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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 April 2010 (22.04.2010)

PCT

(10) International Publication Number
WO 2010/044745 A1

(51) International Patent Classification:
G06T 7/60 (2006.01) *G01N 33/574* (2006.01)
G06T 5/40 (2006.01)

(21) International Application Number:
PCT/SG2008/000397

(22) International Filing Date:
16 October 2008 (16.10.2008)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **NANYANG POLYTECHNIC** [SG/SG]; 180 Ang Mo Kio Avenue 8, Sinagpore 569830 (SG).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WEE, Angela, Jean** [SG/SG]; Blk 484 Admiralty Link, #09-45, Sigapore 750484 (SG). **KONG, Wai, Ming** [SG/SG]; Blk 638C Punggol Drive, #16-463, Singapore 823638 (SG). **TAN, Tsu, Soo** [SG/SG]; Blk 228 Choa Chu Kang Central, #12-111, Singapore 680228 (SG).

(74) Agent: **LAWRENCE Y D HO & ASSOCIATES PTE LTD**; 30 Bideford Road, #02-02, Thongsia Building, Singapore 229922 (SG).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI

[Continued on next page]

(54) Title: PROCESS AND DEVICE FOR AUTOMATED GRADING OF BIO-SPECIMENS

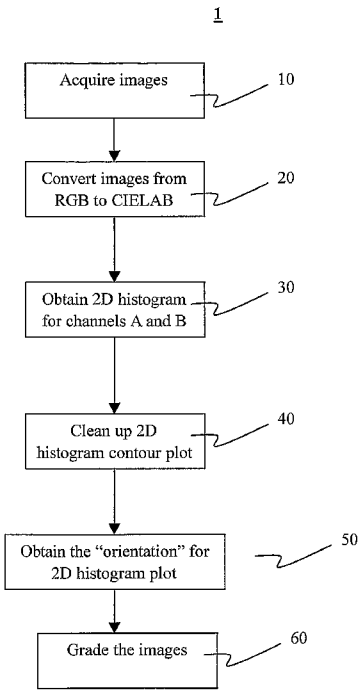


FIG 1

(57) Abstract: The present invention provides a process for automated grading of bio-specimens employing a 2-dimensional histogram plotted from the pixel values of the images of the bio-specimens. The present invention also provides a process of automated grading of bio- specimens by employing a 2-dimensional histogram plotted from the pixel values of the images of the bio-specimens and using the L component of the cell wall pixels to determine the grade of the image. The present invention further provides an electronic device for performing the processes.

WO 2010/044745 A1

(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

PROCESS AND DEVICE FOR AUTOMATED GRADING OF BIO-SPECIMENS

Field of the Invention

[0001] The present invention generally relates to the field of bio-specimen analysis, and more particularly to a process and device for automated grading of bio-specimens.

Background of the Invention

[0002] Many diseases are manifested by their expressions or lost of expressions of one or more genes in the affected tissues or cells; the genes specifically associated with diseases are usually called biomarkers; thus diagnosis of the diseases can be done with the biospecimens taken from patients by analysing the expressions of biomarkers.

[0003] One of the exemplary biomarkers is Her2/neu. The Her2/neu (also known as c-erbB2) oncogene is a cell-surface membrane tyrosine kinase receptor that is overexpressed in 20% to 30% of breast cancer cases and is significantly associated with breast cancer recurrence and death in the last two decades. In addition, breast cancer cases associated with the overexpression of Her2/neu have been known to respond poorly to conventional hormonal therapy, chemotherapy or other cytotoxic treatments.

[0004] Hence, detection of Her2/neu in breast carcinomas is important to determine appropriate therapeutic management and treatment of patients diagnosed with breast cancer. Often, evaluation of the presence of Her2/neu proteins is performed through means of immunohistochemistry (IHC), which has been the expertise of the histopathologist. The sample tissue that is stained using hematoxylin and eosin (H&E) is viewed and assessed under microscopic review, which relies on the identification of stained cell membranes.

[0005] Although protocols are in place for all observers to grade cell samples according to membrane staining, the assessment is still intrinsically subjective and presents variability between observers. As such, an objective quantification would be useful in circumventing the problem of being subjective by the observers. A computer-assisted image analysis system will be beneficial not only in providing such a means of quantification, but can also be used as a means of second opinion. However, current

methods of the automated grading of tissue sample images is tedious and complex, resulting in high use of computing resources and long running time

Summary of the Invention

[0006] One embodiment of the present invention provides a process for automated grading of bio-specimens by employing a 2-dimensional histogram plotted from the pixel values of the images of the bio-specimens. The process comprises the steps of converting an image into CIELAB color space if the image is not in the CIELAB color space, plotting a 2-dimensional histogram with A channel and B channel values from the CIELAB values of each pixel in the image, converting the 2-dimensional histogram into a binary image by scoring 0 for all cells having values 0, and scoring 1 for all cells having values greater than 0, cleaning the binary image by removing small spots and filling of holes, calculating from the binary image the orientation angle of the ellipse that best fits the binary image, and grading the bio-specimens into 4 grades according to pre-determined thresholds based on the orientation angle.

[0007] Another embodiment of the present invention provides a process of automated grading of bio-specimens by employing a 2-dimensional histogram plotted from the pixel values of the images of the bio-specimens and using the L component of the cell wall pixels to determine the grade of the image. The process comprises the steps of converting an image into CIELAB color space if the image is not in the CIELAB color space, plotting a 2-dimensional histogram with A channel and B channel values from the CIELAB values of each pixel in the image, converting the 2-dimensional histogram into a binary image by scoring 0 for all cells having values 0, and scoring 1 for all cells having values greater than 0, cleaning the binary image by removing small spots and filling of holes, obtaining the "orientation" for 2D histogram plot, and checking whether the "orientation" is over a predetermined threshold, then grading the image as Grade 0 if the "orientation" is over the predetermined threshold, extracting cell wall pixels from CIELAB image if the "orientation" is less than the predetermined threshold, obtaining mean CIELAB value (L value) of cell wall pixels, and grading image based on L value.

[0008] Another embodiment of the present invention provides an electronic device comprising an electronic storage medium embedded with an electronically executable

program, and a microprocessor being electronically coupled with the electronic storage medium, executing the embedded program; wherein the program is capable of performing the operations of converting an image into CIELAB color space if the image is not in the CIELAB color space, plotting a 2-dimensional histogram with A channel and B channel values from the CIELAB values of each pixel in the image, converting the 2-dimensional histogram into a binary image by scoring 0 for all cells having values 0, and scoring 1 for all cells having values greater than 0, cleaning the binary image by removing small spots and filling of holes, calculating from the binary image the orientation angle of the ellipse that best fits the binary image, and grading the bio-specimens into 4 grades according to pre-determined thresholds based on the orientation angle.

