COMPUTERS AND BIOMEDICAL RESEARCH 12, 517–544 (1979)

Software Aids for the Analysis of 2D Gel Electrophoresis Images

P. LEMKIN,* C. MERRIL,† L. LIPKIN,* M. VAN KEUREN,† W. OERTEL,‡§
B. SHAPIRO,* M. WADE,* M. SCHULTZ,* AND E. SMITH*

**Image Processing Unit, Division of Cancer Biology and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20014; † Laboratory of General and Comparative Biochemistry, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20014; and ‡ Laboratory of Clinical Science, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20014*

Received November 27, 1978

A computer program, FLICKER, is described for aiding the analysis of two-dimensional gel electrophoresis images. It facilitates the comparison of parts of two or more images by flicker comparison as well as enabling the extraction and measurement of operator-selected regions. Measures of image distortion and the comparison of images are available for testing the validity of biochemical differences between the gels. The measurements available include transforming the density feature into a measure of polypeptide concentration.

1. INTRODUCTION

The development of two-dimensional electrophoresis (EP) over the past 3 years has increased the amount of potentially available information discernable in complex protein mixtures by two orders of magnitude (*1*). Separation of proteins is optimized by utilizing independent separation of proteins in each dimension. Generally isoelectric focusing is used for the first dimension followed by a discontinuous sodium dodecylsulfate–polyacrylamide gel electrophoresis (SDS–PAGE) gel in the second dimension (*2*).

In whole cell lysates one can often see 1000 or more proteins. Each spot corresponds to one protein and if the methods are carefully controlled the patterns are highly reproducible. Catalogs of protein patterns have already been published for *Escherichia coli* proteins during different growth rates (*3*) and in human plasma proteins (*4*).

It is clear that these two-dimensional techniques will play a vital role in unraveling the complex protein patterns in many systems and will allow new observations in experimental mutagenesis studies. Full utilization of this new tool will require rapid accurate methods to analyze these patterns, often for subtle variations.

§ Supported by a grant from the "Studienstiftung des Deutschen Volks," Bonn, West Germany.

0010–4809/79/060517–28$02.00/0

The analytical methods can be divided into two categories. The first involves the problem of visualizing the protein patterns while the second is concerned with the analysis of the visualized patterns. Currently two methods are generally employed for visualization: (1) protein stains (such as Coomassie blue) or a fluorescent stain (such as fluorescamine), and (2) the use of radioactively labeled proteins followed by autoradiography. This study will not be concerned with the problems of pattern generation or visualization, but rather the analysis of the resulting patterns.

The main problem in analyzing gel patterns is to rapidly find changes in proteins from one experiment to the next. That is, is the set of proteins (as characterized by their position in two space on the gel) in gel A the same as in gel B? If it is not, then one must determine which protein has been added or is absent in gel A with respect to gel B. For this problem, the "FLICKER" program has been developed. This program allows the comparison of a number of gels (two at a time) by flickering them at various rates and magnifications. This facilitates interactively finding the spot changes. Once a protein appears to shift in either isoelectric point or molecular weight, it is necessary to prove that this shift is real and not due to movement of other proteins or gel distortion. Other facilities in the program allow the creation of "networks" of spots for comparison of sets of spots between gel images (see below). At times, it is important to determine the relative amounts of a particular protein and for this the program offers a densitometry facility.

Figure 1 shows an example of the type of difficulty encountered in analyzing pairs of gel images. The difference between the polypeptide patterns, in the *E. coli* strain with an amber mutation in the UDP-galactose transferase gene (T gene) and a strain with an amber mutation in the galactose kinase gene (K gene) consists of two polypeptide "shifts." The small shift is not easily apparent when viewing the two images or even when simply flickering them. What was required to detect this change was (1) to get an indication at normal magnification that something might be happening in the spot, and then (2) to view it at a higher magnification with variable flicker rates. The latter step was necessary to really visualize the change. In addition, at the higher magnification, the spot appeared as multiple spots. In an analysis to simply find all well-defined spots in the images, this relatively low-density two-spot cluster might have been counted as a single simple spot.

This is but one example of an intergel difference, rapidly brought out by the computer–human interaction, which would have been incomparably more difficult for the unaided human.

1.1 Analysis of Two-Dimensional Electrophoresis Gels

The analysis of the electrophoresis gel images entails several steps: (1) the location and identification of new or missing protein(s), and (2) the measurement of the amount of protein(s) present.

The first step is complicated by several problems. EP images may contain over 1000 proteins as well as artifacts. The gels may be distorted due to nonuniformity in the electrophoresis process (in complex nonsystematic ways) making exact overlay

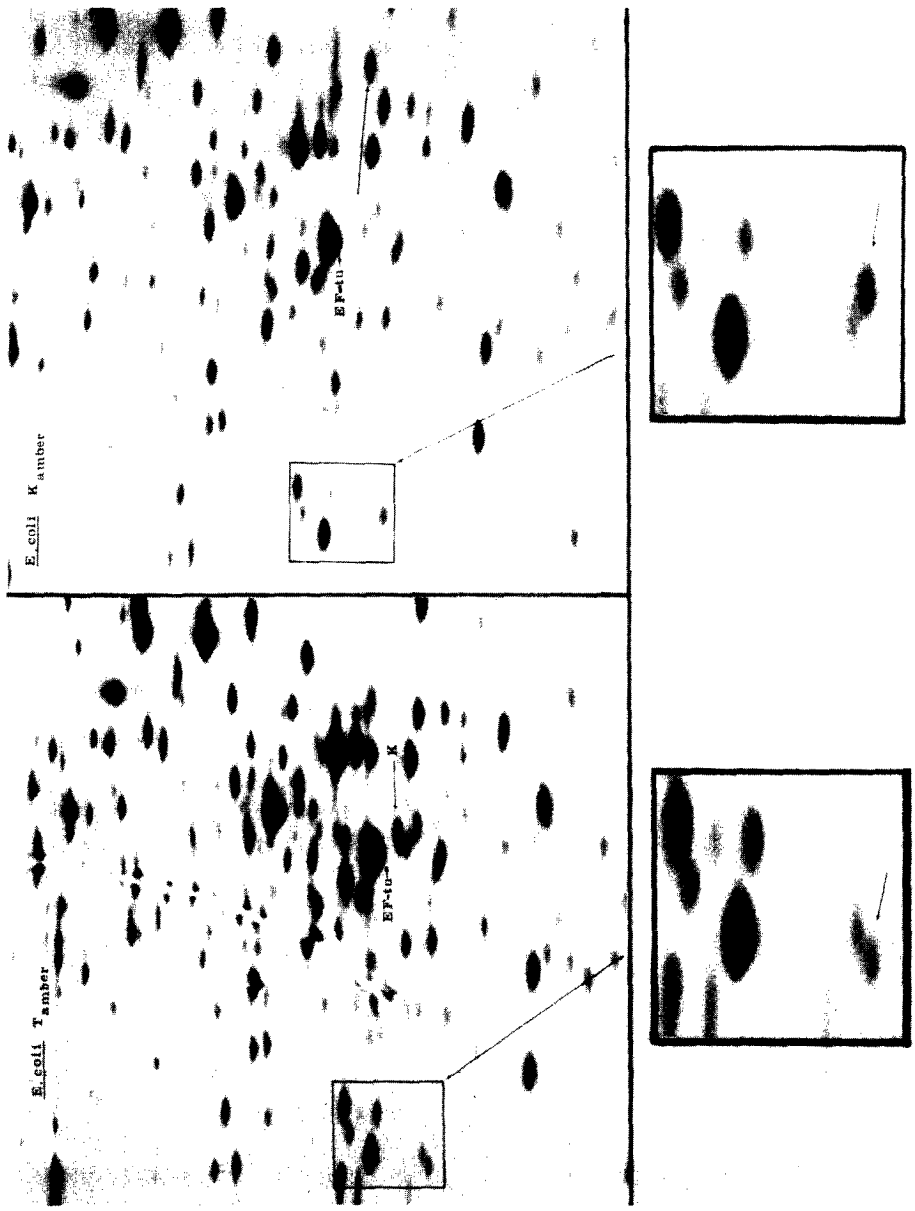

FIG. 1. Electrophoretic polypeptide patterns from *E. coli* strains containing cryptic lambda viruses, with temperature sensitive repressor products (a) T-amber pattern, (b) K-amber pattern. At 41°C the lambda "N" gene causes "escape synthesis" of the galactose operon genes as observed above. The kinase gene (K) is apparent at the initial magnification. Upon further magnification, an aberration in the position of a polypeptide in the strain with an amber mutation in the (T) gene is apparent. The position of the normal gal-kinase polypeptide, as noted in the (a) pattern, has been confirmed by antibody precipitation.

matching of an entire gel difficult. Despite these difficulties, gels may be manually aligned (by overlaying the films) to detect differences between local regions. An EP map for 140 *E. coli* proteins has been constructed by Pederson *et al.* (*3*). They have also proposed a 2D gel-labeling code based on (1) an isoelectric point zone name (zone A is acid and I basic) concatenated with (2) the molecular weight. Having defined this map for many of the proteins, one is able to reliably locate these protein patterns in a given sample even with considerable gel distortion. This may be done manually and in some cases automatically.

