



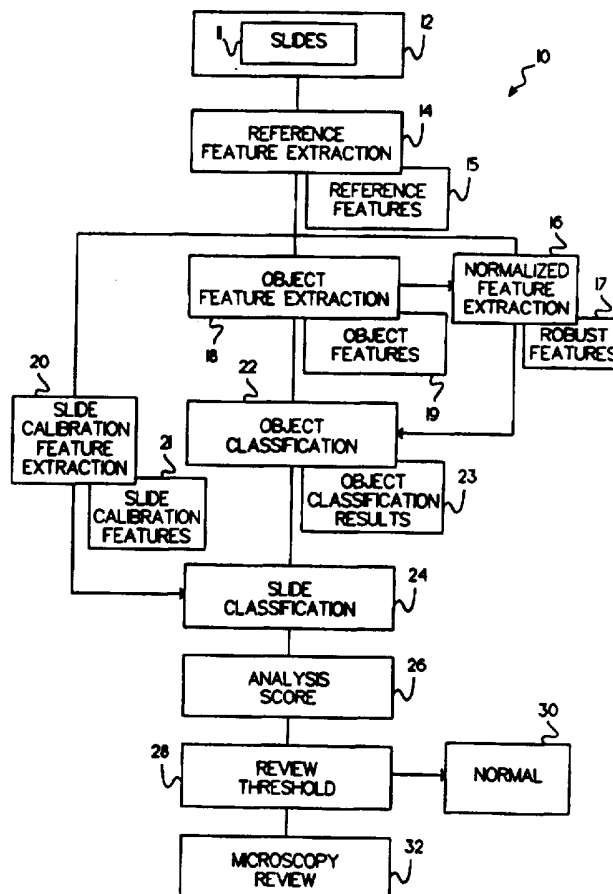
## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |           |                                                                                                                                                                                            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (51) International Patent Classification <sup>6</sup> :<br><b>G06K 9/00</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | <b>A1</b> | (11) International Publication Number: <b>WO 96/09602</b><br>(43) International Publication Date: 28 March 1996 (28.03.96)                                                                 |
| <p>(21) International Application Number: PCT/US95/11378</p> <p>(22) International Filing Date: 8 September 1995 (08.09.95)</p> <p>(30) Priority Data:<br/>08/309,115 20 September 1994 (20.09.94) US</p> <p>(71) Applicant: NEOPATH, INC. [US/US]; 8271 154th Avenue N.E., Redmond, WA 98052 (US).</p> <p>(72) Inventors: LEE, Shih-Jong, J.; 14116 Southeast 46th Street, Bellevue, WA 98006 (US). OH, Seho; 5229 107th Street S.W., Mukilteo, WA 98275 (US). KUAN, Chih-Chau, L.; 18217 N.E. 100th Court, Redmond, WA 98052 (US). ELLISON, Dayle, G.; 13226 N.E. 115th Court, Redmond, WA 98052 (US). BANNISTER, Wendy, R.; 358 North 82nd Street, Seattle, WA 98103 (US).</p> <p>(74) Agent: LEONE, George, A., Sr.; Leone &amp; Moffa, P.A., Suite 1275, 601 Lakeshore Parkway, Minnetonka, MN 55305 (US).</p> |           | <p>(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p><b>Published</b><br/>With international search report.</p> |

(54) Title: BIOLOGICAL ANALYSIS SYSTEM SELF-CALIBRATION APPARATUS

## (57) Abstract

Reference information (15) for a biological slide (11) is obtained. The reference information (15) normalizes the measured object features (19). Calibrated feature measurement (21), not based on absolute measurements, self-adjusts to match the situation of each slide (11), where each slide (11) is characterized by the reference information (15). The reference may be different from slide to slide because of the preparation variations. The calibrated features (21) will not carry the interslide variations. In addition, the reference information (15) provides a good indication of the slide condition such as dark stained, air dried, etc. which can be used as slide features (21) for the specimen classification. No alteration of the current practice of specimen preparation is required. The many slide context dependent features (21) improve the classification accuracy of the objects in a specimen (11).



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |                                       |    |                          |
|----|--------------------------|----|---------------------------------------|----|--------------------------|
| AT | Austria                  | GB | United Kingdom                        | MR | Mauritania               |
| AU | Australia                | GE | Georgia                               | MW | Malawi                   |
| BB | Barbados                 | GN | Guinea                                | NE | Niger                    |
| BE | Belgium                  | GR | Greece                                | NL | Netherlands              |
| BF | Burkina Faso             | HU | Hungary                               | NO | Norway                   |
| BG | Bulgaria                 | IE | Ireland                               | NZ | New Zealand              |
| BJ | Benin                    | IT | Italy                                 | PL | Poland                   |
| BR | Brazil                   | JP | Japan                                 | PT | Portugal                 |
| BY | Belarus                  | KE | Kenya                                 | RO | Romania                  |
| CA | Canada                   | KG | Kyrgyzstan                            | RU | Russian Federation       |
| CF | Central African Republic | KP | Democratic People's Republic of Korea | SD | Sudan                    |
| CG | Congo                    | KR | Republic of Korea                     | SE | Sweden                   |
| CH | Switzerland              | KZ | Kazakhstan                            | SI | Slovenia                 |
| CI | Côte d'Ivoire            | LI | Liechtenstein                         | SK | Slovakia                 |
| CM | Cameroon                 | LK | Sri Lanka                             | SN | Senegal                  |
| CN | China                    | LU | Luxembourg                            | TD | Chad                     |
| CS | Czechoslovakia           | LV | Latvia                                | TG | Togo                     |
| CZ | Czech Republic           | MC | Monaco                                | TJ | Tajikistan               |
| DE | Germany                  | MD | Republic of Moldova                   | TT | Trinidad and Tobago      |
| DK | Denmark                  | MG | Madagascar                            | UA | Ukraine                  |
| ES | Spain                    | ML | Mali                                  | US | United States of America |
| FI | Finland                  | MN | Mongolia                              | UZ | Uzbekistan               |
| FR | France                   |    |                                       | VN | Viet Nam                 |
| GA | Gabon                    |    |                                       |    |                          |

- 1 -

**BIOLOGICAL ANALYSIS SYSTEM SELF—  
CALIBRATION APPARATUS**

5 This invention relates to an automated biological specimen classification system and more particularly to a self-calibrating automated biological specimen classification system.

**BACKGROUND OF THE INVENTION**

10 Conventionally prepared Pap Smear slides vary in presentation considerably from slide to slide. These variations greatly degrade the discrimination potential derivable from cellular features or slide characteristics. It is generally accepted that, the ability of a cellular feature or slide characteristic to indicate normal vs. abnormal is limited by the  
15 great variations in appearance from slide to slide. The major source of variations are specimen sampling and preparation related including sampling tools, fixation methods, staining, and coverslip application, among others. The slide features that vary as a  
20 result include nuclear size, nuclear shape, nuclear density, chromatin patterns, cytoplasm size, cytoplasm density, and textures. The prior art relies on special preparation techniques such as fluid based fixation and special DNA staining to control the  
25 variations. However, they alter the current practice in Pap smear preparations and substantially increase cost.

**SUMMARY OF THE INVENTION**

30 The invention provides a self-calibrating automated biological specimen classifier that adjusts for slide variations.

The self calibration method of the invention establishes reference information for each slide. The reference information for each slide is used to  
35 normalize measured object features. This results in

- 2 -

calibrated feature measurement not based on absolute measurements. The apparatus of the invention self adjusts to match the situation of each slide, where each slide is characterized by the reference  
5 information. In this way, even the reference information may be different from slide to slide because of preparation variations. The calibrated features will not carry the inter-slide variations. In addition, the reference information provides a good  
10 indication of the slide condition such as dark stained, air dried, etc. which may be used as slide features for subsequent specimen classification.

The method provides a practical way to solve the specimen variation problem without the alteration of  
15 the current practice of specimen preparations.

The method generates many slide context dependent features to improve the classification accuracy of the objects in a specimen.

Other objects, features and advantages of the  
20 present invention will become apparent to those skilled in the art through the description of the preferred embodiment, claims and drawings herein wherein like numerals refer to like elements.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

25 To illustrate this invention, a preferred embodiment will be described herein with reference to the accompanying drawings.

Figures 1A, 1B and 1C show the automated cytology screening apparatus of the invention.

30 Figure 2 shows a process flow diagram of the self calibration method of the invention.

Figure 3 shows a cell with a cell nucleus surrounded by a bounding box.

Figure 4 shows a SKIZ boundary.

35 Figure 5 shows a 3x3 dilation residue method of

- 3 -

the invention.

Figure 6 shows one example of a self calibrating biological specimen analysis apparatus.

5 Figure 7 shows another example of a self calibrating biological specimen analysis apparatus.

Figure 8 shows yet another example of a self calibrating biological specimen analysis apparatus.

Figure 9 shows still another example of a self calibrating biological specimen analysis apparatus.