[0009] Another embodiment of the present invention provides an electronic device comprising an electronic storage medium embedded with an electronically executable program, and a microprocessor being electronically coupled with the electronic storage medium, executing the embedded program; wherein the program is capable of performing the operations of converting an image into CIELAB color space if the image is not in the CIELAB color space, plotting a 2-dimensional histogram with A channel and B channel values from the CIELAB values of each pixel in the image, converting the 2-dimensional histogram into a binary image by scoring 0 for all cells having values 0, and scoring 1 for all cells having values greater than 0, cleaning the binary image by removing small spots and filling of holes, obtaining the "orientation" for 2D histogram plot, and checking whether the "orientation" is over a predetermined threshold, then grading the image as Grade 0 if the "orientation" is over the predetermined threshold, extracting cell wall pixels from CIELAB image if the "orientation" is less than the predetermined threshold, obtaining mean CIELAB value (L value) of cell wall pixels; and grading image based on L value.

[0010] The advantages of the present invention include faster computation as number of color clusters can be quickly determined, and lesser total amount of computational resources when processing multiple images as those images with no stained membrane require less complex calculations. Hence, it may be used as a means of high-throughput screening using a tissue microarray. The present invention will be able to grade each sample in a tissue microarray to provide a putative score that can be used in

conjunction with the histopathologist's opinion. In addition, the present invention is objective and is not subjected to fatigue from assessing large sets of images.

[0011] The objectives and advantages of the invention will become apparent from the following detailed description of preferred embodiments thereof in connection with the accompanying drawings.

Brief Description of the Drawings

[0012] Preferred embodiments according to the present invention will now be described with reference to the Figures, in which like reference numerals denote like elements.

[0013] FIG 1 is a flowchart of the automated grading process in accordance with one embodiment of the present invention.

[0014] FIG 2 shows a panel of four typical normal and cancerous tissue samples.

[0015] FIG 3 shows that the normal tissue sample takes a 2D histogram with an ellipse superimpose.

[0016] FIG 4 shows that the cancerous tissue sample takes a 2D histogram with a triangle superimpose.

[0017] FIG 5 is a functional flowchart of the automated grading process utilizing the L, A, and B components in accordance with one embodiment of the present invention.

Detailed Description of the Invention

[0018] The present invention may be understood more readily by reference to the following detailed description of certain embodiments of the invention.

[0019] Throughout this application, where publications are referenced, the disclosures of these publications are hereby incorporated by reference, in their entireties, into this application in order to more fully describe the state of art to which this invention pertains.

[0020] The present invention provides a process for automated grading of bio-specimens employing a 2-dimensional histogram plotted from the pixel values of the images of the bio-specimens. It is to be noted that the format of the pixel values is not

limited to any particular one as long as it is suitable for the application of the principles of the present invention. Briefly, when the CIELAB format is used for illustration, the automated grading process comprises the following operations. If the bio-specimen images are originally obtained in their RGB format, they are first converted into CIELAB format. A 2-dimensional histogram with A channel and B channel values is then plotted from the CIELAB values of each pixel in the image. The 2-dimensional histogram is then converted into a binary image by scoring 0 for all cells having values 0, and scoring 1 for all cells having values greater than 0. The binary image is then cleaned by removing small spots and filling of holes. From the binary image, the orientation angle of the ellipse that best fits the binary image is calculated and used to grade the bio-specimens into 4 grades according to pre-determined thresholds.

[0021] Now referring to FIG 1, there is provided a more detailed description of the automated grading process in accordance with one embodiment of the present invention. As shown in FIG 1, the automated grading process **1** comprises: acquire images **10**, convert images from RGB to CIELAB **20**, obtain 2D histogram for channels A and B **30**, clean up 2D histogram contour plot **40**, obtain the "orientation" for 2D histogram plot **50**, and grade the images **60**, where some of the operations will be described in more detail herein below.

[0022] 1. Convert an image from Red-Green-Blue (RGB) colour space to CIELAB colour space

[0023] The conversion from RGB to CIELAB takes 2 steps: first, convert RGB to XYZ, then convert XYZ to CIELAB. The exemplary codes to convert RGB to XYZ are as follows:

```
var_R = ( R / 255 )          //R from 0 to 255
var_G = ( G / 255 )          //G from 0 to 255
var_B = ( B / 255 )          //B from 0 to 255

if ( var_R > 0.04045 ) var_R = ( ( var_R + 0.055 ) / 1.055 ) ^ 2.4
else                   var_R = var_R / 12.92
if ( var_G > 0.04045 ) var_G = ( ( var_G + 0.055 ) / 1.055 ) ^ 2.4
else                   var_G = var_G / 12.92
if ( var_B > 0.04045 ) var_B = ( ( var_B + 0.055 ) / 1.055 ) ^ 2.4
else                   var_B = var_B / 12.92

var_R = var_R * 100
var_G = var_G * 100
var_B = var_B * 100


X = var_R * 0.4124 + var_G * 0.3576 + var_B * 0.1805
Y = var_R * 0.2126 + var_G * 0.7152 + var_B * 0.0722
Z = var_R * 0.0193 + var_G * 0.1192 + var_B * 0.9505
```

[0024] The exemplary codes to convert XYZ to CIELAB are as follows:

```
var_X = X / ref_X          //ref_X =  95.047
var_Y = Y / ref_Y          //ref_Y = 100.000
var_Z = Z / ref_Z          //ref_Z = 108.883

if ( var_X > 0.008856 ) var_X = var_X ^ ( 1/3 )
else                    var_X = ( 7.787 * var_X ) + ( 16 / 116 )
if ( var_Y > 0.008856 ) var_Y = var_Y ^ ( 1/3 )
else                    var_Y = ( 7.787 * var_Y ) + ( 16 / 116 )
if ( var_Z > 0.008856 ) var_Z = var_Z ^ ( 1/3 )
else                    var_Z = ( 7.787 * var_Z ) + ( 16 / 116 )

CIE-L* = ( 116 * var_Y ) - 16
CIE-a* = 500 * ( var_X - var_Y )
CIE-b* = 200 * ( var_Y - var_Z )
```

[0025] 2. Plot a 2-dimensional histogram with A and B channel values from each pixel of the CIELAB image

[0026] For each pixel in the CIELAB image, the A channel value and the B channel value are plotted on a graph to count the number of times when the A and B values have appeared together in a pixel. This graph will have the A channel on the x-axis, B channel on the y-axis and the count as the z-axis. After plotting all the (A,B) values on the graph, a 2D histogram (a 2-dimension histogram is an XYZ graph) is obtained. The exemplary pseudocodes for obtaining the 2D histogram are as follows:

```
m = width of image;
n = height of image;

for x = 1 to m {
  for y = 1 to n {
    (l,a,b) = get pixel value at (x,y);
    histogram(a,b)=histogram(a,b)+1;
  };
};
```

[0027] Since the 2D histogram keeps a count of the number of times when (a,b) has been plotted together in a pixel, it is a 3-dimensional graph in essence where the 3rd dimension is the count).