A gel image analysis system (*5*) is being used to develop protein maps automatically and interactively on a wide variety of biological materials. Ellipses are fit over the detected spots to approximate their boundaries, an interactive mode using a color display allows comparison of spots between gels. Geometric correction is done locally using a linear interpolation between short-range regions. The corrected images can then be used to produce protein maps containing position and amount of protein. These maps can then be compared between gels.

It is fairly easy to match well-defined protein patterns in two gel images. However, polypeptides may often be clustered so as to confound the edges of the proteins in the clusters. In addition, other protein spots may be very dilute so as to be missed or ambiguous. Proteins may also be arranged in continuous distributions whose beginnings and ends are difficult to define. Thus the problem domain may at times be too complex to fit protein objects with simple curve models. A further complication, when autoradiography is used, is that it is often necessary to view several exposures of the gel autoradiograph in order to cover the dynamic range of the radiation emitted by the radioactive atoms in the labeled polypeptides. Thus dark spots are more easily resolved in low exposures and light spots in longer exposures. Problems arise when light and dark spots overlap. Therefore a composite polypeptide map, consisting of the results of several exposures of the same gel, might be assembled to better describe a given gel.

The gels may be thought of as having been put through a D'Arcy–Thompson type transformation (*6*) where different points on the grid are distorted by a continuously differentiable 2D distortion function. One approach to the EP gel analysis is to remove the distortion (as is currently done with satellite images) and then perform the point-by-point comparison. This correction implies some knowledge of the inverse distortion function. Another approach might be to align the images (as well as possible) in order to detect minimum differences in a given local region, and then compute a 2D distortion function for the two images. A missing protein may be verified by using this technique if one observes low distortion in that particular region of the two gel images being compared.

The commonly used technique of autoradiograph labeling with ^{14}C, ^{35}S, or other isotopes (incorporated into the growth medium) permits approximate stoichiometric labeling of the proteins in the sample (as well as making them visible). The stoichiometry is, as in other forms of autoradiography, at least partially dependent

on the energy distribution of the particle (usually beta) which is emitted. Currently, the protein spots may be quantitated by cutting them out of the gel and measuring their radioactivity by scintillation counting. By measuring the optical density of the autoradiograph of a specific protein sample, it is possible to compute its relative concentration as defined by its radioactivity. Care must be taken however, to make sure that the film is used in the linear portion of the density versus log (exposure) curve. If it is not, or if the image acquisition system (for performing the computer analysis) does not operate in its linear region then the densitometry is invalid. In addition, the self-absorption of emitted particles may also complicate the interpretation of the densitometry to a greater or lesser extent.

There are a number of schemes to identify polypeptides found in the two-dimensional arrays with known polypeptides. First, any peptide can be cut out of the gel and digested with specific enzymes. The fragment pattern of the polypeptide can then be generated by either one- or two-dimensional analytic techniques. This pattern can then be compared with the pattern generated from the known peptide, digested with a similar enzyme. A second scheme utilizes the generation of antibodies with the known peptide. These antibodies can either be used to precipitate the unknown peptide out of the mixture prior to electrophoresis or the precipitated peptide can be resolubilized and electrophoresed with known markers so that its position can be determined. A third method of identifying an unknown peptide requires careful measurement of the peptide's apparent isoelectric point and apparent molecular weight in the electrophoretic gel and comparing these to the values obtained for the purified known polypeptide. Actually, a combination of one or more of these methods is best.

Other methods such as differential precipitation (as in the use of an ammonium sulfate concentration known to precipitate a specific enzyme) may be used (suggested by N. L. Anderson), or the use of different radioactive labels on different amino acids, for example. The spots could be cut out and counted in a scintillation detector for energy spectra to differentiate the labels. The density differences of the sample spots in the two gels can be computed and ratios formed for each spot (suggested by J. Garrells).

As demonstrated in this paper, the use of genetic mutations, particularly mutations which cause the loss of a peptide (such as nonsense mutations), in isogenic strains can also be used to localize a protein in the 2D map.

2. Materials and Methods

2.1. Specimen Preparation

Several ^{14}C labeled autoradiograph EP gel images were obtained from the Laboratory of General and Comparative Biochemistry. The gels are about 146 × 164 mm by approximately 0.8 mm thick and one or two gels are placed on the same 8 × 10 in. sheet of Kodak XRP X-ray film. This permits identical exposure and

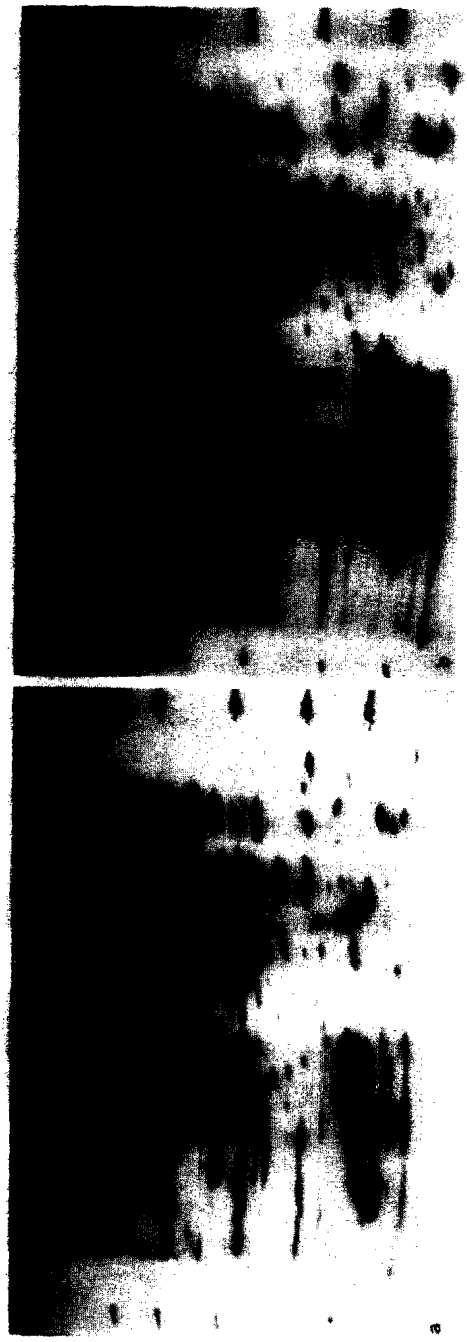

FIG. 2. Two-dimensional electrophoretic gel images. (a) *E. coli* is2+ sample 7. (b) *E. coli* is2 mutant 6SRNA polypeptide sample 8.

development times such that densitometric comparisons may be more favorably made. The 8 × 10 in. negatives also incorporate radiation density standards for calibrating protein concentrations when gels are prepared under conditions leading to valid stoichiometric assumptions.

Figure 2 shows two two-dimensional electrophetic gel images. Figure 2a is *E. coli* with is2+ sample 7 (4893), while Fig. 2b shows *E. coli* is2 mutant 6SRNA polypeptide sample 8 (4977).

A gel is calibrated by recognizing particular marker molecules whose isoelectric point and molecular weight are known and then interpolating the remaining molecules. Pederson (*3*) suggests six marker molecules listed in Table 1. Another six marker molecules are listed in the table used here by Merril. Figure 3 shows a radiation density standard incorporated into the autoradiograph.

TABLE I

MOLECULAR WEIGHT MARKER MOLECULES[a]

Pedersen's standard protein marker molecules	
MW (Daltons)	Molecule
155 000	Beta subunit RNA polymerase
84 000	EF-
65 000	S1
42 000	EF (major protein in *E. coli*)
40 700	Alpha subunit RNA polymerase
13 000	L7 and L12 ribosomal proteins
Merril's standard protein marker molecules[b]	
94 000	Phosphorylase b (rabbit muscle)
67 000	Albumin (bovine serum)
43 000	Ovalbumin (egg white)
30 000	Carbonic anhydrase (bovine erythrocyte)
20 000	Trypsin inhibitor (soybean)
14 000	Alpha-lactalbumin (bovine milk)

[a] Six molecular weight marker molecules suggested by (*3*) which might be used as calibration standards. An additional six molecules used by Merril are also presented. These marker molecules are inserted at either end of the gel during the second SDS stage of processing and serve to mark the molecular weight regions of the 2D gel (as seen in Fig. 2).

