using simple thresholding techniques, light spots may often be lost, and a dark cluster of spots might segment as a single spot.

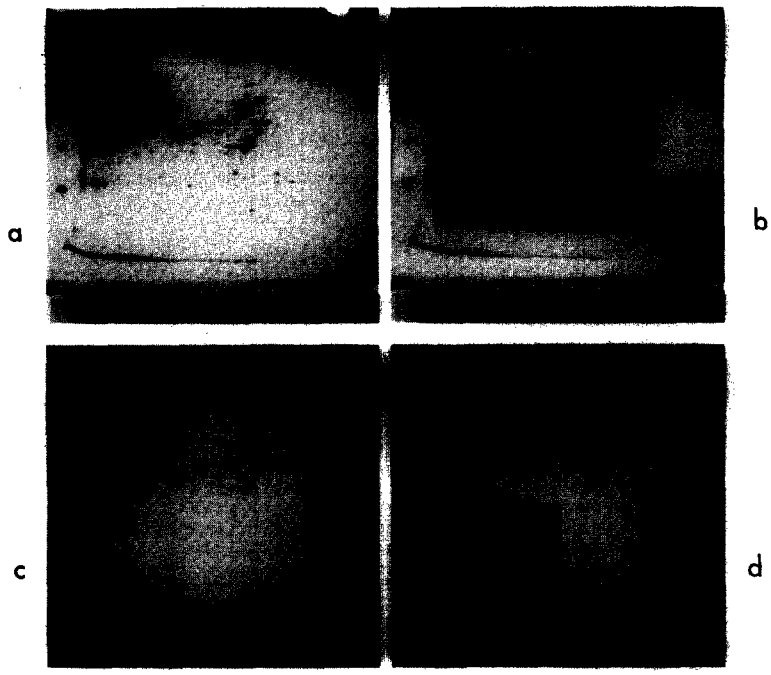

FIG. 2. Typical 2D gel with ND wedge for P388D1 macrophage like cells. (a) Original gel image scanned at 250 μm/pixel, (b) image with computing window noted to define region to segment, (c) the ND wedge sample window used to define wedge calibration, (d) smoothed ND wedge sample gray value histogram with corresponding piecewise linear calibration function on top of it. The ND value and gray value frequency are on the ordinate and gray value on the abscissa.

This increased variability and inhomogeneity of gray scale in spot and background result from gel preparation, digitization error, optical system errors and photographic technique variations. 2D gel preparation is detailed in O'Farrels original paper (*2*). For variations for our system, preparation procedures are given in (*3, 15–19*).

Polypeptides in the gel must be visualized in order to perform an analysis. Currently, at least four methods are used which include: Coomassie blue staining, autoradiography (on radioactively labeled proteins), silver staining (*15, 16*), and fluorescent dyes. Each of these spot detection methods have widely different dynamic ranges and stochiometry, as well as different ranges of effectiveness for different types of biological material.

Care must be taken to make sure that the autoradiographic film is used in the linear portion of the density versus log (exposure) curve. Further, the dynamic range of spot detection may be increased by multiple autographic exposures of the same gel. Intermediate photographs of a developing silver stained gel may be used analogously. The Vidicon or other detector is subject to similar saturation problems.

Spot detection methods have individual noise characteristics which affect the

method of gel analysis. Autoradiographs are subject to a "fog" haze of small pinpoint like noise due to background radiation which is especially a problem for long exposures. The silver stain, because of its extreme sensitivity and the opacity of metallic silver, is particularly subject to saturation and scatter. Saturation using any of the above detection methods presents another problem. Two spots close together might be resolvable if neither are saturated, but in the case of saturation the "valley" between them effectively disappears. This is because spots will saturate in their centers but not in their margins.

Because some spots will be recorded as saturated, it would be useful to know which ones and furthermore to be able to track these spots throughout the entire analysis process. This is important, because spots saturating in one gel might not do so in another so that substitute measurements could be made, as for example, in the case of multiply exposed autoradiographs.

Segmentation Model

In any locally determined (e.g., nonthresholding) image segmentation, some explicit or implicit model of the pictorial objects is necessary. The segmentation algorithm is an embodiment of some of the ideas of the underlying spot model. Spot extraction methods previously reported use one spot model or another to aid the process (*1, 5–7*). A first order model is the triple (x, y, d) consisting of the spot's centroid in the cartesian space of the image and its total integrated density (a measure of polypeptide concentration). This triple appears adequate for many types of multiple gel analyses where the object of the analyses is to measure the amounts of polypeptides present rather than the much more complex task of resynthesis of images. Segmentation is a method of spot extraction which results in obtaining this triple, as well as other features. We will present our spot segmentation algorithm below. It is based on a shape and relative density independent model and takes into account the realities of touching and overlapping spots.

Role of Segmenter in Overall System

As shown in the multiple gel analysis procedure flow diagram (Fig. 1) the segmenter is applied immediately following gel image acquisition. To handle large numbers of gels it is important that the segmentation procedure be made as automatic as possible with minimum manual intervention.

Although we present here a specific spot extractor which is able to handle a wide variety of spot shapes, density and cluster morphology, any spot extractor generating an ordered list of spot triples (x, y, d) could be used in the first stage of the GELLAB analysis. Here (x, y) are the centroid coordinates of the spot and d, its total integrated density, corresponding to the amount of polypeptide present in the gel. This latter correspondence is of course strictly true if and only if the method of spot development is based on a stochiometric reaction with absorbance characteristics that follow Beer's Law. Deviations from stochiometry and Beer's Law will reduce reliability and precision of

comparability. However, it may not necessarily preclude order of magnitude comparability which may be sufficient for some tasks.

Parameterization of the segmentation algorithm allows a wide variety of gel stains to be handled and produces different types of output which can be put to varied uses.

2. Segmentation

The Segmentation Algorithm

The segmentation algorithm is a sequence of procedures applied necessarily to a locally averaged image. The first of these is the digital analog of the spatial second derivative; it is used to construct an image consisting of the centers of spots. Second derivative (cf. Fig. 3) information delimits the extent of outward propagation to result in an algorithmic limit on individual spot extent. Spot candidate generation with this algorithm is parameter independent. The decision function which separates out noises spots is adjusted by user defined parameters. Auxiliary information required by the segmenter is obtained from the accession file. The segmentation algorithm, SG2DRV, is illustrated in flow chart form in Fig. 4. To illustrate the various steps of the segmenter, the actual

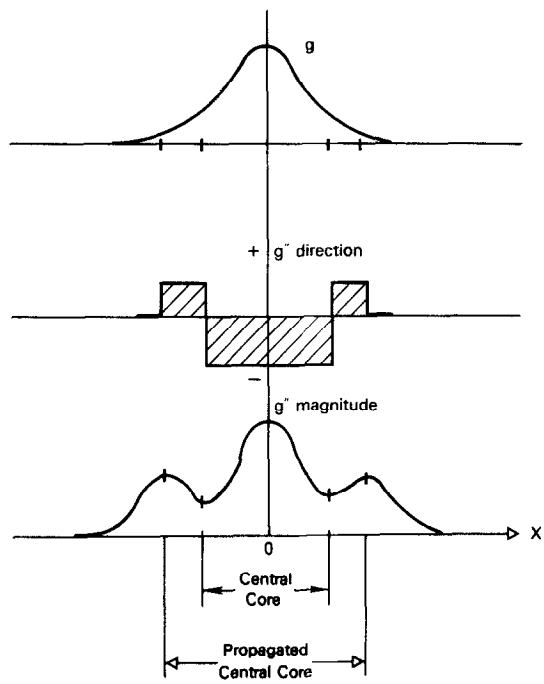

FIG. 3. One-dimensional representation of a Gaussian like function g (a) and its second derivative direction (b) and magnitude (c) functions. In the central core region, the direction of g'' is <0 and changes sign in the propagated central core region. The outer extent of the propagated central core region is denoted by a second maxima in the g" magnitude function.