[0028] 3. Convert the 2-dimensional histogram to a binary image

[0029] For every cell(or pixel) in the histogram, if the cell value is greater than 0, the pixel in the binary image will be given a value of 1. Else, the pixel in the binary image

will be given a value of 0. The exemplary pseudocodes for conversion of the 2D histogram into the binary image are as follows:

```
for x = 1 to m {
 for y = 1 to n {
  If(histogram(x,y)>0)=binary_image(x,y)=1;
  else binary_image(x,y)=0;
 };
};
```

[0030] 4. Clean up the binary image

[0031] The background noise from the binary image is eliminated by removing of small spots and filling up holes. The removal of small spots is done by erosion operation followed by dilation operation. In one embodiment, the erosion of image is done by using disk of diameter 5 pixels, and the dilation of image by disk of diameter 5 pixels. The filling up of holes in the image is done by dilation operation and follow by erosion operation. In one embodiment, the erosion of image is done by using disk of diameter 5 pixels, and the dilation of image by disk of diameter 5 pixels.

[0032] 5. Obtain the Orientation angle

[0033] From the cleaned binary image, the orientation of the ellipse that best fits the image is calculated by the equations (1) and (2) below:

$$Orientation\ angle = 0.5 \times tan^{-1}\left|\frac{2\mu_{1,1}}{\mu_{2,0}-\mu_{0,2}}\right| \quad (1)$$

$$\mu_{p,q}=\sum_{(x,y)\in R}\sum(x-\bar{x})^p(y-\bar{y})^q) \quad (2)$$

[0034] Where p,q=0,1,2 and region R represents the image that contains N pixels (N is the total number of pixels in region R), x and y are the 2-dimensional coordinate variables and

$$\bar{x}=\frac{1}{N}\sum_{(x,y)\in R}\sum x \quad (3)$$

$$\bar{y}=\frac{1}{N}\sum_{(x,y)\in R}\sum y \quad (4)$$

[0035] 6. Use the orientation angle to grade the images into 4 grades

[0036] If orientation angle > threshold 1, image is grade 0; if threshold 2 < orientation angle ≤ threshold 1, image is grade 1; if threshold 3 < orientation angle ≤ threshold 2, image is grade 2; if orientation angle ≤ threshold 3, image is grade 3. The threshold for each grade is determined empirically, e.g., threshold 1 = 10.0 degree, threshold 2 = -20.65 degree, and threshold 3 =-25.29 degree.

[0037] FIG 2 shows a panel of four typical normal and cancerous tissue samples. FIG 3 shows that the normal tissue sample takes a 2D histogram with an ellipse superimpose; FIG 4 shows that the cancerous tissue sample takes a 2D histogram with a triangle superimpose.

[0038] The present invention further provides a process of automated grading of bio-specimens by employing a 2-dimensional histogram plotted from the pixel values of the images of the bio-specimens and using the L component of the cell wall pixels to determine the grade of the image. One of advantages of the automated grading process is that it does not require additional information such as cell count and scores of individual cells for grading.

[0039] Now referring to FIG 5, there is provided a functional flowchart of the automated grading process utilizing the L, A, and B components in accordance with one embodiment of the present invention. The automated grading process **100** comprises: locate images in a database **101**, open the images **102**, convert the images from RGB to CIELAB **103**, obtain 2D histogram for channels A and B **104**, clean up 2D histogram contour plot **105**, obtain the "orientation" for 2D histogram plot **106**, check whether the "orientation" is over a predetermined threshold **110**, grade the image as Grade 0 if the "orientation" is over the predetermined threshold **120**, extract cell wall pixels from CIELAB image if the "orientation" is less than the predetermined threshold **130**, obtain mean CIELAB value of cell wall pixels **140**, check grade of image based on L value **150**, grade the image as Grade 1 if the L value is no less than 195 **160**, grade the image as Grade 2 if the L value is no less than 173 but less than 195 **170**, and grade the image as Grade 3 if the L value is less than 173 **180**. The predetermined threshold is determined empirically by measuring the angles from the images collected; the value is dependent on the CIELAB values of the images and may differ for a different set of images. The similar operations having been described above will not be repeated herein.

[0040] The extraction of cell wall pixels from CIELAB images **130** uses simple mathematical calculations. Briefly, 3 seeds with pre-assigned values are used to represent the background, nucleus and cell wall; for example, (-128, 0) for background, (0, -128) for nucleus, and (127,127) for cell wall respectively in (A,B) format. Then, the distance between each pixel against each of these seeds is measured to determine which distance is the shortest; the seed whose distance to the pixel is the shortest is the classification assigned to that pixel.

[0041] The operation of obtaining the mean CIELAB value of cell wall pixels **140** receives the pixels representing the cell wall from the extraction of cell wall pixels from CIELAB images **130**; then the CIELAB values of all pixels representing the cell wall is averaged to give the mean CIELAB value (L) of cell wall pixels.

[0042] If the mean L is >= 195 then the image is scored as 1

[0043] If the mean L is >= 173 and L < 195 then the image is scored as 2

[0044] If the mean L is < 173 then the image is scored as 3

[0045] The present invention further provides a process for automated grading of bio-specimens that first employs a 2-dimensional histogram plotted from the pixel values of the images of the bio-specimens for distinguishing negative (e.g., normal) bio-specimens from positive (e.g., cancer) ones, and then grades the positive bio-specimens with subsequent operations including various classification, segmentation and measurement. Briefly, the automated grading process uses the 2D histogram described above to quickly separate score 0 samples (i.e., negative ones) from samples of other scores (i.e., positive ones). Finally, the stain of the cell membranes of the resultant cells is measured to determine the final score of the sample.

[0046] The present invention further provides an electronic device that includes an electronic storage medium for storing the programs that are capable of performing the processes as discussed above, and a microprocessor being electronically coupled with the storage medium. The electronic device includes, but not limited to, computers, desktops and laptops.

[0047] The following examples are provided herein for illustrating the principles of the present invention.

[0048] Examples

[0049] The Her2/neu (also known as c-erbB2) oncogene is a cell-surface membrane tyrosine kinase receptor that is overexpressed in 20% to 30% of breast cancer cases and has been significantly associated with breast cancer recurrences and deaths occurred in the last two decades. While one protocol is in place for all pathologists to grade cell samples according to membrane staining, the assessment is intrinsically subjective and presents variability between observers.