[b] Pharmacia kit—Pharmacia Fine Chemicals, division of Pharmacia, Inc., Piscataway N.J.

2.2. Image Acquisition

The autoradiographs were scanned using the buffer memory monitor image processing system, BMON2, of the National Cancer Institute's Image Processing

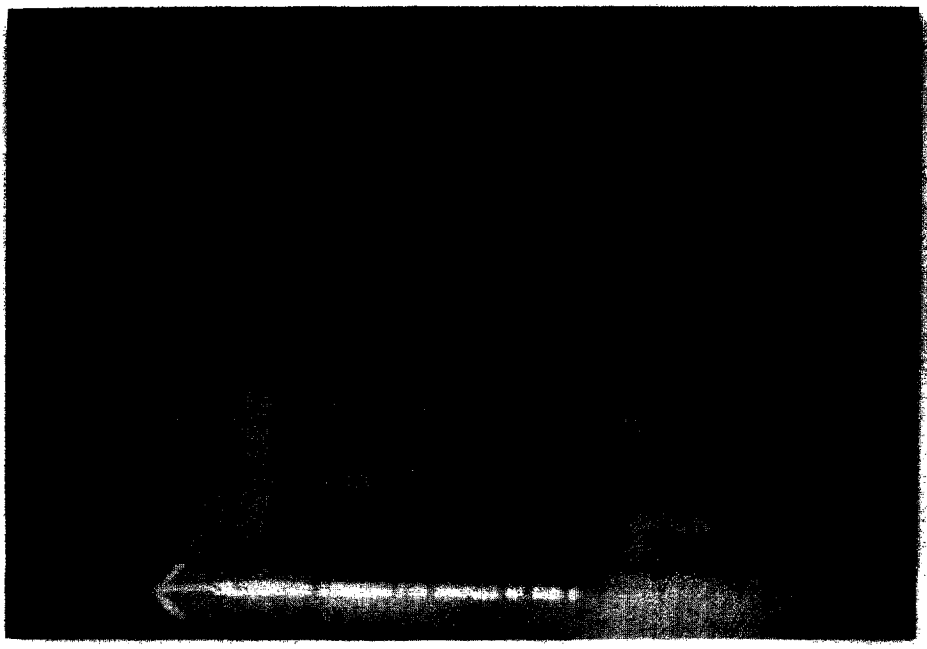

FIG. 3. Radiation density standard, consisting of six calibrated radiation samples, adjacent to gel 7 in autoradiograph image.

Unit (*7, 8*) running on the real time picture processor, RTPP, (*9–11*). Currently, the base computer of the RTPP is a Digital Equipment Corporation PDP8e with 32K words of core memory, four disk pack drives and two nine-track 800BPI magnetic tape drives. An Imanco Quantimet 720 (with shading corrector) uses a vidicon (or optionally, a Plumbicon) TV camera. Both cameras have 720-line resolution and the ability to rotate. Images may be acquired directly from the camera into a set of hardware image buffer memories. These image memories are also used for display (on the Quantimet display) and/or image transformation by the PDP8e computer. The Vidicon camera is coupled to a Nikon-N auto 1:2 28-mm close-up lens routinely set to f8 with the autoradiograph film mounted 69 cm away and backlighted on a Picker type 240096 light box. This light box has fairly uniform illumination.

Functionally there are four 512 × 512 picture element (pixel) gray scale images available to the system. Two of these are displayable and are denoted the "high" and the "low" images. Either the high or the low image may be displayed at any one time. The other two images are used to back up copies of the data in the high and low images. Each digitized 512 × 512 pixel image consists of four smaller image buffer memories each containing 256 × 256 pixels with 256 possible gray levels (although fewer are seen due to the lower dynamic range of the TV display). The dynamic range of the image processing system is white (for which the gray level is digitized at the value 0) to black (255). The RTPP hardware permits positioning the buffer

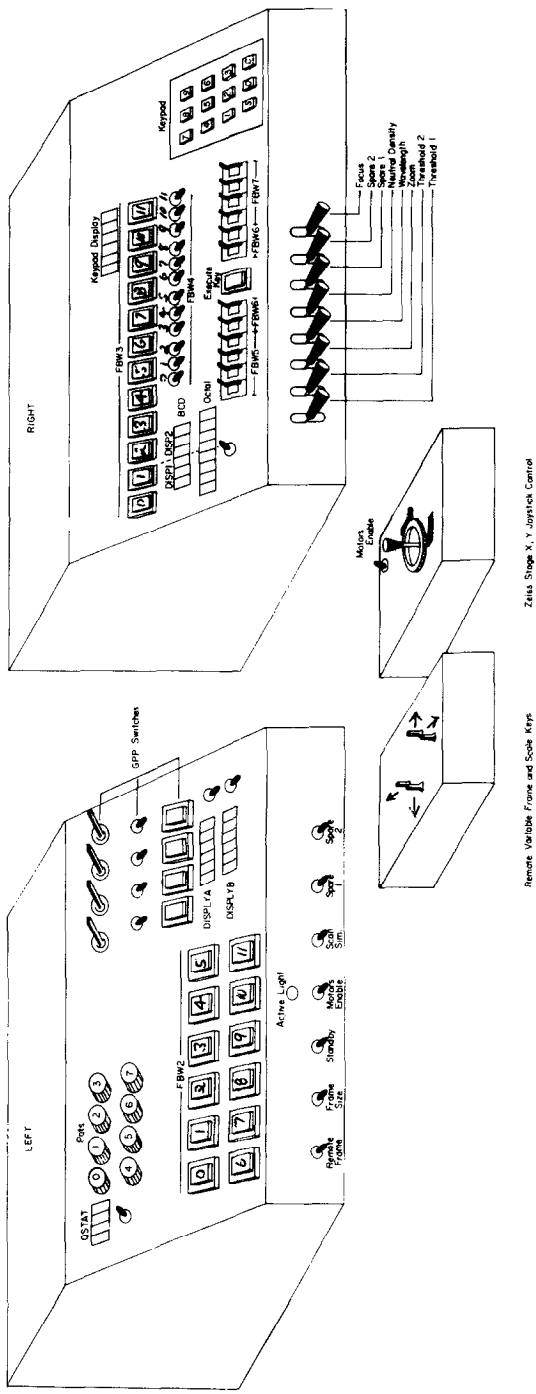

FIG. 4. RTPP control desk keys used in the FLICKER program: the PEN-command pushbutton keys (FBW2), FLICKER command pushbutton keys (FBW3), toggle switches (FBW4), potentiometers POT0–POT7, frame and scale position keys F&S, and the keypad.

memories independently at arbitrary locations on an 880 × 720 pixel screen area. This special characteristic of the RTPP system will be used to advantage for the two-dimensional gel analysis problem. The effective magnification of the image is about 0.25 mm/pixel (using the current lens and light box distance, although this is easily changed using alternative lenses and distances). Images were acquired using the BMON2 system and saved on magnetic tape for later analysis with the FLICKER (program) function of the BMON2 system.

The RTPP has an interactive keyboard shown in Fig. 4. This consists of among other components: 12 FLICKER-COMMAND pushbutton keys; 12 PEN-COMMAND pushbutton keys; 12 toggle switches; a moveable variable-size frame overlay for the TV display (denoted the variable frame and scale or F&S) which is manually adjustable using position keys which are read by the computer; a GrafPen (registered trademark of Science Accessories Corporation) spark tablet with TV cursor for pointing or drawing boundaries of objects; a keypad for entering numbers; and 8 knob-controlled potentiometers (denoted by POT0, POT1, . . ., POT7) which may be read by the computer.

2.3. Description of the FLICKER Program

The FLICKER program written in OS8 FORTRAN II runs under the BMON2 image processing system and is used to aid the analysis of 2D electrophertic gels. It operates in two basic modes: flicker mode for polypeptide difference detection and concentration measurement mode.

Images to be analyzed may be acquired three ways: (1) directly using the FLICKER program, (2) by preloading into the image buffer memories from various sources including magnetic tape or disks before running the FLICKER program, or (3) read in from image disks when used with the accession numbering system (see below).

2.3.1. Flicker subsystem

The flicker subsystem facilitates the comparison of parts of the two images by alternating the display of a low-image memory with a high-image memory. Various operations related to this function are relegated to the FLICKER-COMMAND keys. The underlying TV camera video image may be viewed (UNPOST command key) or either of the two acquired images may be viewed (via POST high-image or POST low-image command keys).

Pressing the FLICKER pushbutton command key to invoke flickering causes the high and low images to be displayed alternatively every 0.1 sec. The images are located on the TV screen in the logical coordinate system (LCS) (discussed in (*8*)). The sense of the LCS is such that $(x, y) = (0, 0)$ is the upper left-hand corner and (880, 720) is the lower right-hand corner. Because the stored 2D gel images are 512 × 512 pixels in size, they occupy a subregion of the available LCS region. During

flickering of the images, the F&S position keys are used to position the high image while the low image remains fixed in the center of the screen.

