(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/31321 A2

(51) International Patent Classification[7]: G01N 27/00

(21) International Application Number: PCT/US00/41276

(22) International Filing Date: 17 October 2000 (17.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/425,434 22 October 1999 (22.10.1999) US

(71) Applicant: ORINCON TECHNOLOGIES, INC. [US/US]; 9363 Towne Centre Drive, San Diego, CA 92121 (US).

(72) Inventors: THOMPSON, Alan, W.; 3019 Pueblo Street, Carlsbad, CA 92009 (US). BROTHERTON, Thomas, W.; 12687 Benavente Way, San Diego, CA 92129 (US).

(74) Agent: MEADOR, Terrance, A.; Gray Cary Ware & Freidenrich, LLP, 401 B Street, Suite 1700, San Diego, CA 92101-4297 (US).

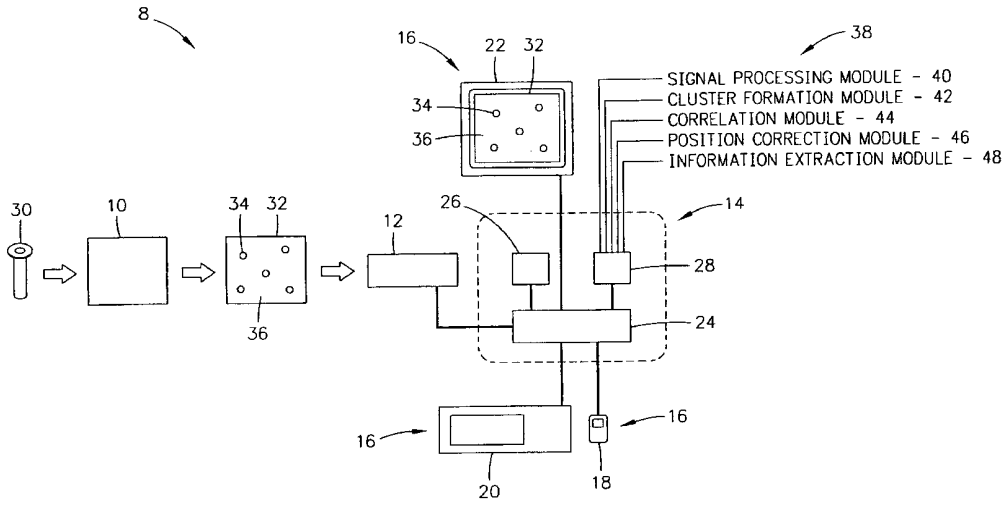
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

[Continued on next page]

(54) Title: INFORMATION EXTRACTION FROM GEL ELECTROPHORESIS IMAGES



(57) Abstract: A method of analyzing gel electrophoresis images is disclosed. The method includes accessing data from an experimental image and a reference image developed under different experimental conditions. The difference in the experimental conditions causes displaced regions of the experimental image to occupy a different position than the displaced regions would occupy if experimental image were developed under the same experimental conditions as the reference image. The method also includes determining a corrected position for at least a portion of the displaced regions. The corrected position is the approximate position that a displaced region of the experimental image would occupy if experimental image were developed under the same experimental conditions as the reference image.

WO 01/31321 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INFORMATION EXTRACTION FROM GEL ELECTROPHORESIS IMAGES

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was developed under grant number 2 R44 CA62469-02A1 from the National Cancer Institute.

BACKGROUND

1. Field of the Invention

The invention relates generally to extraction of information from two dimensional gel electrophoresis images. In particular, the invention relates to comparison of gel electrophoresis images developed under different experimental conditions.

2. Background of the Invention

Two dimensional gel electrophoresis is a fast growing technique for separation, identification, and quantification of proteins in biological samples. The selection of proteins present in a single sample can indicate the physiological activity in the source of the sample and can even indicate the health of a patient. As a result, identifying the proteins present in a biological sample can be used for medical diagnosis and treatment.

One dimensional gel electrophoresis has been successful in separating and identifying proteins in biological samples having a low number of proteins. However, many biological samples have hundreds and even thousands of different proteins. Accordingly, one dimensional gel electrophoresis is not appropriate for many biological samples.

Two dimensional gel electrophoresis permits separation of a larger number of proteins. Two dimensional gel electrophoresis includes dissolving the biological

sample in a liquid and allowing the biological sample to diffuses through a gel such as polyacrylamide. The gel has a molecular weight gradient in one direction and an electrical charge gradient in a perpendicular direction. As a result, proteins having different molecular weights or different charges settle in different portions of the gel while proteins having similar molecular weight and charge congregate in a similar portion of the gel. The proteins are stained so congregations of protein are visible. The result is a two dimensional gel electrophoresis image consisting of spots against the gel background.

Information about a biological sample can be developed by comparing gel electrophoresis images developed from different biological sample. For instance a first gel electrophoresis image developed from a first biological sample may have a spot which is not present on a second gel electrophoresis image developed from a second biological sample. The absence of the spot on the second gel electrophoresis image may indicate that the protein responsible for the presence of the spot on the first gel electrophoresis image is not present in the second biological sample.

Comparison of two dimensional gel electrophoresis images is often difficult due to the number of experimental variables which can affect the appearance of a two dimensional gel electrophoresis image. For instance, the appearance of a two dimensional gel electrophoresis image can be affected by variations in the kind and consistency of polyacrylamide gel, buffers (variations in pH), detergents, equipment, ion effects, extraneous nucleic acids, extraneous charged species, isoelectric focusing, temperature, current flow and staining artifacts. A difference in one or more of these variables can cause one gel electrophoresis image to have results which are warped relative to another gel electrophoresis image. This warp even occurs when identical biological samples are used to develop the images. As a result, different experimental conditions can make two dimensional gel electrophoresis images difficult and even impossible to compare.

For the above reasons, there is a need for an improved methods of analyzing gel electrophoresis images.

SUMMARY OF THE INVENTION

The invention relates to a method of analyzing gel electrophoresis images. The method includes accessing data from an experimental image and a reference image developed under different experimental conditions. The difference in the experimental conditions causes displaced regions of the experimental image to occupy a different position than the displaced regions would occupy if experimental image were developed under the same experimental conditions as the reference image. The method also includes determining a corrected position for at least a portion of the displaced regions. The corrected position is the approximate position that a displaced region of the experimental image would occupy if experimental image were developed under the same experimental conditions as the reference image.

Another embodiment of the invention includes a method for identifying regions of the experimental image and regions of the reference image which would occupy approximately the same position if the experimental image and the reference image were developed under similar experimental conditions. The method includes accessing data from an experimental image and a reference image. The reference image includes spots which are correlated with spots of the reference image in that the correlated spots result from the presence of the same proteins in biological samples used to develop the reference image and the experimental image. The method also includes selecting at least a portion of the spots in the experimental image and at least a portion of the spots in the experimental image. A potentially correlated spot from the reference image is identified for each spot selected from the experimental image. Then a potentially correlated spot is identified from the experimental image for each spot selected from the reference image.

The method can further include identifying correlated pairs. Each correlated pair includes a spot from the reference image and a spot from the experimental image. The spot from the experimental image is identified as being potentially correlated to

the spot from the reference image and the spot from the reference image is identified as being potentially correlated with the spot from the experimental image. Another embodiment of the method includes selecting a test spot from the experimental image, N of the spots from the experimental image and M of the spots from the reference image. The test spot and the N spots are moved such that the test spot occupies a position similar to the position occupied by one of the M spots. The degree of correlation between the positions of the N spots and the positions of the M spots is measured.

Another embodiment of the method includes providing a composite image which includes a reference image and one or more correction markers indicating the corrected position of one or more spots from the experimental image.

Another embodiment of the invention includes determining a corrected position for each region of the experimental image and constructing a corrected experimental image from the determined corrected positions. The method also includes providing a composite image consisting of the reference image and the corrected experimental image.

The invention also relates to a signal bearing medium including machine readable instructions which are executable by a processing apparatus to perform the methods according to the present invention. Further, the invention also includes a system and a processing unit for performing the methods according to the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram illustrating a system for extracting information from a two dimensional gel electrophoresis image.

Figure 2 is a digital two dimensional gel electrophoresis image.

Figure 3A is an illustrative representation of a reference gel electrophoresis image having defined clusters.

Figure 3B is an illustrative representation of an experimental gel electrophoresis image having defined clusters.

Figure 3C is a composite image including the image of Figure 3A superimposed on the image of Figure 3B.

Figure 3D illustrates criteria for including pixels in a cluster.

Figure 4A illustrates selection of a target cluster, N of the nearest experimental clusters and M of the nearest reference clusters from the composite image of Figure 3C.

Figure 4B illustrates the target cluster and the N experimental clusters translated such that the target cluster is positioned over one of the M reference clusters. The positions of the N experimental clusters are poorly correlated with the positions of the M reference clusters.

Figure 4C illustrates the target cluster and the N experimental clusters translated such that the target cluster is positioned over one of the M reference clusters. The positions of the N experimental clusters are highly correlated with the positions of the M reference clusters.

Figure 4D illustrates a translation vector extending from each experimental cluster to a potentially correlated reference cluster.

Figure 4E illustrates a translation vector extending from each reference cluster to a potentially correlated experimental cluster.

Figure 4F illustrates a translation vector extending from each experimental cluster to the correlated reference cluster.

Figure 5A illustrates a grid positioned over the experimental image.

Figure 5B illustrates the correction vector associated with each node of the grid.

Figure 6A is a composite image constructed from a reference image and a corrected experimental image.

Figure 6B is a close-up of a composite image including a reference image and markers indicating the position occupied by spots in eight different experimental images.

Figure 6C is Figure 6B with the marks positioned at the corrected position of each spot in the experimental image.

Figures 7A-7B together embody a process flow for a method of extracting information from two dimensional gel electrophoresis images.

Figures 8A-8B together embody a process flow illustrating a method for assigning pixels which define spots on a gel electrophoresis image into clusters.

Figures 9A-9B together embody a process flow illustrating a method for identifying regions of the experimental image and the reference image which would occupy approximately the same position is the experimental image and the reference image were developed under similar experimental conditions.

Figure 10 is a process flow illustrating a method for determining a correlation score which measure the degree of correlation between the position of a target cluster and a test cluster.

Figures 11A-11B together embody a process flow illustrating a method for determining the approximate position that a region of the experimental image would occupy if the experimental image were developed under the same experimental conditions as the reference image.

Figure 12 is a graphic user interface for displaying information from two dimensional gel electrophoresis images to a user.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The invention relates to a method of analyzing gel electrophoresis images. The method includes obtaining an experimental image and a reference image developed under different experimental conditions. The difference in the experimental conditions causes displaced regions of the experimental image to

occupy a different position than the displaced regions would occupy if the experimental image and the reference image were developed under similar experimental conditions. The method also includes determining a corrected position for at least a portion of the displaced regions. The corrected position is the approximate position that a displaced region of the experimental image would occupy if the experimental image and the reference image were developed under similar experimental conditions.

The displaced regions of the experimental image can include one or more spots on the experimental image. Accordingly, corrected positions can be determined for the spots on the experimental image. Once a corrected position has been determined, a correction mark can be displayed on the reference image. Each correction mark illustrates where a spot would have been positioned if the experimental image and the reference image were developed under similar experimental conditions. As a result, the positions of the spots on the experimental image and the spots on the reference image can be directly compared without the effects of the different experimental conditions.

A corrected experimental image can also be created by determining a corrected position for each region of the experimental image. The corrected experimental image is constructed from the corrected positions. A composite image can then be created from the corrected experimental image and the reference image. For instance, the composite image can include the experimental image superpositioned over the reference image to provide a direct comparison of the corrected experimental image and the reference image. This comparison is effectively a comparison of the experimental image and the reference image without the effects of the differences in experimental conditions.

Figure 1 is a block diagram illustrating a system 8 for analyzing two dimensional gel electrophoresis images. The system 8 includes a gel electrophoresis component 10, a scanner 12 and a processing unit 14 in communication with user interfaces 16 such as a mouse 18, keyboard 20 and a monitor 22.

The gel electrophoresis component 10 includes equipment suitable for developing a two dimensional gel electrophoresis image from biological samples. Suitable equipment for the gel electrophoresis component 10 includes the ISO-DALT system produced by L.S.B.. The scanner 12 is suitable for scanning and digitizing the two dimensional gel electrophoresis images produced by the gel electrophoresis component 10. Suitable scanners 12 include, but are not limited to, the ScanJet manufactured by Hewlett Packard. The processing unit 14 receives the output from the scanner 12. A suitable processing unit 14 includes, but is not limited to, PCs and workstations.

The processing unit includes a processor 24, one or more memories 26 and a signal bearing medium 28 including machine readable instructions executable by a processing apparatus. The processor 24 is preferably a microprocessor. The memory 26 can be any memory device or combination of memory devices suitable for read/write operations such as storing images and data developed during execution of code. Suitable signal bearing media include, but are not limited to optical discs such as a compact disk (CD), CD-ROM, CD-R (a recordable CD-ROM that can be read on a CD-ROM drive), CD-RW (multiple-write CD), CD-E (recordable and erasable CD), or DVD (digital video disc). Alternatively, instead of, or in addition to an optical disc, the signal bearing medium 28 may include one or more of the following: a magnetic data storage diskette (floppy disk), a Zip disk, DASD storage (e.g., a conventional "hard drive" or a RAID array), magnetic tape, RAM, electronic read-only memory (e.g., ROM, EPROM, or EEPROM), paper punch cards, or transmission media such as digital and/or analog communication links.

In some instances, the signal bearing medium 28 is positioned outside or remote from the processing unit 14. For instance, the signal bearing medium 28 may be part of, or may be connected to, a server computer that is connected to a computer network, in order to make the machine-readable code available to other computers. The network may be a local area network (LAN), a wide area network (WAN), or any other type of network. This arrangement enables one or more other computers

connected to the network to copy the machine-readable code from the signal bearing medium 28 that is part of, or connected to, the (server) computer, to a signal bearing medium 28 that is part of, or connected to, the computer that is downloading the code. This may be accomplished, for example, by connecting computers from one or more networks, over the Internet.

In other instances, the signal bearing medium 28 may be part of, or may be connected to, a computer that is operating a bulletin board system (BBS), which can be accessed by other computers. This arrangement enables one or more computers to connect to the BBS, and copy the machine-readable code from the signal bearing medium 28 that is part of, or connected to, the computer that is operating the BBS, to a signal bearing medium 28 that is part of, or connected to, a computer that is downloading the machine-readable code.

During operation of the system 8, the biological sample 30 is delivered to the gel electrophoresis component 10 and a two dimensional gel electrophoresis image 32 is developed. The two dimensional gel electrophoresis image 32 is scanned and digitized. The processing unit 14 displays the digital two dimensional gel electrophoresis image 32 on the monitor 22 and stores it in the memory 26 as a bitmap. Figure 2 illustrates a digital two dimensional gel electrophoresis image 32 displayed on a monitor 22. The image includes a plurality of spots against a background 36. As described above, the spots 34 result form stained proteins congregating in different regions of the gel according to the their molecular weight and charge.

Digital two dimensional gel electrophoresis images 32 are formed from a plurality of pixels. Pixel refers to a region of an image and not to the individual segments of a monitor which light up to form the image. Each pixel is associated with a particular region of the two dimensional gel electrophoresis image 32. Further, each pixel in the image is assigned an intensity level between 0 and 255. The spots are defined by clusters of pixels having an increased intensity. In contrast, the background 36 is defined by pixels having lighter intensities. Accordingly, the pixel

clusters correspond to regions of the two dimensional gel electrophoresis image 32 where proteins are congregated.

As illustrated in Figure 1, the signal bearing medium 28 can include one or more modules 38 such as a signal processing module 40, a cluster formation module 42, a correlation module 44, a position correction module 46 and an information extraction module 48. These modules can be performed independently of one another or can be performed sequentially.

The signal processing module 40 includes instructions for preparing a digital two dimensional gel electrophoresis image 32 for further processing. The digital two dimensional gel electrophoresis image 32 can be filtered to remove extreme pixel intensity values, smoothed to enhance continuity between adjacent pixels and normalized to enhance the uniformity of the background 36. Suitable filtering algorithms include, but are not limited to, outlier rejection algorithms. Suitable smoothing algorithms include, but are not limited to, a two dimensional, linear, low pass filter. Suitable normalization algorithms include, but are not limited to a multi-pass background estimation algorithm with signal rejection.

The cluster formation module 42 includes instructions for identifying clusters of pixels which define the spots 34 on a digital two dimensional gel electrophoresis image 32. Since clusters refer to the collection of pixels which define a spot 34, the term spot 34 and cluster can often be used interchangeably.

Figure 3A illustrates a first digital two dimensional gel electrophoresis image where the pixels have been assigned to clusters. The pixels at the perimeter of each cluster are shown as solid lines. Figure 3B illustrates a second digital two dimensional gel electrophoresis image where the pixels at the perimeter of the clusters are shown as broken lines. The first two dimensional gel electrophoresis image is referred to as the first image 32A and the second two dimensional gel electrophoresis image is referred to as the second image 32B. The biological samples 30 used to develop the first and second images 32B include the same proteins.

Figure 3C is a composite gel electrophoresis image including the first image 32A superpositioned over the second image 32B. Although the first and second image 32B were developed using the same proteins, regions of the first and second images 32B are not aligned as indicated by the lack of alignment between the clusters. These regions which are not aligned are referred to as displaced regions below. In addition to being unaligned certain of the displaced regions are better aligned than others. This feature is illustrated by the existence of certain cluster which are nearly aligned while other clusters are badly aligned. The uneven degree of alignment across the images indicates that the second image 32B is warped relative to the first image 32A. Accordingly, simply translation the first image 32A relative to the second image 32B will not align all of the clusters. As described above, the lack of alignment and warping results from the first and second digital images being developed under different experimental conditions.