[0050] Process I

[0051] Image Acquisition

[0052] The images used were acquired through microscopic imaging of normal breast tissues as well as the breast tissues of varying grades of cancer. The tissue sections usually provide images with a dimension of 1300 pixels by 1030 pixels; the images are then exported and saved in TIFF format in Red-Green-Blue (RGB) color space. These images are manual focused such that they are optimized for study and classification by the histopathologist. The tissues were immunohistochemically stained such that cell walls of normal breast tissue are blue while that of abnormal breast tissue are of varying degrees of orange and brown.

[0053] FIG 2 shows typical images of normal breast tissue (a) and different stages of cancerous breast tissues (b)-(d), where the nuclei of cells take up a blue stain, and the Her-2/neu is immunohistochemically stained of varying degrees of orange and brown. From visual observations, normal breast tissue cells do not exhibit strong straining vis-à-vis abnormal breast tissue cells. Strong contrast is usually lacking in normal breast tissue cells while abnormal breast tissue cell walls usually exhibit varying contrasts of orange, which determines the score to be assigned to the sample. For visual classification, a score of 0 is assigned to a sample when no observable staining occurs or when there is less than 10% membranous staining of the tissue cells; score of 1 is given when there is faint or barely perceptible membrane staining in greater than 10% of cells and the cells are stained only in part of the sample; score of 2 is given when there is weak to moderate complete membrane staining in greater than 10% of the tissue cells; and the maximum score of 3 is assigned when there is strong complete membrane staining in greater than 10% of the tissue cells.

[0054] Convert an image from Red-Green-Blue (RGB) colour space to CIELAB colour space

[0055] The acquired images in the RGB format were converted into the ones in the CIELAB format according to the operations as discussed above.

[0056] Plot a 2-dimensional histogram with A and B channel values from each pixel of the CIELAB image

[0057] The 2D histograms for images in the CIELAB format were plotted with A and B channel values according to the operations as discussed above.

[0058] Convert the 2-dimensional histogram to a binary image

[0059] The 2D histograms were converted into the binary images according to the operations as discussed above.

[0060] Clean up the binary image

[0061] The binary images were cleaned up according to the operations as discussed above.

[0062] Obtain the Orientation angle

[0063] The orientation of the ellipse that best fits the binary images was calculated according to the operations as discussed above.

[0064] Use the orientation angle to grade the images into 4 grades

[0065] The normal and cancer samples were graded based on their orientation angles according to the operations as discussed above.

[0066] Results

[0067] A total of 42 samples were obtained and graded, of which 3 samples were score-0 images, 12 score-1 images, 11 score-2 images, and 16 score-3 images. The grading process returned 100% accuracy for the score-0 images, 83.33% accuracy for score-1 images, 81.81% for score-2 images, and 87.50% for score-3 images. Overall, 36 out of 42, or 85.71% of the tissue sample images were scored correctly. The results are summarized in Table 1 below.

[0068] Table 1. Summary of grading results using A and B components

Actual score count	3	12	11	16

Automated score count	3	10	9	14

[0069] Process II

[0070] The automated grading process utilizing the L, A, and B components was applied to samples following the operations as described above. The predetermined threshold for the "orientation" for 2D histogram plot was empirically obtained as 10. For all samples, graded the image as Grade 0 if the "orientation" is over 10, or extracted cell wall pixels from CIELAB image if the "orientation" is less than the 10. Then, graded the images based on their L values, grading the image as Grade 1 if the L value was no less than 195, grading the image as Grade 2 if the L value was no less than 173 but less than 195, and grading the image as Grade 3 if the L value was less than 173.

[0071] A total of 42 samples were subjected to the automated grading process using L, A, and B components as described above. Of the 42 samples, 3 samples were score-0 images, 12 score-1 images, 11 score-2 images, and 16 score-3 images. This represented 100% accuracy for the score-0 images, 91.67% accuracy for score-1 images, 100% for score-2 images, and 100% for score-3 images. Overall, 41 out of 42, or 97.6% of the tissue sample images were scored correctly. The results are summarized in Table 2 below.

[0072] Table 2. Summary of grading results using L, A and B components

Actual score count	3	12	11	16
Automated score count	3	11	11	16

[0073] One of the advantages of the present invention is that it is able to quickly sieve score-0 images from the others. This is made possible using the angle between the x-axis and the major axis of the ellipse within the 2-dimensional histogram to determine the score-0 images.

[0074] The segmentation of the stained membranes into the different scores is made possible using existing images and determining a suitable cut-off value. With little samples in the learning process, a simply averaging of the values might suffice. However, better statistical methods may be required with more samples and a wider range of values. The application described here does not include the derivation of segmentation values as many methods are available, including those that implement artificial intelligence, neural networks and hidden markov models.

[0075] Importantly, the application provides an objective and consistent means of evaluating tissue sample images without being subjected to human errors and perceptions.

[0076] While the present invention has been described with reference to particular embodiments, it will be understood that the embodiments are illustrative and that the invention scope is not so limited. Alternative embodiments of the present invention will become apparent to those having ordinary skill in the art to which the present invention pertains. Such alternate embodiments are considered to be encompassed within the spirit and scope of the present invention. Accordingly, the scope of the present invention is described by the appended claims and is supported by the foregoing description.

CLAIMS

What is claimed is:

1. A process for automated grading of bio-specimens employing a 2-dimensional histogram plotted from the pixel values of the images of the bio-specimens, said process comprising the steps of:

converting an image into CIELAB color space if the image is not in the CIELAB color space;

plotting a 2-dimensional histogram with A channel and B channel values from the CIELAB values of each pixel in the image;

converting the 2-dimensional histogram into a binary image by scoring 0 for all cells having values 0, and scoring 1 for all cells having values greater than 0;

cleaning the binary image by removing small spots and filling of holes;

calculating from the binary image the orientation angle of the ellipse that best fits the binary image; and

grading the bio-specimens into 4 grades according to pre-determined thresholds based on the orientation angle.