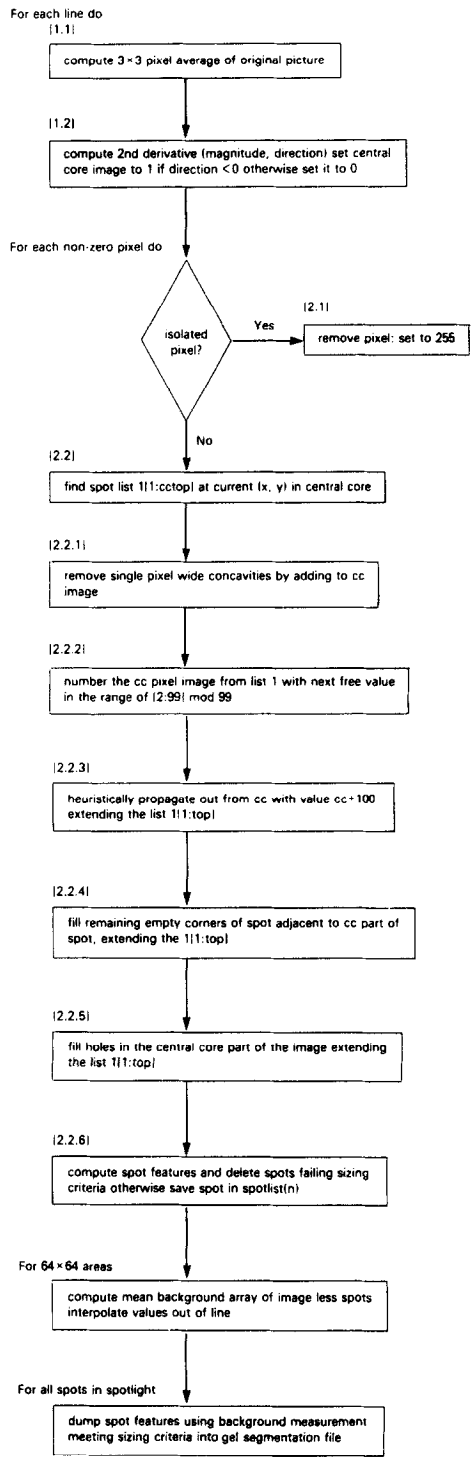

FIG. 4. **Block diagram of the 2D-gel segmentation procedure. Substeps k performed during passes 1 and 2 (through the image) are denoted by 1.k and 2.k, respectively. The central core image is computed during pass 1 while pass 2 processes each spot to completion.**

gray values in the image calculations will be presented for a small window enclosing several touching spots. A digitized version (pixel values explicit) of a small portion of an *E.coli* gel is shown in Fig. 5a to illustrate the stepwise operation of the segmenter.

Principle

Let g be an image gray scale function whose mode, median and mean are all more or less central with respect to the extrema. Let its second derivative be g''. The region adjacent to the center of a spot has a negative g'' direction. Beyond the region where the direction of g'' changes sign, there is a second smaller peak in the magnitude of g''. Our segmentation algorithm is based on finding these two maxima in two dimensions. The approximation to the boundary is operationally defined to be the second maxima in the g'' magnitude function.

Smoothing

The original image is first smoothed using a 3×3 convolution filter proposed by (*5*):

$$\begin{matrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1. \end{matrix}$$

It is applied over the entire picture, pixel by pixel in a top down TV-raster fashion by multiplying each 3×3 pixel neighborhood by the 3×3 filter (pixel for pixel) and dividing the total by 16. The result is saved as an *averaged* image. This filter removes enough of the high spatial frequency noise so the second derivative analysis algorithm may be successfully applied. Yet, it does not distort the spot shapes to any noticeable degree. The actual spot density measurements are made on the original image.

Central core and Magnitude Second Derivative Images

The second derivative is computed as (dx^2, dy^2) using the following difference formulas (*20*):

$$dx^2 = \begin{matrix} 0 & 0 & 0 \\ 1 & -2 & 1, \\ 0 & 0 & 0 \end{matrix} \qquad dy^2 = \begin{matrix} 0 & 1 & 0 \\ 0 & -2 & 0. \\ 0 & 1 & 0 \end{matrix}$$

The *magnitude* image of g'' is approximated by the sum of the absolute values of dx^2 and dy^2. The direction image is not actually computed. Instead a *central core* image pixel is defined as having a "1" where $(dx^2 < 0)$ and $(dy^2 < 0)$, and being "0" everywhere else. Figure 5b shows the initial central core image numeric data. Both the magnitude and central core images are computed during the first raster scan through the image.