Figure 3D is a graph which illustrates suitable criteria for including a pixel in a cluster. The y-axis is the pixel intensity and the x-axis is the number of pixels away from the most intense pixel of a cluster. The cluster is started by identifying the most intense pixel on the image which has not been assigned to a cluster or to the background. The identified pixel is labeled the cluster maximum as illustrated in Figure 3D. Moving away from the cluster maximum along the x-axis, each pixel is included in the cluster until one of the pixels has an intensity less than the intensity minimum threshold. Once a pixel has an intensity less than the intensity minimum threshold, the only pixels included in the cluster are pixels with an intensity less than the intensity of the previous pixel. Once the intensity of the pixels falls below a background intensity threshold, the pixels are assigned to the background. The process is then repeated. Applying these criteria to Figure 3D, pixels 1-26 would be included in a first cluster, pixels 27-40 would be included in a second cluster and pixels 41-43 would be included in the background.

The intensity maximum threshold and/or the background intensity threshold can be an absolute value but is preferably a function of the cluster maximum's

intensity. For instance, the intensity minimum threshold could be 90% of the cluster maximum intensity and the background intensity threshold can be 10% of the cluster maximum's intensity.

The correlation module 44 includes instructions for identifying regions of the first and second image 32B which would occupy approximately the same position on their respective images if the first and second images 32B were developed under similar experimental conditions. A protein would occupy approximately the same position on the first and second image 32B in the absence of the different experimental conditions. However, the protein would not necessarily occupy the exact same position because of the effects of different proteins in the biological samples used to develop the images, because the protein may have different concentrations in the biological samples used to develop the images, etc. Accordingly, a spot which results from a particular protein would occupy approximately the same position in multiple images if each image were developed in the absence of differences in experimental conditions. Hence, the correlation module 44 seeks to identify spots which result from the presence of the same protein. A spot on the first image which result from the presence of the same proteins as a spot on the second image is said to be correlated with the spot from the second image.

The regions identified by the correlation module can be one or more pixels which make up a spot. For instance the identified region can be the most intense pixel defining the associated spot. The most intense pixel most likely correlates with the region of the spot having the highest concentration of protein. The identified region can also be the pixel positioned at the intensity weighted centroid of the spot. Further, the identified region can also include all the pixels which make up the entire spot. During the following discussion, the identified region will be the most intense pixel defining a spot, i.e. the most intense pixel in a cluster.

During the correlation module 44, one of the images is compared against the other image. As a result, one of the images serves as a reference image 32A and the other image serves as an experimental image 32B. The titles of reference image 32A

and experimental image 32B do not have any significance. For instance, any gel electrophoresis image stored in the memory 26 can be assigned the role of reference image 32A and one or more of the remaining images can be assigned to role of experimental image 32B. Alternatively, the reference image 32A can be an image which is specially prepared to serve as a reference image 32A. For instance, the reference image 32A can be an image where the protein associated with each spot has been positively identified. Further, the proteins in a prepared reference image 32A can be selected to match the proteins expected to be present in a biological sample 30 being tested. When a prepared reference image 32A is used, the processing unit 14 can obtain the reference image 32A from a signal bearing medium 28. Accordingly, prepared reference images 32A which are suitable for certain applications can be commercially obtained, downloaded, etc. For the purposes of the following discussion, the second image 32B illustrated in Figure 3B will serve as an experimental image 32B and the first image 32A illustrated in Figure 3A will serve as the reference image 32A. The clusters on the experimental image 32B are referred to as experimental clusters 50 and the clusters on the reference image 32A are referred to as reference clusters 52.

As illustrated in Figure 4A, the correlation module 44 includes selecting a neighborhood of experimental clusters 50 including a target cluster 54 and the N nearest experimental clusters 56. The module also includes selecting a neighborhood of reference clusters 52 including the M reference clusters 58 closest to the position of the target cluster 54. N can be any number up to the total number of spots in the experimental image and is preferably 4. Similarly, M can be any number up to the total number of spots in the reference image and is preferably 9. M is preferably larger than N.

As illustrated in Figure 4B, one of the M reference clusters 58 is selected as a test cluster 60. The experimental clusters 56 and the target cluster 54 are translated according to a translation vector V. The translation vector V originates at the position of the target cluster 54 and extends to the position of the test cluster 60. As a result,

translation of the target cluster 54 according to the translation vector positions the target cluster 54 over the test cluster 60.

A correlation score is calculated for each translation of the target cluster 54 over a test cluster 60. The correlation score measures the degree of correlation between the position of each of the N experimental clusters 56 with the positions of N of the M reference clusters 58. Starting with the experimental cluster 50 nearest the target cluster 54, the distance to the nearest reference cluster, d_1, is determined. The determined distance is saved and that reference cluster 52 is tagged as used and is unavailable for further calculations until the next translation. The process is repeated for the remaining experimental clusters 56. The correlation score is then calculated according to Equations 1 which approximates the degree of error associated with translating the target cluster 54 over the test cluster 60. Accordingly, a low value of E indicates a higher degree of correlation. In Equation 1, d_i is the distance between the ith one of the N clusters and the nearest available one of the M clusters. Figure 4B illustrates d_1-d_4 for each of the N clusters. a_{ie} is the amplitude of the ith one of the N clusters and a_{ir} is the amplitude of the available reference cluster closest to the ith one of the N clusters. The amplitude of a cluster is the sum of the intensities of each pixel included in the cluster. C_1 and C_2 are scaling constants. A suitable value of C_1 is .1 and a suitable value of C_2 is .25 when the image is scaled to be between –1 and 1 in both the x and y directions and the amplitudes are scaled to be between 0 and 1.

$$E = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{d_i^{\,2}}{C_1^{\,2}} + \frac{(a_{ie} - a_{ir})^2}{C_2^{\,2}}\right] \qquad \textit{Equation } 1$$

The target cluster 54 is translated over each of the M reference clusters 58. Accordingly, each of the M clusters serves as a test cluster 60. A correlation score is determined for each translation. The translation providing the highest degree of correlation is identified and is illustrated in Figure 4C. The distances between the N experimental clusters 56 and the nearest of the M reference clusters 58 is small.

Accordingly, error E is small. As a result, the correlation score would indicate a high degree of correlation between the test cluster and the target cluster. The high degree of correlation indicates an increased probability that the target cluster and the test cluster result from the presence of the same protein in the biological samples used to develop both the experimental image and the reference image. Since the combination of the target cluster and the test cluster of Figure 4C provide the correlation score indicating the highest degree of correlation, these clusters are labeled a first potentially correlated pair.

The above process is repeated using each of the experimental clusters 50 as the target cluster 54. Accordingly, each experimental cluster 50 is a member of a first potentially correlated pair. Each first potentially correlated pair is associated with a correlation score and a translation vector as illustrated in Figure 4D. The translation vectors originate at an experimental cluster 50 and extends to the first potentially correlated reference cluster 52. As illustrated, the reference clusters 52 can be a member of more than one potentially correlated pair. To identify which correlated pair is correct, a inverse process is performed.

The inverse process includes repeating the above process with the experimental image 32B temporarily switched to the reference image 32A and the reference image 32A temporarily switched to the experimental image 32B. Accordingly, the inverse process produces second potentially correlated pairs. Figure 4E illustrates each second potentially correlated pair and the associated translation vector.

Correlated pairs are next determined by comparing the first potentially correlated pairs with the second potentially correlated pairs. When a first potentially correlated pair includes the same experimental cluster 50 and reference cluster 52 as a second potentially correlated pair, that experimental cluster 50 and reference cluster 52 become a correlated pair. A correlated pair consists of an experimental cluster and a reference cluster which would occupy approximately the same position in the their respective images in the absence of different experimental conditions during the

development of the images. As illustrated in Figure 4F, the translation vector associated with each correlated pair originates at the experimental cluster 50 and extends to the reference cluster 52 of the correlated pair. The inverse process step results in some portion of the experimental clusters 50 and reference clusters 52 not being assigned to a correlated pair.

The position correction module 46 includes instructions for determining a corrected position for at least a portion of the displaced regions in the experimental image 32B. The corrected position is an approximate position that a displaced region would occupy on the experimental image 32B if the experimental image 32B were developed under the same experimental conditions as the reference image 32A.

The position correction module 46 includes positioning a grid 62 over the experimental image 32B as illustrated in Figure 5A. The grid 62 preferably has J horizontal lines 64 and K vertical lines 66. J and K are preferably ten. The intersection of each grid 62 line defines a node 68. A correction vector Vc is determined for each node 68. The correction vector Vc defines the approximate position each node 68 would occupy if the experimental image 32B were developed under the same experimental conditions as the reference image 32A. The correction vectors Vc are determined by identifying the translation vectors which originate closest to each node 68. For instance, Figure 5A illustrates a node and the three nearest translation vectors, V_1, V_2, V_3. The correction vector Vc is then determined as a composite of the identified translation vectors. A suitable composite includes, but is not limited to, an average of the identified translation vectors and an average of the identified translation vectors weighted by the correlation score associated with each translation vector. Figure 5B illustrates the correction vector Vc determined for each node 68. This illustration makes the warp caused by the differences in experimental conditions to be visually apparent.

Once a correction vector has been determined for each node, a correction vector can be determined for any pixel or region of the experimental image by interpolating, and in some cases extrapolating, between the correction vectors at the

nodes of the grid. Accordingly, a correction vector can be determined for each of the spots on the experimental image by determining correction vectors for each of the pixels in the spot. Alternatively, a correction vector can be determined for each of the spots on the experimental image by determining correction vectors for a pixel of the spot and presuming all the other pixels making up the spot move according to the determined correction vector.

The information extraction module can include formation of a composite image including information from the experimental image and the reference image. For instance, a corrected experimental image can be created by determining a corrected position for each region of the experimental image including the background regions of the experimental image. The corrected experimental image is created from the corrected positions. A composite image can then be constructed from the corrected experimental image and the reference image. For instance, Figure 6A illustrates the experimental image superpositioned over the reference image. The composite image allow a direct comparison of the experimental image and the reference image without the effects of different experimental conditions.

As an alternative to creating the corrected experimental image, a correction mark indicating the position of spots on the experimental image can be placed on the reference image. For instance, Figures 6B and 6C show a portion of a composite image having marks positioned against a reference image background. Diamonds indicate the positions of the spots on the reference image. Circles indicate the position of spots on four experimental images generated from one type of sample and x s indicate the position of spots on four experimental images generated from another type of sample. For instance, the x s may be derived from rats having cancer while the circles may be derived from rats without cancer.

Although only a mark is shown to indicate the location of each spot, a selection of pixels that indicate the location of a spot can also be employed. Each selection of pixels can have approximately the same size and shape as the represented spot. The selection of pixels can be used by itself or each mark can be shown

positioned with the associated selection of pixels. When a selection of pixels are used in conjunction with marks, the marks can indicate the intensity weighted center of the spot. The selection of pixels can be used to show locations of spots in one or more of the experimental images and/or the reference image.

Figure 6B shows the uncorrected positions of the spots in the experimental images against a reference image background. Lines are optionally used to show the correlation between the spots from the experimental images with the spots from the reference image. As illustrated, many of the spots are not correlated with any spot from the reference image.

Figure 6C is the same portion of the reference image illustrated in Figure 6B. However, the marks now show the corrected positions of the spots in the experimental images and are accordingly correction marks. As illustrated, the correction marks are now clearly correlated with a cluster in the reference image. Further, many of the spots which were not labeled as being correlated in Figure 6B are now visibly correlated with a particular spot. Running the correlation module a second time on the data from this image would correlate these previously uncorrelated marks with the visibly correlated spots from the reference image. Further, a diamond is present which is not correlated with any marks from the experimental image. The diamond is caused be a protein being present in the biological sample used to develop the reference image which is not present in the biological samples used to develop the experimental image. Figure 6B does not clearly illustrate that the spots from the experimental images are not correlated with this lone diamond. Accordingly, the composite image of Figure 6C allows extraction of additional information from the two dimensional gel electrophoresis image. Further, if one of the diamonds was clearly correlated with circles but not with x s would indicate that one sample type included a protein that the other sample type did not.

Figures 7A-7B together embody a process flow for a method of analyzing two dimensional gel electrophoresis images. At start block 102 input is received from the user interface 16. The input includes selection of an experimental image 32B and a

reference image 32A from a plurality of digital two dimensional gel electrophoresis images stored in the memory 26.

After receiving the input, the signal processing module 40 is entered. At query block 104 it is determined whether signal processing has occurred for the selected experimental image 32B and the selected reference image 32A. When the inquiry is negative, a suitable filtration algorithm is applied to the unprocessed experimental image 32B and/or the unprocessed reference image 32A at process block 106. The filtration algorithm removes extreme pixel intensity values. Suitable filtering algorithms include, but are not limited to, outlier rejection algorithms. At process block 108 a suitable smoothing algorithm is applied to the unprocessed experimental image 32B and/or the unprocessed reference image 32A to enhance continuity between adjacent pixels. Suitable smoothing algorithms include, but are not limited to, a two dimensional, linear, low pass filter. At process block 110 a suitable normalization algorithm is applied to the unprocessed experimental image 32B and/or the unprocessed reference image 32A to enhance the uniformity of the image background 36. Suitable normalization algorithms include, but are not limited to a multi-pass background estimation algorithm with signal rejection. At process block 112, the unprocessed experimental image 32B and/or the unprocessed reference image 32A is tagged as having been processed.

After completion of the signal processing, the cluster formation module 42 is entered. Similarly, the cluster formation module 42 is entered when the inquiry at inquiry block 102 is positive. At inquiry block 114, an inquiry is made whether clusters have been identified in the experimental image 32B and the reference image 32A. When the inquiry is negative, the unclustered experimental image 32B and/or reference image 32A is processed to assign pixels to clusters which define the spots on the images at process block 116. Figures 8A-8B are a process flow illustrating a method of assigning pixels to clusters and to background. The identified clusters are stored in the memory 26. At process block 118, the unclustered experimental image 32B and/or reference image 32A is tagged as being clustered.

After completion of cluster formation, the correlation module 44 is entered. Similarly, the correlation module 44 is entered when the inquiry at inquiry block 114 is positive. At process block 120, regions of the experimental image and the reference image are identified which would occupy approximately the same position if the experimental image were developed under the same experimental conditions as the reference image. Specifically, clusters which would occupy approximately the same position in the absence of differences in experimental conditions are identified. Figures 9A-9B together embody a process flow for a method of identifying these correlated clusters.

After completion of the correlation module 44, the position correction module 46 is entered. At process block 122 a corrected position is determined for one or more regions of the experimental region. The corrected position is the approximate position that a region of the experimental image would occupy on the experimental image if experimental image were developed under the same experimental conditions as the reference image. Corrected positions can be determined for each spot on the experimental image or for a portion of each spot on the reference image.

After completion of the position correction module the information extraction module is entered. At process block 124, a composite image is created. As described above, the composite image can include the reference image and a plurality of correction marks as illustrated in Figure 6A. Alternatively, the composite image can include a corrected experimental image 32B superimposed on the reference image as illustrated in Figure 5B. The composite image can be displayed on the monitor 22 so the corrected experimental image 32B can be visually compared to the reference image 32A.

At process block 126, the correlation module 44 is repeated using the corrected positions of the spots. Specifically, the method used in process block 120 can also be used at process block 126, however, the corrected positions of the spots are substituted for the positions of the spots in the experimental image. An increased number of correlated clusters are likely to be found at process block 126 than were

found at process block 120 because the positions of the spots in the experimental image 32B has been corrected to compensate for the experimental differences. As a result, a larger number of clusters will likely be assigned to a correlated pair.

At process block 130, information is extracted from the composite image using the results of process block 126. The information can include identification of proteins which are present in the biological sample used to develop the reference image but not present in the biological sample used to develop the experimental image and vice versa. These situations can be detected when one image includes a particular spot that the other does not. This condition will often exist when a cluster is not assigned to any correlated pair.

The information extracted can also include information concerning the quantity of a protein in the experimental image 32B as compared to the quantity of the protein in the reference image 32A. This analysis can be performed by measuring the relative size and intensities of correlated spots.

The method illustrated in the process flow of Figures 7A-7B can also be performed with a plurality of experimental images 32B. Accordingly, the composite images can include information form a plurality of experimental images 32B and a single reference image 32A.

Figures 8A-8B together embody a process flow illustrating a method for assigning pixels to clusters and to background. The method of cluster formation beings at start block 140 where input is received from the user interface 16 or retrieved from the memory 26. The input can include a digital image which can be an experimental image 32B or an reference image 32A. The user input also includes a background intensity threshold and an intensity minimum threshold which are illustrated in detail with respect to Figure 3C.

At process block 142 the most intense pixel which is available for inclusion in a cluster is identified and tagged a cluster maximum. At process block 144, a new cluster is created. The cluster maximum is assigned to the new cluster and made unavailable. At process block 146 an available pixel adjacent to the perimeter of the

cluster is identified. At inquiry block 148, an inquiry is made whether the identified pixel has an intensity less than the lower threshold intensity. When the inquiry is positive the selected pixel is tagged unavailable at process block 150. When the query is positive, inquiry block 152 is accessed. At inquiry block 152 an inquiry is made whether the identified pixel has an intensity less than the intensity of the nearest pixel included in the cluster. When the inquiry is positive, the identified pixel is included in the cluster and made unavailable at process block 154. When the inquiry is negative, inquiry block 156 is accessed. At inquiry block 156 an inquiry is made whether any pixel between the identified pixel and the cluster maximum has an intensity less than the intensity minimum threshold. When the inquiry is negative the identified pixel is assigned to the cluster and tagged as unavailable at process block 154.