10 **DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT**

Now refer to Figures 1A, 1B and 1C which show a schematic diagram of the apparatus of the invention. The apparatus of the invention comprises an imaging system 502, a motion control system 504, an image processing system 536, a central processing system 540, and a workstation 542. The imaging system 502 is comprised of an illuminator 508, imaging optics 510, a CCD camera 512, an illumination sensor 514 and an image capture and focus system 516. The image capture and focus system 516 provides video timing data to the CCD cameras 512, the CCD cameras 512 provide images comprising scan lines to the image capture and focus system 516. An illumination sensor intensity is provided to the image capture and focus system 516 where an illumination sensor 514 receives the sample of the image from the optics 510. In one embodiment of the invention, the optics may further comprise an automated microscope. The illuminator 508 provides illumination of a slide. The image capture and focus system 516 provides data to a VME bus 538. The VME bus distributes the data to an image processing system 536. The image processing system 536 is comprised of field-of-view processors 568. The images are sent along the image bus 564 from the image capture and focus system 516. A central processor 540 controls

15  
20  
25  
30  
35

- 4 -

the operation of the invention through the VME bus 538. In one embodiment the central processor 562 comprises a 68030 CPU. The motion controller 504 is comprised of a tray handler 518, a microscope stage controller 520, a microscope tray controller 522, and a calibration slide 524. The motor drivers 526 position the slide under the optics. A bar code reader 528 reads a barcode located on the slide 524. A touch sensor 530 determines whether a slide is under the microscope objectives, and a door interlock 532 prevents operation in case the doors are open. Motion controller 534 controls the motor drivers 526 in response to the central processor 540. An Ethernet communication system 560 communicates to a workstation 542 to provide control of the system. A hard disk 544 is controlled by a Sun Spark Classic workstation 550. A tape drive 546 is connected to the Sun Spark Station as well as a modem 548, a monitor 552, a keyboard 554, and a mouse pointing device 556. A printer 558 is connected to the ethernet 560.

In a presently preferred embodiment of the invention, the system disclosed herein is used in a system for analyzing cervical pap smears, such as that shown and disclosed in U.S. Patent Application Serial No. 07/838,064, entitled "Method For Identifying Normal Biomedical Specimens", by Alan C. Nelson, et al., filed February 18, 1992; U.S. Patent Application Serial No. 08/179,812 filed January 10, 1994 which is a continuation in part of U.S. Patent Application Serial No. 07/838,395, entitled "Method For Identifying Objects Using Data Processing Techniques", by S. James Lee, et al., filed February 18, 1992; U.S. Patent Application Serial No. 07/838,070, now U.S. Pat. No. 5,315,700, entitled "Method And Apparatus For Rapidly Processing Data Sequences", by Richard S.

- 5 -

Johnston, et al., filed February 18, 1992; U.S. Patent Application Serial No. 07/838,065, filed 02/18/92, entitled "Method and Apparatus for Dynamic Correction of Microscopic Image Signals" by Jon W. Hayenga, et al.; and U.S. Patent Application Serial No. 08/302,355, filed September 7, 1994 entitled "Method and Apparatus for Rapid Capture of Focused Microscopic Images" to Hayenga, et al., which is a continuation-in-part of Application Serial No. 07/838,063 filed on February 18, 1992 the disclosures of which are incorporated herein, in their entirety, by the foregoing references thereto.

The present invention is also related to biological and cytological systems as described in the following patent applications which are assigned to the same assignee as the present invention, filed on September 20, 1994 unless otherwise noted, and which are all hereby incorporated by reference including U.S. Patent Application Serial No. 08/309,118, to Kuan et al. entitled, "Field Prioritization Apparatus and Method," U.S. Patent Application Serial No. 08/309,061, to Wilhelm et al., entitled "Apparatus for Automated Identification of Cell Groupings on a Biological Specimen," U.S. Patent Application Serial No. 08/309,116 to Meyer et al. entitled "Apparatus for Automated Identification of Thick Cell Groupings on a Biological Specimen," U.S. Patent Application Serial No. 08/308,992, to Lee et al. entitled "Apparatus for Identification and Integration of Multiple Cell Patterns," U.S. Patent Application Serial No. 08/309,063 to Lee et al. entitled "A Method for Cytological System Dynamic Normalization," U.S. Patent Application Serial No. 08/309,248 to Rosenlof et al. entitled "Method and Apparatus for Detecting a Microscope Slide Coverslip," U.S. Patent Application

- 6 -

Serial No. 08/309,077 to Rosenlof et al. entitled  
"Apparatus for Detecting Bubbles in Coverslip  
Adhesive," U.S. Patent Application Serial No.  
08/309,931, to Lee et al. entitled "Cytological Slide  
5 Scoring Apparatus," U.S. Patent Application Serial No.  
08/309,148 to Lee et al. entitled "Method and  
Apparatus for Image Plane Modulation Pattern  
Recognition," U.S. Patent Application Serial No.  
08/309,250 to Lee et al. entitled "Apparatus for the  
10 Identification of Free-Lying Cells," U.S. Patent  
Application Serial No. 08/309,209 to Oh et al.  
entitled "A Method and Apparatus for Robust Biological  
Specimen Classification," U.S. Patent Application  
Serial No. 08/309,117, to Wilhelm et al. entitled  
15 "Method and Apparatus for Detection of Unsuitable  
Conditions for Automated Cytology Scoring."

It is to be understood that the various processes  
described herein may be implemented in software  
suitable for running on a digital processor. The  
20 software may be embedded, for example, in the central  
processor 540.

Now refer to Figure 2 which shows the slide  
sorting data flow of the method of the invention. A  
biological specimen such as a pap smear is loaded onto  
25 a slide processing system at step 12. The system  
processes the slides 11 and generates an analysis  
score 26. The slides 11 are sorted into normal 30 or  
microscopy review 32 categories based on a threshold  
28 of the analysis score.

30 The decision logic is:

**IF** analysis score < normal threshold **THEN** normal  
**ELSE** microscopy review

The self calibration strategy starts with an  
initial scan 12 of a slide 11 to establish the  
35 reference features 15. The slide 12 is then re-

- 7 -

scanned 33 after the reference features 15 are established in step 14. During the re-scanning process, object features 19 are extracted in step 18. The extracted object features 19 are calibrated by the reference features 15 to generate a set of robust features 17 in step 16. The robust features 17 along with the original object features 19 are used to perform object classification in step 22. The object classification results 23 are accumulated for the whole slide 11. A set of slide calibration features 21 are extracted from the reference features 15 in step 20. The extracted slide calibration features 21, combined with the accumulated object features 23, are then used by the slide classification step 24 to generate the anomaly score 26 also called the analysis score 26 for a slide.

The reference feature extraction screening method proceeds as follows. A slide 11 is pre-screened 12 to extract the reference features 15. The slide 11 is then re-screened 33 and the reference features 15 along with the object features 19 are used for object classification 22. In a preferred embodiment of invention, the pre-screening process scans up to a maximum number of reference cells or a maximum number of FOV's whichever comes first.

Reference cell characteristics are defined as follows. The characteristics for good reference cells include high prevalence, easily detectable, and good correlation to changes in objects of interest. The free-lying squamous intermediate cells are selected as reference cells for Pap smear screening. Those skilled in the art will recognize that other reference cell characteristics may be used.

#### Feature Descriptions

float compactness -  $\text{perimeter}^2 / \text{area}$

- 8 -

float area2 - Same feature as area, except the area of interest (labeled nuclear region) is first eroded by a 3x3 structure element (1-pixel wide).

5 float mod\_N\_C\_ratio - A ratio between the nuclear area and the cytoplasm area is calculated. The cytoplasm for each nuclei is determined by taking only the cytoplasm area that falls inside of a SKIZ boundary between all nuclei objects. The area of the cytoplasm is the number of the cytoplasm pixels that are in the  
10 SKIZ area corresponding to the nuclei of interest. The edge of the image is treated as an object and therefore creates a SKIZ boundary as shown in Figure 4.

15 cell\_semi\_isolated - This feature can take on two values: 0.0 and 1.0 (1.0 indicates the nucleus is semi-isolated). A nucleus is determined to be semi-isolated when the center of its bounding box is a minimum euclidian pixel distance from all other nuclei  
20 (center of their bounding boxes). Figure 3 shows a cell with a cell nucleus surrounded by a bounding box. The minimum distance that is used as a threshold is stored in the global floating point variable *AlgSemilsolatedDistanceMin* (default is 50.0).

25 A number of features are designed to measure nuclear textures, the original image is blurred using a 5x5 binomial filter. A residue is created with the absolute magnitude differenced between the original and the blurred image.

30 The original image nucleus is assumed to contain information not only about the nucleus, but also about background matter. The gray level recorded at each pixel of the nucleus summarize the optical density of all matter in the vertical column that contains the

- 9 -

particular nucleus pixel. In other words, if the nucleus is located in a cytoplasm which itself is located in a mucus stream, the gray level values of the nucleus may reflect not only the nuclear matter, but also the cytoplasm and mucus in which the nucleus lies. To try to measure features of the nucleus without influence of the surroundings and to measure the nucleus surroundings, two regions have been defined around the nucleus. Two regions have been defined because of a lack of information about how much area around the nucleus is enough to identify what is happening in proximity to the nucleus.

The two regions are rings around each nucleus. The first ring expands 5 pixels out from the nucleus (box 7x7 and diamond 4) and is designated as the "small" ring. The second region expands 15 pixels out from the nucleus (box 15x15 and diamond 9) and is called the "big" ring.

float big\_blur\_ave - Pixel average of blur residue image over the big ring area.

float nuc\_blur\_ave - Pixel average of blur residue image over the nuclear area.

float nc\_score\_alt\_r4 - The median value of the 8 connect distance transform of the cytoplasm mask is found inside the 3x3 dilation residue of the nuclear mask. Figure 5 shows a 3x3 dilation residue method of the invention. This value is always an integer since the discrete probability density function always crosses 0.5 at the integer values. Call this value as Y. The n feature is : nuclear area/ (3.14\*Y\*Y).