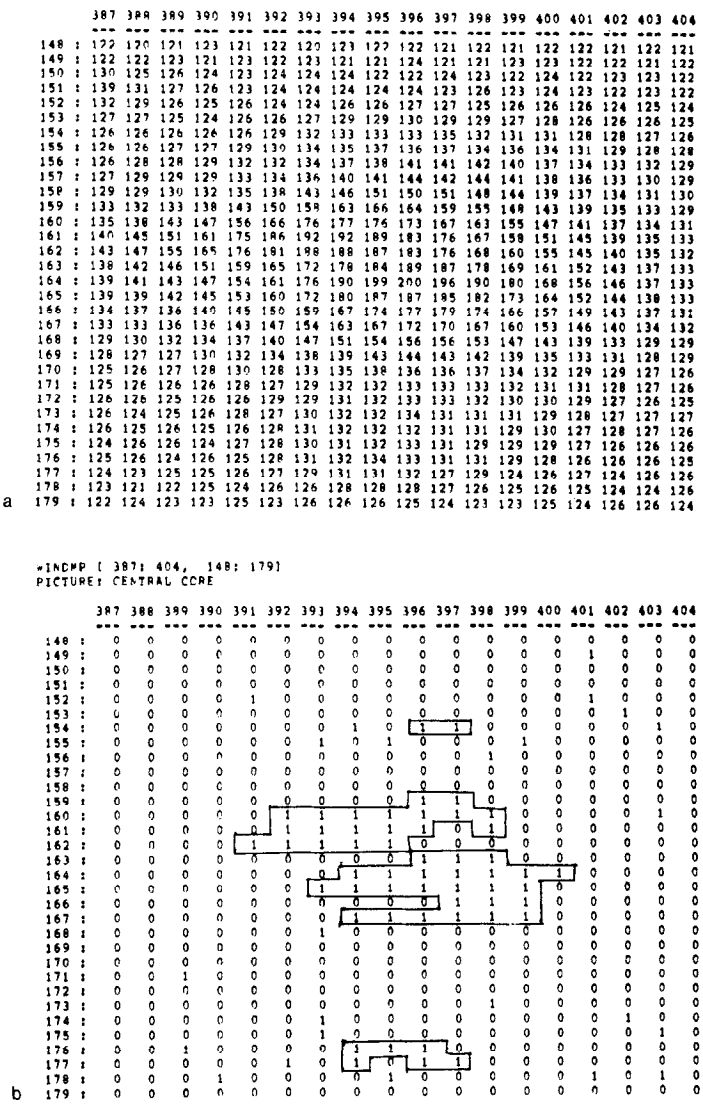

FIG. 5. Pixel gray value arrays of (a) original gray scale subregion of an *E.coli* gel image, (b) corresponding initial central core image (negative direction of second derivative) subregion, (c) corresponding final propagated central core image subregion. In this last image (c), central core values range from 2 to 99. Associated propagated central core values range from 102 to 199. Spots not meeting sizing criteria are marked for deletion by having their values set to 254. Values of 255 are isolated spots. The unrestricted sizing parameter limits ar area [1 : 500], density [0 : 400], spot density range [0 : 2.7].

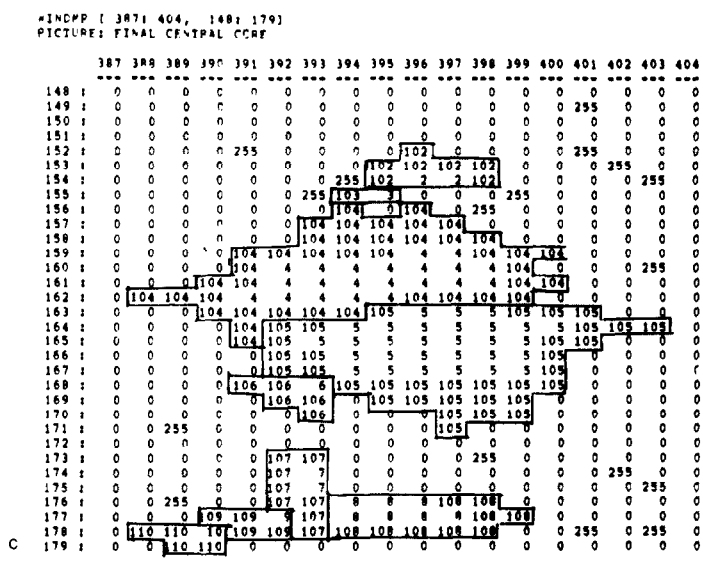

FIG. 5—*Continued.*

Extracting a Spot

In the second raster pass through the image only those pixels coded as "1" are processed. Isolated pixels (4-neighbor unconnected) are marked for deletion by setting them to a 255 code in step [2.1]. Otherwise, each time the program encounters a 1 code, it forces a spot pixel list (SPL) to be computed as in step [2.2]. The algorithm uses a push down stack to keep track of all unexpanded pixels. An unexpanded pixel is one found as a neighbor of an expanded pixel but whose 4-neighbors have not been checked. Each unexpanded pixel is expanded and checked to determine whether any of its 4-neighbor pixels have a 1 code and are not already in the SPL. Neighboring pixels so identified are put into the push down stack while the pixel being investigated is saved in the SPL. The entire spot is then processed to completion using the SPL which will grow as the spot is propagated to the region approximated by the second derivative magnitude function's second local maximum.

Removing Concavities

Single pixel wide artifactual concavities which occasionally occur are removed in step [2.2.1] by checking each SPL pixel p for the following four conditions and filling in the 0 valued pixel in the central core image if any one of them occurs. The SPL is also updated. In each case of the mask, a "0" and "1" must occur and a "—" means "don't care."

```
1 0 1        1 1 —        — — —        — 1 1
1 p 1  or    0 p —  or    1 p 1  or    — p 0
— — —        1 1 —        1 0 1        — 1 1.
```

Numbering Spots in the Central Core Image

In the central core image itself, the spot is then numbered (step [2.2.2]) with the next free number in the range of [2 : 99] modulo 99. It is very unlikely but in an extremely densely populated spot image, it is possible for two adjacent spots to have the same value. This notation problem is easily solved by alternative coding schemes.

Propagating a Spot from the Central Core

The numbered spot is then propagated with the value (C + 100) from the central core (C) of the spot to the propagated central core region in step [2.2.3]. This propagation from a central core edge point is performed in the 4-neighbor directions until it is terminated based on various constraints. The SPL is updated with the new pixels as well. The heuristic termination conditions are:

1. The second derivative magnitude is increasing (starting 1 pixel out from the central core), outward from the central core indicating the second local maxima.
2. The second derivative magnitude outward from the central core has the same value twice in a row indicating a noisy edge.
3. The propagation would impinge on another central core pixel.
4. The propagation would extend beyond the computing window.
5. The propagation would impinge on an isolated pixel.
6. The gray value outward from the central core is increasing instead of decreasing indicating that the spot is overlapping a much larger spot.

Corner Filling a Spot

This type of propagation sometimes forms small rectangular empty corner regions in the four corners of the spot. These can be filled with propagated central core values in step [2.2.4]. Both 0 and 255 (isolated pixel) 45-degree corner values are candidates for filling if the central pixel is its central core. The four cases are as follows with C being the central core value, N being the propagated central core value, E being either 0 or 255, and '—' meaning "don't care."

```
E N —        — N E        — — —        — — —
N C —  or    — C N  or    — C N  or    N C —
— — —        — — —        — N E        E N —.
```

Hole Filling the Central Core of a Spot

In a small number of cases of very large possibly saturating spots, the center of the spot may not be detected as such and thus not segmented. This problem is repaired in step [2.2.5] by filling any artifactual holes in the central core region. The leftmost and rightmost x-coordinates for each line of the central core are found and any 0s in the central core image between these points are changed to central core values.

Spot Features and Initial Sizing Test

After the SPL is completed, it consists of the pixels in the central core and propagated central core. Various features are computed using density values mapped from the original image. A preliminary sizing is performed to remove most of the background noise spots in step [2.2.6]. A 254 code is also placed in each deleted spot pixel in the propagated central core image. Figure 5c shows the final propagated central core gray scale numeric data in the window currently being segmented.