When the inquiry at inquiry block 156 is positive, inquiry block 158 is accessed. Similarly, inquiry block 158 is accessed once the identified pixel is made unavailable at process blocks 150 and 154. At inquiry block 158, the process is returned to process block 146 until there are no more pixels available adjacent to the perimeter of the cluster. At process block 160, an intensity weighted centroid is optionally calculated for the cluster. At process block 162 additional data are stored in the memory 26. The additional data includes the pixels, cluster maximum and centroid associated with the cluster. At inquiry block 164 the process is returned to process block 142 until additional pixels are not available. The process terminates at end block 166.

Figures 9A-9B together embody a process flow illustrating a method for finding regions of the experimental image and the reference image which would occupy approximately the same position if the experimental image and the reference image were developed under the same experimental conditions. Specifically, clusters on the experimental image 32B are correlated with clusters on a reference image 32A.

The process begins at start block 180 where input is received from the user interface 16 or retrieved from the memory 26. The input includes an experimental

image 32B and reference image 32A. The input can also include lists of pixels which are assigned to clusters on the experimental image 32B called experimental clusters 50 and lists of pixels which are assigned to clusters on the reference image 32A called reference clusters 52. The input can also include the position of the cluster maximum and centroid associated with each experimental cluster 50 and each reference cluster 52.

At process block 182 an experimental cluster 50 is selected to be a target cluster 54. At process block 184, N experimental clusters 56 which are positioned closest to the target cluster 54 are selected. At process block 186, M reference clusters 58 which are positioned closest to the position of the target cluster 54 are also selected. M is preferably greater than or equal to N. Further, N is preferably 4 while M is preferably 9. As described above, Figure 4A pictorially illustrates the selection of a target cluster 54, N experimental clusters 56 and M reference clusters 58.

At process block 188 one of the M reference clusters 58 is selected to serve as a test cluster 60. At process block 190 a translation vector is calculated. The translation vector is the vector which would translate the target cluster 54 over the test cluster 60. Specifically, the translation vector can position the cluster maximum for the target cluster 54 and the test cluster 60 at the same position. Alternatively, the translation vector can position the centroid of the target cluster 54 and the test cluster 60 in the same location. At process block 192 the target cluster 54 and the N experimental clusters 56 are translated according to the translation vector. As described above, Figure 4B pictorially illustrates of a target cluster 54 and N experimental clusters 56 translated according to a translation vector. At process block 194 a correlation score is determined. The correlation score measures the quality of the correlation between the positions of the N experimental clusters 56 with the positions of N of the M reference clusters 58. Figures 11A-11B together embody a process flow illustrating a suitable method for determining a correlation score. The correlation score and translation vector are each associated with the test cluster 60 to

which the target cluster 54 was transferred and are stored in the memory 26 accordingly.

At inquiry block 196, the process is returned to process block 188 until each of the M reference clusters 58 has served as a test cluster 60. At process block 198 the correlation scores are examined to identify the test cluster 60 which provided the translation with the highest degree of correlation. The identified test cluster 60 and the target cluster 54 are tagged a potentially correlated pair at process block 200. At process block 202 the identity of the clusters in the potentially correlated pair is stored along with the associated translation vector and correlation score. As described above, Figure 4D pictorially illustrates each potentially correlated pair and the associated translation vector.

At inquiry block 204, the process is returned to process block 182 until each of the experimental clusters 50 has served as a target cluster 54. At process block 206, process blocks 182 through 204 are repeated with the experimental image 32B serving as the reference image 32A and the reference image 32A serving as the experimental image 32B. At process block 208, the potentially correlated pairs developed from process blocks 182 through 204 are compared with the potentially correlated pairs developed at process block 206. When a potentially correlated pair from process blocks 182 through 204 includes the same experimental cluster 50 and reference cluster 52 as a potentially correlated pair from process block 206, that experimental cluster 50 and reference cluster 52 is labeled a correlated pair at process block 210. The process terminates at end block 212 by recording each correlated pair, the associated correlation score and translation vector into the memory 26. As described above, Figure 4F pictorially illustrates correlation pairs and their associated translation vector.

Figure 10 is a process flow illustrating a method of calculating a correlation score. The process begins at start block 230 where input is received from the user interface 16 or retrieved from the memory 26. The input includes the position of the target cluster 54, the N experimental clusters 56, the M reference clusters 58 the test

cluster 60. At process block 232 the experimental cluster 50 closest to the target cluster 54 is selected. At process block 234 the distance between the selected cluster and the nearest reference cluster 52 is determined. The determined distance is associated with the selected experimental cluster 50 and stored accordingly in the memory 26 at process block 236. At process block 238, the selected experimental cluster 50 and the nearest reference cluster 52 are made unavailable for further selection.

At inquiry block 240 the process is returned to process block 232 until each of the N experimental clusters 56 has been selected. At process block 242 the distances stored at process block 236 are used in Equation 1 to determine the correlation score E. The process ends at end block 242 by storing the correlation score in the memory 26.

Figures 11A-11B together embody process flows illustrating a method for determining a corrected position for regions of the experimental image. The process begins at start block 260 where input is received from the user interface 16 or retrieved from the memory 26. The input can include the experimental image 32B, the correlated pairs and the associated translation vectors. At process block 262 a grid 62 is positioned over the experimental image 32B as is pictorially illustrated in Figure 5A. The grid 62 preferably includes 10 vertical lines 66 and 10 horizontal lines 64. At process block 264 a node 68 of the grid 62 is selected. At process block 266, each experimental cluster 50 which is a member of a correlated member and is positioned near the selected node 68 is identified. As described above, Figure 5A pictorially illustrates the identified experimental clusters 50 positioned near a node 68 and their associated translation vector. At process block 268, the translation vector associated and correlation score associated with each identified correlated pair is accessed. At process block 270 a correction vector Vc is determined for the node 68 by finding an average of the translation vectors.

At inquiry block 272, the process is returned to process block 264 until a correction vector Vc has been determined for each node 68 of the grid 62. As

described above, Figure 5B pictorially illustrates the correction vector Vc for each node 68 of a grid 62 positioned over an experimental image 32B.

At process block 274, a correction vector Vc is determined for one or more pixels of the experimental image 32B. The correction vectors Vc for various pixels are determined by interpolating between the correction vectors Vc for the nodes 68 which are adjacent to each pixel. At process block 276, a corrected position is determined for each of the pixels by translating each of the pixels in the experimental image 32B by the associated correction vector Vc. For instance, by adding the translation vector to the position of the associated pixel. The process ends at end block 278 by storing the corrected experimental image 32B in the memory 26.

Figure 12 illustrates a graphic user interface 500 which is suitable for displaying two dimensional gel electrophoresis images to a user. The graphic user interface includes one or more composite image region 502 where a composite image is displayed. Further, the graphic user interface includes two or more individual image regions 504 where the experimental images and/or reference images which make up the composite image are displayed. Actions taken by the user on the composite image region 502 can also occur on the individual image regions 504. Similarly, actions taken on an individual image regions 504 can also occur on the composite image region 502. For instance, a user can zoom in on a portion of on the composite image and the same portion remaining images will be zoomed in on. Similarly, the user can place a mark such as an X on one of the images and the mark will appear in the same position of the remaining images. Additionally, the user can choose a portion of an image to be centered within a region and the same portion of the remaining images will also be centered. Accordingly, this user interface allows a user to identify which image is responsible for features on the composite image, etc.

Other embodiments, combinations and modifications of this invention will occur readily to those of ordinary skill in the art in view of these teachings. Therefore, this invention is to be limited only by the following claims, which include

all such embodiments and modifications when viewed in conjunction with the above specification and accompanying drawings.

CLAIMS

1. A method of analyzing two dimensional gel electrophoresis images, comprising:

accessing data from an experimental image and a reference image, the reference image including spots which are correlated with spots of the reference image in that the correlated spots result from the presence of the same proteins in biological samples used to develop the reference image and the experimental image;

identifying a potentially correlated spot from the reference image for at least a portion of the spots in the experimental image; and

identifying a potentially correlated spot from the experimental for at least a portion of the spots in the reference image.

2. The method of claim 1, further comprising:

identifying correlated pairs which include a spot from the reference image and a spot from the experimental image, the spot from the experimental image identified as being potentially correlated to the spot from the reference image and the spot from the reference image identified as being potentially correlated with the spot from the experimental image.

3. The method of claim 1, wherein identifying a potentially correlated spot from the reference image includes:

selecting a test spot from among the spots of the experimental image, N of the spots from the experimental image and M of the spots from the reference image;

relocating the test spot and the N spots such that the test spot occupies a position similar to the position occupied by one of the M spots; and

measuring a degree of correlation between the positions of the N spots and the positions of the M spots.

4. The method of claim 3, wherein the N selected spots are the N spots which are closest to the test spot.

5. The method of claim 3, wherein the M selected spots are the M spots which are closest to position of the test spot.

6. A method of analyzing two dimensional gel electrophoresis images, comprising:

accessing data from a reference image and an experimental image, each image including a plurality of spots;

selecting a test spot from the experimental image, N of the spots from the experimental image and M of the spots from the reference image;

relocating the test spot and the N spots such that the test spot occupies a position similar to the position occupied by one of the M spots; and

measuring the degree of correlation between the positions of the N spots and the positions of the M spots.

7. The method of claim 6, wherein the N selected spots are the N spots which are closest to the test spot.

8. The method of claim 6, wherein the M selected spots are the M spots which are closest to position of the test spot.

9. A signal bearing medium including machine readable instructions executable by a processing apparatus to perform a method of analyzing two dimensional gel electrophoresis images, the method comprising:

accessing data from an experimental image and a reference image, the reference image including spots which are correlated with spots of the reference

image in that the correlated spots result from the presence of the same proteins in biological samples used to develop the reference image and the experimental image;

identifying a potentially correlated spot from the reference image for at least a portion of the spots in the experimental image;

identifying a potentially correlated spot from the experimental for at least a portion of the spots in the reference image.

10. The medium of claim 9, wherein the method further comprises:

identifying correlated pairs, each potentially correlated pair including a spot from the reference image and a spot from the experimental image which are identified as being potentially correlated to one another.

11. The medium of claim 9, wherein identifying a potentially correlated spot from the reference image includes:

selecting a test spot from among the selected spots of the experimental image, N of the spots from the experimental image and M of the spots from the reference image;

relocating the test spot and the N spots such that the test spot occupies a position similar to the position occupied by one of the M spots; and

measuring a degree of correlation between the positions of the N spots and the positions of the M spots.

12. The medium of claim 11, wherein the N selected spots are the N spots which are closest to the test spot.

13. The method of claim 11, wherein the M selected spots are the M spots which are closest to position of the test spot.

14. A signal bearing medium including machine readable instructions executable by a processing apparatus to perform a method of analyzing two dimensional gel electrophoresis images, the method comprising:

accessing data from a reference image and an experimental image, each image including a plurality of spots;

selecting a test spot from the experimental image, N of the spots from the experimental image and M of the spots from the reference image;

relocating the test spot and the N spots such that the test spot occupies a position similar to the position occupied by one of the M spots; and

measuring the degree of correlation between the positions of the N spots and the positions of the M spots.

15. The medium of claim 14, wherein the N selected spots are the N spots which are closest to the test spot.

16. The medium of claim 14, wherein the M selected spots are the M spots which are closest to position of the test spot.

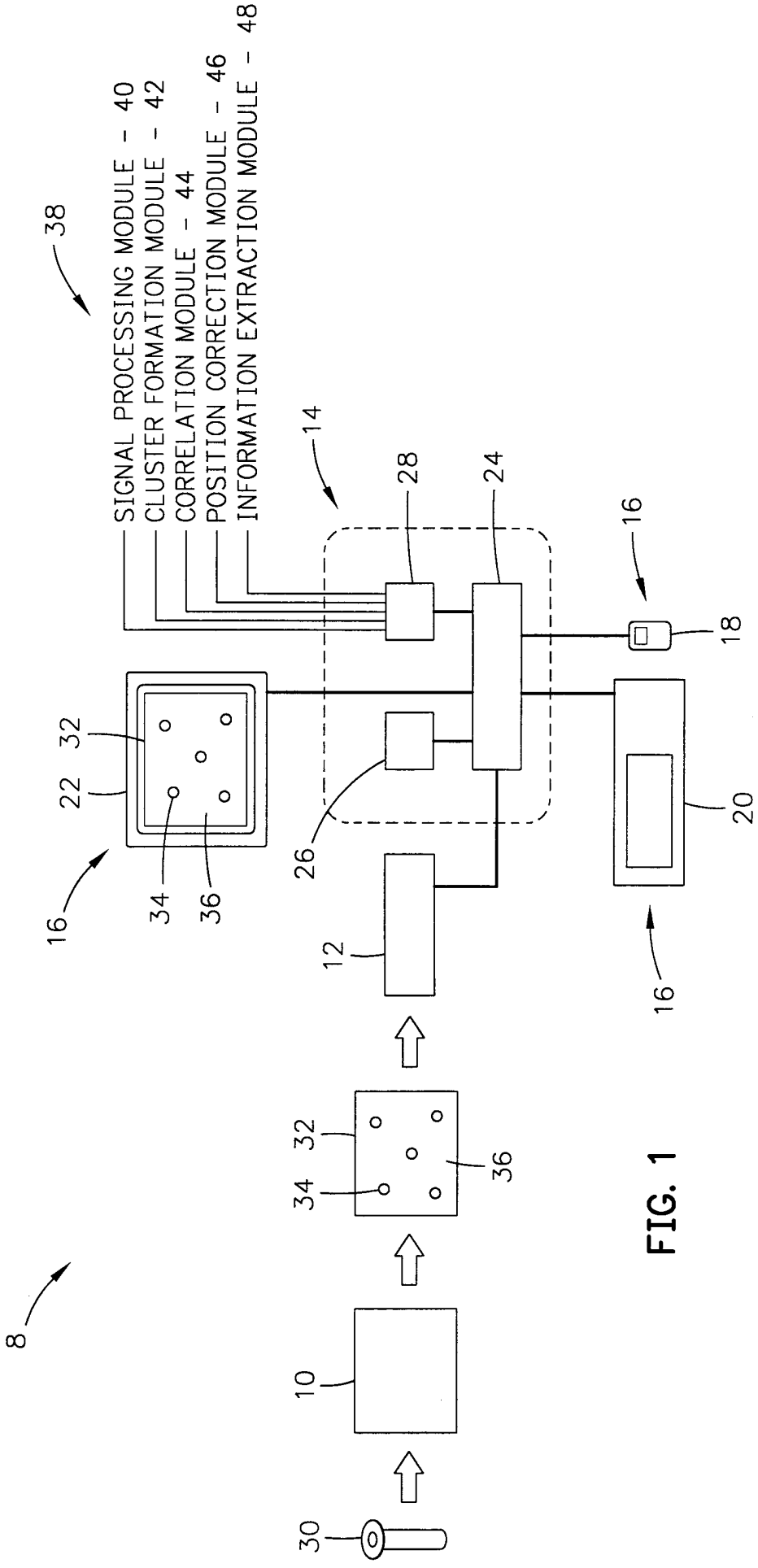

8
10
30
12
14
16
18
20
22
24
26
28
32
34
36
38
SIGNAL PROCESSING MODULE - 40
CLUSTER FORMATION MODULE - 42
CORRELATION MODULE - 44
POSITION CORRECTION MODULE - 46
INFORMATION EXTRACTION MODULE - 48