As it is sometimes desirable to leave the system flickering without having to continually press the FLICKER pushbutton key, toggle switch 0 may also be used to invoke continuous flickering. It may be desirable at times to slow the flicker rate in order to verify polypeptide differences possibly detected with the fast flicker rate. This facility is also built into the system. Toggle switch 1 changes the flicker rate (when flickering) from 10 frames/sec to a variable rate determined from potentiometer 0 (denoted POT0). This pot controlled rate varies from 10 to 0.017 frames/sec. An alternative variable rate mode is available during flickering using toggle switch 2. This mode controls the length of time the low image is displayed according to POT0 and the length of time the high image is displayed according to POT1. The variable rate facility has been found to be very useful in visualizing polypeptides with subtle differences (especially density differences) and will be discussed later in more detail.

Images may be acquired from the TV camera (512 × 512 pixel images) into either low- (GET LOW image command key) or high- (GET HIGH image command key) image memory at the current F&S position in the LCS. Normally, when toggle switch 3 is down, the frame size is set to 512 × 512 pixels, whereas it may be set to a 256 × 256 pixel size by setting the switch up. The 512 × 512 pixel F&S window is thus used to frame that part of the image to be acquired. When images are acquired, they are backed up into additional (nondisplayable) image memory for possible later use (for example, in RESTORing the images after MARKing, DRAWing, or LABELing polypeptides in them (see below)).

Alternatively, images previously acquired on magnetic tape (using the BMON2 "PIXMTA" function and naming conventions (*7*, *8*)) may be copied to scratch disks (1, 2, and/or 3) and from there randomly read into the currently posted image. The READ IMAGE command key requests an image accession number XXXX.E to be entered into the keypad (as E0XXXX). It then reads the corresponding image into the currently posted and its backup images. The accession number is defined by sequentially numbering gels with a four-digit number XXXX. Multiple exposures of a given gel are given sequential digits E (defaulting to value 1 if there is only one gel).

The correspondence between accession numbers and the images corresponding to them is made in an auxiliary ASCII data descriptor file GEL.DA. Table II illustrates a typical GEL.DA descriptor file. All of the images specified in a descriptor file may not be able to be stored on the three RTPP scratch disks (which can hold a maximum of 27 images). Therefore, attempting to read an image with a valid accession number but which is not stored on a disk will cause the system to notify the user which files to transfer to a disk from a specific nine-track magnetic tape. Thus the user is told *where* the missing image may be found. This automatic record keeping is especially useful when a large number of images are to be processed.

The Quantimet TV display has a seven-digit numeric display at the upper right

TABLE II

EXAMPLE OF GEL.DA DATA DESCRIPTOR FILE

```
ACCESS. #/BIRTHDAY/RACE & SEX/DATE OF EXPR/EXPR #/CULTURE REAGENT/
  INTRVL BEFR LABLING/LABLING ISOTOPE/DURTN LABEL/DURTN EXPSR/
  DIAGNOSIS FILE #/MAGTAPE #/OPT. BACKUP MAGTAPE #/CAMERA,
  LENS, DISTANCE*

0001.1/JOE AVERAGE/WM/12-22-25/21-09-78/#5/MEDIA 199/24 HRS/C14/3 HRS/
  1 WEEKS/NORMAL
  A0001/R123/R777/VIDICON, 28MM F8, 69CM*
0002.1/ED AVERAGE/WM/11-11-30/21-09-78/#5/MEDIA 199/24 HRS/C14/3 HRS/
  1 WEEKS/NORMAL
  A0005/R123/R777/VIDICON, 28MM F8, 69CM*
0003.1/MS. JACK SPRAT/WF/3-4-19/21-09-78/#5/MEDIA 199/24 HRS/C14/3 HRS/
  1 WEEKS/ABNORMAL
  A0009/R123/R777/VIDICON, 28MM F8, 69CM*
0004.1/JACK SPRAT/WM/4-5-23/21-09-78/#5/MEDIA 199/24 HRS/C14/3 HRS/
  1 WEEKS/ABNORMAL
  A0013/R123/R777/VIDICON, 28MM F8, 69CM*
0001.2/JOE AVERAGE/WM/12-22-25/21-09-78/#5/MEDIA 199/24 HRS/C14/3 HRS/
  3 WEEKS/NORMAL
  A0017/R123/R777/VIDICON, 28MM F8, 69CM*
0002.2/ED AVERAGE/WM/11-11-30/21-09-78/#5/MEDIA 199/24 HRS/C14/3 HRS/
  3 WEEKS/NORMAL
  A0021/R123/R777/VIDICON, 28MM F8, 69CM*
0003.2/MS. JACK SPRAT/WF/3-4-19/21-09-78/#5/MEDIA 199/24 HRS/C14/3 HRS/
  3 WEEKS/ABNORMAL
  A0025/R123/R777/VIDICON, 28MM F8, 69CM*
0004.2/JACK SPRAT/WM/4-5-23/21-09-78/#5/MEDIA 199/24 HRS/C14/3 HRS/
  3 WEEKS/ABNORMAL
  A0029/R123/R777/VIDICON, 28MM F8, 69CM*
          .
          .
          .
```

hand corner. It contains (while in flicker mode) the accession number of the image currently loaded into the posted image. The accession number XXXX.E is shown as 0E0XXXX. The high-image number is initially 0222222 and the low-image number is initially 0111111. Thus it is possible to check at any time which images are currently loaded into the image buffer memories (using the READ IMAGE command). When in pen mode and the pen is active, the least significant three-digits of the numeric display contains the gray value of the currently posted image pixel pointed to by the GrafPen.

A portion of an image may be magnified two or four times from its backed up copy using the ZOOM command key. A 256 × 256 pixel subregion of the currently posted image is specified by the position of the upper left-hand corner of the F&S

which may be positioned anywhere on the screen. The F&S may be temporarily set to a 256 × 256 size for selecting a subregion to zoom by putting toggle switch 3 up and to 64 × 64 with switch 8 up. The zoom algorithm expands each pixel in the original image into a 2 × 2 or 4 × 4 pixel square copy. Thus the 256 × 256 pixel subimage zoomed by 2× or the 64 × 64 pixel subimage zoomed by 4× fills the 512 × 512 pixel image. It should be noted that the zoomed image is created from the backup copy of the currently displayed image but replaces the currently posted image.

Thus command keys associated with image acquisition, posting and unposting and flickering are grouped together. These FLICKER-COMMAND subsystem operations are listed in Table III.

TABLE III

FLICKER/COMMAND KEY MENU

Key menu	Function
0	FLICKER between the low and high images
1	UNPOST from the TV the currently posted image (if any was posted)
2	POST the low image (unposting the high if it was posted)
3	POST the high image (unposting the low if it was posted)
4	RESET standard F & S and high and low image positions, print FLICKER-COMMAND key menu, backup the currently defined distance matrix
5	TOGGLE between PEN-COMMAND and FLICKER-COMMAND mode printing the options (Table III or IV) currently available
6	RESTORE backed up high and low image from secondary image memories
7	GET LOW image at F & S window and back it up in secondary image memories
8	GET HIGH image at F & S window and back it up in secondary image memories
9	READ IMAGE specified by the accession number XXXX.E entered into the keypad from the image disks into the currently posted image and its backup image
10	ZOOM by 2× the 256 × 256 pixel region denoted by the uper left hand corner of the F & S window
11	EXIT to the BMON2 image processing system

2.3.2. Pen subsystem

In addition to the set of FLICKER-COMMAND pushbutton keys, there is a subsystem called the PEN-COMMAND pushbutton keys. This subsystem is entered and left by pressing the TOGGLE pen subsystem command key. The appropriate menu (Table III for FLICKER- and Table IV for PEN-command mode) is printed on the teletype to show the options currently available.

The PEN-COMMAND subsystem, listed in Table IV, offers various GrafPen related interative operations. These include marking points in an image, comparing distances between points and between successive images, extracting objects by encircling them with the pen, defining a calibration standard for use in measuring features, and labeling text.