Background Computation

A background density correction is then performed during a third pass through the image using an algorithm similar to that described in (7). The computing window is broken up into 64 × 64 pixel subimages. A histogram is computed for each of these subimages of the averaged image masked by the *logical complement* of the propagated central core image (i.e., what is left *after* the preliminary sized spots are removed). Each of the up to 64 histograms is then smoothed and the first major peak found. The standard deviation of the left side of the first peak (assumed to be the background peak) is computed and used as an estimate of background peak standard deviation. If the peak exceeds its 4-neighbor average by a predetermined threshold (currently 10 gray values), the peak is replaced by this average.

Computing Corrected Density (D') and Secondary Sizing Test

The features which at present may be used to determining acceptance of a spot include: spot area (in square pixels), total spot density, range of pixel OD seen in the spot. The last is useful for eliminating noise in the image. The corrected spot density D' is then computed from D taking the background estimate into account. Those spots which meet the criteria are output into the gel segmentation file (GSF). Those spots failing the density sizing criteria are deleted as before by having 254 placed in the central core image of each pixel. The gray value numeric data for the final output image window is shown in Fig. 6a with 9 spots segmented.

It is possible for the spot feature sizing parameters to be changed to more restrictive ranges to eliminate some of the smaller noise objects. Figure 6b shows the three extracted spots remaining for the more restrictive sizing which

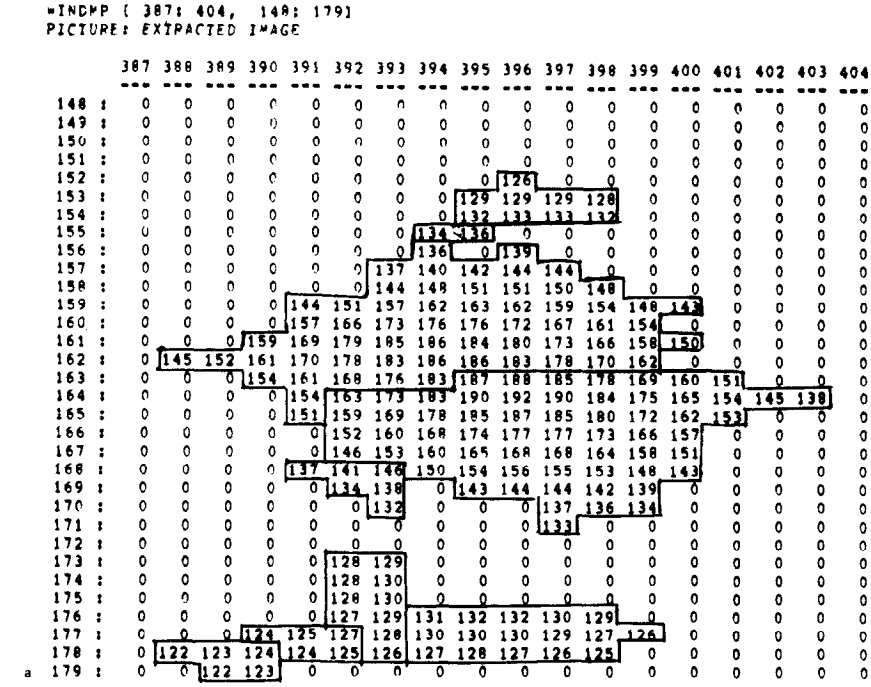

FIG. 6. Pixel gray value arrays of (a) final image extracted from original image of window of *E.coli* gel image using unrestricted sizing to accept all spots, (b) final image extracted from original image of window of *E.coli* gel image using more restrictive parameter limits set to: area [10.00 : 2000.00] pixels, density [0.10 : 500.00] OD units, and spot density range [0.05 : 2.70] OD.

is normally used for autoradiographs. Table II shows part of a gel segmentation file for a typical gel.

3. Results

In this section, we show some image segmenter output which illustrates the

TABLE II

Examples of Gel Segmentation File

```
SG2PRV : Version March 16, 1980 - 3:25PM
Today's date is 03/17/1980,  04:21:49 PM
User: [33,2]
Gel Segmentation File is: P10094.GSF
0094,1/P388D1/24 HRS TOTAL/-/12-4-78/#A40/FISCHER'S/3:10, 10%/
0 HRS/C14/24 HRS/1 MONTH/ASBESTOS MACROPHAGE CONTROL/
[00041/R602/--NONE--/VIDICON-MAN,28MM F8,69CM/LIPKIN#
 38 65 96 120 139 156 173 186 197 206 213 222 0 0 0 0 47 441 100 374
S#ITCHES: /CTLCORE
Window [47:441,100:374]
Area sizing limits ( 10.00: 2000.00)
Density sizing limits ( .10: 500.00)
Density range sizing limits ( .05: 2.70)
Saving central core image in [33,2]C00041.PIX
Saving output image in [44,2]Z00041.PIX
Mean background matrix in ND (std dev)
 .06(.03) .06(.03) .07(.03) .07(.03) .08(.03) .08(.03) .09(.03)
 .06(.03) .06(.03) .07(.03) .07(.03) .08(.03) .08(.03) .09(.03)
 .06(.03) .07(.03) .07(.03) .07(.03) .08(.03) .08(.03) .09(.03)
 .06(.03) .06(.03) .07(.03) .07(.03) .08(.03) .08(.03) .09(.03)
 .06(.03) .07(.03) .07(.03) .07(.03) .08(.03) .09(.03) .09(.03)
 CC#   1 M.E.R[ 62: 68,  101: 106] D.R.=[.19: .40] D/A= .326 MnB= .061
  1st MOM[64.52, 103.38] A=    22  D=  7.17 D'=  5.82 (D'/totalD')%=   .22%
  Sx=  1.43 Sy=  1.20 Sxy=   .88     V=  4.91
 CC#   2 M.E.R[117: 122, 105: 108] D.R.=[.17: .23] D/A= .207 MnB= .061
  1st MOM[119.18,106.22] A=    18  D=  3.72 D'=  2.62 (D'/totalD')%=   .10%
  Sx=  1.28 Sy=  1.00 Sxy=   .74     V=  2.05
 CC#   3 M.E.R[ 60: 68,  106: 110] D.R.=[.20: .41] D/A= .333 MnB= .061
  1st MOM[62.72, 107.72] A=    23  D=  7.67 D'=  6.26 (D'/totalD')%=   .24%
  Sx=  1.78 Sy=  1.07 Sxy=   .89     V=  5.52
        .
        .
        .
 CC# 395 M.E.R[192: 196, 363: 368] D.R.=[.06: .12] D/A= .048 MnB= .061
  1st MOM[193.21, 365.67] A=    18  D=  1.59 D'=   .49 (D'/totalD')%=   .02%
  Sx=  1.09 Sy=  1.27 Sxy=   .86     V=  1.15
Total of    395 accepted D spots accumulated density= 3416.21, area= 12969
Total of    395 accepted D' spots accumulated density= 2623.66, area= 12969
Total of   5589 omitted spots accumulated density= 4472.37, area= 48610
Omitted/Accepted density =   131%
FINISHED! The GSF is P10094.GSF
Real TIME =00:28:11
CPU TIME  =00:10:08,   35.942%
```