FIG. 1

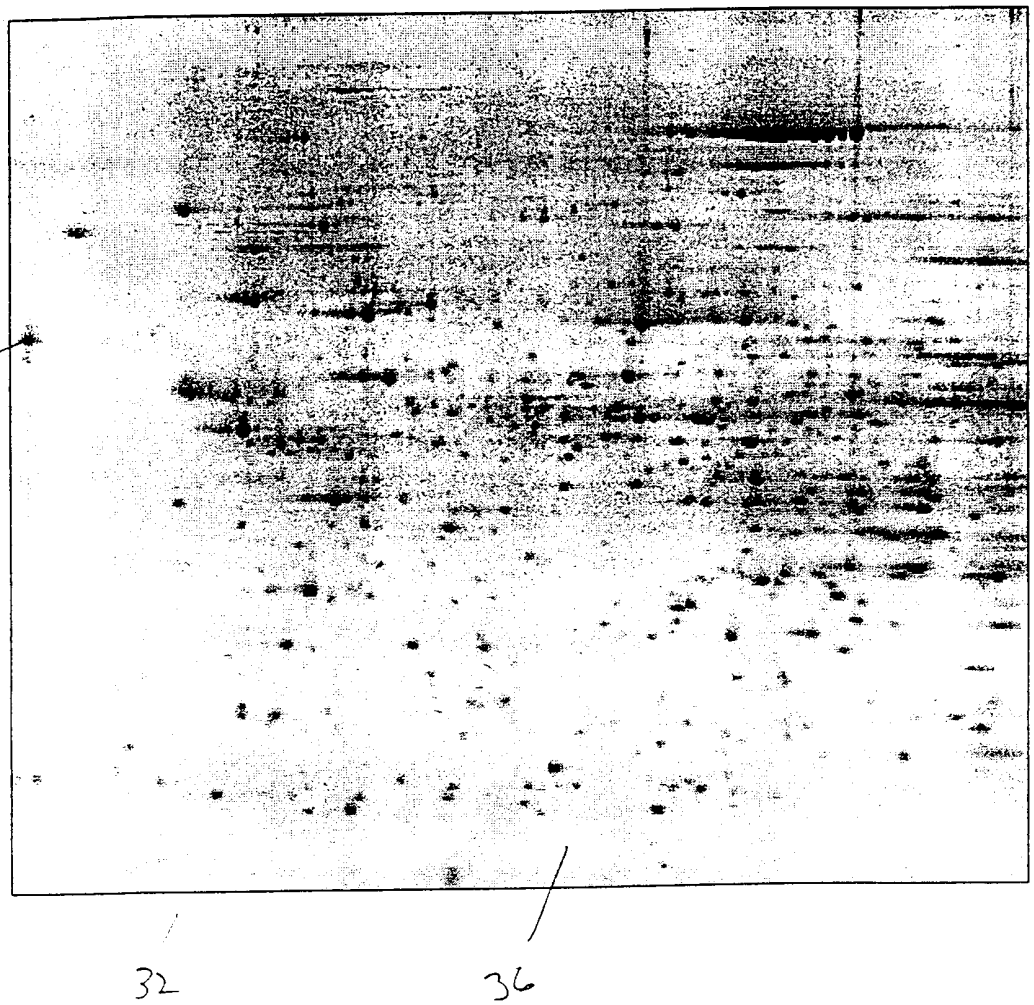


Figure 2

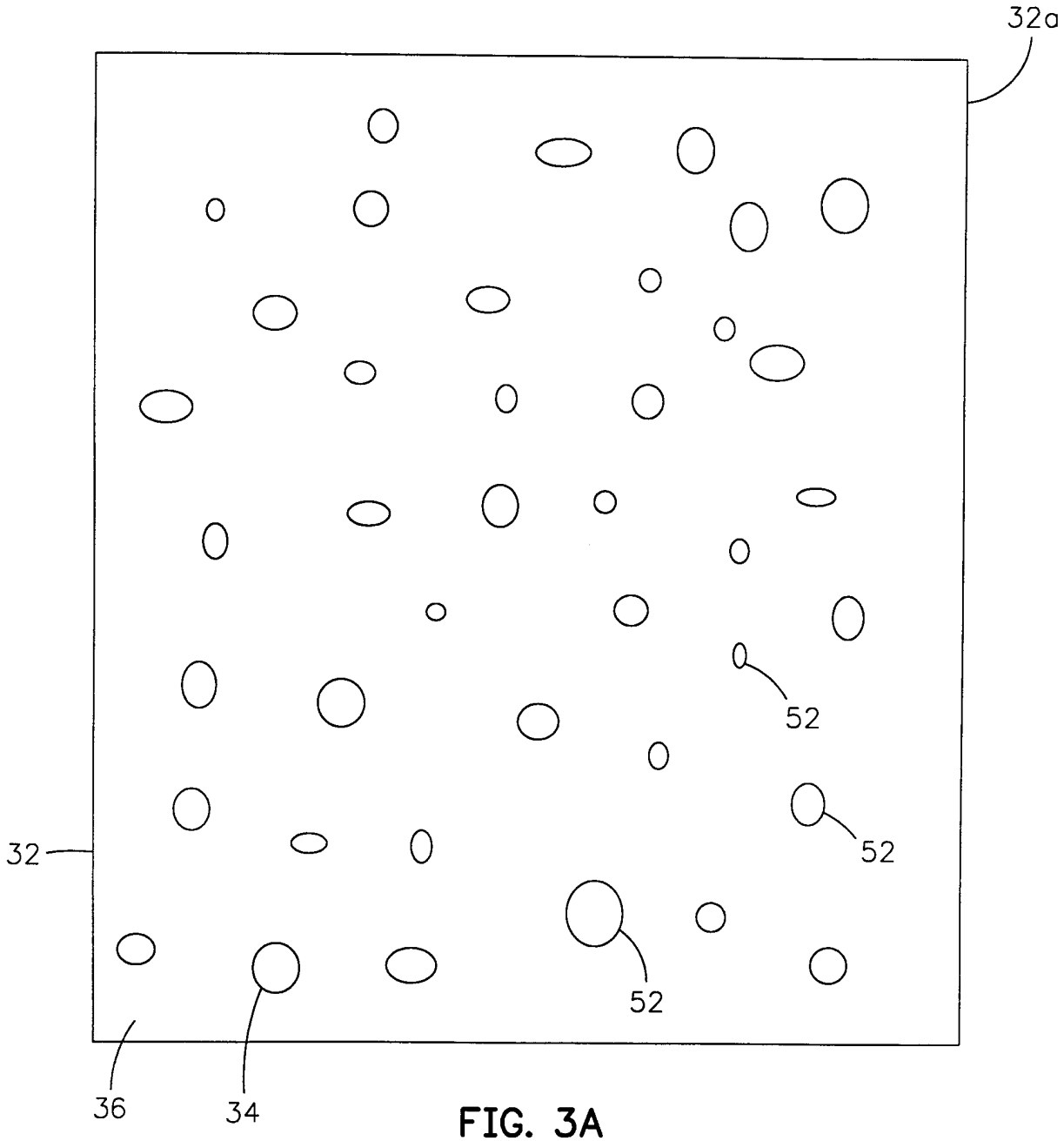


FIG. 3A

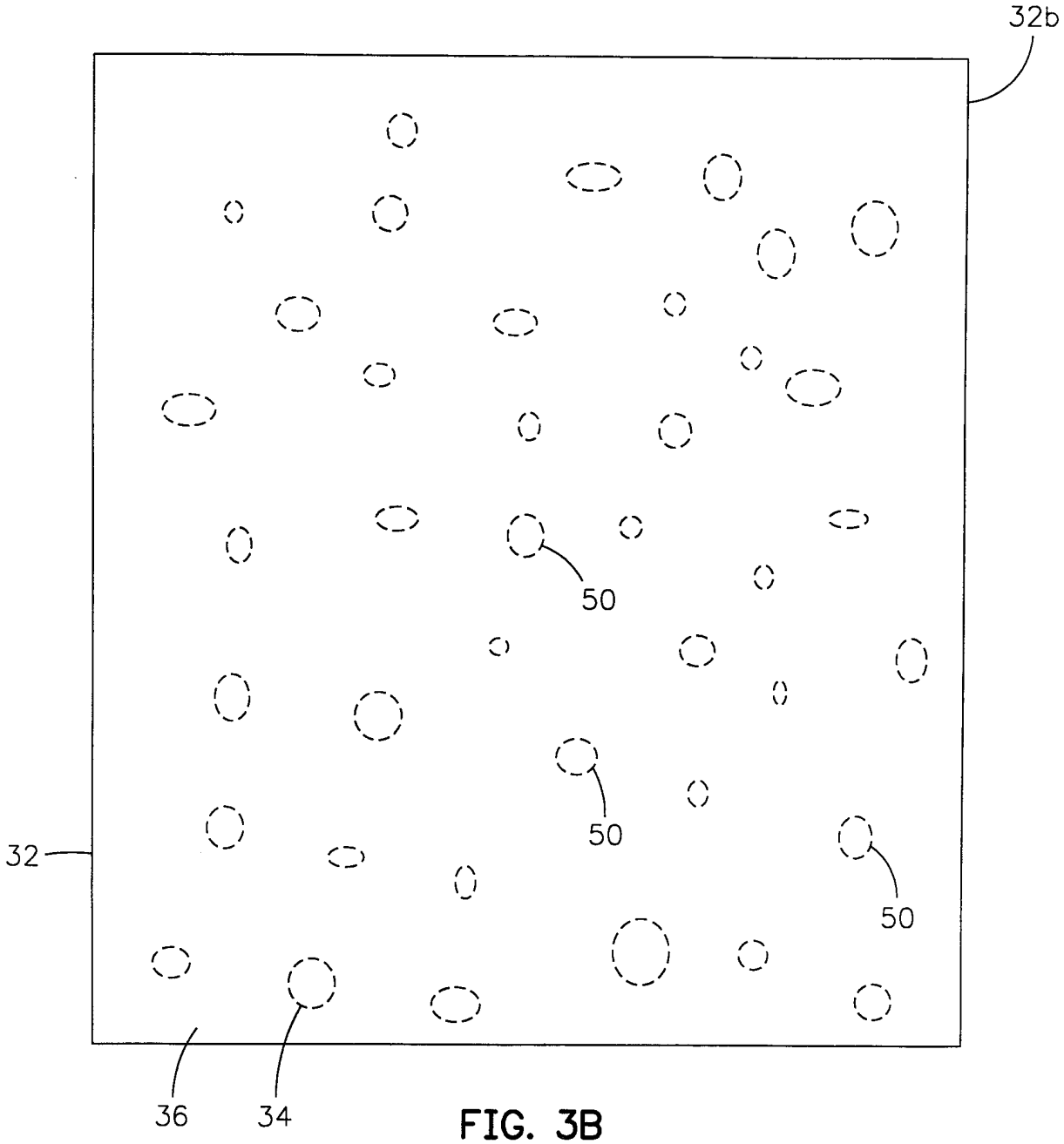


FIG. 3B

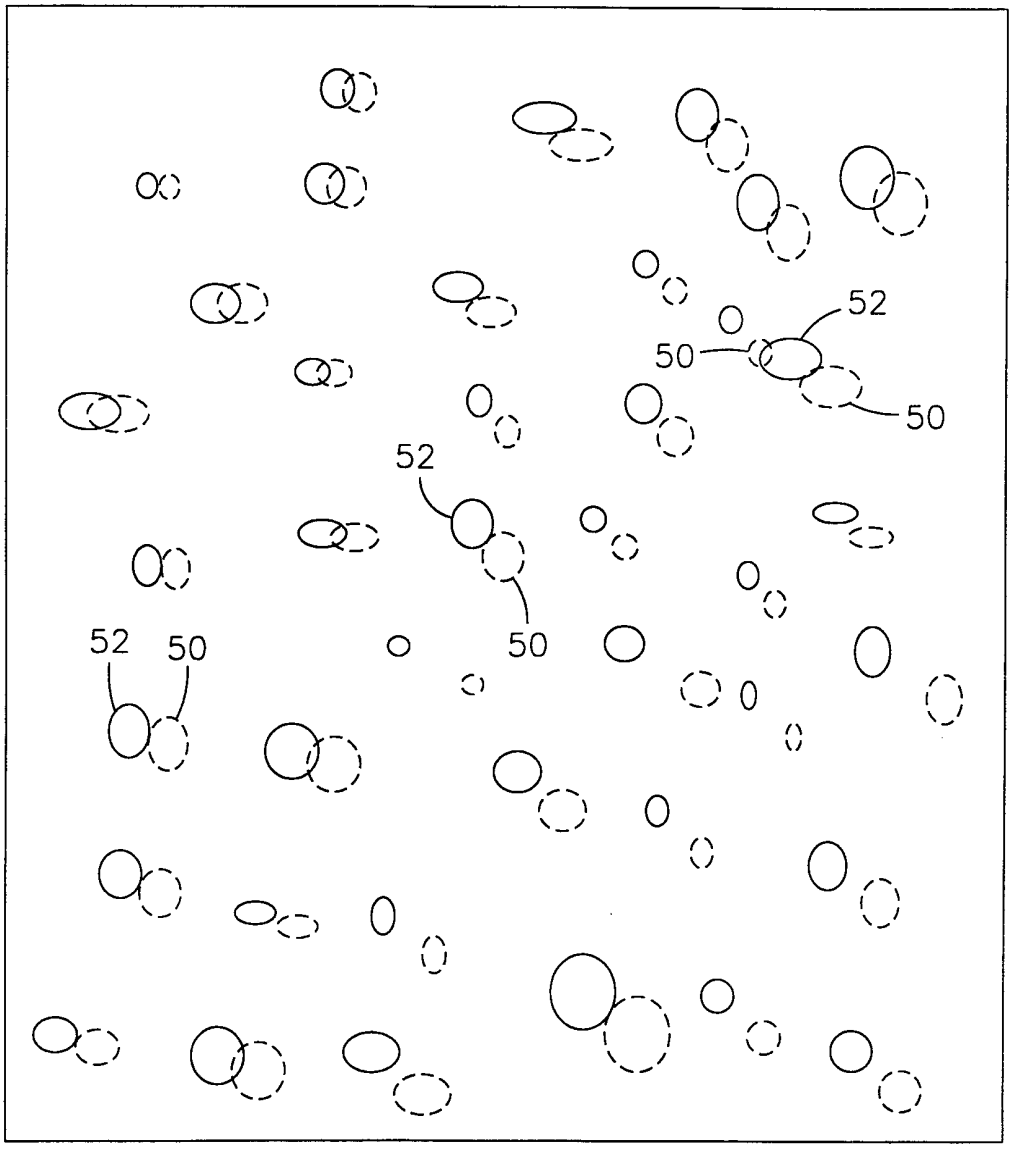


FIG. 3C

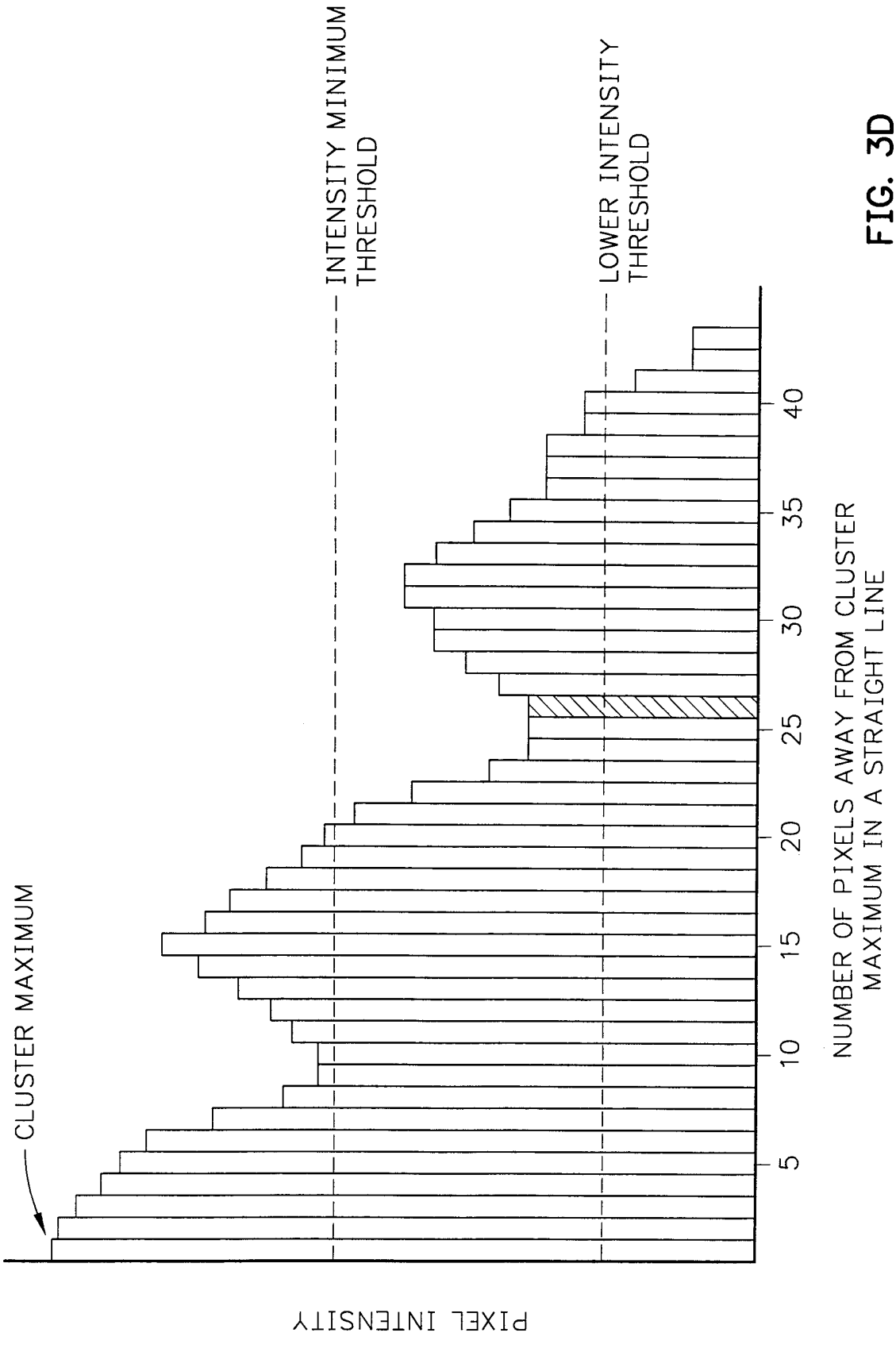

INTENSITY MINIMUM THRESHOLD
LOWER INTENSITY THRESHOLD
CLUSTER MAXIMUM
PIXEL INTENSITY
NUMBER OF PIXELS AWAY FROM CLUSTER MAXIMUM IN A STRAIGHT LINE
5
10
15
20
25
30
35
40