2. The process of claim 1, wherein if the image is in Red-Green-Blue (RGB) colour space, the step of converting the image into CIELAB colour space comprises:

converting RGB to XYZ first, and then converting XYZ to CIELAB;

wherein the codes to convert RGB to XYZ are as follows:

```
var_R = ( R / 255 )        //R from 0 to 255
var_G = ( G / 255 )        //G from 0 to 255
var_B = ( B / 255 )        //B from 0 to 255

if ( var_R > 0.04045 ) var_R = ( ( var_R + 0.055 ) / 1.055 ) ^ 2.4
else                   var_R = var_R / 12.92
if ( var_G > 0.04045 ) var_G = ( ( var_G + 0.055 ) / 1.055 ) ^ 2.4
else                   var_G = var_G / 12.92
if ( var_B > 0.04045 ) var_B = ( ( var_B + 0.055 ) / 1.055 ) ^ 2.4
else                   var_B = var_B / 12.92

var_R = var_R * 100
var_G = var_G * 100
var_B = var_B * 100
```

```
X = var_R * 0.4124 + var_G * 0.3576 + var_B * 0.1805
Y = var_R * 0.2126 + var_G * 0.7152 + var_B * 0.0722
Z = var_R * 0.0193 + var_G * 0.1192 + var_B * 0.9505
```

and wherein the codes to convert XYZ to CIELAB are as follows:

```
var_X = X / ref_X          //ref_X =  95.047
var_Y = Y / ref_Y          //ref_Y = 100.000
var_Z = Z / ref_Z          //ref_Z = 108.883

if ( var_X > 0.008856 ) var_X = var_X ^ ( 1/3 )
else                    var_X = ( 7.787 * var_X ) + ( 16 / 116 )
if ( var_Y > 0.008856 ) var_Y = var_Y ^ ( 1/3 )
else                    var_Y = ( 7.787 * var_Y ) + ( 16 / 116 )
if ( var_Z > 0.008856 ) var_Z = var_Z ^ ( 1/3 )
else                    var_Z = ( 7.787 * var_Z ) + ( 16 / 116 )

CIE-L* = ( 116 * var_Y ) - 16
CIE-a* = 500 * ( var_X - var_Y )
CIE-b* = 200 * ( var_Y - var_Z )
```

3. The process of claim 1, wherein the operation of plotting a 2-dimensional histogram with A and B channel values from each pixel of the CIELAB image comprises the steps of:

for each pixel in the CIELAB image, the A channel value and the B channel value are plotted on a graph to count the number of times when the A and B values have appeared together in a pixel, where the graph has the A channel on the x-axis, B channel on the y-axis and the count as the z-axis; and

after plotting all the (A,B) values on the graph, a 2D histogram (a 2-dimension histogram is an XYZ graph) is obtained;

wherein the pseudocodes for obtaining the 2D histogram are as follows:

```
m = width of image;
n = height of image;

for x = 1 to m {
  for y = 1 to n {
    (l,a,b) = get pixel value at (x,y);
    histogram(a,b)=histogram(a,b)+1;
  };
};
```

thereby, since the 2D histogram keeps a count of the number of times when (a,b) has been plotted together in a pixel, it is a 3-dimensional graph in essence where the 3rd dimension is the count.

4. The process of claim 1, wherein the operation of converting the 2-dimensional histogram to a binary image comprises the steps of:

for every cell(or pixel) in the histogram, if the cell value is greater than 0, the pixel in the binary image is given a value of 1; else, the pixel in the binary image is given a value of 0;

wherein the pseudocodes for conversion of the 2D histogram into the binary image are as follows:

```
for x = 1 to m {
 for y = 1 to n {
  If(histogram(x,y)>0)=binary_image(x,y)=1;
  else binary_image(x,y)=0;
 };
};
```

5. The process of claim 1, wherein the operation of cleaning up the binary image comprises the steps of:

eliminating the background noise from the binary image by removing of small spots and filling up holes.

6. The process of claim 5, wherein the removal of small spots is done by erosion operation followed by dilation operation; and the filling up of holes in the image is done by dilation operation and follow by erosion operation.

7. The process of claim 1, wherein the operation of calculating the orientation angle comprises the steps of:

from the cleaned binary image, the orientation of the ellipse that best fits the image is calculated by the equations (1) and (2) below:

$$Orientation\ angle = 0.5 \times tan^{-1}\left|\frac{2\mu_{1,1}}{\mu_{2,0}-\mu_{0,2}}\right| \quad (1)$$

$$\mu_{p,q}=\sum_{(x,y)\in R}\sum(x-\bar{x})^p(y-\bar{y})^q) \quad (2)$$

Where p,q=0,1,2 and region R represents the image that contains N pixels (N is the total number of pixels in region R), x and y are the 2-dimensional coordinate variables and

$$\bar{x}=\frac{1}{N}\sum_{(x,y)\in R}\sum x \quad (3)$$

$$\bar{y}=\frac{1}{N}\sum_{(x,y)\in R}\sum y \quad (4)$$

8. The process of claim 1, wherein the operation of grading the images comprises the steps of:

If orientation angle > threshold 1, image is grade 0; if threshold 2 < orientation angle ≤ threshold 1, image is grade 1; if threshold 3 < orientation angle ≤ threshold 2, image is grade 2; if orientation angle ≤ threshold 3, image is grade 3; wherein the threshold for each grade is determined empirically.

9. A process of automated grading of bio-specimens by employing a 2-dimensional histogram plotted from the pixel values of the images of the bio-specimens and using the L component of the cell wall pixels to determine the grade of the image, said process comprising the steps of:

converting an image into CIELAB color space if the image is not in the CIELAB color space;

plotting a 2-dimensional histogram with A channel and B channel values from the CIELAB values of each pixel in the image;

converting the 2-dimensional histogram into a binary image by scoring 0 for all cells having values 0, and scoring 1 for all cells having values greater than 0;

cleaning the binary image by removing small spots and filling of holes;

obtaining the "orientation" for 2D histogram plot, and checking whether the "orientation" is over a predetermined threshold, then grading the image as Grade 0 if the "orientation" is over the predetermined threshold;

extracting cell wall pixels from CIELAB image if the "orientation" is less than the predetermined threshold;

obtaining mean CIELAB value (L value) of cell wall pixels; and

grading image based on L value.

10. The process of claim 9, wherein the operation of converting the image into CIELAB colour space comprises:

converting RGB to XYZ first, and then converting XYZ to CIELAB;

wherein the codes to convert RGB to XYZ are as follows:

```
var_R = ( R / 255 )          //R from 0 to 255
var_G = ( G / 255 )          //G from 0 to 255
var_B = ( B / 255 )          //B from 0 to 255

if ( var_R > 0.04045 ) var_R = ( ( var_R + 0.055 ) / 1.055 ) ^ 2.4
else                   var_R = var_R / 12.92
if ( var_G > 0.04045 ) var_G = ( ( var_G + 0.055 ) / 1.055 ) ^ 2.4
else                   var_G = var_G / 12.92
if ( var_B > 0.04045 ) var_B = ( ( var_B + 0.055 ) / 1.055 ) ^ 2.4
else                   var_B = var_B / 12.92

var_R = var_R * 100
var_G = var_G * 100
var_B = var_B * 100


X = var_R * 0.4124 + var_G * 0.3576 + var_B * 0.1805
Y = var_R * 0.2126 + var_G * 0.7152 + var_B * 0.0722
Z = var_R * 0.0193 + var_G * 0.1192 + var_B * 0.9505
```

and wherein the codes to convert XYZ to CIELAB are as follows:

```
var_X = X / ref_X          //ref_X =  95.047
var_Y = Y / ref_Y          //ref_Y = 100.000
var_Z = Z / ref_Z          //ref_Z = 108.883

if ( var_X > 0.008856 ) var_X = var_X ^ ( 1/3 )
else                    var_X = ( 7.787 * var_X ) + ( 16 / 116 )
if ( var_Y > 0.008856 ) var_Y = var_Y ^ ( 1/3 )
else                    var_Y = ( 7.787 * var_Y ) + ( 16 / 116 )
if ( var_Z > 0.008856 ) var_Z = var_Z ^ ( 1/3 )
else                    var_Z = ( 7.787 * var_Z ) + ( 16 / 116 )

CIE-L* = ( 116 * var_Y ) - 16
CIE-a* = 500 * ( var_X - var_Y )
CIE-b* = 200 * ( var_Y - var_Z )
```