Illustration of part of the gel segmentation file for a ^{14}C labeled P388D1 macrophage like cells autoradiograph gel ACC# 94.1. Both segmenter parameters and some of the spot feature list data are presented. CC is connected component number, MnB is mean background density for a spot, A is spot area, D is uncorrected total spot density and corrected density D' is D-(A) (MnB). D/A is the mean spot density and (D'/Total D') % is D' expressed as a percentage of total gel spot density. First MOM is the spot centroid while D.R. is the density range of pixels seen in the spot. Sx, Sy, and Sxy are the standard deviation and covariance for the propagated central core region with V being the Gaussian volume estimate of this region. Density values are given in OD calibrated in terms of the associated ND wedge in the image.

wide range of effectiveness of the algorithm. Other results of segmentation are deferred to the last paper in this series (*9*). The segmentation algorithm appears applicable to a wide range of gel magnifications and densities, i.e., autoradiographs of varying exposures, varying spot detection modalities, etc. It is capable of resolving touching spots and other image complications over a wide range of conditions.

Examples are shown in Fig. 7. The separation during segmentation of the touching spots in the window of the *E.coli* gel was performed adequately and all spots (as confirmed by visual inspection of the set of related images) including very light and touching spots were extracted. By relaxing the parameter limits slightly, most of the noise spots in the image were removed without deleting minor spots. Spot features including the required (x, y, d) triple are computed for each spot accepted. The maximum pixel OD value found for each spot is also recorded for possible later use in tracking saturated spots.

The segmenter has been applied to various types of gels of different types of material scanned at different magnifications with satisfactory results. At least 300 gels have been segmented using this program. Figure 7 is a composite photograph showing gel images before and after segmentation. Figures 7a and c are of a PHA stimulated human lymphocyte gel ^{35}S autoradiograph scanned at 250 and 170 μm/pixel, respectively. Figure 7e is a P388D1 macrophage like cell ^{14}C autoradiograph gel scanned at 250 microns/pixel. Figure 7g is a normal RBC silver stained image scanned from 120 size film with approximately 250 μm/pixel resolution. As can be seen in Figs. 7b,d,f,h, the segmentation of spots were successful in a vast majority of cases. The number of spots segmented in these four gels were 602, 1081, 700, and 556, respectively.

We have empirically determined a set of operating parameters for various classes of gels. The parameters currently used for the 250 μm/pixel autoradiographs are: total spot area range [10–2000] square pixels, total spot density range [0.1–500] OD, spot pixel density range [0.05–2.7] OD. For the 170 μm/pixel autoradiographs, the total density minimum limit is increased to at least 1.0 OD. For silver stained gels at 250 μm/pixel, the total density minimum is 1.0 OD and the minimum spot pixel density range is 0.15 OD.

Typical times for segmenting 250 μm/pixel resolution gels on the DECSYSTEM-20 are on the order of 8 to 10 min/gel for gels with 400 to 1000 spots. The 250 μm/pixel gel images have the computing window set to about two-thirds to one-half the area of the gel because of the need to include the ND wedge in the image. These computing times increase when performed over the full 512×512 pixel image for the 170 μm/pixel gel images. Changes currently under way in the picture I/O package (cf. IO2NEW in the Appendix) should speedup the segmentation algoritm.

4. Discussion

Earlier, we presented the requirements for a gel image segmenter in the context of both this particular problem domain and at the same time showing its

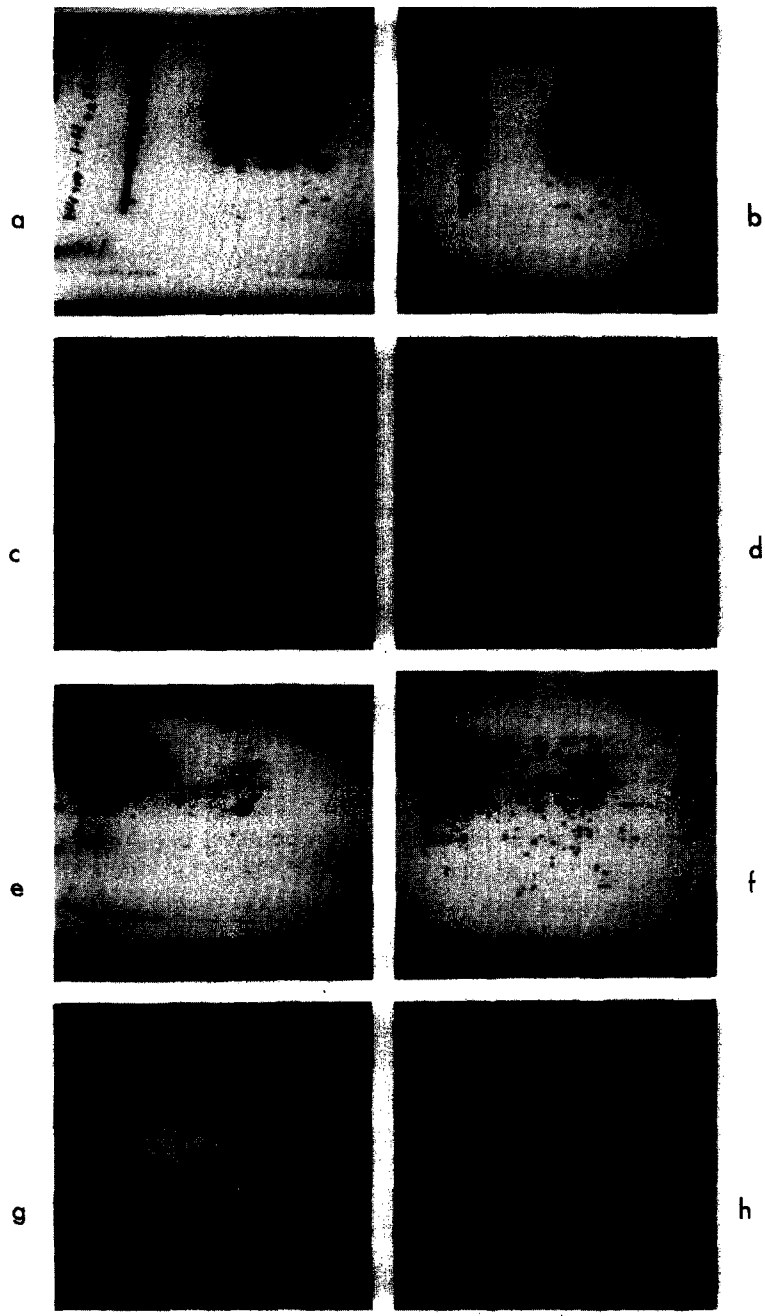

FIG. 7. Composite photograph showing four different gels before and after segmentation has been performed. (a) PHA stimulated lymphocyte ^{35}S autoradiograph scanned at 250 μm/pixel, (b) resultant segmentation of (a), (c) same PHA stimulated lymphocyte ^{35}S autoradiograph scanned at 170 μm/pixel, (d) resultant segmentation of (c), (e) P388D1 macrophage like cell ^{14}C autoradiograph scanned at 250 microns/pixel, (f) resultant segmentation of (e), (g) silver stained normal RBC gel scanned at about 250 μm/pixel, (h) resultant segmentation of (g).

relation to digital image processing in general. The segmenter we have adopted for 2D gels is only one of many applicable to this problem. We will show reasons for selection of our approach after consideration of some other segmenters and spot extractors which have been applied to gel images.