A reference cell classifier detects the reference cells. The reference cell classifier uses stable

- 10 -

features such as staining independent or relative features for classification. In addition, the classifier is designed to maintain high specificity and avoid false positives. In this way, the detected  
5 objects are most likely true reference cells. In a preferred embodiment of the implementation, the squamous intermediate cells in a Pap smear may be detected by the following rule:

If CSEMISOL > 0.5 and MDNCRATO <= 0.073325 and  
10 NCSCATr4 <= 0.15115 and NBLURAVE > 4.6846 and BGBLUAVE <= 4.5655 and AREA2 > 96.5 and COMPACT <= 10.2183

Where CSEMISOL > 0.5 requires the cell be an isolated one;

MDNCRATO <= 0.073325 and NCSCATr4 <= 0.15115 are  
15 two rules which require the nuclear to cytoplasm size and radius ratio be lower than a limit (need to make sure the cytoplasm size are large with respect to nuclear size);

NBLURAVE > 4.6846 requires the nucleus to have  
20 sufficient texture;

BGBLUAVE <= 4.565 requires the cytoplasm area to be relatively smooth;

area2 > 96.6 requires the nucleus to be large enough;

25 and COMPACT <= 10.2183 requires the nuclear shape be round.

An object is classified as a reference cell only when all conditions are met. Note that no density features are used for squamous intermediate cell  
30 detection. This is because density is not a stable feature, it varies from slide to slide.

Important features of each detected reference cells are saved. The features include:

- 11 -

- Morphological features such as nuclear size and shape
- Density features such as nuclear Integrated Optical Density (IOD), nuclear mean optical density (OD), cytoplasm mean OD
- Texture features such as standard deviation of nuclear OD, average edge gradient of nucleus, etc.
- Nuclear to cytoplasm contrast features such as OD difference, edge contrast, etc.

Slide reference features can be computed from the saved reference cell features. In a preferred embodiment of implementation, the slide reference features are: Means ( $m$ ), medians ( $m_0$ ), standard deviations ( $\sigma$ ), and 10-90% spreads ( $s_0$ ) of the features over all reference cells saved for the whole slide.

Calibrated feature extraction proceeds as follows. During the re-screening of the slide, significant objects are detected and object features are extracted. The extracted object features are calibrated by a feature combination strategy which combines object features with reference features in certain mathematical forms. Slide reference features are slide staining or squamous intermediate cell nuclear size related features as described above. In a preferred embodiment of implementation, six types of feature combination formulas are used. The unified formula of these six combinations is:

$$NewFeature = \frac{f - \alpha g}{\beta}$$

where  $\alpha$  is evaluated to maximize the discrimination power of the resulting *NewFeature* and  $\beta$  is a calibrated factor. The forms  $f$  and  $g$  for six types of combinations are listed in the following table.

- 12 -

5

| Combination Type | f                             | g                               |
|------------------|-------------------------------|---------------------------------|
| Type 1           | s                             | r                               |
| Type 2           | s/r                           | 1/4                             |
| Type 3           | s/r <sup>2</sup>              | r <sub>1</sub> /r <sub>2</sub>  |
| Type 4           | $\frac{\log(s)  r_1 }{ r_2 }$ | $\frac{\log(r_1)  r_1 }{ r_2 }$ |
| Type 5           | $\frac{\log(s)}{\log(r)}$     | $\frac{1}{\log(r)}$             |
| Type 6           | log(s)                        | log(r)                          |

where  $s$  is object feature and  $r$ ,  $r_1$  and  $r_2$  are slide reference features. Note that where  $\log(x)$  is indicate  $\log(|x|+0.001)$  is used; where  $x=s, r, r_1$  for the singularity of the logarithmic function, and we set  $r_2=10^{-6}$  if  $r_2 < 10^{-6}$ . For Type 2, we set  $r = 0.01 * \text{sgn}(r) * R$  if  $|r| < 0.01 * R$  where  $R$  is maximum of  $|r|$  for all training data where the function  $\text{sgn}(\cdot)$  is

$$\text{sgn}(x) = \begin{cases} 1; & x \geq 0 \\ -1; & x < 0 \end{cases}$$

15 In the same way, for Type 5, we set  $\log(r) = 0.01 * \text{sgn}(\log(r)) * \log(\hat{R})$  if  $|\log(|r|+0.001)| < 0.01 * \log(\hat{R})$  where  $\log(\hat{R})$  is maximum of  $|\log(|r|+0.001)|$  for all training data.

20 As shown above, there are a variety of combinations of the calibrated features. Feature selection methods such as the stepwise discriminate analysis (see "Numerical Evaluation of Cytologic Data: VII. Multivariate Significance Tests," by Bartels P.H., Analyt Quant Cytol 3:1-8, 1981.) can be used to  
25 select the features having the highest discriminate power. The calibrated features along with the original object features are used to train the object classifier.

- 13 -

In a preferred embodiment of the invention, the classifiers were constructed through a training process as described in "A Processing Strategy for Automated Papanicolaou Smear Screening" by Lee et al. in the Analytical and Quantitative Cytology and Histology, Vol. 14 No. 5, pp. 415-425, October 1992. The Classification And Regression Trees (CART) (see "Classification And Regression Trees", by Breiman et al., Belmont, California, Wadsworth, 1984) is used as the basis for the classifiers.

Slide calibration features are also extracted from the reference features. These features can be used to normalize slide features. In a preferred embodiment of the invention, the slide calibration features include:

- slide reference features;
- number of reference cells detected from the whole slide;
- number of reference cells detected from the whole slide/number of objects detected from the whole slide;
- | mean - medians | for each reference feature;
- | mean - medians | / standard deviation for each reference feature;
- | 10 -90% spread - standard deviation | for each reference feature; and
- | 10 -90% spread - standard deviations | / standard deviation for each reference feature.

Object classification results along with the slide calibration features can be used for slide classification. The slide classifier can again be constructed through a training process as described in

- 14 -

"A Processing Strategy for Automated Papanicolaou Smear Screening" by Lee et al. in the Analytical and Quantitative Cytology and Histology, Vol. 14 No. 5, pp. 415-425, October 1992. The Classification And Regression Trees (CART) (See "Classification And Regression Trees", by Breiman et al., Belmont, California, Wadsworth, 1984.) can be used as the basis for the classifiers. The output of the slide classifier is the analysis score or anomaly score.

Refer now to Figure 6, one example of a self calibrating biological specimen analysis apparatus is shown. The apparatus includes a reference feature extractor 42 to detect reference features 46 in at least one biological specimen image 40, wherein the reference feature extractor has a reference feature output. An object feature extractor 44 detects object features in at least one biological specimen, wherein the object feature extractor has an object feature output. A calibrated feature extractor 48 is connected to the reference feature output and the object feature output, wherein the calibrated feature extractor has a calibrated feature output. An object classifier 52 is connected to the object feature output and the calibrated feature output. The object classifier 52 has a object classification result output. A slide feature extractor 54 is connected to the object classification result output, wherein the slide feature extractor has a slide feature output. A slide calibration feature extractor 50 is connected to the reference feature output, wherein the slide calibration feature extractor 50 has a slide calibration feature output. A slide classifier 56 is connected to the slide calibration feature output and the slide feature output, wherein the slide classifier has an analysis score output. An analysis score

- 15 -

thresholder 58 is connected to the analysis score output, having a normal indicator output 62 and a review indicator output 60. In one embodiment the biological specimen may be prepared by the Papanicolaou method.

Refer now to Figure 7, another example of a self calibrating biological specimen analysis apparatus is shown. The apparatus includes a reference feature extractor 68 to detect reference features in at least one biological specimen image 64, wherein the reference feature extractor has a reference feature output. An object feature extractor 66 detects object features in at least one biological specimen, wherein the object feature extractor has an object feature output. An object classifier 70 is connected to the object feature output and the calibrated feature output. The object classifier 70 has a object classification result output 72.

Refer now to Figure 8, another example of a self calibrating biological specimen analysis apparatus is shown. The apparatus includes a reference feature extractor 78 to detect reference features in at least one biological specimen image 74, wherein the reference feature extractor has a reference feature output. An object feature extractor 76 detects object features in at least one biological specimen, wherein the object feature extractor has an object feature output. A calibrated feature extractor 80 is connected to the reference feature output and the object feature output, wherein the calibrated feature extractor has a calibrated feature output 82.

Refer now to Figure 9, another example of a self calibrating biological specimen analysis apparatus is shown. The apparatus includes a reference feature extractor 88 to detect reference features in at least

- 16 -

one biological specimen image 84, wherein the reference feature extractor has a reference feature output. A slide calibrated feature extractor 92 is connected to the reference feature output, wherein the  
5 slide calibrated feature extractor has a calibrated feature output 96. An object classifier 86 is connected to the object feature output and the calibrated feature output. The object classifier 86 has a object classification result output. A slide  
10 feature extractor 90 is connected to the object classification result output, wherein the slide feature extractor has a slide feature output 94.