TABLE IV

PEN-COMMAND KEY MENU

Key menu	Function
0	DRAW—enable the GrafPen for drawing a boundary (boundary drawing mode) in the currently posted image (high or low) by erasing any previous drawing
1	ERASE the currently drawn boundary backwards from the last point drawn if boundary drawing mode in active
2	CONNECT the current set of marked points with white lines in the currently posted image
3	DISTANCE—Print all labeled points (x, y, density, image name), then compute and print the distance matrix D_{n+1} of all points taken two at a time
4	DIFFERENCE—Compute and print relative difference matrix RD of current distance matrix D_{n+1} and previous distance matrix D_n
5	CLOSE the boundary being drawn, then compute the maximal hull of the boundary and print features of the image inside of the boundary
6	MARK—Enable the MARK mode after which pressing the pen records up to 18 labeled points in the currently posted image
7	CALIBRATE GEL—Compute the density standard calibration by acquiring samples q through 6 and then compute the calibration regression line. The plot of log (counts/min) against log (total density) is then drawn
8	LABEL NAME—Draw the associated image name information (if it exists) at the bottom of the currently posted image
9	LABEL TEXT—Draw the specified text (entered from the teletype) preceded by a line drawing (previously started using the DRAW key) in the currently posted image
10	LABEL TEXT BOTH is the same as LABEL TEXT but draws the image into both images
11	CALIBRATE WEDGE—Calibrate the TV camera in terms of NBS ND step wedge

The four submodes of the pen system include: marking spots, boundary drawing, image labeling, and density calibration modes. They are now presented in greater detail.

Mark mode. Using the GrafPen in MARK command key mode, a posted image may be labeled with up to 18 white sequential letters "A" through "P". The label is shown within a white 14 × 14 pixel square. The next sequential letter is position and displayed each time the pen tip is pressed in this mode.

The (x, y) positions of the labeled points are saved in a list. As points are generated, the upper triangular matrix D of the absolute Euclidian distances between all points in the list is generated using Eq. (1.1). The matrix may be printed (as well as the point parameters $[x, y]$) using the DISTANCE command key. After a difference matrix, D_n, is generated for one posted image n, the matrix may be saved and the marked images restored from the unmarked copies of the image using the RESTORE command key. Alternatively, the RESET command key may be used to save D_n but not to restore the images. This increments the counter n. After acquiring D_n and D_{n+1}, the upper triangular relative difference matrix, RD, may be computed as in Eq. [1.2] and printed using the DIFFERENCE command key

$$d(i,j) = ((x_i - x_j)^2 + (y_i - y_j)^2)^{1/2}, \qquad \text{all labels } i,j, \qquad [1.1]$$

$$rd(i,j,n+1) = d(i,j,n+1) - d(i,j,n), \qquad \text{all labels } i,j. \qquad [1.2]$$

It is possible to draw the transitive closure graph (i.e., "network") of the set of marked polypeptides (which is equivalent to the distance matrix D_{n+1}). This is done using the CONNECT command key which draws and displays white lines between all marked points superimposed on the original image points. Comparing the graphs between two gels gives a rough visual idea of the directions and extents of distortions between the two images.

Boundary drawing mode. Before discussing boundary mode operations we first define the property "maximal hull." The maximal hull is similar to the convex hull

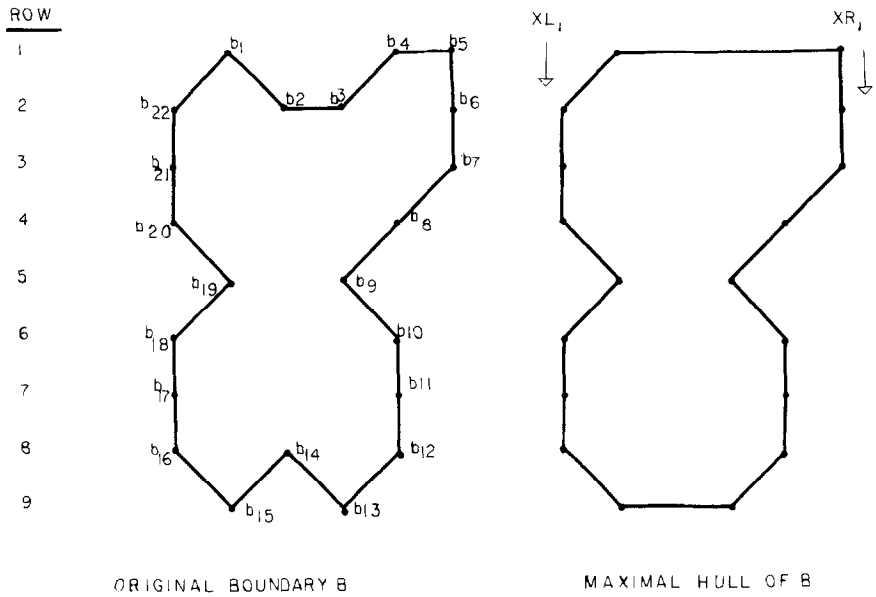

FIG. 5. Illustration of the maximal hull of an object concave on all four sides. The maximal hull is an approximation to an object obtained by finding the maximum width of the object at each line i ($X_L i$ and $X_R i$).

but is asymmetric. However, for drawing convex boundaries around objects it is simpler to compute and not subject to error. The maximal hull is defined by equations [2.1]–[2.2] and is illustrated in Fig. 5. Let a boundary B be a circular sequence of points $(b_1, b_2, \ldots, b_p)$ or $((x_1, y_1), (x_2, y_2), \ldots, (x_p, y_p))$. Then the left and right endpoint arrays (X_L and X_R) are initialized by

$$\begin{aligned} X_L(y) &= \text{infinity}, \qquad \text{all } y, \\ X_R(y) &= -1, \qquad \text{all } y. \end{aligned} \tag{2.1}$$

Then the endpoint arrays are defined from the boundary points as,

$$\begin{aligned} X_L(y_i) &= \text{Min}(X_i, X_L(y_i)), \qquad \text{all } b_i, \\ X_R(y_i) &= \text{Max}(X_i, X_R(y_i)), \qquad \text{all } b_i. \end{aligned} \tag{2.2}$$

A boundary drawing mode, activated by the DRAW command key, is alternatively available which permits drawing the maximal hull around a region. The CLOSE command key closes the partially drawn boundary and prints an identifier number in "|...|" followed by various features. The identifier is normally incremented by one each measurement. However, it may be optionally specified as a three-digit number from the keypad.

Given the computed endpoint arrays X_L and X_R, various features may be computed. These include: polypeptide area, total nominal density, total density/area, perimeter, perimeter-squared/area, counts-minute of sample (if calibrated—see below), minimum enclosing rectangle around the polypeptide sample, the first-order moments, and the absolute range of pixel density in the sample. This last feature is useful for checking on the possible saturation of the image acquired by the TV system. By total nominal density of a region we mean the sum of the individual pixel relative densities (if calibrated using the CALIBRATE WEDGE command—see below—if not it uses gray values) for that region.

The actual total nominal density measurement must be corrected for the background density. This is done by estimating the background equivalent density for the same area of the sample by [3.1]–[3.2]. Let D_i and A_i be total density and total area features of for a polypeptide sample i, D'_{sample} is the corrected density

$$d = \frac{D_{\text{background}}}{A_{\text{background}}}, \tag{3.1}$$

$$D'_{\text{sample}} = D_{\text{sample}} - (d \times A_{\text{sample}}). \tag{3.2}$$

Image labeling mode. The associated image name of an image loaded using the READ command key may be drawn in the bottom of the currently posted image using the LABEL NAME command key.

It is often useful to label a region with a line drawing (possibly point to or encircling a spot) with associated text. The LABEL TEXT command key is used instead of the CLOSE command key *after* entering the line drawing using the DRAW command key. It requests text from the operator via the teletype and then draws it to the right of the last point drawn with the GrafPen.

The LABEL TEXT BOTH command key performs the same operation as LABEL TEXT but draws the same line drawing in both images. The drawing is done in the unposted image at the relative position *previously determined by flicker alignment.* In addition, two requests for text are given so that different labels may be specified for the two labeled images. This feature is especially useful for labeling regions in gels where polypeptides are absent.

Drawing in a posted image destroys the unposted image (which actually contains the line drawing). Thus the unposted image normally cannot be used later without first doing a RESTORE operation. Optionally, the system can save the unposted image before using it for drawing. This mode of operation swaps the unposted image to disk 3 when required and restores it when needed.

Density calibration mode. The total nominal density measurement (made with the DRAW/CLOSE command keys) may be mapped to a counts/minute value if several known radiation density calibration samples are included in the exposure of the autoradiograph negative (as illustrated in Fig. 3). Alternatively, total nominal density measurements may be made using a special mode of operation. This special mode, although correcting for background density using Eq. [3.2], does not report results in counts/minute. Rather it reports total nominal density.

The normal calibration standard we use has six approximately circular spots of known radiation intensity. The radiation density standard values are listed in Table V with sample 1 being the normal background of the film. The FLICKER program also permits arbitrary standard density values to be entered instead of our default values.