4.1. Some Other Gel Segmenters

A large number of techniques have been brought to bear on the detection and extraction of objects from images (*20, 21*). Some of these, using global techniques such as thresholding, fail because of inhomogenity of the gel image. Others using local algorithms globally applied have been much more successful.

A threshold segmentation technique (*7*) uses the analysis of the density histogram to find the mean background value of the image and then thresholds the image at 0.05 to 0.10 OD above this value to detect faint spots. Darker conglomerates of spots are segmented using this technique into a single "spot." Each spot is tested to find maxima and minima regions within it to determine whether it should be iteratively split into subspots. Spots found in this way are then expanded to the region defined by a derivative of the gray scale data.

Another segmentation technique (*1, 22*) reported earlier detects spots by scanning first in a raster direction and then orthogonally to find spot maxima. Ellipses are automatically fit over the detected spots to approximate their boundaries and density information measured. Other techniques using spot extraction by convolution techniques are being developed by this group.

Another technique of spot detection is based on finding spots in the image starting at the darkest pixel in the image and then removing fitted spots before continuing the search (*5*). Using the assumed property of the spots that they are relatively Gaussian in x and skewed Gaussian in y, the algorithm fits parametric curves to the spots and then estimates the volume density from the parameters. The segmenter seems to give quite good results with synthesized gel image (from the Gaussian parameters) being quite similar to the original image (personal communication, P. Jansson). Output consists of density rank ordered spot lists consisting of spot centroid, density and parametric spot description information.

Spots may be detected by assembling line segments of spots from successive lines using a procedure called "chain assembly" (*6*). The chains are smoothed and Gaussian curves fit to estimate the spot. If the spot meets sizing criteria, its features are saved and it is removed from the image chain list.

4.2. A Distribution-free Density Independent Segmenter

This algorithm uses the central core model of a spot. The central core defined as the region of negative slope of the second derivative function of the image. Such a region will only occur within the first minimum of g'' surrounding the spot's center. This model seems to work robustly on real gel data, failing only

on those few spots of such a huge extent or saturation that no simple model exists for the spot. It is obviously independent of assumptions as to exact spatial distributions of density as well as of orientation.

Because shape of the spots corresponds to the physical diffusion process used in their generation, no true boundaries are present. Manual measurements made at what observers define as the boundary have resulted in up to 50% error in total spot density. We have chosen to algorithmically define a spot as its propagated central core which is found to be reproducible.

A problem common to many gels is artifactual streaking. One may attempt to remove streaks before processing. Alternatively, the spots may be segmented in the context of the streaks. We performed some experiments using a notch filter (*23*) to remove streaks using a 12 × 12 pixel window. We found that although streaks were effectively removed from the image, the edges of spots were "chipped" away, thus distorting the spot as a function of its size and shape. The central core algorithm finds spots regardless of whether the streak is present and therefore streak removal by preprocessing is not necessary.

The central core algorithm finds a peak if it is present and if there is sufficient resolution (in both the spatial and density domains of the image) will resolve overlapping peaks. The cases where it fails can be understood if one looks at what is computed for g''. g'' is approximated over a 3 × 3 region using the difference formulas given above. This discrete approximation cannot resolve spatial position differences less than about five pixels. Thus gels scanned at higher resolution (170 μm/pixel) show fewer unresolved touching spots after segmentation than gels scanned at lower resolution (250 microns/pixel).

Two overlapping saturating spots will also sometimes be unresolved. This is because the plateau effect occasioned by saturation obscures the second peak in g'' which is necessary for spot separation. If one wishes to keep track of known saturated spots it is necessary to employ a different morphologic analysis. Subsequently, attempts might be made to estimate the spots' edges on the basis of boundary curvature type of information. Although in general these spots should not be analyzed in the same way, this situation can occur in gels where the remainder of the spots are valid. Because some spots will be recorded as saturated, it would be useful to know which ones are and further, to track these spots throughout the entire analysis process. Spots saturating in one gel might not do so in another so that substitute measurements could be made in the case of multiply exposed autoradiographs. Saturation and its relation to normalization will be discussed further in (*9*).

Light spots will be detected by the central core algorithm. In this context, the averaging phase of the algorithm is essential for removing the high frequency spatial noise in order for the central cores to be computed correctly for light spots.

Because of gel loading, temperature and ampholine/acrylamide variations, spots may be somewhat distorted so that an idealized shape of a spot may be rarely found. The segmenter works well for such spots because their extent is defined by the second maximum of the g'' magnitude over all of the spots' edge

rather than by the standard deviations of x and y in the Gaussian model. Changes in the spot orientation due to gel rotation are also easily handled by the current algorithm since the entire spot edge is analyzed.

The absolute density of the spot detector (e.g., stain) varies globally among gels due to various physical processes. It is thus necessary to normalize spots in each gel in order to compare them among gels. Several options are available (*9*). The segmenter performs one type of normalization which is useful for well-segmented gels. In addition to reporting each spots' total density (D) and its background corrected density (D'), it also reports D' divided by the sum of D' for all spots accepted expressed as a percentage. These features are illustrated in Table II.

Multiple scans of the same gel at two different resolutions provide an opportunity to investigate differences in segmenter output. Manual use of PIXODT (cf. Appendix) leads to the conclusion that the few instances where spots were incorrectly merged were due to either (a) a lack of spatial resolution (spots were too close) or (b) gray scale resolution (spots were close to or at saturation). Overall, the correlation was very good with most spots being segmented correctly.

5. Conclusion

A shape independent algorithm for the segmentation of a 2D electrophorectic gel image into a spot list has been presented. Based on central core propagation in accordance with local region changes in gray scale value this algorithm is successful in extracting spots under a wide variety of gel conditions. It is thus useful as a first stage processor (after image digitization) in the analysis of 2D gels. Because there is no need for user interaction during segmentation, a series of gels may be batch processed with increased efficiency.

After segmentation, each gel is represented by a very large list of spot features. The next phase is the comparison of the same spot within a number of gels. We will give an algorithm for the pairing of spots between two gels (*8*). This pairing is a prerequsite for the analysis of multiple gels where one seeks correspondences for the values of a particular spot among a set of gels (*9*).