The invention has been described herein in considerable detail in order to comply with the Patent  
15 Statutes and to provide those skilled in the art with the information needed to apply the novel principles and to construct and use such specialized components as are required. However, it is to be understood that the invention can be carried out by specifically  
20 different equipment and devices, and that various modifications, both as to the equipment details and operating procedures, can be accomplished without departing from the scope of the invention itself.

What is claimed is:

- 17 -

## CLAIMS

1. A self calibrating biological specimen analysis apparatus comprising:
  - 5 (a) imaging means (502) for obtaining at least one image (40) of a biological specimen, wherein the imaging means (502) has an image output (40);
  - 10 (b) a calibration feature extraction means (14, 18), connected to receive the image output (40), for generating at least one calibrated feature output (15, 19);
  - 15 (c) an object classifier (22) connected to the calibrated feature output (15, 19) wherein the object classifier (22) has an object classification result output (23); and
  - (d) a slide classifier (24) connected to the object classification result output (23), wherein the slide classifier (24) has an analysis score output (26).
- 20 2. The self calibrating biological specimen analysis apparatus of claim 1 wherein the calibration feature extraction means (14, 18) further comprises an object feature extractor (18) and a reference feature extractor (14), wherein the  
25 object feature extractor (18) is connected to receive the image output (40), and the object feature extractor (18) has an object feature extractor output (19), and wherein the reference feature extractor (14) is connected to receive  
30 the image output (40), and the reference feature extractor (14) has an reference feature extractor output (15).
3. The self calibrating biological specimen analysis

- 18 -

apparatus of claim 2 wherein the object classifier (22) is further connected to the object feature extractor output (19).

4. The self calibrating biological specimen analysis apparatus of claim 2 wherein the slide classifier (24) further comprises a slide calibration feature extractor (20) connected to the reference feature extractor output (15), wherein the slide calibration feature extractor (20) has a slide calibration feature output (21).
5. The self calibrating biological specimen analysis apparatus of claim 1 further comprising an analysis score thresholder (28) connected to the analysis score output (26), wherein the analysis score thresholder (28) compares the analysis score output (26) to a review threshold, wherein the analysis score thresholder (28) has a normal indicator output (30) and a review indicator output (32).
6. The apparatus of claim 1 wherein the biological specimen is prepared by the Papanicolaou method.
7. The apparatus of claim 1 wherein the biological specimen is a gynecological specimen.
8. The apparatus of claim 1 further comprising a computer processor (562).
9. The biological specimen analysis self calibrating apparatus of claim 1 further comprising an automated microscope (511).

- 19 -

10. A self calibrating object classifier apparatus (Fig. 7) comprising:
- 5 (a) a reference feature extractor (68) to detect reference features in at least one biological specimen image (64), wherein the reference feature extractor (68) has a reference feature output;
  - 10 (b) an object feature extractor (66) to detect object features in the at least one biological specimen image (64), wherein the object feature extractor (66) has an object feature output; and
  - 15 (c) an object classifier (70) connected to the object feature output and the reference feature output, wherein the object classifier (70) has a object classification result output (72).
11. A self calibrating feature extractor apparatus (Fig. 8) comprising:
- 20 (a) a reference feature extractor (78) to detect reference features in at least one biological specimen image (74), wherein the reference feature extractor (78) has a reference feature output;
  - 25 (b) an object feature extractor (76) to detect object features in at least one biological specimen image (74), wherein the object feature extractor (76) has an object feature output; and
  - 30 (c) a calibration feature extractor (80) connected to the reference feature output and the object feature output, wherein the calibration feature extractor (80) has a calibrated feature output (82).

- 20 -

12. A self calibrating slide feature extraction apparatus (Fig. 9) comprising:
- 5 (a) a reference feature extractor (88) to detect reference features in at least one biological specimen image (84), wherein the reference feature extractor (88) has a reference feature output;
  - 10 (b) an object classifier (86) connected to receive the at least one biological specimen image (84), having an object classifier output;
  - 15 (c) a slide feature extractor (90) connected to the object classifier output, wherein the slide feature extractor (90) has a slide feature output (94); and
  - 20 (d) a slide calibration feature extractor (92) connected to the reference feature output, wherein the slide calibration feature extractor (92) has a slide calibration feature output (96).
13. A self calibrating biological specimen analysis apparatus comprising:
- 25 (a) a reference feature extractor (42) to detect reference features (46) in at least one biological specimen image (40), wherein the reference feature extractor (42) has a reference feature output;
  - 30 (b) an object feature extractor (44) to detect object features in at least one biological specimen (40), wherein the object feature extractor (44) has an object feature output;
  - (c) a calibrated feature extractor (48) connected to the reference feature output and the object feature output, wherein the

- 21 -

- calibrated feature extractor (48) has a calibrated feature output;
- 5 (d) an object classifier (52) connected to the object feature output and the calibrated feature output, wherein the object classifier (52) has a object classification result output;
- 10 (e) a slide feature extractor (54) connected to the object classification result output, wherein the slide feature extractor (54) has a slide feature output;
- 15 (f) a slide calibration feature extractor (50) connected to the reference feature output, wherein the slide calibration feature extractor (50) has a slide calibration feature output;
- 20 (g) a slide classifier (56) connected to the slide calibration feature output and the slide feature output, wherein the slide classifier (56) has an analysis score output (58); and
- 25 (h) an analysis score thresholder (58), connected to the analysis score output (58), having a normal indicator output (62) and a review indicator output (60).
14. The apparatus of claim 13 wherein the biological specimen (40) is a prepared by the Papanicolaou method.
- 30 15. The apparatus of claim 13 wherein the biological specimen (40) is a gynecological specimen.
16. A method of calibrating a biological specimen analysis system (Fig. 6) comprising the steps of:

- 22 -

- 5 (a) detecting reference features in a biological specimen (42);
- (b) detecting object features from the biological specimen (44);
- 10 (c) detecting calibrated features from the reference features and object features (48);
- (d) classifying objects in the biological specimen based on the calibrated features and object features to provide classification results (52);
- (e) detecting slide calibration features from the reference features (50);
- 15 (f) extracting slide features from the object classification results (54); and
- (g) classifying biological specimen slides (56) as either normal (62) or needing review (60) based on the slide calibration features and slide features.
- 20 17. The apparatus of claim 16 wherein the biological specimen (40) is prepared by the Papanicolaou method.
18. The apparatus of claim 16 wherein the biological specimen (40) is a gynecological specimen.
- 25 19. A method of feature calibration comprising the steps of:
- (a) detecting reference features in a biological specimen (42); and
- (b) calibrating the reference features to provide calibrated features (48).
- 30 20. A method of detecting a reference cell in a biological specimen comprising:

- 23 -

- 5 (a) detecting objects in a biological specimen image that resemble isolated cells (44);  
(b) extracting staining independent features from the isolated cells (48);  
(c) classifying the staining independent features to detect reference cells (52).

21. The method of claim 20 further comprising the steps of:

- 10 (a) detecting a cytoplasm size and radius (54);  
(b) computing a nuclear to cytoplasm size and radius ratio (50); and  
(c) determining if the radius ratio is lower than a predetermined limit (56).

15 22. The method of claim 20 further including the steps of:

- (a) determining if a cytoplasm size is large with respect to nuclear size (56);  
(b) determining if a nucleus has sufficient texture (56);  
20 (c) determining if a cytoplasm area is relatively smooth (56);  
(d) determine that the nucleus is large enough (56); and  
25 (e) determine that the nuclear shape is nearly round (56).

23. The method of claim 20 wherein the reference cell (40) is a squamous intermediate cell.

30 24. A calibrated feature extraction process for a screened biological slide comprising the steps of:

- (a) screening the slide reference features (78);

- 24 -

- (b) re-screening the slide;  
(c) detecting objects on the slide (76);  
(d) extracting object features (76); and  
(e) calibrating the object features by combining  
5 the object features and slide reference features (82).
25. The method of claim 24 wherein the slide reference features (15) comprise slide staining, reference cell nuclear to cytoplasm density contrast, and nuclear size related features.  
10
26. The method of claim 24 further comprising a computer processor (562) to process feature combination formulas.
27. The method of claim 26 wherein the feature combination formulas comprise:  
15
- $$NewFeature = \frac{f - \alpha g}{\beta}$$
28. The method of claim 27 wherein  $\alpha$  is evaluated to maximize the discrimination power of the resulting *NewFeature* and  $\beta$  is a calibrated factor; and wherein the forms *f* and *g* for six  
20 types of combinations are listed in the following table:

- 25 -

| Combination Type | f                             | g                               |
|------------------|-------------------------------|---------------------------------|
| Type 1           | s                             | r                               |
| Type 2           | s/r                           | 1/4                             |
| Type 3           | s/r <sup>2</sup>              | r <sub>1</sub> /r <sub>2</sub>  |
| Type 4           | $\frac{\log(s)  r_1 }{ r_2 }$ | $\frac{\log(r_1)  r_1 }{ r_2 }$ |
| Type 5           | $\frac{\log(s)}{\log(r)}$     | $\frac{1}{\log(r)}$             |
| Type 6           | log(s)                        | log(r)                          |

where  $s$  is object feature (19) and  $r$ ,  $r_1$  and  $r_2$  are slide reference features (15); and wherein  $\log(x)$  indicates  $\log(|x|+0.001)$  is used; where  $x=s, r, r_1$  for the singularity of a logarithmic function, and we set  $r_2=10^{-6}$  if  $r_2 < 10^{-6}$ ; and wherein for Type 2, we set  $r = 0.01 * \text{sgn}(r) * R$  if  $|r| < 0.01 * R$  where  $R$  is maximum of  $|r|$  for all training data where the function  $\text{sgn}(\cdot)$  is