TABLE V

RADIATION CALIBRATION STANDARDS

Calibration sample	Counts/minute (CM_i)
1	5.10×10^1
2	1.00×10^2
3	7.00×10^2
4	6.19×10^3
5	5.80×10^4
6	6.00×10^5

[a] Radioactive ^{14}C labeled compounds covering a range of concentrations were deposited on the gel after its creation but before autoradiograph exposure. Writing on the gel was done using india ink with a ^{14}C compound added.

In normal operation, the CALIBRATE GEL command key computes the density standard calibration for samples q through 6 where the starting sample number q may be defined at run time (the default is 1). The first sample, whether in normal or special mode, is a blank background region. The CLOSE command key is used to enter each GrafPen encircled sample's total nominal density denoted by the drawn boundary. For normal operatin, a linear regression is applied to denisty samples $q + 1$ to 6 to fit a linear equation in log–log space of the log (counts/minute) as a function of the log (total sample density) given in [4]. Note that the background sample is not included in the regression but is used for computing the corrected density D' given by [3.2]. The coefficients of the linear equation M and B are given in [5.1]–[5.2]

$$\log(CM_i) = (M \log(D'_i)) + B. \qquad [4]$$

Let lcm be $\log(CM_i)$ and ld be $\log(D'_i)$ for the ith sample, then

$$M = \frac{(k(\text{SUM}(lcm\ ld)) - (\text{SUM}\ lcm)\ (\text{SUM}\ ld))}{(k(\text{SUM}\ ld^2) - (\text{SUM}\ ld)^2)}, \quad [5.1]$$

$$B = \frac{(\text{SUM}\ lc) - \text{M}(\text{SUM}\ ld)}{k} \quad [5.2]$$

The regression coefficient, r, is given in (6).

$$r = \frac{(k(\text{SUM}(lcm\ ld)) - (\text{SUM}\ ld)^2)}{(k(\text{SUM}\ ld^2) - (\text{SUM}\ ld)^2)\ (k(\text{SUM}\ lcm^2) - (\text{SUM}\ lcm)^2))^{1/2}}. \quad [6]$$

After the regression line is computed, its plot, along with the calibration samples labeled as numbers $q + 1$ to 6 inside of squares, is displayed. At this point in the program, arbitrary regions of the gel images (taken on the same autoradiograph negative) may be measured in counts/minute using the boundary drawing mode (DRAW and CLOSE command keys). Because the system removes background density from a sample using Eq. [3.2], excess background density surrounding an enclosed sample is not counted in the total density. However, this area is reported in the area measurement.

Since the full calibration procedure leaves the calibration plot in the previously unposted image, the operator should post it again or restore both images before making measurements by using the RESTORE command key.

The TV camera measures light intensity in terms of gray value. This feature should be mapped to a neutral density (ND) value relative to a given standard. The standard selected is the National Bureau of Standards ND-step wedge type 1009. This wedge consists of 21-ND steps on a film emulsion which cover the range of 0.05 to 3.0 ND. This range of course is much greater than the dynamic range of either a Plumbicon (0.0 to about 1.3 ND) or a vidicon (0.0 to about 2.0 ND).

The CALIBRATE WEDGE command calibrates the TV camera–lens–illumination system in terms of neutral density. The operator positions a 20-pixel-high F&S window over the NBS standard step wedge. The wedge is horizontally oriented in the bottom of the gel image with the 0.05-ND step on the left. The right edge of the F&S window is positioned to include the highest neutral density value in the wedge expected from the gel data. The gray level histogram of the wedge image is then computed and the histogram maxima (corresponding to the neutral density wedge values) extracted. This set of corresponding gray levels and neutral density values $\{G_i, ND_i\}$ is then used to generate a piecewise linear interpolation to the neutral density function of gray level relative to the standard step wedge. The gray level histogram and piecewise linear interpolation function are then plotted at the bottom of the image as a function of gray value. Polynomials were also fit to the data and correspond very well to the piecewise linear model in the midrange of the gray

levels where gel data is expected to be taken. All gel images to be measured should be acquired at the same time that the neutral density step wedge image is acquired. Furthermore, an additional neutral density wedge image might be acquired at the end of the image acquisition session and compared with the initial wedge image to determine if any measureable drift in the TV system has occurred.

3. Results

3.1. Use of Image Flickering to Find Polypeptide Differences

The gel images shown in Fig. 2 were flickered. Sample 8 was moved while sample 7 was held stationary. By moving one image against the other, the TV display (with less than about 0.5-sec phosphor decay time) eye–brain system performs a type of differencing of the images. Prominently shaped sets of polypeptides were used in the alignment process which terminated when a local set of polypeptides (around six or so) was optimally aligned and (a) no differences were noted or (b) one or more gross differences were noted characterized by a blinking of the additional polypeptide(s). This flickering technique cannot of course be illustrated here.

A set of objects in optimal registration appears to expand and shrink depending on the size and shape differences of the same objects in the two images. Images slightly out of registration will yield a motion in the direction of the misregistration indicating to the operator how to align the images. Images greatly out of alignment appear as a confusion of images with no obvious alignment clues offered. In such a situation, the operator turns off or slows down the flickering to view the images one at a time in an attempt to find landmarks. These are then used to bring the images into approximate registration before resuming the higher rate of flickering. This procedure allows approximate alignment so that the eye–brain system can detect the direction of change to optimize the alignment.

After a suspected difference has been found using the fast flickering mode, it is useful to slow down the flicker rate to about 1 frame/sec in order to verify the existence of a polypeptide difference. This is done (as discussed before) by invoking the variable rate flickering from POT0. After determining which image has the additional polypeptide, the difference may be further enhanced using the differential flicker rates (using POT0 and POT1). The gel image without the additional polypeptide is displayed longer than the gel image with it. This gives the operator time to verify the absence of the polypeptide from the former gel.

Figures 6a and b show an electronically zoomed (by a magnification of 2×) region where there is a polypeptide difference between samples 7 and 8. Note the adjacent polypeptide positions. These may be labeled for comparison between the images.

3.2. Use of Image Distance and Difference Matrices

The region around the missing polypeptide was investigated using the distance and difference matrices D and RD. Figure 7a and b show the labeled polypeptides for the

FIG. 6. Region of the image where there is a polypeptide difference. (a) Sample 8 with additional polypeptide zoomed 2×, (b) sample 7 with missing polypeptide zoomed 2×.

two images in Figs. 6a and b. Figures 7c and d show the upper left-hand quadrant of Figs. 7a and b zoomed again by a magnification of 2×.

Tables VI and VII show the (x, y) coordinate pairs and distance matrices D_1 and D_2 of samples 8 and 7, respectively, while Table VIII shows the difference matrix RD

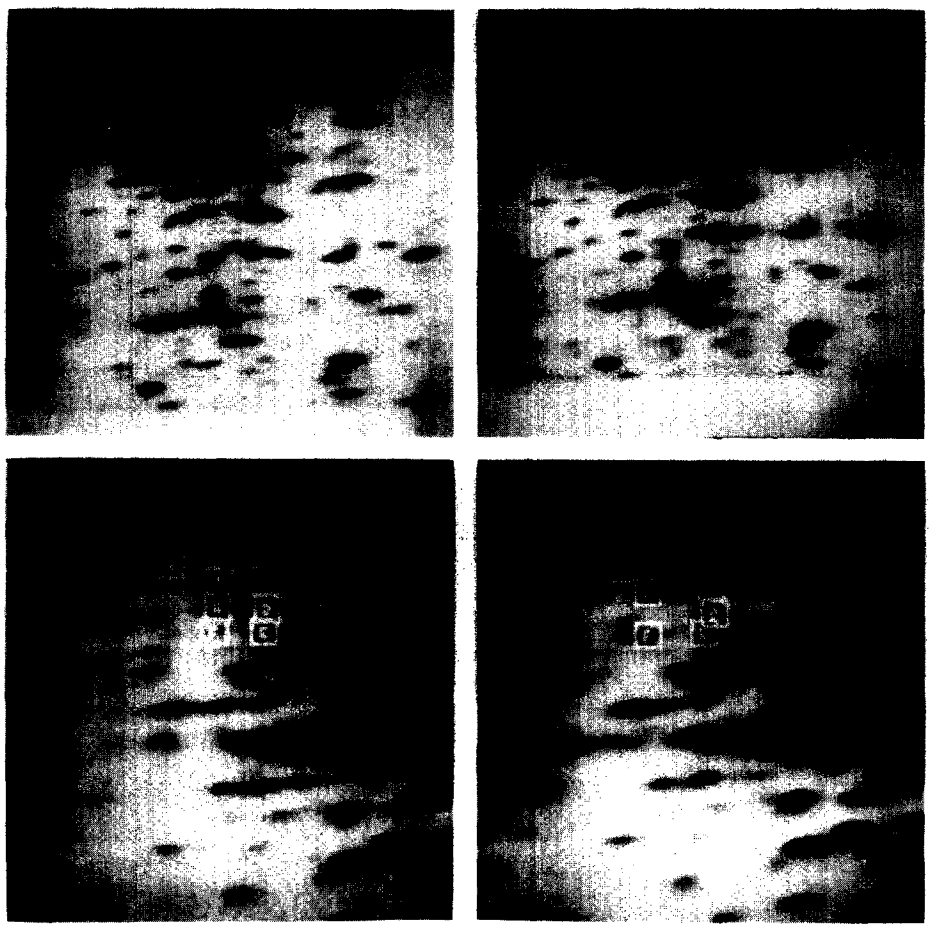

FIG. 7. Labeled region of the 2× zoomed images where there is a polypeptide difference. (a) Sample 8 with additional polypeptide (b) sample 7 with missing polypeptide, (c) upper left quadrant of (a) zoomed by additional 2×, (d) upper left quadrant of (b) zoomed by additional 2×.

of $(D_2 - D_1)$. The difference matrix shows minor distortion differences between polypeptides A–F except for a minor shift in polypeptides A and F. Thus the new polypeptide G is not a mislabeling of one of the adjacent polypeptides due to local distortions.