Appendixes

A.1. SAIL as Language for Building System

The set of GELLAB programs SG2DRV (reported here), CMPGEL (*8*), and CGEL, MARKGEL, SEERSPOT (*9*) are written in the SAIL programming language (*24*). This language is currently implemented only on DECSYSTEM-10 and DECSYSTEM-20 computers. It has distinct advantages in its ease of algorithm expression, macro expansion, string, list, set and associative processing and record structure operations. Since SAIL strongly encourages structured programming it is an ideal environment in which to implement a set of complex interacting algorithms. Although the programs could have been

written in FORTRAN, it would have been at an excessive price in terms of time and clarity.

A.2. Implementing the Segmenter on a Small Machine

The segmenter, although currently implemented on the DECSYSTEM-20 a medium size computer, could be implemented on a small machine with some modifications. Resultant spot lists are kept in core memory but could be moved to a file without too great a loss in efficiency. A frame buffer interactive display system might be required to set up the computing window and ND wedge calibrations. However, if the entire image was scanned and an alternative wedge calibration procedure employed, then interactive display hardware is not absolutely essential. We believe, however, that such a display is an integral part of any gel analysis system—at the very least for checking segmented images.

A.3. BMON20 Environment for GELLAB

A system called BMON20, initially mentioned in (*11*), is used to invoke and control the GELLAB set of programs and associated image and nonimage data bases. It will be the subject of a later publication.

BMON20 is a sharable image-processing distributed-monitor-system. It may invoke, through user requests, a wide variety of picture processing operations. The GELLAB set of programs is only a subset of the BMON20 system. Each user has, in his disk file area, one or more substate files which describe the current environment of the picture operators being used. The GELLAB system has a substate called GEL.STA which includes among other parameters:

GEL.ID data base accession file.
Current gel accession number (ACC#).
Landmark set data base file.
E-spot list data base file.
Computing window coordinates.
Current ND wedge nd values for GEL.ID accession file.
ND wedge gray value calibration for current gel ACC#.
Current gel picutre file corresponding to gel ACC#.
Representative gel ACC.
List of standard proteins in the representative Gel.
Picture disk area name.
SG2DRV total area sizing limits.
SG2DRV total density sizing limits.
SG2DRV density range sizing limits.

Some of these parameters will be described in the later GELLAB papers.

IO2NEW Picture Pager. Since the segmenter program performs I/O on five 512 × 512 8-bit pixel images, an efficient picture I/O package is required. A

TABLE III

SAMPLE OF PIXODT PICTURE DEBUGGER OUTPUT

```
PIXODT :Version March 21, 1980 - 11:45AM
Commands
--------
<number> <CR> replace old pixel contents at current address.
/ reopen current pixel location.
+ or R open (right) pixel address.
- or L open previous (left) pixel address.
U open pixel address up 1 line.
D open pixel address down 1 line.
N print 3x3 pixel neighborhood of previous (x,y) address.
* print 18x20 pixel window centered at previous (x,y) address.
C print 18x20 pixel window with ULH corner at previous (x,y) address.
M toggle pixel print out mode between decimal (default) and octal.
E exit saving changed image.
P toggle the write-enable protection of the image (initially protected)
H print this list of commands
Edit user image file: C00565.BIX
 x,y picture address: 345,223
/ 94  N
*INDMP [ 336: 353,  213: 233]
       336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353
       --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 213 : 254 254   0   0 255   0   0   0   0   0   0   0   0   0   0   0 254 254
 214 : 254 254 254   0   0   0   0   0   0   0 255   0   0   0   0   0 254 254
 215 : 254 254 254 254 254 255   0   0 255   0   0   0   0   0 133   0   0   0
 216 : 254 254 254 254 254   0   0   0   0   0   0   0   0 133 133   0   0   0
 217 :   0 254 254 254 254   0   0   0   0   0   0   0 133 133 133   0   0   0
 218 :   0 254 254 254 254   0   0   0   0   0   0 133 133 133 133 133 254 254
 219 :   0   0   0   0   0   0   0   0 194   0 133 133  33  33  33 133 254 254
 220 :   0 255   0   0   0   0   0   0 194 194   0 133  33  33  33 133 254 254
 221 :   0   0 255   0   0   0   0 194 194 194 194 133  33  33  33 133 254 254
 222 :   0 255   0   0   0   0 194 194  94  94  94  94 194  33  33 133   0   0
 223 : 254   0   0   0 158   0 194  94  94  94 194 194 133 133 133 159 159
 224 : 254   0   0 158 158 194 194 194  94  94  94 194 194   0 133   0   0   0
 225 :   0   0   0 158  58  58  58 194 194 194 194 194 254 254   0   0   0   0
 226 :   0   0 158 158  58  58  58   0 194   0   0   0 254 254   0   0 255   0
 227 :   0   0   0 158  58  58  58  58 194   0 254 254 254   0   0   0   0   0
 228 :   0 254 254 254 158  58  58  58 158   0 254 254   0 255   0   0   0 255
 229 :   0   0 254 254 158  58  58 158 158 254 254 254 254 254   0   0   0   0
 230 :   0   0 254 254 254 158 158 158   0 254 254   0 254 254   0   0   0 255
 231 : 172 172   0 254 254 153   0 154 154 254   0   0   0   0   0   0   0   0
 232 : 172 172   0   0   0 153   0 154 154   0   0   0   0   0   0   0   0   0
 233 : 172 172 172   0 153 153 154 154 154 154   0   0   0   0   0   0   0   0
```

The PIXODT picture debugger is a software tool useful for checking progress of image processing algorithms to aid their development. The list of operations available in PIXODT is listed in the top of the table with some typical output illustrated in the bottom of the table. The user responses are underlined.

picture pager subroutine package called IO2NEW was written in SAIL to support the segmenter program. I/O on the DECSYSTEM-10 or DECSYSTEM-20 disk is done in multiples of blocks where a block is 128 36-bit words. Since four 8-bit pixels are packed into one word, a 128-word block holds one 512-pixel image line. Thus a scheme to page (i.e., to transfer from fast core memory to/from a slower disk memory) image lines can be set up corresponding to paging disk blocks.

The IO2NEW procedure package keeps a working set of lines of the active pictures in core at any one time. The algorithm ages pictures and lines (on a least recently used basis), and uses this age parameter to determine which

image to page out (if it was written into, i.e., "dirty") when a free image is required. The furthest line from the new line in an image is paged out of memory (if it was "dirty") when space is required for the new line.

Each picture has an associative map of 512 registers, one for each line. These determine whether the requested line is already in core before trying to page it in. Lines are only written out when necessary (when they are "flushed out" by the program and at its termination). The number of lines kept in core is a compile time parameter. The segmenter works well with the picture page maintaining 20 lines/picture in core but is currently set to 40 lines/picture for greater ease in handling large spots. The I/O data modes include pixel, line unpacked, and line packed, all of which are used in various parts of the segmenter in order to optimize data flow.