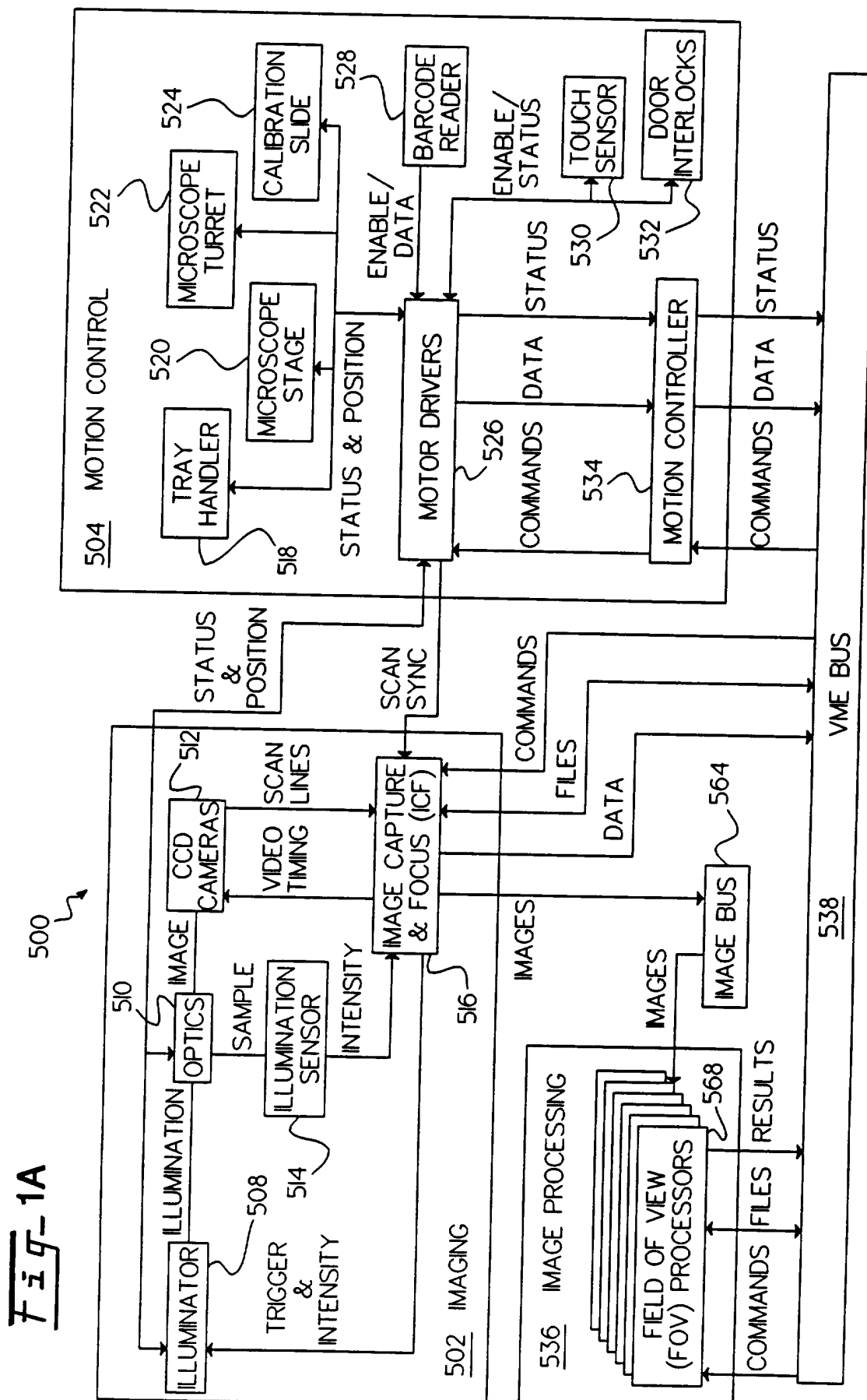
$$\text{sgn}(x) = \begin{cases} 1; & x \geq 0 \\ -1; & x < 0 \end{cases}$$

and wherein, for Type 5, we set  $\text{Log}(r) = 0.01 * \text{sgn}(\log(r)) * \log(\hat{R})$  if  $|\log(|r|+0.001)| < 0.01 * \log(\hat{R})$  where  $\log(\hat{R})$  is maximum of  $|\log(|r|+0.001)|$  for all training data.

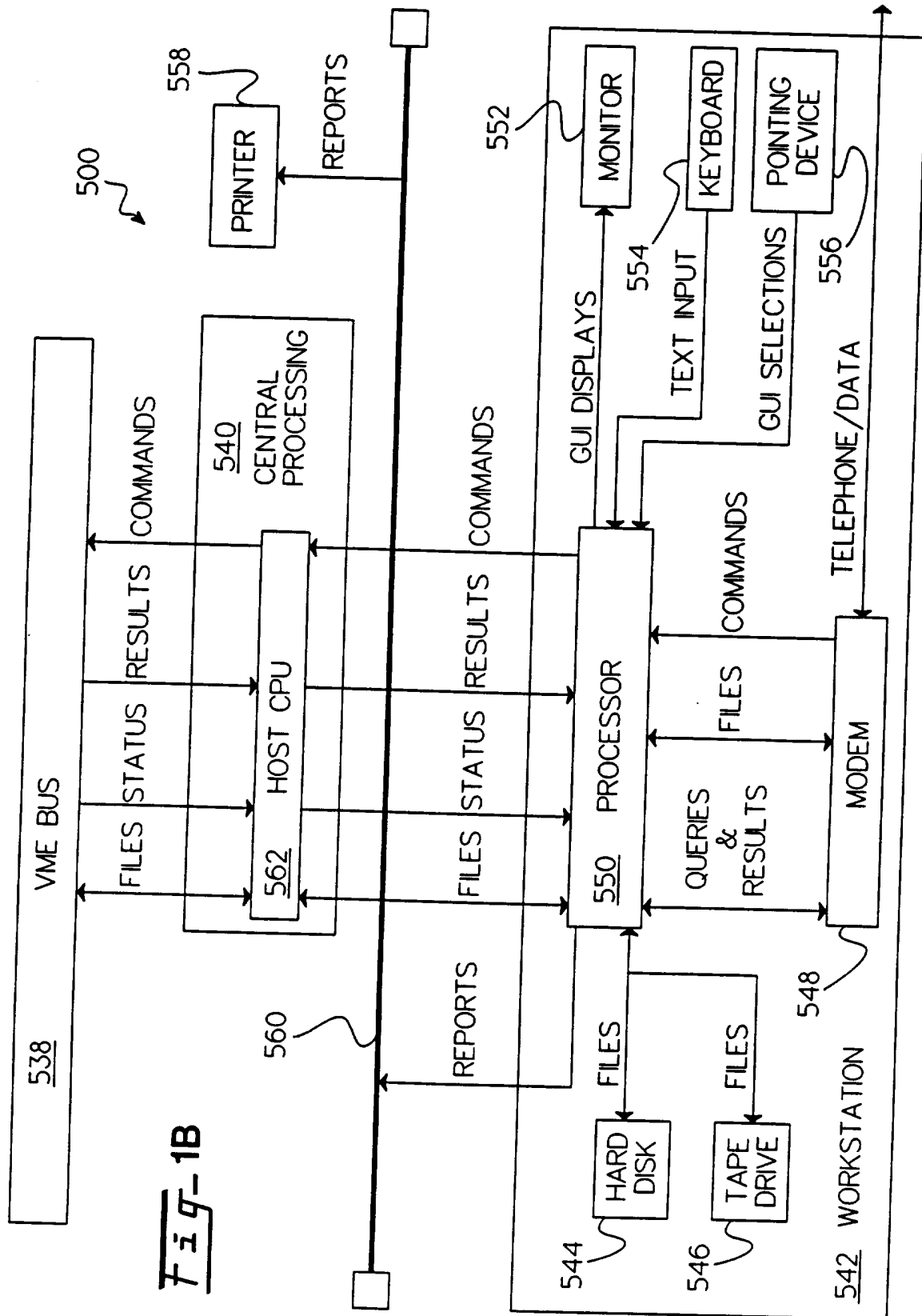
29. The method of claim 24 wherein the features (21) are selected by stepwise discriminate analysis.

30. The method of claim 24 wherein the calibrated features (82) and the original object features (72) are used to train an object classifier.