The relative distortion of a region in two gels may be estimated visually using the following procedure. First mark a central point of interest in one of the images. Then

TABLE VI

DISTANCE MATRIX D_1 OF LABELED POLYPEPTIDES IN SAMPLE 8

Polypeptide	x (in pixels)	y (in pixels)
A	98	68
B	123	78
C	150	77
D	152	92
E	150	105
F	123	104
G	124	91 (New polypeptide*)

Distance matrix D_1 between polypeptide pairs (in pexels)

	A	B	C	D	E	F	G
A	0	26	52	59	63	43	34
B	0	0	27	32	38	26	13
C	0	0	0	15	28	38	29
D	0	0	0	0	13	31	28
E	0	0	0	0	0	27	29
F	0	0	0	0	0	0	13
G	0	0	0	0	0	0	0

TABLE VII

DISTANCE MATRIX D_2 OF LABELED POLYPEPTIDES IN SAMPLE

Polypeptide	x (in pixels)	y (in pixels)
A	83	62
B	98	74
C	131	70
D	133	86
E	129	96
F	96	97

Distance matrix D_2 between polypeptide pairs (in pexils)

	A	B	C	D	E	F
A	0	19	48	55	57	37
B	0	0	33	37	38	23
C	0	0	0	16	26	44
D	0	0	0	0	10	38
E	0	0	0	0	0	33
F	0	0	0	0	0	0

TABLE VIII

DIFFERENCE MATRIX $RD\,(D_2 - D_1)$

Difference matrix of $(D_2 - D_1)$ (in pixels)						
	A	B	C	D	E	F
A	0	−7	−4	−4	−6	−6
B	0	0	6	5	0	−3
C	0	0	0	1	−2	6
D	0	0	0	0	−3	7
E	0	0	0	0	0	6
F	0	0	0	0	0	0

draw the connected graph using the CONNECT command. Do this for the other image as well. Then flicker the two images aligning the central marked points. The expansion and contraction of the lines indicates the direction of the local distortions between the regions. Figure 8a and b show the connected graphs for the two samples. Figure 8c shows a time exposure taken of the flickered connected graph images after alignment of the central marked point.

3.3. Use of Densitometry to Measure Polypeptide Concentrations

The radiation density standard calibration image in Fig. 3 was used to calibrate the system with standard samples 2 through 6 used in the regression. The regression analysis produced an (M, B) parameter pair (from Eq. [7.1]–[7.2] of (2.82,–15.26) with a regression coefficient value of 0.966. The regression plot is shown in Fig. 9.

TABLE IX

CONCENTRATIONS OF THE SEVEN POLYPEPTIDES

Polypeptide	Sample 7	Sample 8 (×1000 counts/min)
A	23.2	60.3
B	19.5	30.4
C	10.9	63.5
D	5.9	41.4
E	4.1	20.2
F	3.1	10.1
G	–	34.2

[a] Concentrations of the seven polypeptides A–G measured after calibration. These measurements have a monotonic relationship between gel images. Exact comparison is difficult because sample preparation was not performed to ensure stoichiometry.

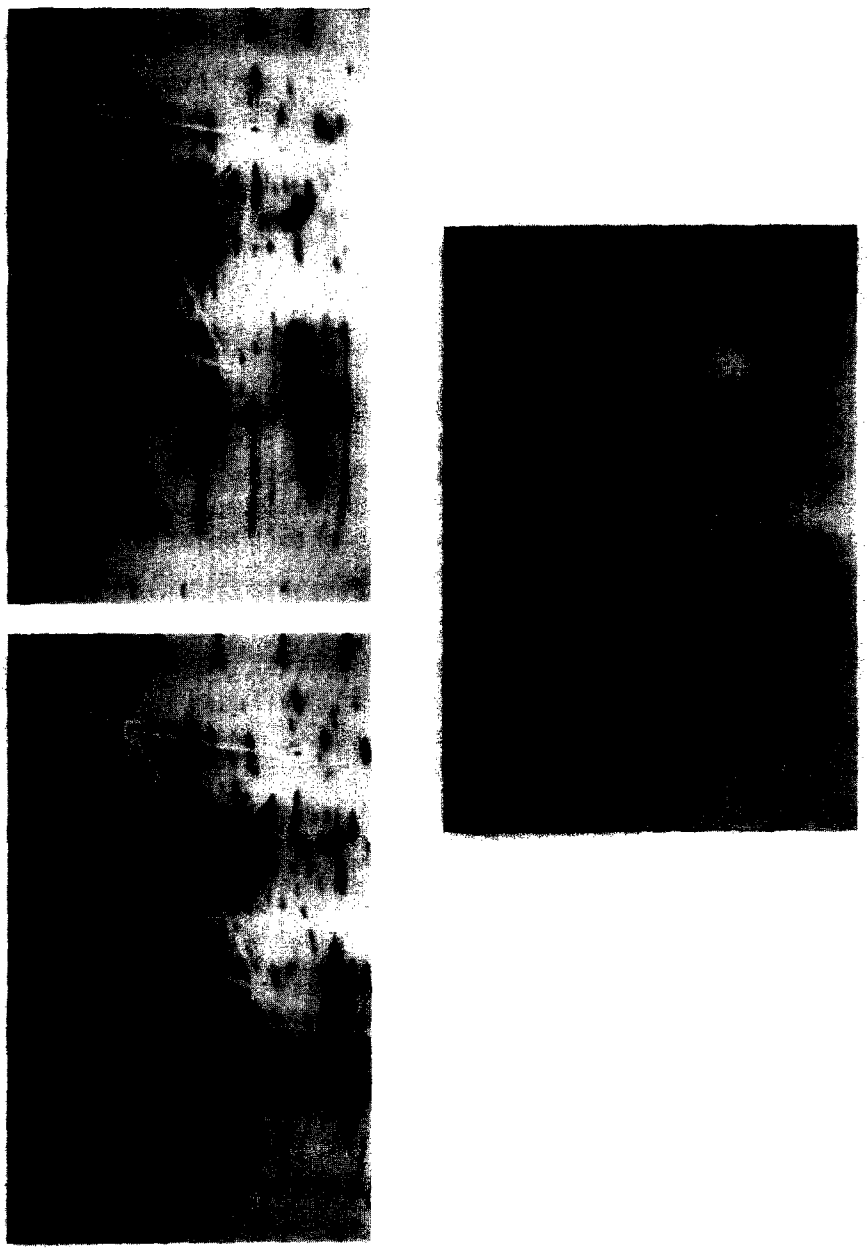

FIG. 8. Connected landmark "marked" points for complete graph of the two gel images. (a) Five connected points in sample 8, (b) same five points connected for sample 7, (c) 1-sec time exposure of 1/10-sec flickering images (a) and (b) centered at middle point.