FIG. 3D

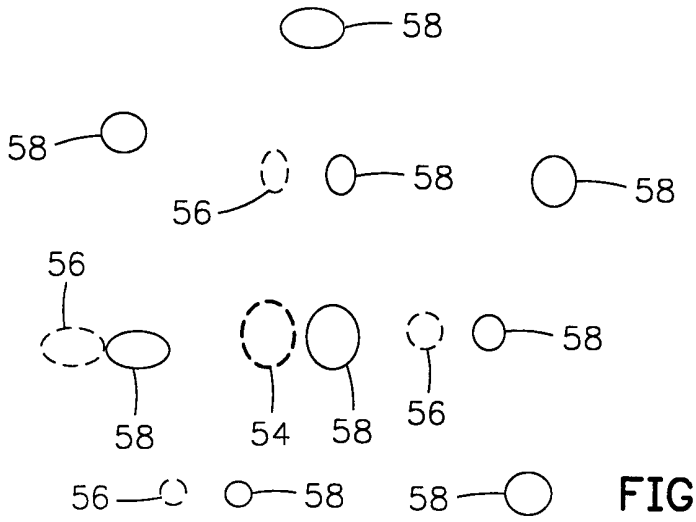


FIG. 4A

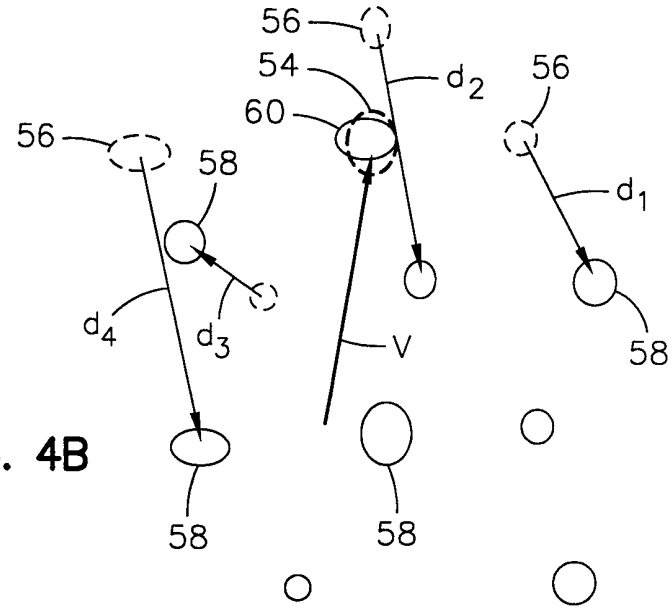


FIG. 4B

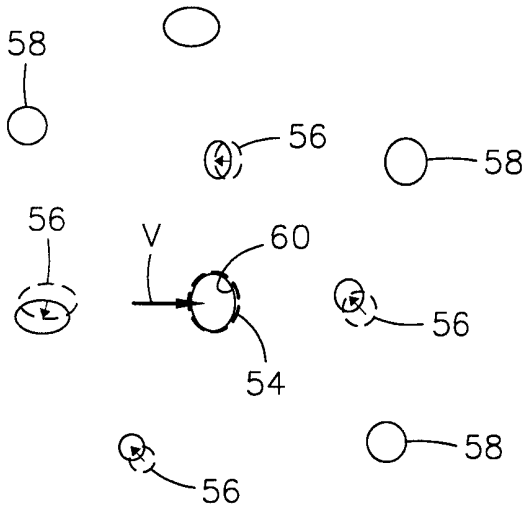


FIG. 4C

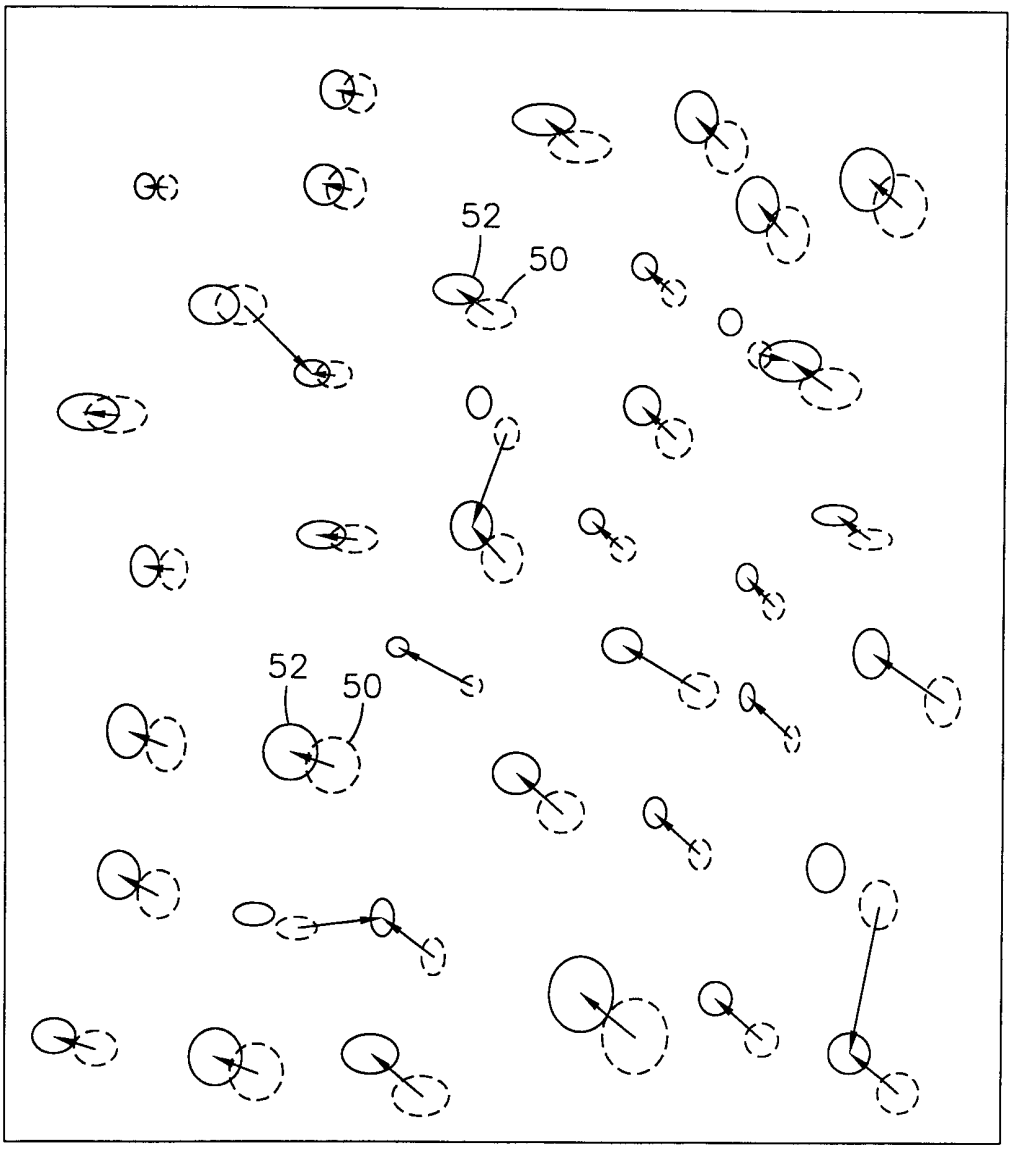


FIG. 4D

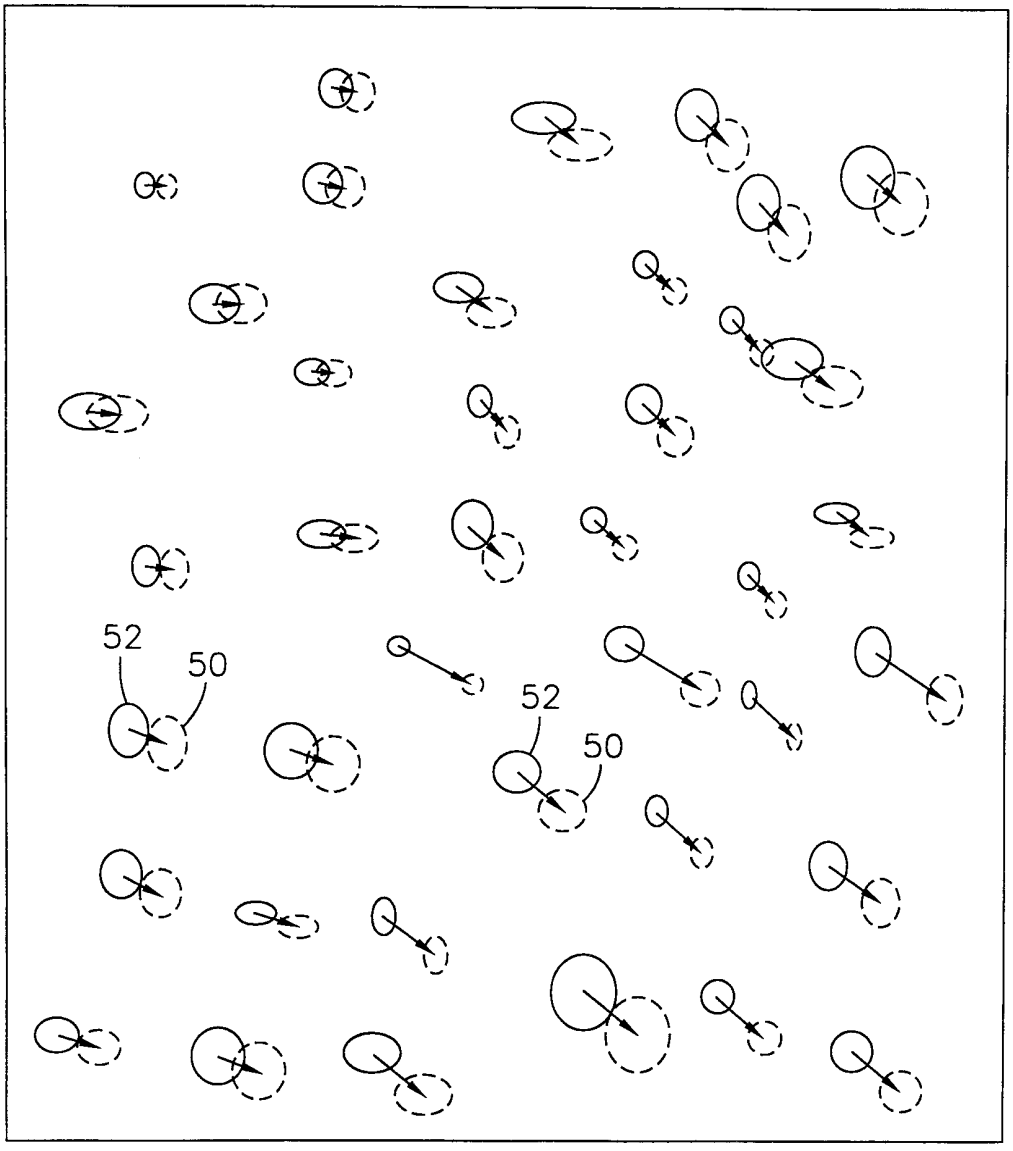


FIG. 4E

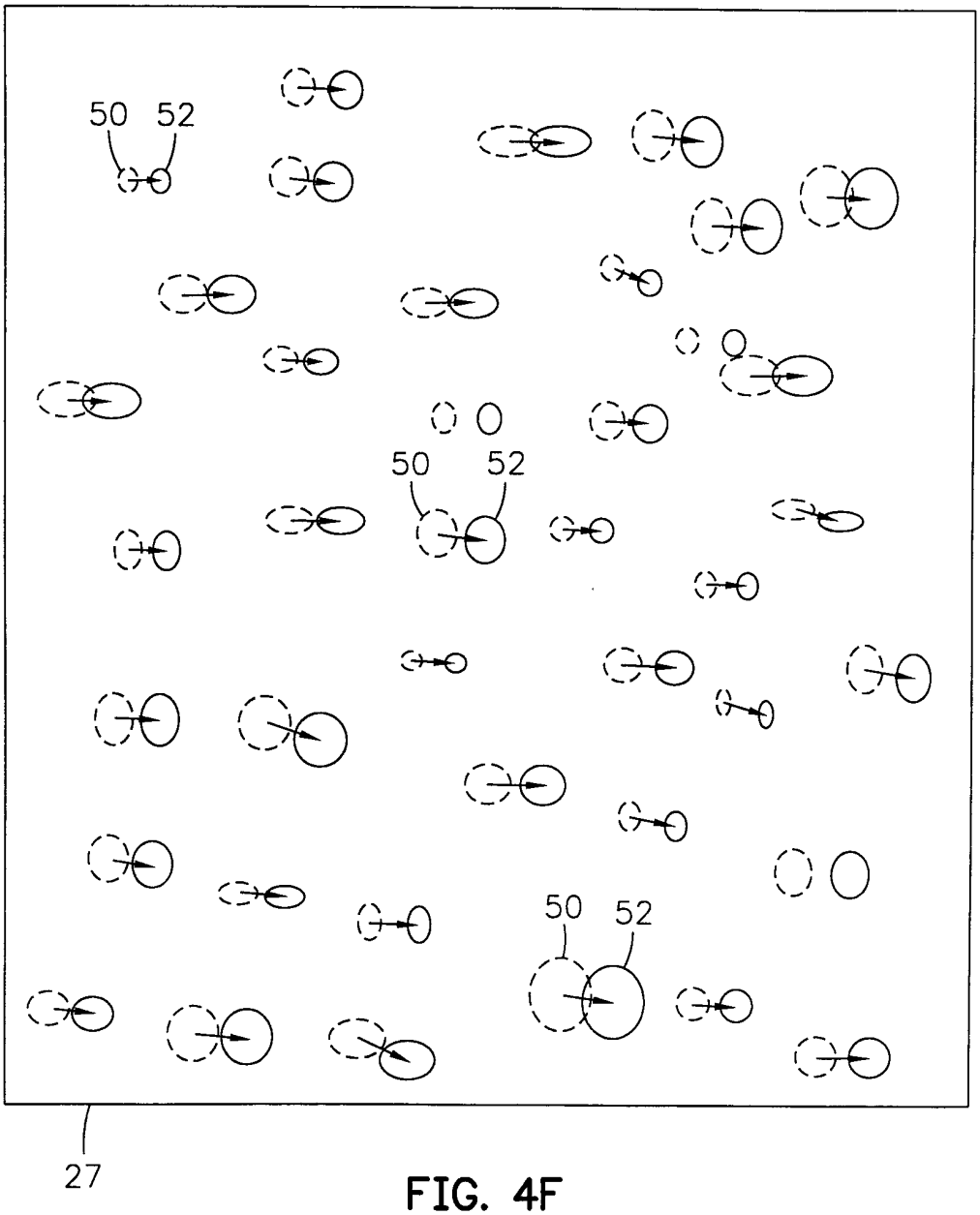


FIG. 4F

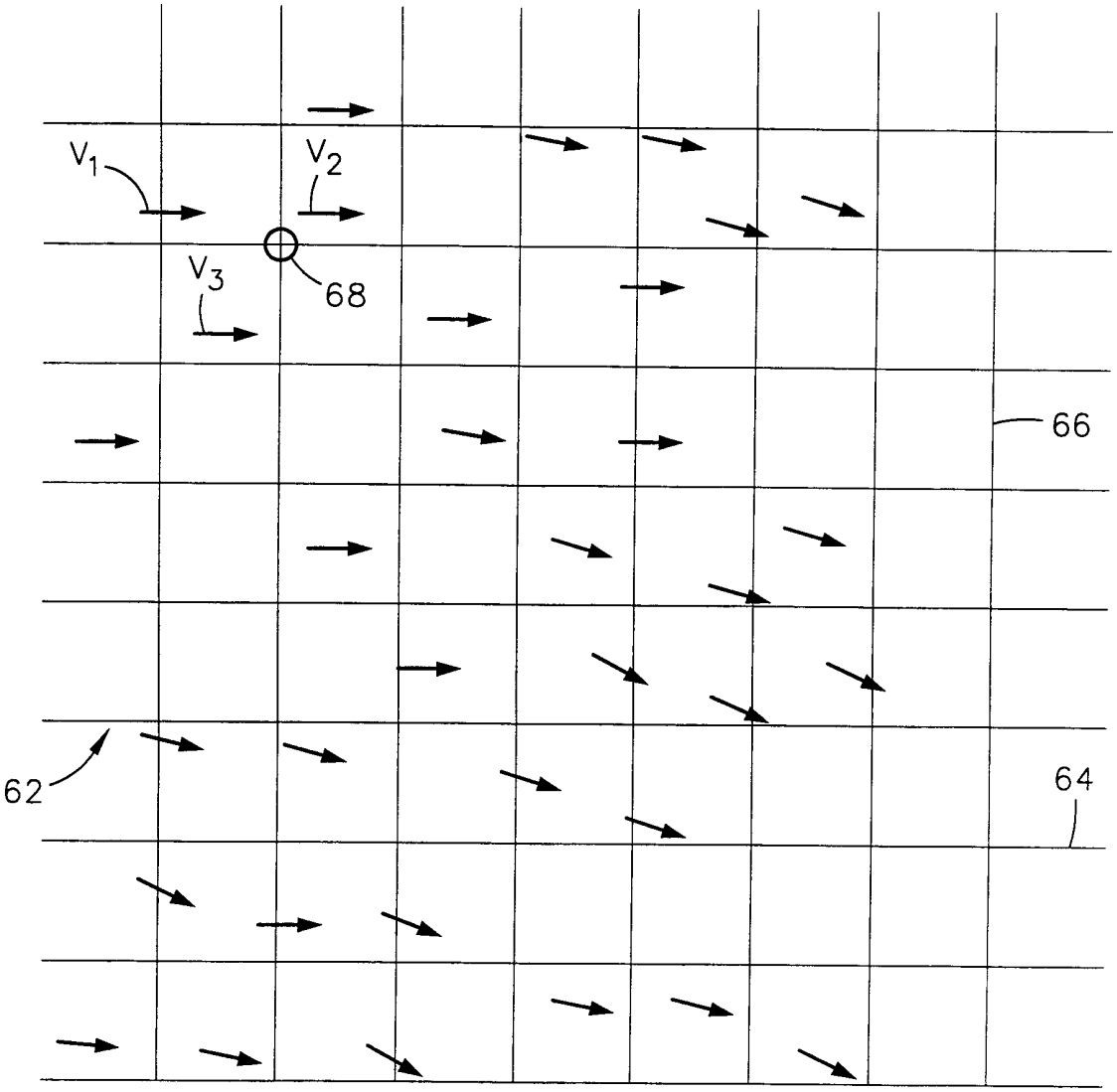


FIG. 5A

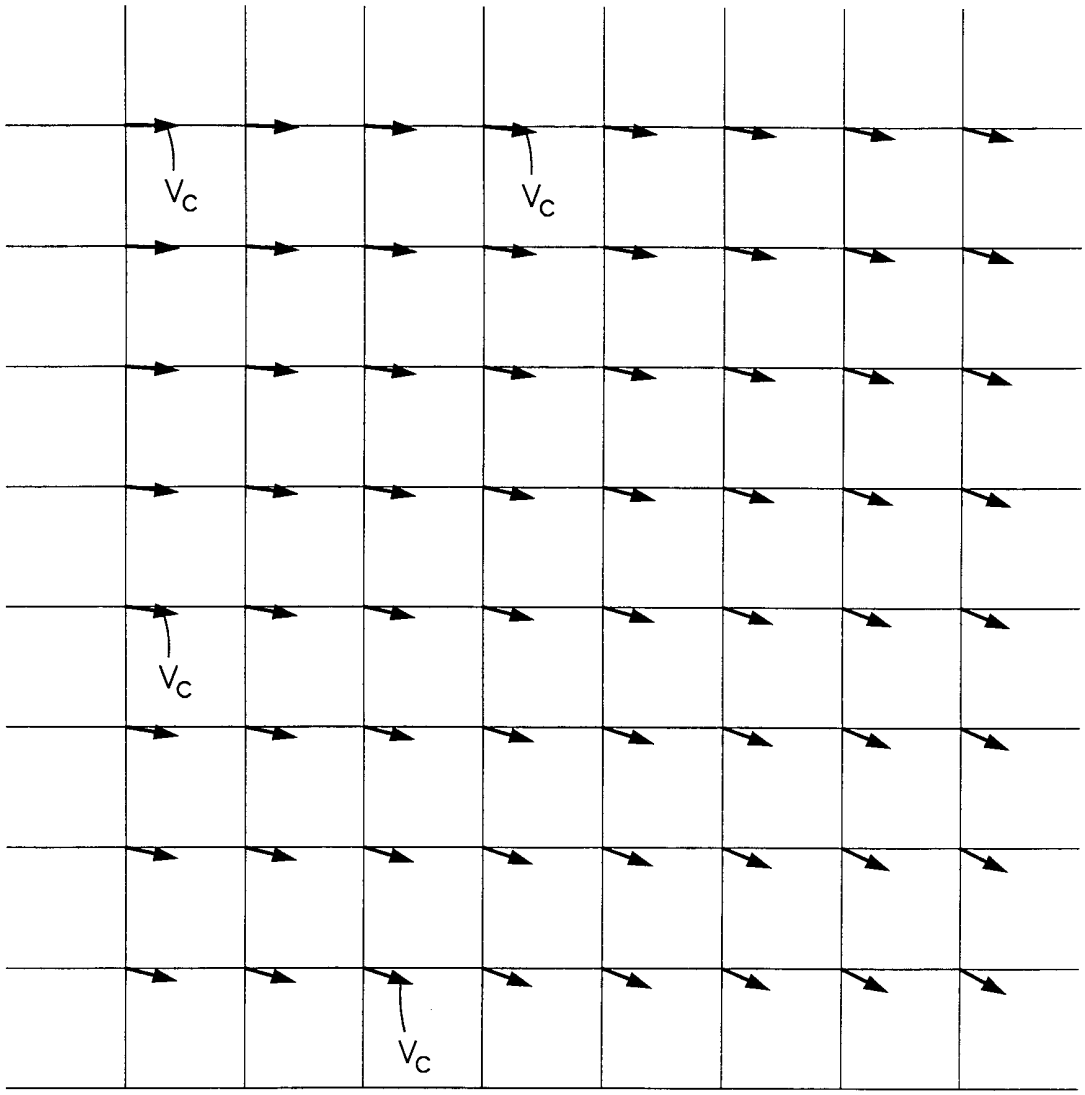


FIG. 5B

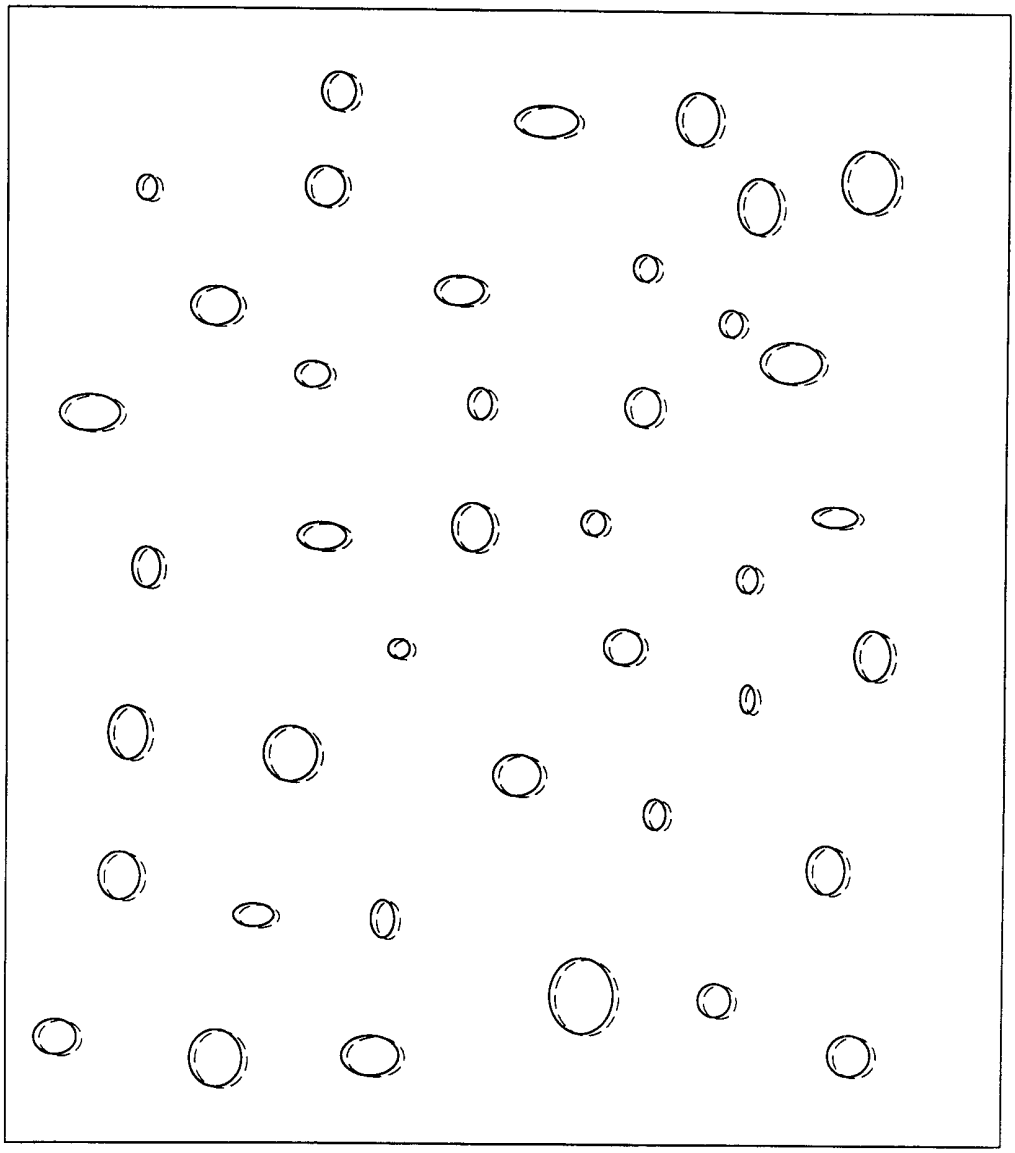


FIG. 6A