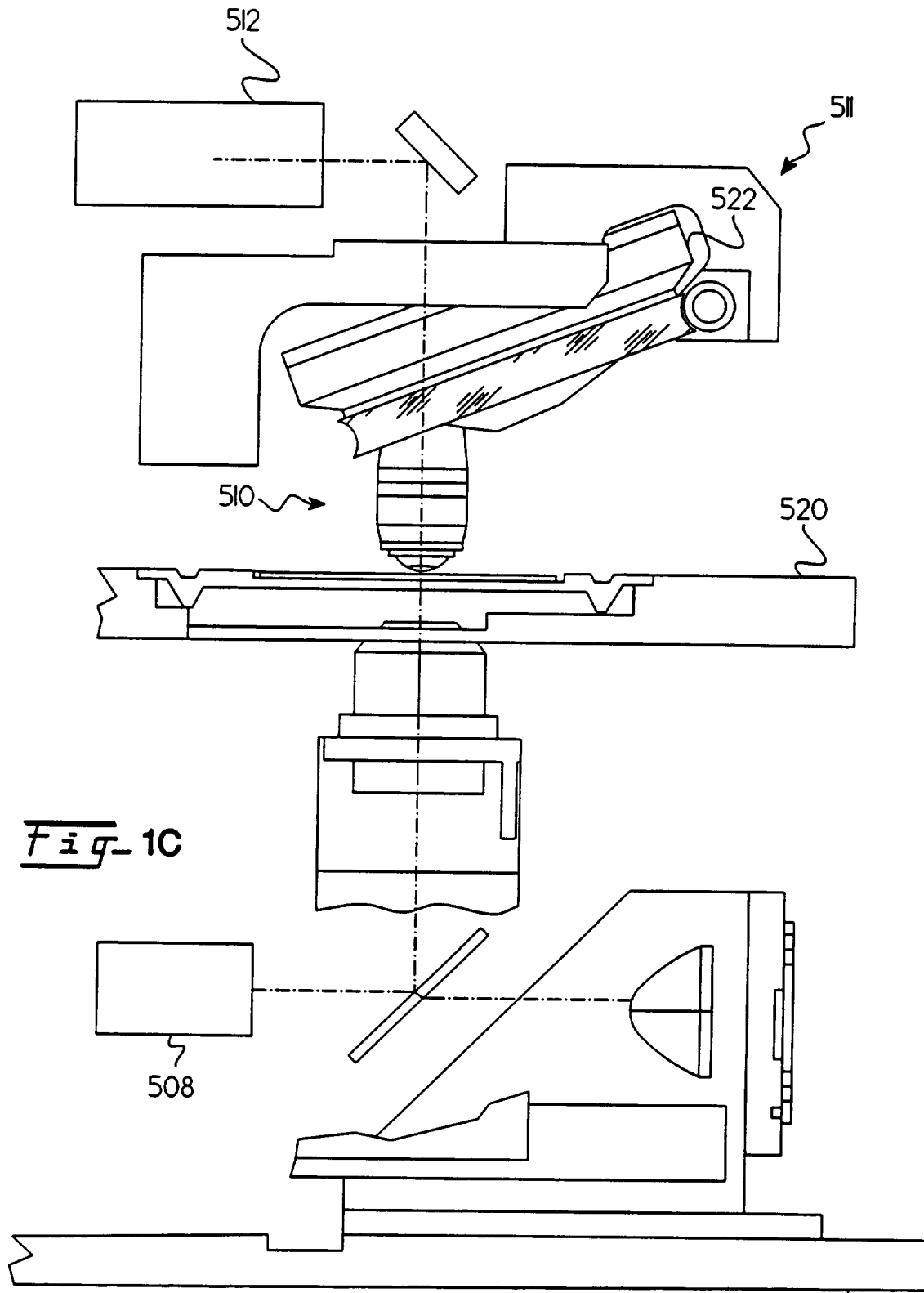
1/10



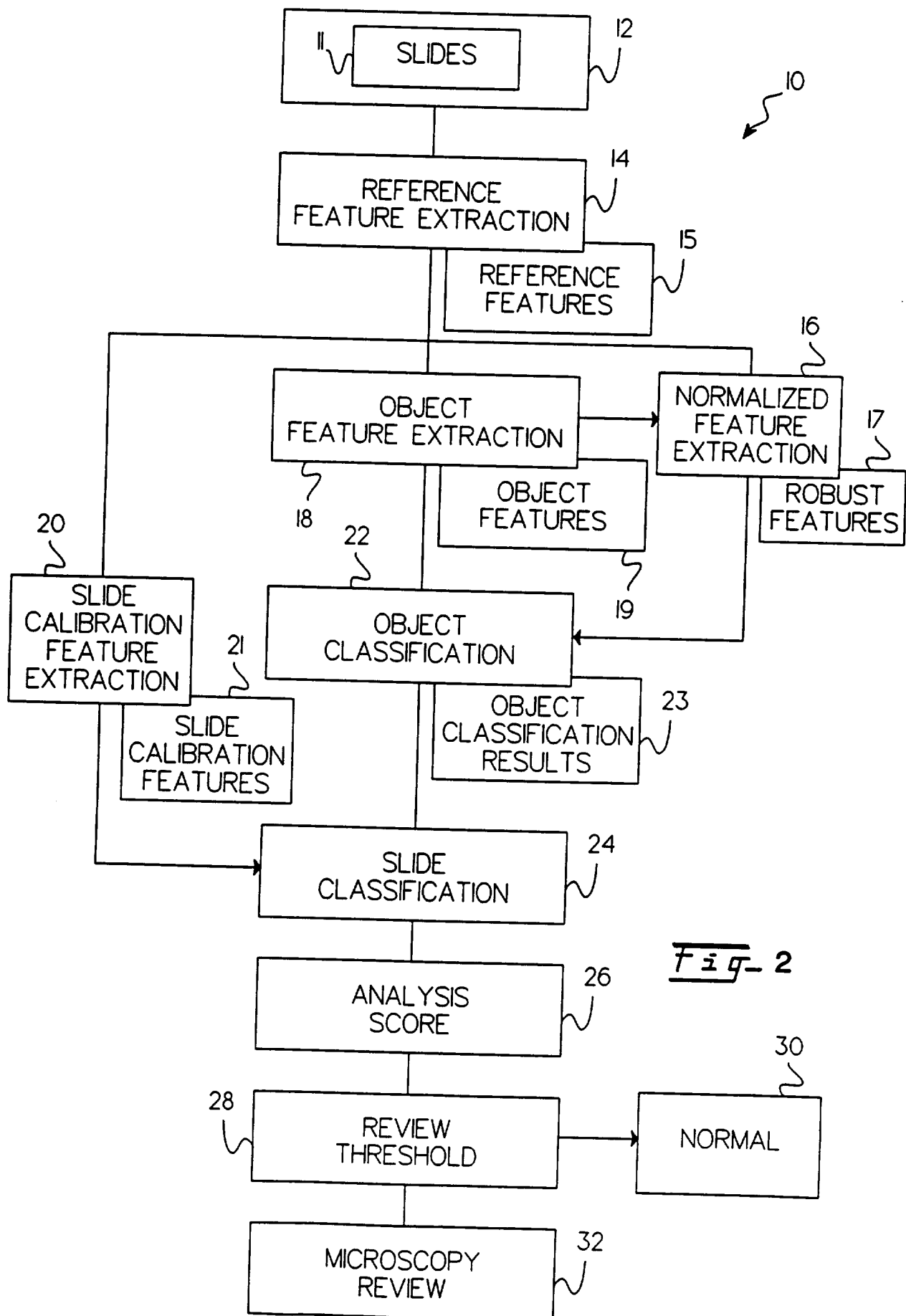
2 / 10



3 / 10



4 / 10

Fig. 2

5 / 10

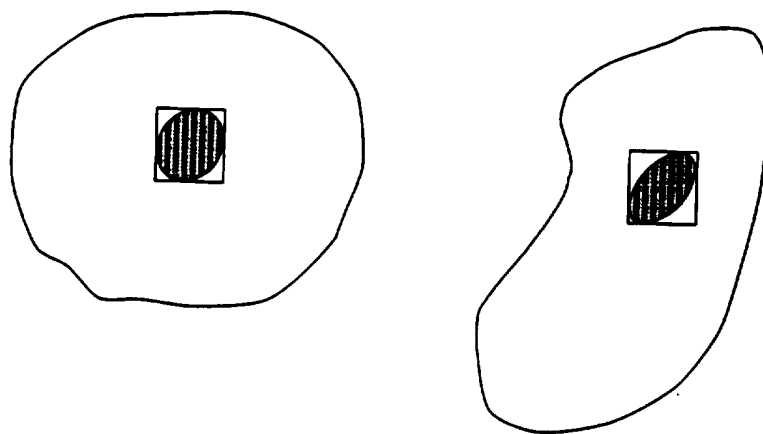


Fig- 3

6 / 10