If more I/O channels are required by the pager, the oldest picture is flushed out and the channel freed for use by the new picture. Existing pictures may be read or written (if they are declared "writable"). Declaring a new picture creates a zeroed image file. Since there is a total of 16 I/O channels available to the user in TOPS-10, the user might wish to allocate a smaller number of these channels to the pager so that other program I/O can be performed.

Although the IO2NEW I/O package is written in SAIL and is not directly exportable to non-SAIL environments, it should not be too difficult to implement on another machine, one with the ability to random access its disk and provide each user several I/O channels.

Improvements in the efficiency of the pager algorithm and increases in its speed are now in progress. A bottleneck in the present version is each line requires a separate disk I/O request. By paging groups of lines together, groups that correspond to a disk cluster (area of the disk which is physically allocated sequentially), we expect to increase its speed.

PIXODT Picture Debugger. A picture debugger was built in order to aid in the debugging of both the IO2NEW.SAI picture pager and the segmenter program SG2DRV.SAI. Its list of commands and sample ouput are illustrated in Table III. It enables the user to interactively examine and change individual picture elements (pixels) of an image.

ACKNOWLEDGMENTS

The constant help afforded by Morton Schultz, Bruce Shapiro, and Earl Smith, our colleagues in the Image Processing Unit, has been invaluable. Our collaborators Carl Merril and David Goldman of NIMH and Eric Lester (formerly of NCI, now at University of Chicago Medical School) have provided stimulating ideas and critical evaluation of the methodology as it has developed. Particular thanks are due to Eric Lester for help in empirically determining the effectiveness of the segmenter, especially in the practical user context.

REFERENCES

1. ANDERSON, N. G., AND ANDERSON, N. L. Molecular anatomy, Behring Inst. Symposium 1977. *Mitt. Behring Inst. Sympos.* **63**, 169 (1979).

2. O'FARRELL, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250,** 4007 (1975).
3. LEMKIN, P., MERRIL, C., LIPKIN, L., VAN KEUREN, M., OERTEL, W., SHAPIRO, B., WADE, M., SCHULTZ, M., AND SMITH, E. Software aids for the analysis of 2D gel electrophoresis images. *Comp. Biomed. Res.* **12,** 517 (1979).
4. LIPKIN, L. E., AND LEMKIN, P. F. Data base techniques for multiple PAGE (2D GEL) analysis. *Clinical Chemistry* **26,** 1403 (1980).
5. LUTIN, W. A., KYLE, C. F., AND FREEMAN, J. A. Quantitation of brain proteins by computer-analyzed two-dimensional electrophoresis, *in* "Electrophoresis '78" (N. Catsimpoolas, Ed.), pp. 93–106. Elsevier/North-Holland, Amsterdam/New York, 1978.
6. GARRELS, J. I. Two-dimensional gel electrophoresis and computer analysis of proteins synthesized by clonal cell lines. *J. Biol. Chem.* **254,** 7961 (1979).
7. BOSSINGER, J., MILLER, M. J., KIEM-PHING, V., GEIDUSCHEK, P., AND XUONG, N. Quantative analysis of two-dimensional electrophetograms. *J. Biol. Chem.* **254,** 7986 (1979).
8. LEMKIN, P., AND LIPKIN, L. GELLAB: A computer system for 2D gel electrophoresis analysis. II. Spot pairing. *Comput. Biomed. Res.* **14,** in press (1981).
9. LEMKIN, P., AND LIPKIN, L. GELLAB: A computer system for 2D gel electrophoresis analysis. III. Multiple gel analysis, submitted for publication.
10. LEMKIN, P. "Buffer Memory Monitor System for Interactive Image Processing," NCI/IP Technical Report #21b, Nat. Tech. Info. Serv. PB278789 (listing PB278789) (1978).
11. LEMKIN, P., AND LIPKIN, L. BMON2-A distributed monitor system for biological image processing. *Comput. Prog. in Biomed.* **11,** 21 (1980).
12. CARMAN, G., LEMKIN, P., LIPKIN, L., SHAPIRO, B., SCHULTZ, M., AND KAISER, P. A real time picture processor for use in biological cell identification. II. Hardware Implementation. *J. Histochem. Cytochem.* **22,** 732 (1974).
13. LEMKIN, P., CARMAN, G., LIPKIN, L., SHAPIRO, B., SCHULTZ, M., AND KAISER, P. A real time picture processor for use in biological cell identification. I. System design. *J. Histochem. Cytochem.* **22,** 725 (1974).
14. LEMKIN, P., CARMAN, G., LIPKIN, L., SHAPIRO, B., AND SCHULTZ, M. Real Time Picture Processor—Description and Specification." NCI/IP Technical Report #7a, Nat. Tech. Info. Serv. PB269600/AS (1977).
15. MERRIL, C., SWITZER, R. C., AND VAN KEUREN, M. L. Trace polypeptides in cellular extracts and human body fluids detected by two-dimensional electrophoresis and a highly sensitive silver stain, *Proc. Nat. Acad. Sci. USA* **76,** 4335 (1979).
16. GOLDMAN, D., MERRIL, C. R., AND EBERT, M. H. Two-dimensional gel electrophoresis of human cerebrospinal fluid proteins I. *Clin. Chem.*, in press.
17. LESTER, E. P., LEMKIN, P., LIPKIN, L. E., AND COOPER, H. L. Two-dimensional electrophorectic analysis of protein synthesis in resting and growing lymphocytes *in Vitro*, *J. Immun.* **126,** 1428 (1981).
18. LESTER, E. P., LEMKIN, P., COOPER, H. L., AND LIPKIN, L. E. Computer-assisted analysis of two-dimensional electrophoresis of human peripheral blood lymphocytes, *Clin. Chem.* **26,** 1397 (1980).
19. LEMKIN, P., LIPKIN, L., MERRIL, C., AND SHIFFRIN, S. Protein abnormalities in macrophages bearing asbestos. *Envir. Health. Perspect.* **34,** 75 (1980).
20. ROSENFELD, A. "Picture Processing by Computer," Academic Press, New York, 1969.
21. ROSENFELD, A., AND KAK, A. "Digital Picture Processing," Academic Press, New York, 1977.
22. ANDERSON, N. G., ANDERSON, N. L., AND TOLLAKSEN, S. L. Proteins of human urine. I. Concentration and analysis by two-dimensional electrophoresis. *Clin. Chem.* **25,** 1199 (1979).
23. LIPKIN, L., LEMKIN, P., SHAPIRO, B., AND SKLANSKY, J., Preprocessing of electron micrographs of nucleic acid molecules for automatic analysis by computer. *Comput. Biomed. Res.* **12,** 279–289 (1979).
24. REISER, J. F., "SAIL," Stanford University Artificial Intelligence Laboratory memo AIM-289, August, 1976. Also available from U.S. Dept. Commerce. Nat. Tech. Inform. Serv. No. AD-A045-102, Springfield, Va., 1976.