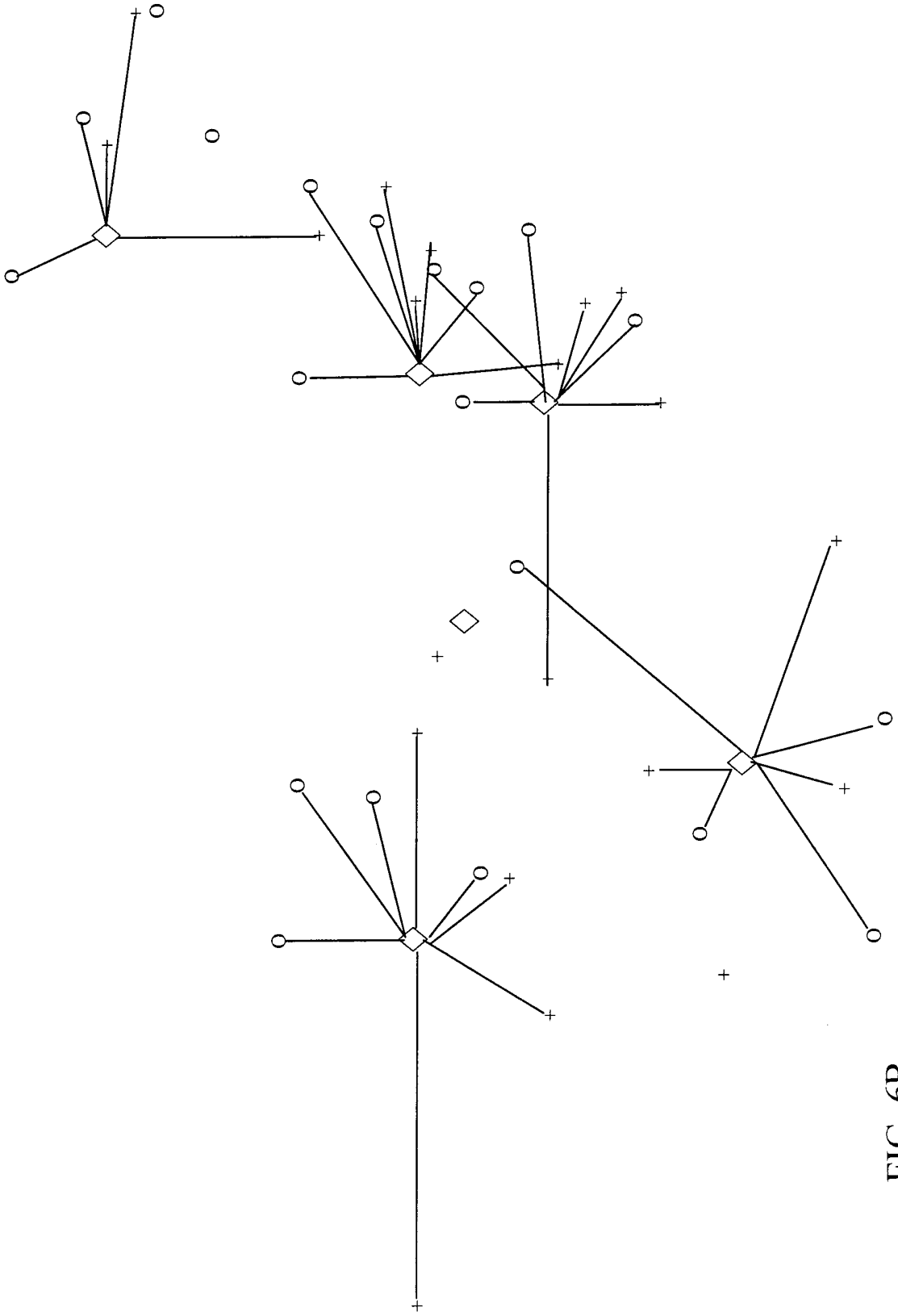


FIG. 6B

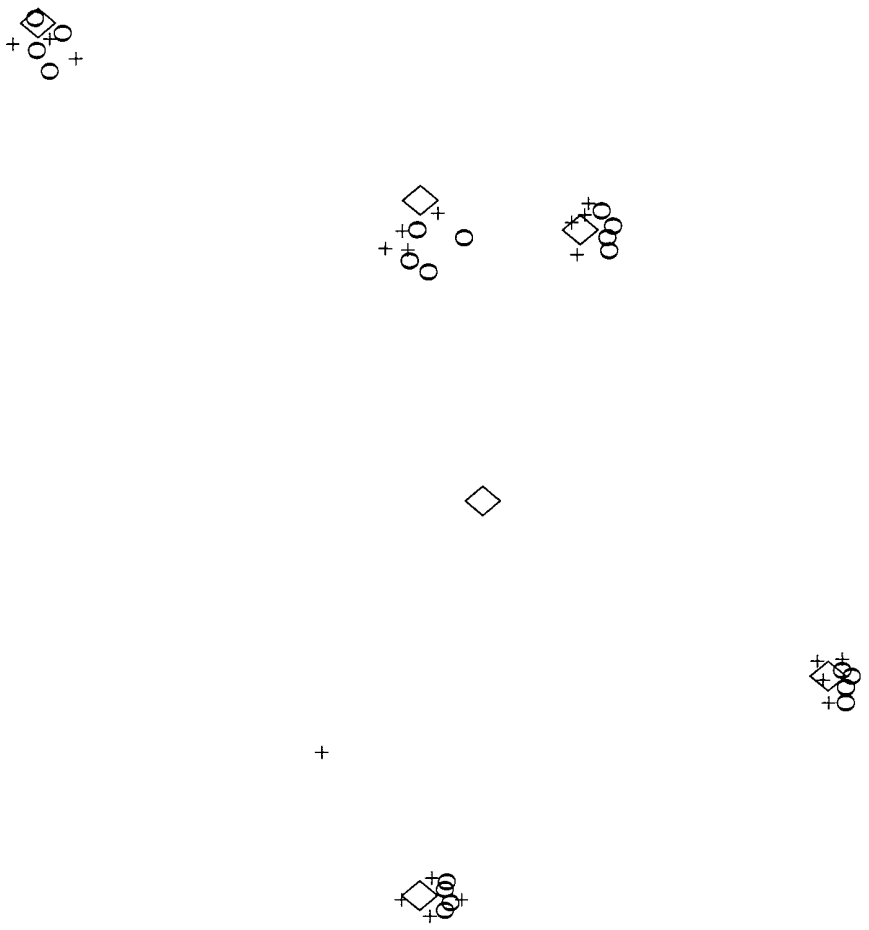


FIG. 6C

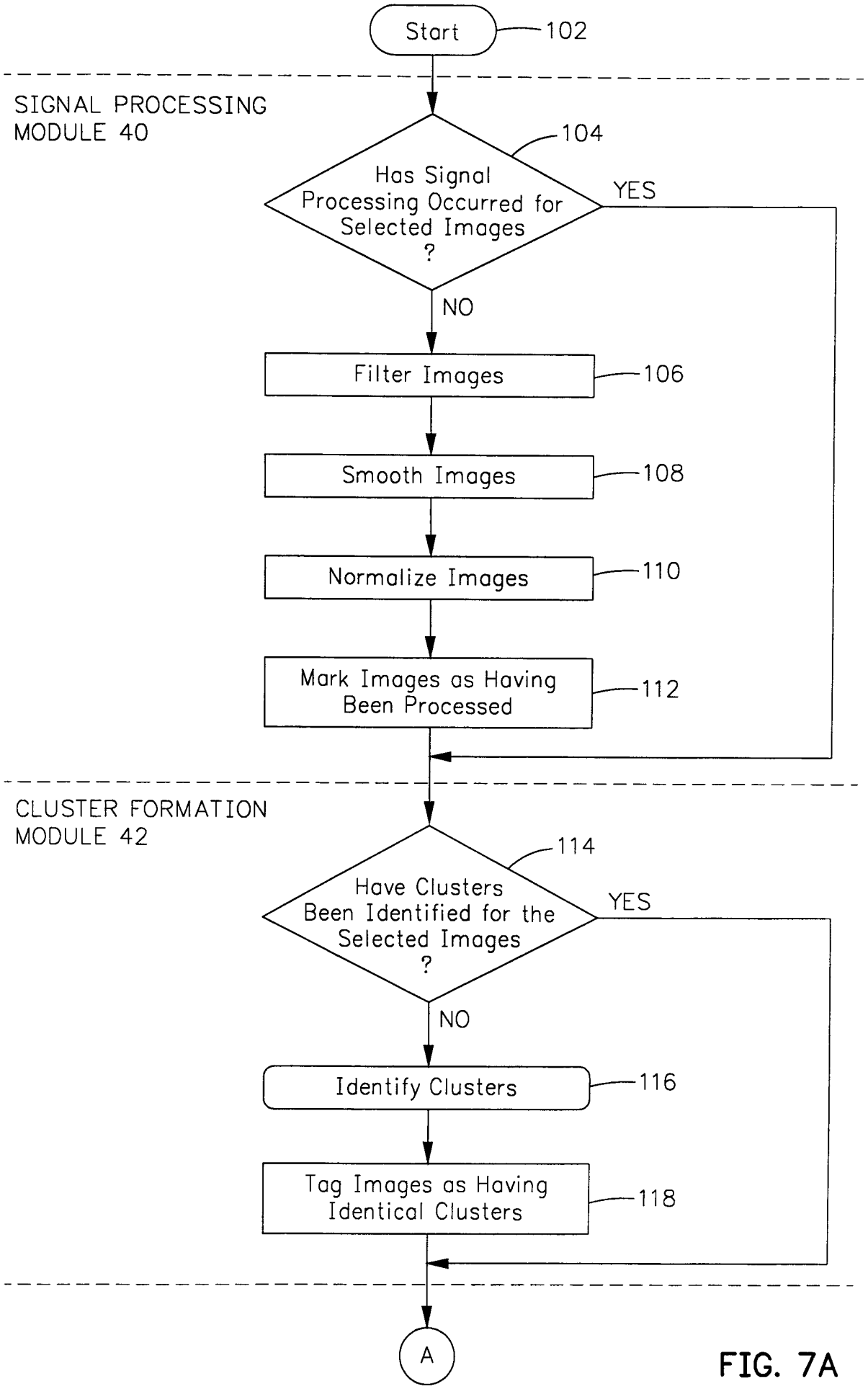

Start
102
SIGNAL PROCESSING MODULE 40
Has Signal Processing Occurred for Selected Images ?
104
YES
NO
Filter Images
106
Smooth Images
108
Normalize Images
110
Mark Images as Having Been Processed
112
CLUSTER FORMATION MODULE 42
Have Clusters Been Identified for the Selected Images ?
114
YES
NO
Identify Clusters
116
Tag Images as Having Identical Clusters
118
A