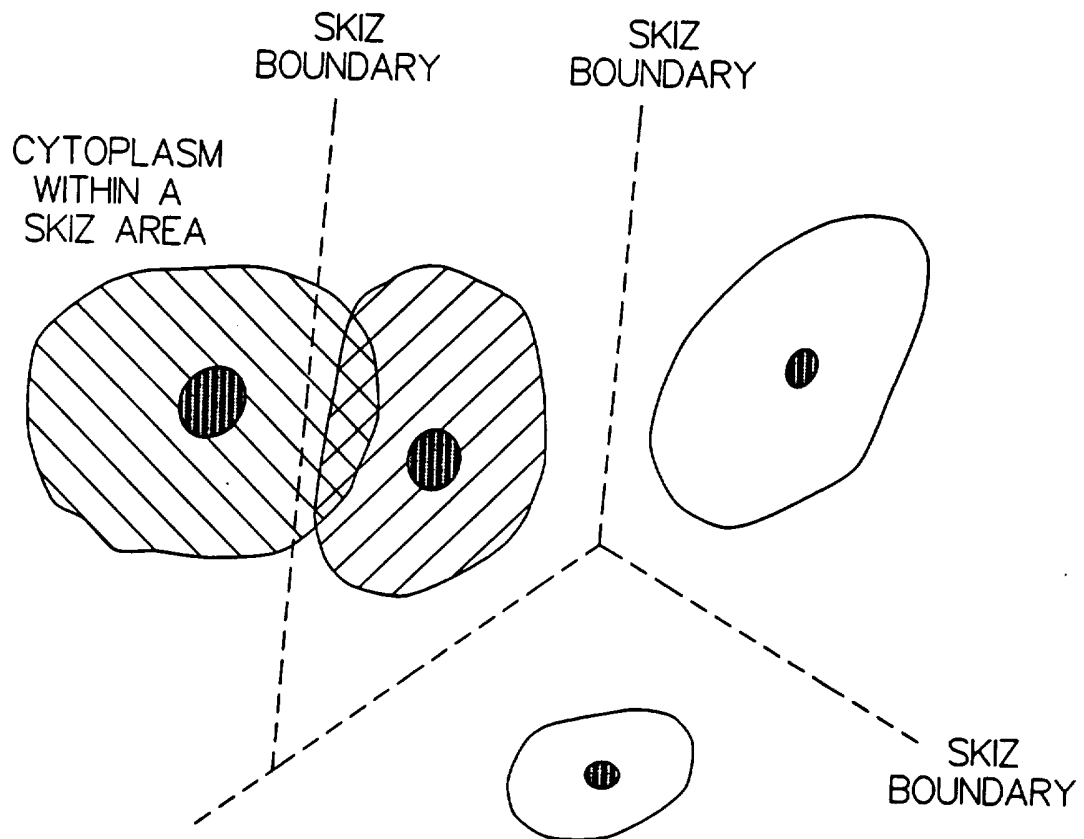


Fig- 4

7/10

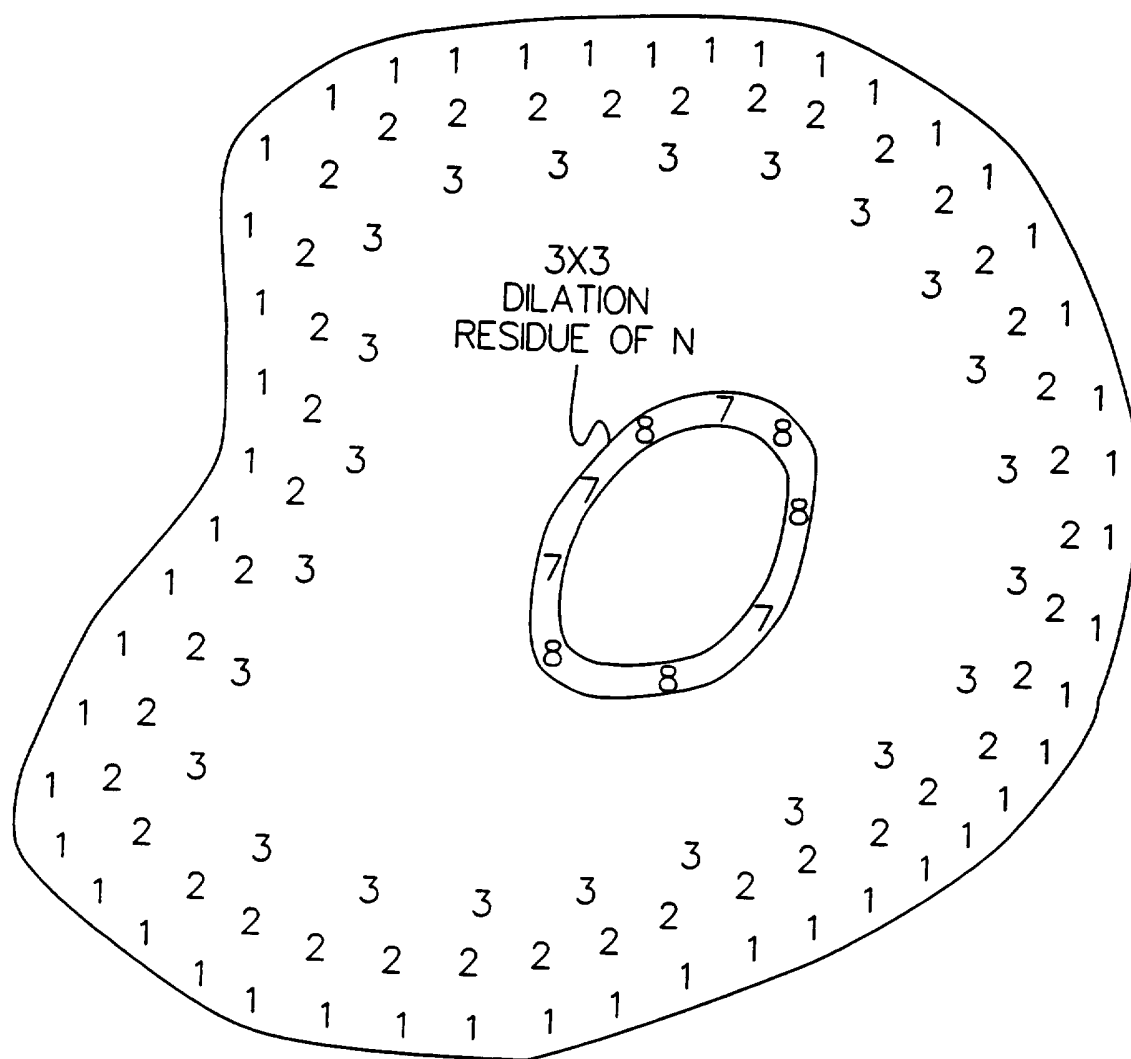
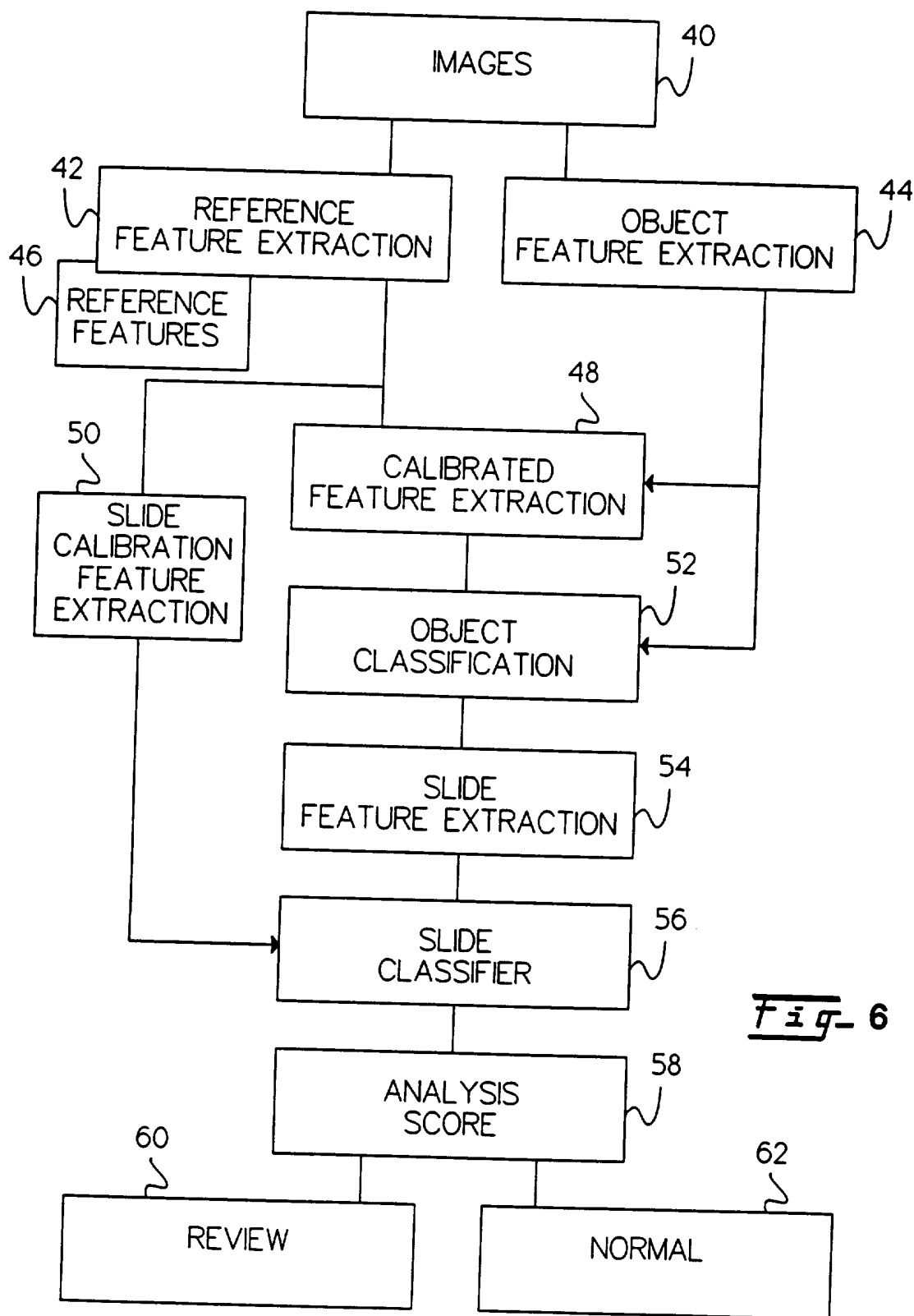
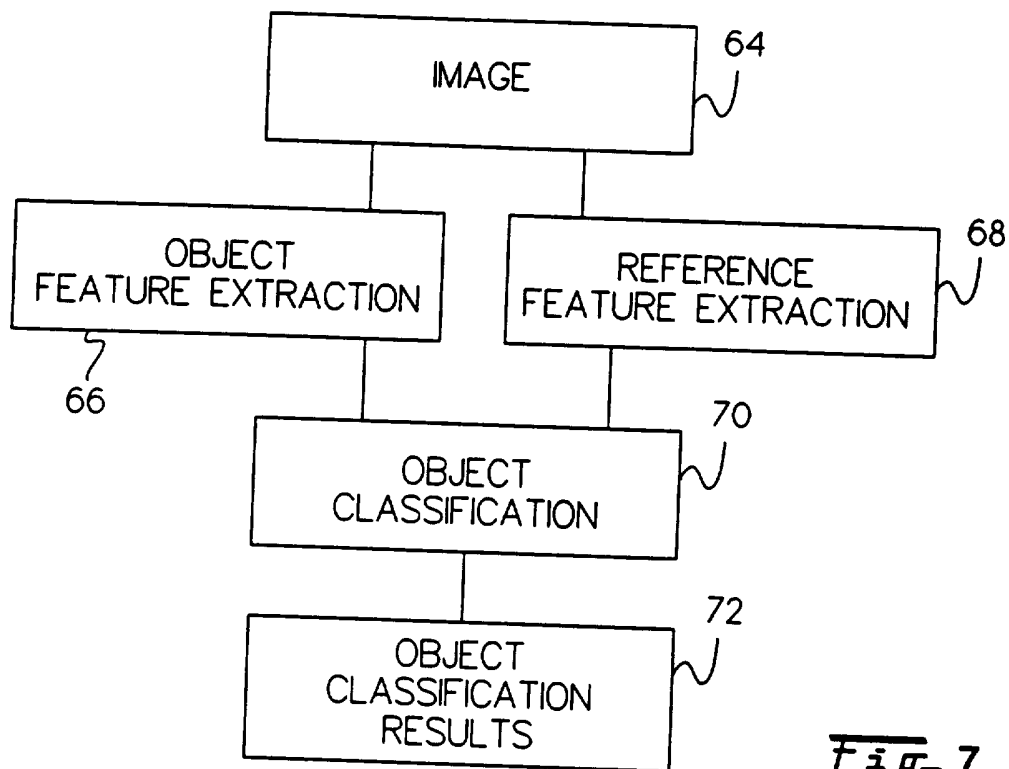
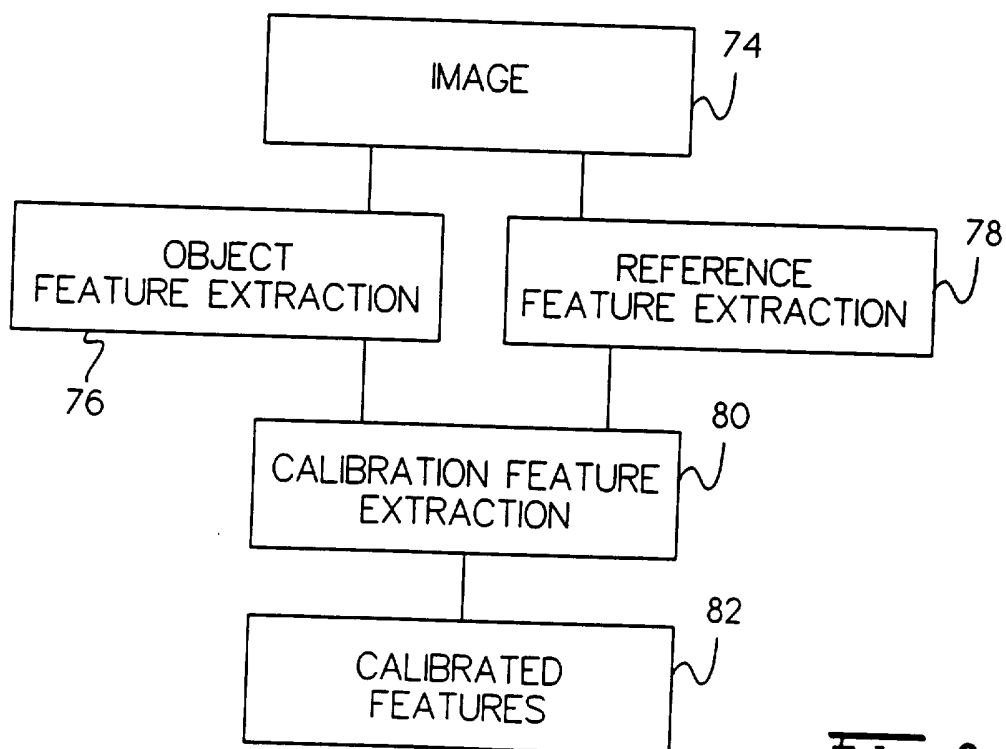