11. The process of claim 9, wherein the operation of plotting a 2-dimensional histogram with A and B channel values from each pixel of the CIELAB image comprises the steps of:

for each pixel in the CIELAB image, the A channel value and the B channel value are plotted on a graph to count the number of times when the A and B values have appeared together in a pixel, where the graph has the A channel on the x-axis, B channel on the y-axis and the count as the z-axis; and

after plotting all the (A,B) values on the graph, a 2D histogram (a 2-dimension histogram is an XYZ graph) is obtained;

wherein the pseudocodes for obtaining the 2D histogram are as follows:

```
m = width of image;
n = height of image;

for x = 1 to m {
 for y = 1 to n {
  (l,a,b) = get pixel value at (x,y);
  histogram(a,b)=histogram(a,b)+1;
 };
};
```

thereby, since the 2D histogram keeps a count of the number of times when (a,b) has been plotted together in a pixel, it is a 3-dimensional graph in essence where the 3rd dimension is the count.

12. The process of claim 9, wherein the operation of converting the 2-dimensional histogram to a binary image comprises the steps of:

for every cell(or pixel) in the histogram, if the cell value is greater than 0, the pixel in the binary image is given a value of 1; else, the pixel in the binary image is given a value of 0;

wherein the pseudocodes for conversion of the 2D histogram into the binary image are as follows:

```
for x = 1 to m {
 for y = 1 to n {
  If(histogram(x,y)>0)=binary_image(x,y)=1;
  else binary_image(x,y)=0;
 };
};
```

13. The process of claim 9, wherein the operation of cleaning up the binary image comprises the steps of:

eliminating the background noise from the binary image by removing of small spots and filling up holes.

14. The process of claim 13, wherein the removal of small spots is done by erosion operation followed by dilation operation; and the filling up of holes in the image is done by dilation operation and follow by erosion operation.

15. The process of claim 9, wherein the operation of obtaining the "orientation" comprises the steps of:

from the cleaned binary image, the orientation of the ellipse that best fits the image is calculated by the equations (1) and (2) below:

$$Orientation\ angle = 0.5 \times tan^{-1}\left|\frac{2\mu_{1,1}}{\mu_{2,0}-\mu_{0,2}}\right| \quad (1)$$

$$\mu_{p,q}=\sum_{(x,y)\in R}\sum(x-\bar{x})^p(y-\bar{y})^q) \quad (2)$$

Where p,q=0,1,2 and region R represents the image that contains N pixels (N is the total number of pixels in region R), x and y are the 2-dimensional coordinate variables and

$$\bar{x}=\frac{1}{N}\sum_{(x,y)\in R}\sum x \quad (3)$$

$$\bar{y}=\frac{1}{N}\sum_{(x,y)\in R}\sum y \quad (4)$$

16. The process of claim 9, wherein the operation of extracting cell wall pixels from CIELAB images comprises the steps of:

using 3 seeds with pre-assigned values are used to represent the background, nucleus and cell wall; and

measuring the distance between each pixel against each of these seeds to determine which distance is the shortest; wherein the seed whose distance to the pixel is the shortest is the classification assigned to that pixel.

17. The process of claim 9, wherein the operation of obtaining the mean CIELAB value of cell wall pixels comprises the steps of:

receiving the pixels representing the cell wall from the extraction of cell wall pixels from CIELAB images; and

averaging the CIELAB values of all pixels representing the cell wall to give the mean CIELAB value (L) of cell wall pixels.

18. An electronic device comprising:

an electronic storage medium embedded with an electronically executable program; and

a microprocessor being electronically coupled with the electronic storage medium, executing the embedded program;

wherein the program is capable of performing the operations of:

converting an image into CIELAB color space if the image is not in the CIELAB color space;

plotting a 2-dimensional histogram with A channel and B channel values from the CIELAB values of each pixel in the image;

converting the 2-dimensional histogram into a binary image by scoring 0 for all cells having values 0, and scoring 1 for all cells having values greater than 0;

cleaning the binary image by removing small spots and filling of holes;

calculating from the binary image the orientation angle of the ellipse that best fits the binary image; and

grading the bio-specimens into 4 grades according to pre-determined thresholds based on the orientation angle.

19. An electronic device comprising:

an electronic storage medium embedded with an electronically executable program; and

a microprocessor being electronically coupled with the electronic storage medium, executing the embedded program;

wherein the program is capable of performing the operations of:

converting an image into CIELAB color space if the image is not in the CIELAB color space;

plotting a 2-dimensional histogram with A channel and B channel values from the CIELAB values of each pixel in the image;

converting the 2-dimensional histogram into a binary image by scoring 0 for all cells having values 0, and scoring 1 for all cells having values greater than 0;

cleaning the binary image by removing small spots and filling of holes;

obtaining the "orientation" for 2D histogram plot, and checking whether the "orientation" is over a predetermined threshold, then grading the image as Grade 0 if the "orientation" is over the predetermined threshold;

extracting cell wall pixels from CIELAB image if the "orientation" is less than the predetermined threshold;

obtaining mean CIELAB value (L value) of cell wall pixels; and

grading image based on L value.