FIG. 7A

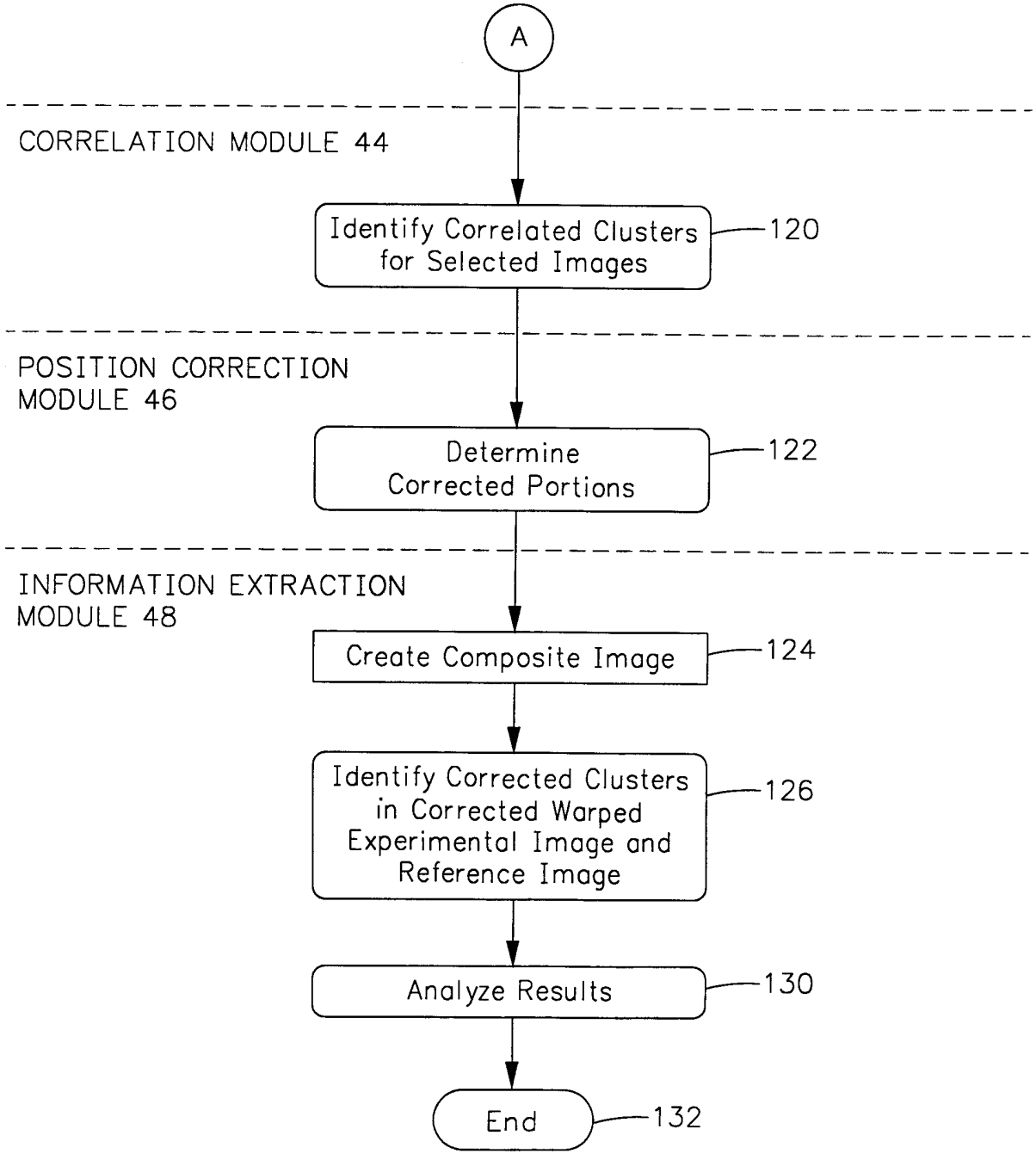

A
CORRELATION MODULE 44
Identify Correlated Clusters for Selected Images
120
POSITION CORRECTION MODULE 46
Determine Corrected Portions
122
INFORMATION EXTRACTION MODULE 48
Create Composite Image
124
Identify Corrected Clusters in Corrected Warped Experimental Image and Reference Image
126
Analyze Results
130
End
132

FIG. 7B

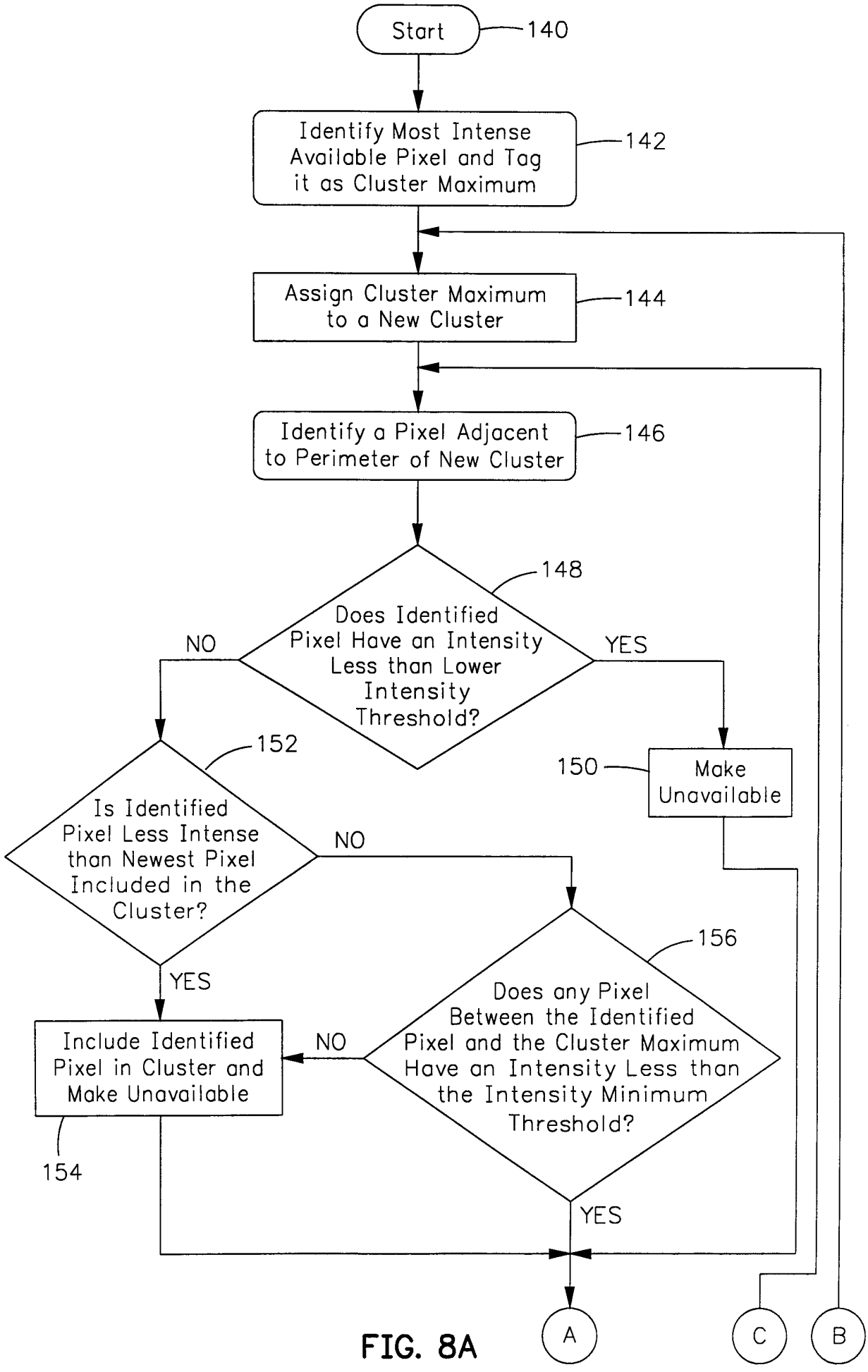

Start
140
Identify Most Intense Available Pixel and Tag it as Cluster Maximum
142
Assign Cluster Maximum to a New Cluster
144
Identify a Pixel Adjacent to Perimeter of New Cluster
146
148
Does Identified Pixel Have an Intensity Less than Lower Intensity Threshold?
NO
YES
150
Make Unavailable
152
Is Identified Pixel Less Intense than Newest Pixel Included in the Cluster?
NO
YES
156
Does any Pixel Between the Identified Pixel and the Cluster Maximum Have an Intensity Less than the Intensity Minimum Threshold?
NO
YES
Include Identified Pixel in Cluster and Make Unavailable
154
A
C
B

FIG. 8A

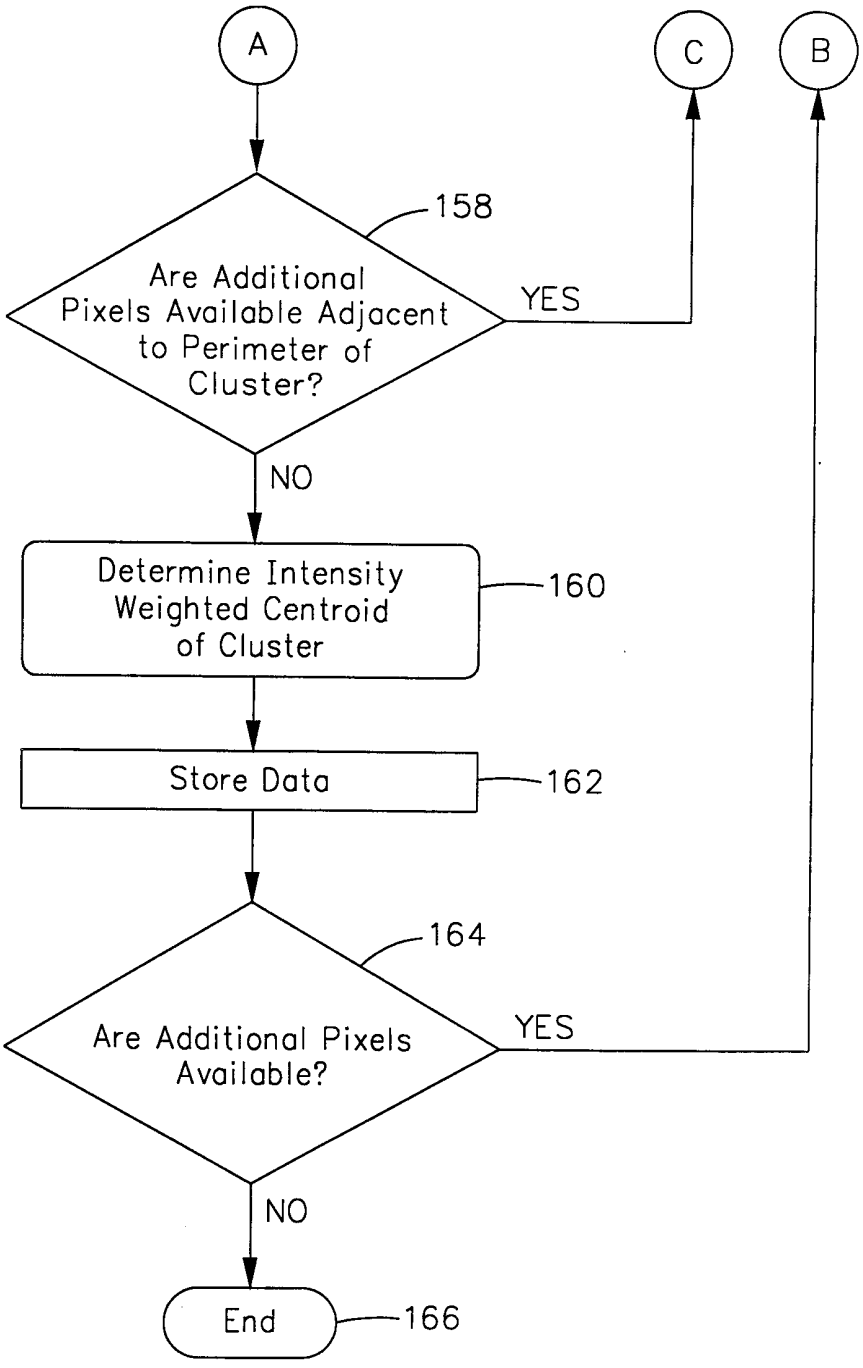

A
C
B
158
Are Additional Pixels Available Adjacent to Perimeter of Cluster?
YES
NO
Determine Intensity Weighted Centroid of Cluster
160
Store Data
162
164
Are Additional Pixels Available?
YES
NO
End
166

FIG. 8B

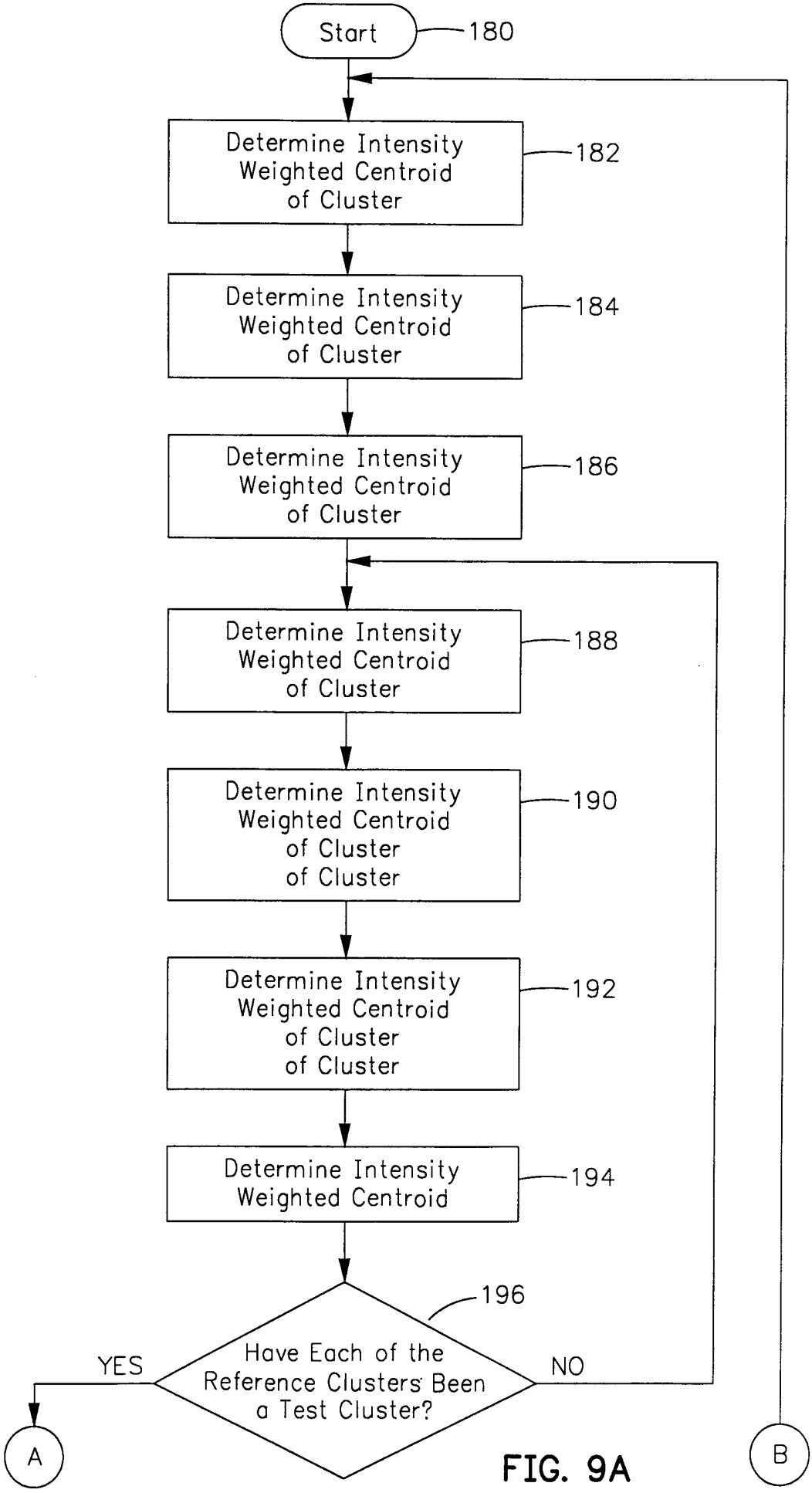

Start
180
Determine Intensity Weighted Centroid of Cluster
182
Determine Intensity Weighted Centroid of Cluster
184
Determine Intensity Weighted Centroid of Cluster
186
Determine Intensity Weighted Centroid of Cluster
188
Determine Intensity Weighted Centroid of Cluster of Cluster
190
Determine Intensity Weighted Centroid of Cluster of Cluster
192
Determine Intensity Weighted Centroid
194
Have Each of the Reference Clusters Been a Test Cluster?
196
YES
NO
A
B