Fig- 5

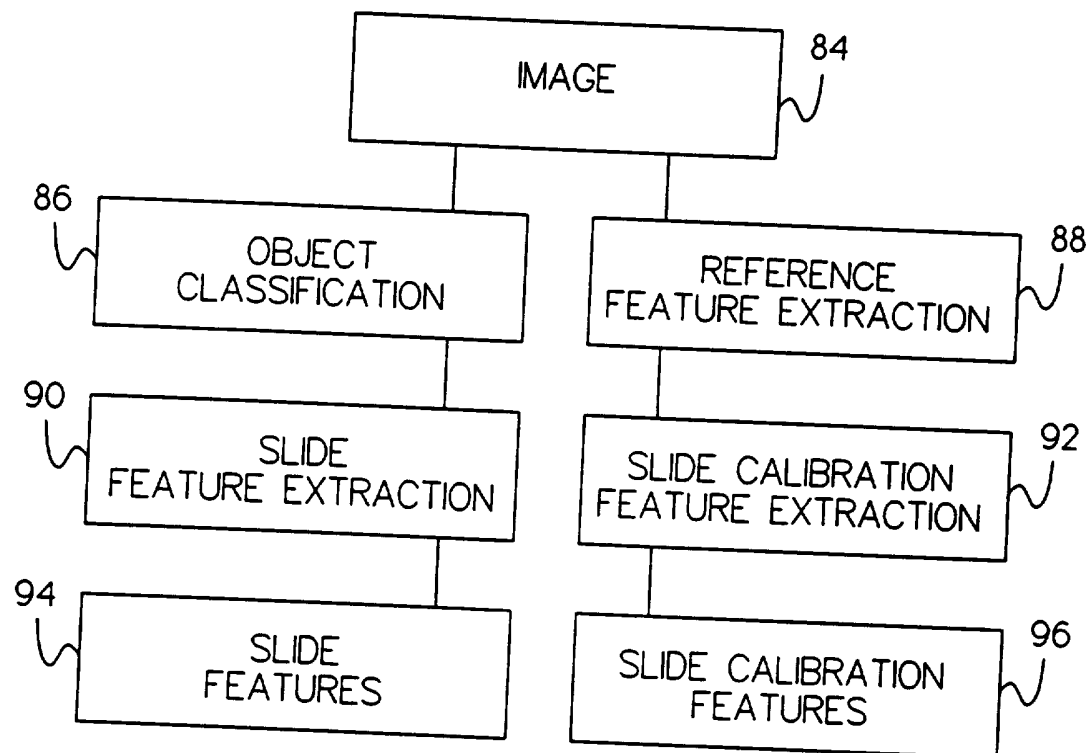
8/10

Fig. 6

9/10

Fig- 7Fig- 8

10/10

Fig. 9

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/11378

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :GO6K 9/00

US CL :382/133

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 382/133, 128, 134, 224, 317; 395/924; 364/413.01, 413.07, 413.08, 413.1, 497

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
none

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
none

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                     | Relevant to claim No. |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Y         | US, A, 5,018,209 (BACUS) 21 MAY 1991, Abstract, Figs. 2-14, 16, 19, column 2, line 58 to column 3, line 4 and column 20, line 46 to column 22, line 60 | 1-30                  |
| Y         | US, A, 5,163,095 (KOSAKA) 10 November 1992, Abstract, Figures 2, 5, 9, column 7, lines 5-12                                                            | 1-30                  |
| Y         | US, A, 5,257,182 (LUCK ET AL.) 26 October 1993, Abstract, Figures 1-3, column 8, lines 27-28.                                                          | 9                     |



Further documents are listed in the continuation of Box C.



See patent family annex.

|                                                                                                                                                                         |                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Special categories of cited documents:                                                                                                                                  |                                                                                                                                                                                                                                                  |
| *A* document defining the general state of the art which is not considered to be part of particular relevance                                                           | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention                                              |
| *E* earlier document published on or after the international filing date                                                                                                | *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone                                                                     |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *O* document referring to an oral disclosure, use, exhibition or other means                                                                                            | *Z* document member of the same patent family                                                                                                                                                                                                    |
| *P* document published prior to the international filing date but later than the priority date claimed                                                                  |                                                                                                                                                                                                                                                  |

Date of the actual completion of the international search

13 NOVEMBER 1995

Date of mailing of the international search report

27 DEC 1995

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

JOSEPH MANCUSO

Telephone No. (703) 305-8503