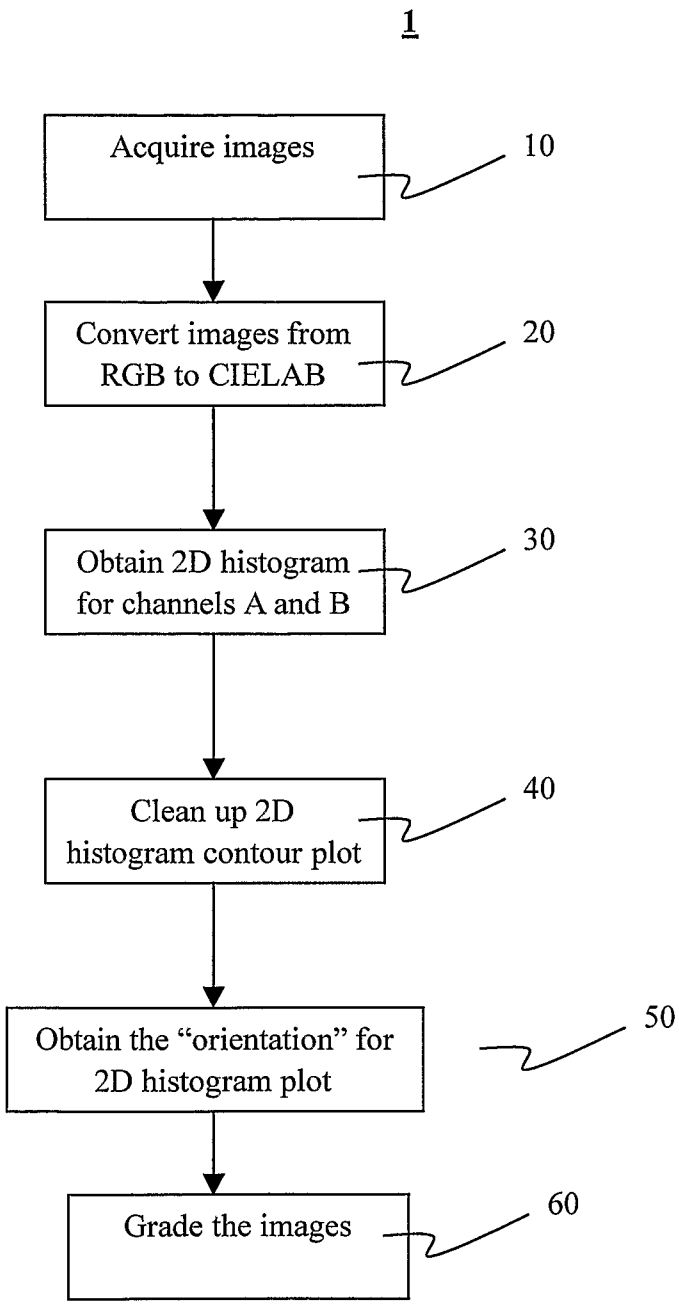
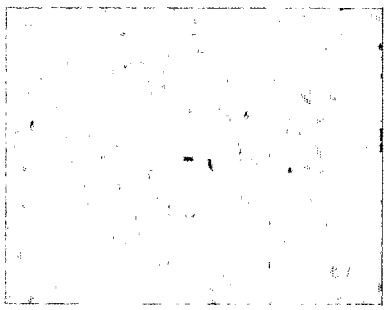
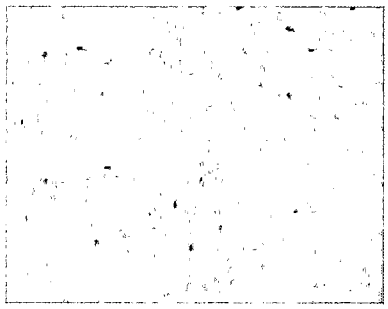

1
Acquire images
10
Convert images from RGB to CIELAB
20
Obtain 2D histogram for channels A and B
30
Clean up 2D histogram contour plot
40
Obtain the "orientation" for 2D histogram plot
50
Grade the images
60

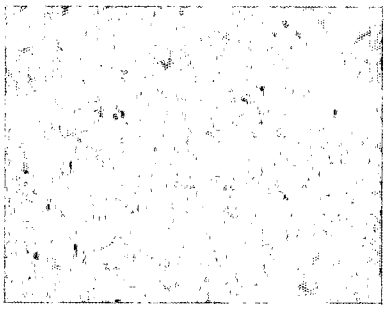
FIG 1



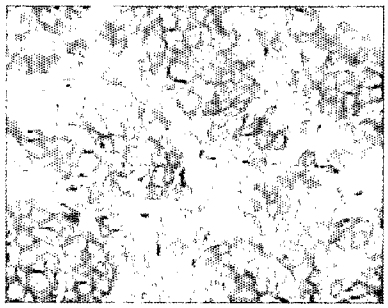
(a)



(b)

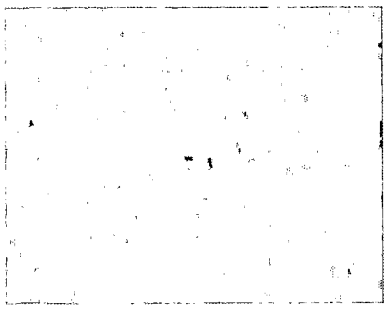


(c)



(d)

FIG 2

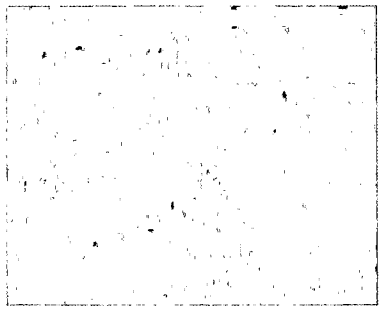


Breast normal tissue sample

(L11,A11,B11)	(L12,A12,B12)	(L13,A13,B12)	(L14,A14,B14)
(L21,A21,B21)	(L22,A22,B22)	(L23,A23,B22)	(L24,A24,B24)
(L31,A31,B31)	(L32,A32,B32)	(L33,A33,B32)	(L34,A34,B34)
(L41,A41,B41)	(L42,A42,B42)	(L43,A43,B42)	(L44,A44,B44)
(L51,A51,B51)	(L52,A52,B52)	(L53,A53,B52)	(L54,A54,B54)

CIELAB triplet values at each pixel of the

FIG 3



Breast Carcinoma tissue sample

(L11,A11,B11)	(L12,A12,B12)	(L13,A13,B12)	(L14,A14,B14)
(L21,A21,B21)	(L22,A22,B22)	(L23,A23,B22)	(L24,A24,B24)
(L31,A31,B31)	(L32,A32,B32)	(L33,A33,B32)	(L34,A34,B34)
(L41,A41,B41)	(L42,A42,B42)	(L43,A43,B42)	(L44,A44,B44)
(L51,A51,B51)	(L52,A52,B52)	(L53,A53,B52)	(L54,A54,B54)