FIG. 9A

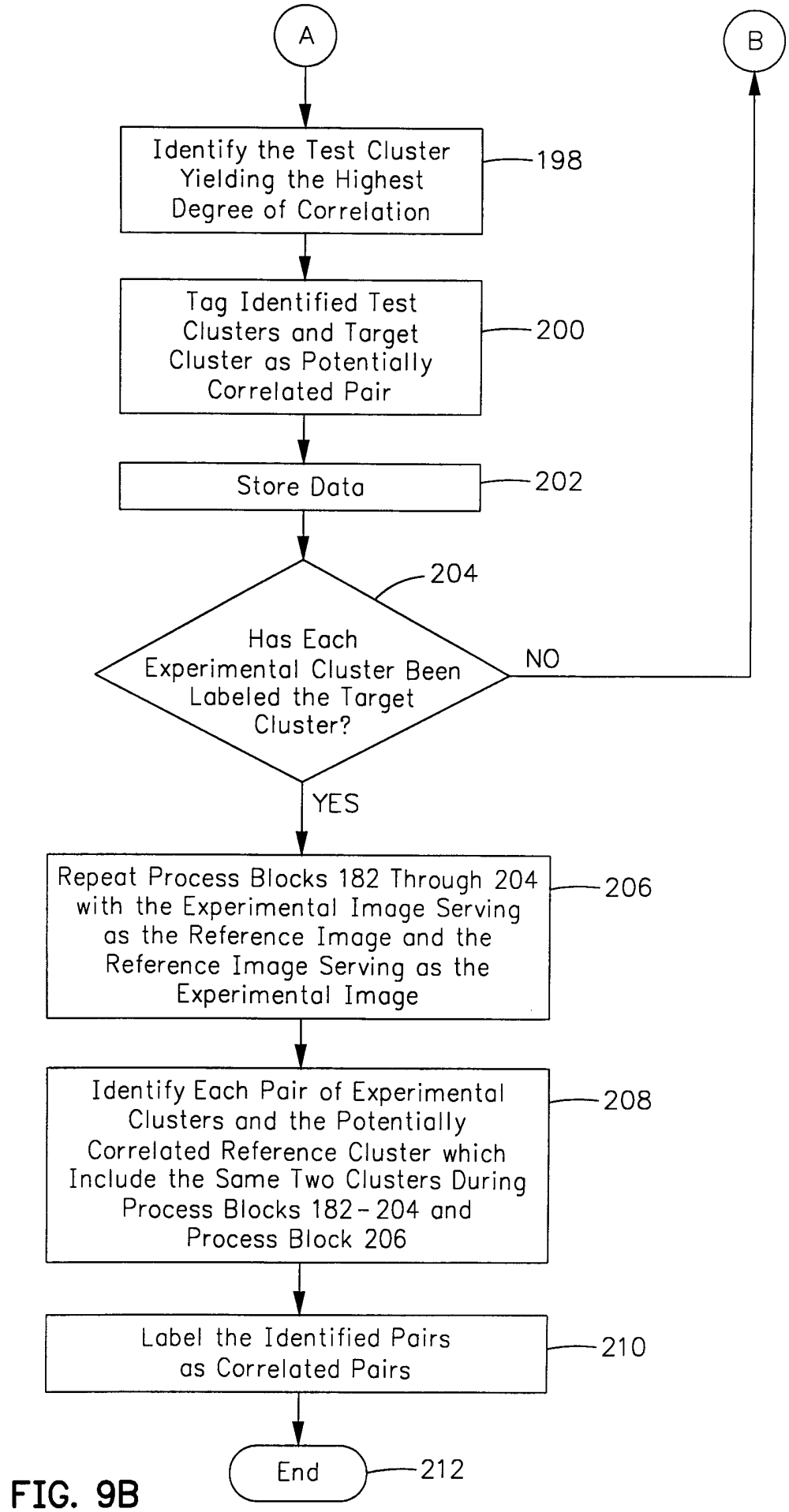

A
B
Identify the Test Cluster Yielding the Highest Degree of Correlation
198
Tag Identified Test Clusters and Target Cluster as Potentially Correlated Pair
200
Store Data
202
204
Has Each Experimental Cluster Been Labeled the Target Cluster?
NO
YES
Repeat Process Blocks 182 Through 204 with the Experimental Image Serving as the Reference Image and the Reference Image Serving as the Experimental Image
206
Identify Each Pair of Experimental Clusters and the Potentially Correlated Reference Cluster which Include the Same Two Clusters During Process Blocks 182 - 204 and Process Block 206
208
Label the Identified Pairs as Correlated Pairs
210
End
212

FIG. 9B

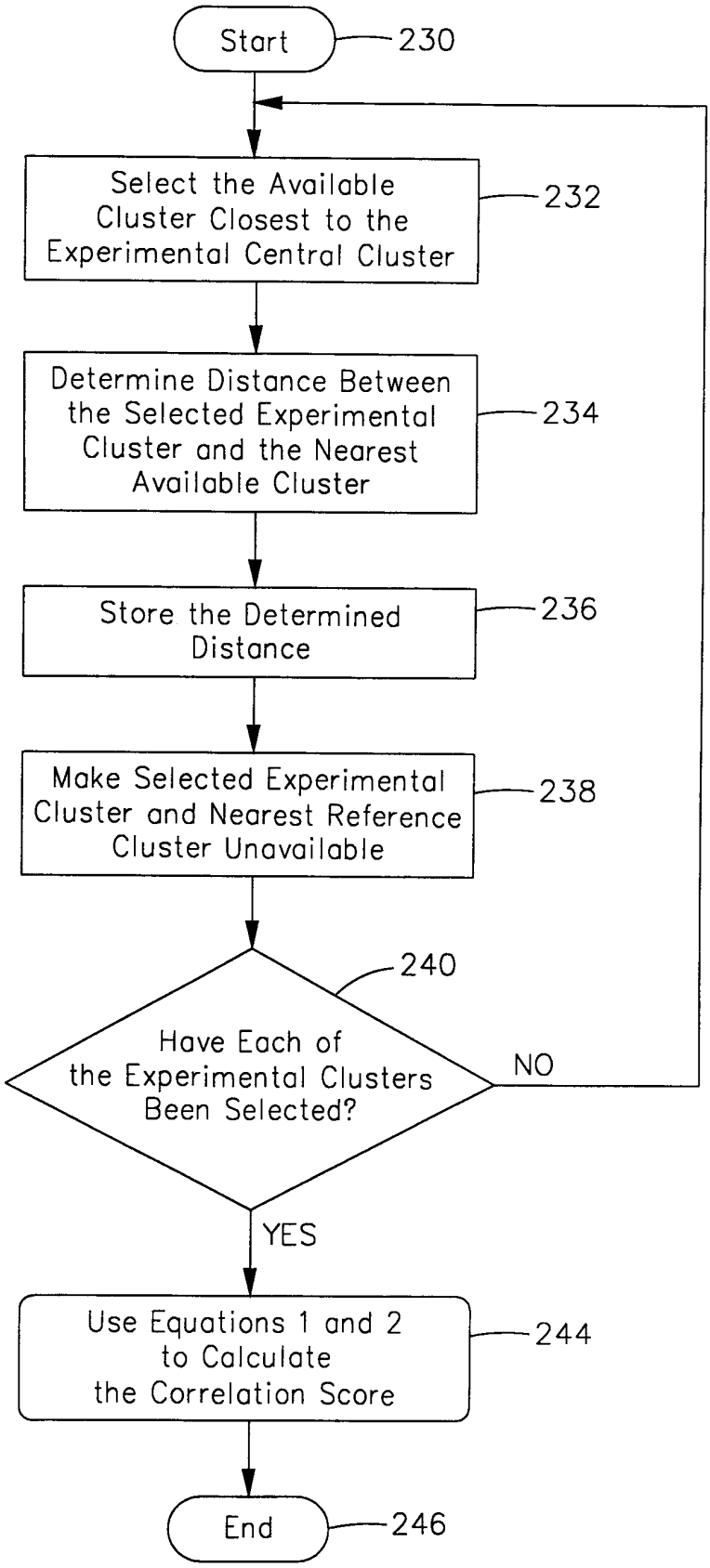

Start
230
Select the Available Cluster Closest to the Experimental Central Cluster
232
Determine Distance Between the Selected Experimental Cluster and the Nearest Available Cluster
234
Store the Determined Distance
236
Make Selected Experimental Cluster and Nearest Reference Cluster Unavailable
238
240
Have Each of the Experimental Clusters Been Selected?
NO
YES
Use Equations 1 and 2 to Calculate the Correlation Score
244
End
246

FIG. 10

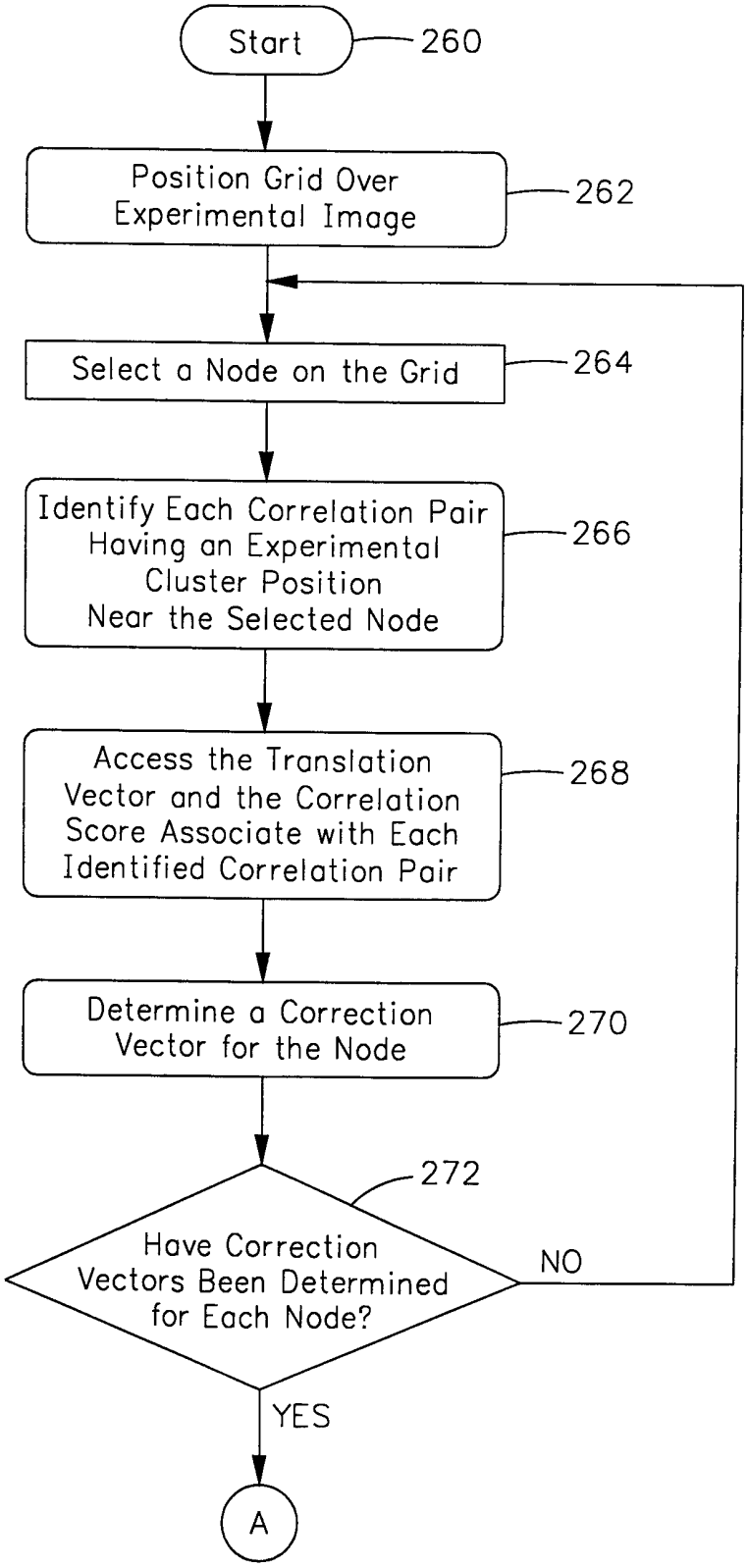

Start
260
Position Grid Over Experimental Image
262
Select a Node on the Grid
264
Identify Each Correlation Pair Having an Experimental Cluster Position Near the Selected Node
266
Access the Translation Vector and the Correlation Score Associate with Each Identified Correlation Pair
268
Determine a Correction Vector for the Node
270
272
Have Correction Vectors Been Determined for Each Node?
NO
YES
A

FIG. 11A

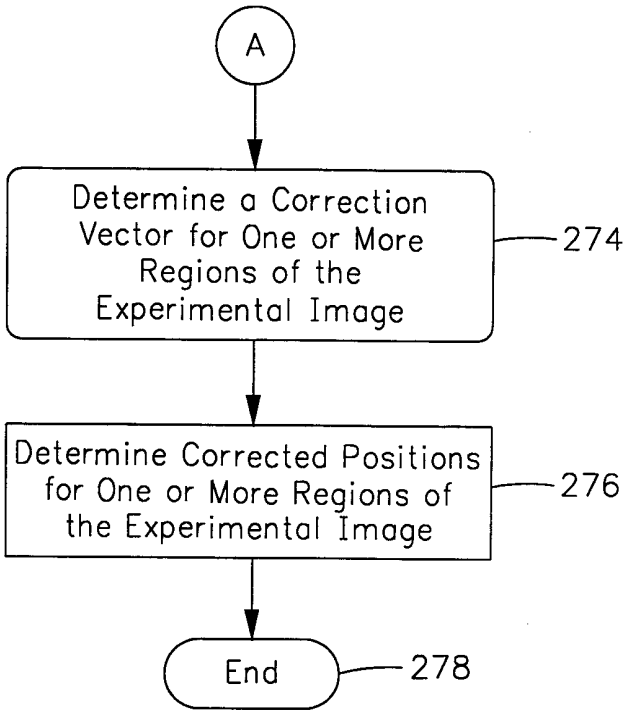

A
Determine a Correction Vector for One or More Regions of the Experimental Image
274
Determine Corrected Positions for One or More Regions of the Experimental Image
276
End
278

FIG. 11B

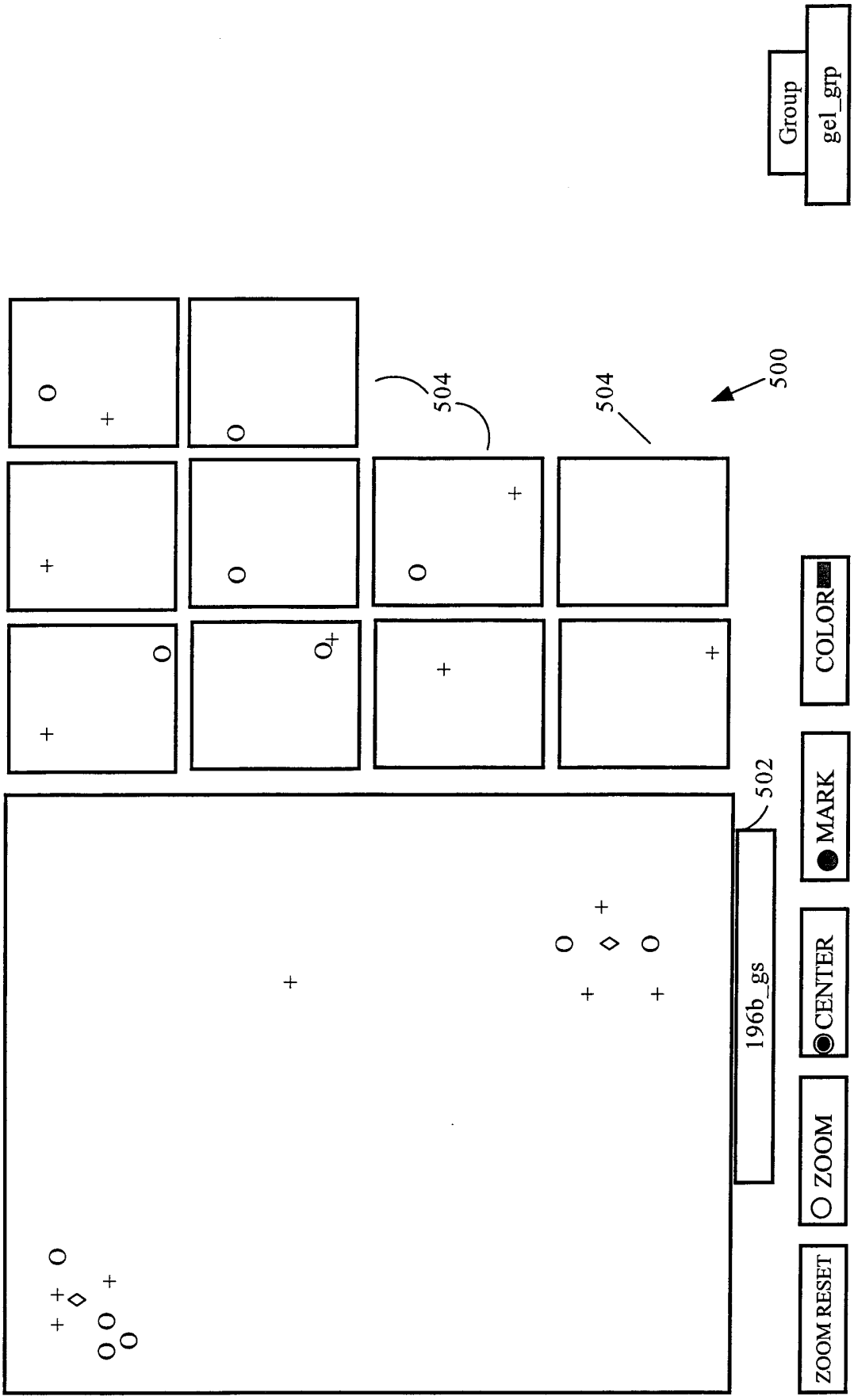

Group
gel_grp
500
504
504
502
196b_gs
COLOR
MARK
CENTER
ZOOM
ZOOM RESET

FIGURE 12