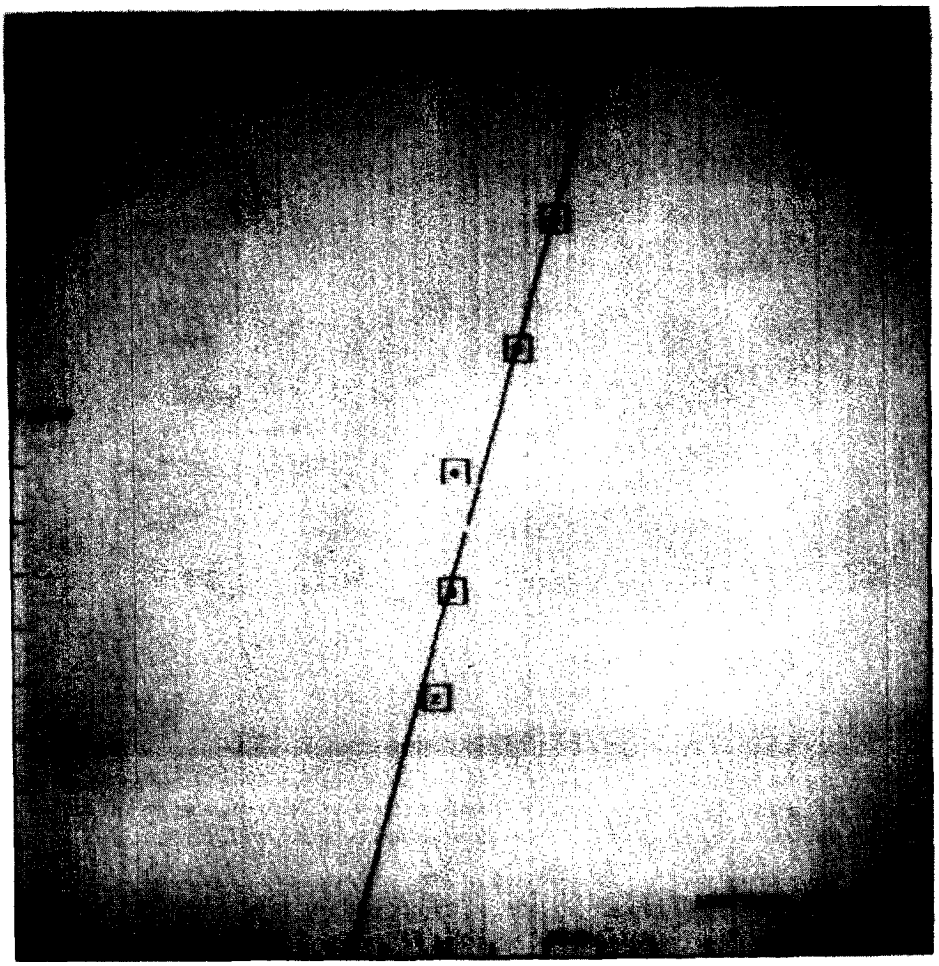

FIG. 9. Calibration plot of regression on calibration data for radiation density standard samples 1 through 6.

Figure 10 shows the boundary drawn around a polypeptide during feature extraction. Table IX lists the concentrations for polypeptides A through G for the two gel images measured after the calibration. Because the sample preparation method used was not stoichiometric, the same polypeptides do not have the same concentrations in the two samples. However, they do have a monotonic relationship.

4. DISCUSSION

The FLICKER system as discussed in this paper is an evolving tool. At this stage it is primarily concerned with interactive detection of polypeptide spot differences between gel pairs. However, it is apparent that additional tools are needed.

FIG. 10. Circled boundary of polypeptide to be measured. Features for this polypeptide to be measured. Features for this polypeptide include: counts/minute = 68.0 Kcounts/min, drawn sample = 297 square pixels, total nominal density' = 26057, drawn sample perimeter = 172 pixels.

The problem of polypeptide identification, boundary generation and regional densitometry is well recognized in other areas of image processing and presents no really unique aspects in the processing of 2D electrophoretic gels. Just as analytic (e.g., cytologic) substrate quantitation via digital densitometry is dependent on antecedent stoichiometric staining (rarely attained), the same constraints apply to the present problem.

All known methods of polypeptide detection (staining fluorescence, radioautography) produce background noise. This plus the fact that a polypeptide spot is distributed in two space (due to variances) rather than point-like, makes algorithmic boundary generation (and spot area definition) a more than trivial problem. These various sources of difficulty combined with limited computer resources have restricted our first efforts to manual definition of spot boundaries (using techniques such as the DRAW options described). There is some reason for guarded optimism with respect to automatic spot boundary definition for densitometry, as, for example, if we are not generally faced with examples of spots with internal structure—i.e.,

spots composed of more than one polypeptide. The analogous problem is cytology has led to greater difficulty and is not yet generally solved. Since any estimation of an overlapping spot's boundary incorporates material from the polypeptide(s) adjacent to it, such estimates will have errors difficult to reliably estimate and correct.

The dynamic range of the visualization of an autoradiograph gel is insufficient to discern both very high and very low concentrations (of polypeptides) in a single exposure. It would be useful to construct a composite image of a range of exposures. Since the same gel is used to make all exposures, the alignment problem is minimized. The alternatives are (1) an extended gray scale image could be constructed to reflect the sum of a number of gels taking the exposure of a given gel by a multiplicative scale factor, or (2) spots extracted from each gel could be merged. The latter method would seem to be the most amenable to automation.

It would be useful to interpolate molecular weight and isoelectric focus values directly from the image. Thus an extracted polypeptide could be labeled directly by its "apparent" (*p*I, MW) indices rather than working in pixels. Such a facility would require a set of known landmark polypeptides which have been previously calibrated, as in (*3*, *5*). Problems with resolution, nonuniform spacing between standard polypeptides, and geometric distortion may add some difficulty to this procedure.

It would be desirable to incorporate, into the interactive system, a polypeptide location map of the current biological system being investigated. Knowing the locations of previously identified polypeptides would enable (1) better understanding of a particular subregion, (2) better estimation of how much distortion a given gel exhibits, and (3) eventual cataloging of entire regions. Several workers (*3*, *5*) suggested the description of a polypeptide's position in the gel by (linearly) interpolating the position of the unknown polypeptide to that of known marker polypeptides.

The geometric correction from gel image to another based on aligning local landmarks and correcting locally would aid considerably in comparing gels. Two approaches may be taken: that of local geometric corrections for small regions using simple linear interpolation techniques (*12*) or global geometric correction involving more complex nonlinear interpolations over larger regions. Unless an adequate number of landmark polypeptides is available a global model would be required.

Flickering images at high rates requires the use of raster image frame buffers which may be changed on alternate frames. The RTPP hardware is not unique in this respect in that various frame buffer hardware is currently available commercially. However, the other central feature required for interactive flickering—the ability to quickly reposition the image frame buffers within the viewing area of the display on every display cycle—is unique. This feature is implemented by adding another level of memory address mapping to the image frame buffer display hardware and would be a useful feature if offered commercially.

The ability to interactively label polypeptides *and* regions with line drawings and text is invaluable in documenting polypeptide differences detected by flickering.

There is a need to compare more than two images in an interactive session. For example, in searching for a polypeptide difference, it is useful to compare several control with several abnormal gels. This can be done by comparing two images at a time, as long as there is the ability to easily switch between pairs of images. In the FLICKER program, this is implemented by keeping these images on a set of random access disks and using an accession number driven paging scheme to access the images upon request.

Further enhancement of the record keeping would be useful. This might include keeping track of the occurrence or absence of spots and their positions in particular gels. The data base file, GEL.DA, (previously discussed) could be searched using additional software to generate a set of accession numbers of images to be compared. For example, it might match all control gels meeting age, set, or other criteria with all corresponding abnormals. This list of accession numbers could be combinatorially expanded to drive the FLICKER program or other gel comparison program.

The program FLICKER, described above, provides an initial interactive tool for the detection and measurement of individual polypeptide differences between comparable EP gels.

REFERENCES

1. O'FARRELL, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250,** 4007 (1975).
2. LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680 (1970).
3. PEDERSEN, S., BLOCH, P. L., REEH, S., AND NEIDHARDT, F. C. Patterns of protein synthesis in *E. coli*: A catalog of the amount of 140 individual proteins at different growth rates. *Cell* **14,** 179 (1978).
4. ANDERSON, N. L., AND ANDERSON N. G., High resolution two-dimensional electrophoresis of human plasma proteins. *Proc. Nat. Acad. Sci. USA* **74,** 5421 (1977).
5. ANDERSON, N. G., AND ANDERSON, N. L., Molecular anatomy. In "Behring Inst. Symposium 1977." Mitt. Behring Inst., 1979.
6. THOMPSON, D. H., "Growth and Form." Vols. 1, 2, Cambridge Univ. Press, New York, 1959.
7. LEMKIN, P. "Buffer Memory Monitor System for Interactive Image Processing." NCI/IP Technical Report No. 21b, Nat. Tech. Inform. Serv. PB278789 (listing PB278790), 1978.
8. LEMKIN, P., AND LIPKIN, L. BMON2—Buffer memory monitor system for interactive image processing of biological images. In preparation.
9. 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).
10. 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).
11. LEMKIN, P., CARMAN, G., LIPKIN, L., SHAPIRO, B., AND SCHULTZ, M. "Real Time Picture Processor—Description and Specification." NCI/IP Technical Report No. 7a, Nat. Tech. Inform. Serv. PB269600/AS, 1977.
12. ROSENFELD, A., AND KAK, A. "Digital Picture Processing." Academic Press, New York, 1977.