CIELAB triplet values at each pixel of the

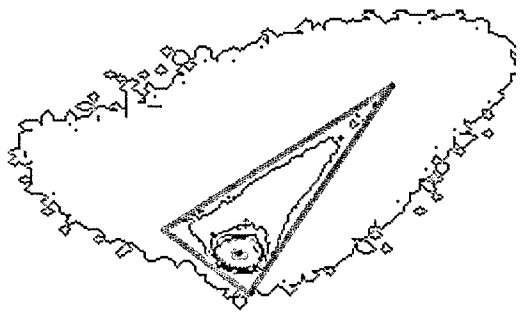


FIG 4

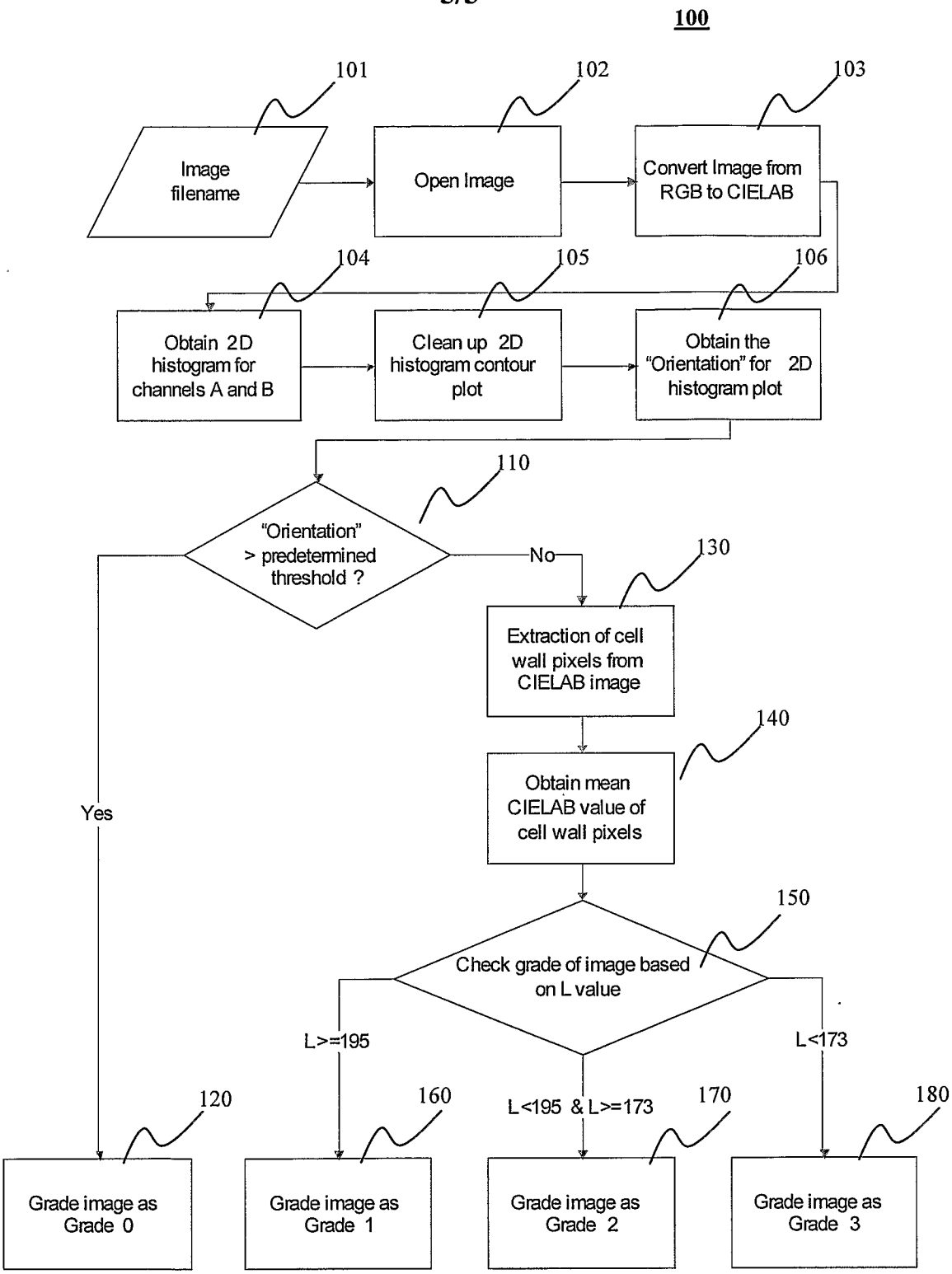

100
101
Image filename
102
Open Image
103
Convert Image from RGB to CIELAB
104
Obtain 2D histogram for channels A and B
105
Clean up 2D histogram contour plot
106
Obtain the "Orientation" for 2D histogram plot
110
"Orientation" > predetermined threshold ?
No
Yes
130
Extraction of cell wall pixels from CIELAB image
140
Obtain mean CIELAB value of cell wall pixels
150
Check grade of image based on L value
L>=195
L<195 & L>=173
L<173
120
Grade image as Grade 0
160
Grade image as Grade 1
170
Grade image as Grade 2
180
Grade image as Grade 3

FIG 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2008/000397

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.

G06T 7/60 (2006.01) ***G06T 5/40*** (2006.01) ***G01N 33/574*** (2006.01)

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)
G06T/ec/ic

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPI, EPODOC: (IMAGE W (ANALYSIS OR PROCESS+)) OR (AUTOMAT+ 2D GRAD+) OR (CANCER W GRAD+) ; CANCER+ OR TUMO?R+ OR BIO_SPECIMEN? OR CARCINOMA? OR TISSUE; ORIENTAT+ OR ELONGAT+ OR DIRECTION; COLO?R OR CIELAB OR PIXEL OR STAIN+ OR HEMATOXYLIN OR EOSIN

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TABESH A. *et. al.* "Tumor Classification in Histological Images of Prostate Using Color Texture" Conference on Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar, Oct.-Nov. 2006, Pages: 841 - 845 Digital Object Identifier 10.1109/ACSSC.2006.354868 Downloaded from IEEE Xplore Whole document	
A	CHHERAWALA Y. *et. al.* "Food Grading/Sorting Based on Color Appearance trough Machine Vision: the Case of Fresh Cranberries" Information and Communication Technologies, 2006. ICTTA '06. 2nd; 24-28 April 2006; Volume: 1, Pages: 1540-1545 ISBN: 0-7803-9521-2 Downloaded from IEEE Xplore Whole document	

[X] Further documents are listed in the continuation of Box C [X] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search
14 November 2008

Date of mailing of the international search report
24 NOV 2008

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
E-mail address: pct@ipaustralia.gov.au
Facsimile No. +61 2 6283 7999

Authorized officer
FIROOZEH RABBANI
AUSTRALIAN PATENT OFFICE
(ISO 9001 Quality Certified Service)
Telephone No : +61 2 6283 2287

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2008/000397

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 1997/24693 A (COGNEX CORPORATION) 10 July 1997 Whole document	
A	WO 2008/032317 A (SIEMENS COMPUTER AIDED DIAGNOSIS LTD) 20 March 2008 Whole document	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/SG2008/000397

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member	
WO	9724693	US	5845007
WO	2008032317	US	2008069421

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX