

METHOD AND SYSTEM FOR ANALYSING A STAINED BIO-SPECIMEN

5 FIELD OF INVENTION

The present invention broadly relates to a method and system for analysing a stained bio-specimen.

10

BACKGROUND

Many diseases are manifested by the expressions or the absence of expressions of one or more genes in the affected tissues or cells. In addition, the genes specifically associated with diseases are usually called biomarkers. Thus, diagnosis of diseases can be done with the bio-specimens taken from patients by analyzing the expressions of biomarkers.

20 One of the biomarkers is Her2/neu. The Her2/neu (also known as c-erbB2) oncogene is a cell-surface membrane tyrosine kinase receptor that is over-expressed in about 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 over-expression of Her2/neu have been known to respond poorly to conventional hormonal therapy, chemotherapy or other cytotoxic treatments.

Advances in oncology have made possible newer modalities of therapy that require precise characterization of tumours for effective treatment. In tandem with this development, the understanding of tumour biology has also improved tremendously with the advent of novel molecular genetics techniques. Increasingly, clinicians and patients will demand more information about a tumour's behaviour than what a traditional histopathological assessment can provide. Apart from having

prognostic implications, genetic information about a neoplasm may direct a specific and effective mode of treatment.

Immunohistochemistry or IHC has been traditionally used to stain and to
5 analyse cancer samples. While samples with strongly stained membranes are able to provide a somewhat conclusive finding, samples that are moderately stained usually require further analysis. In this instance, more advanced techniques are required to accurately determine if a given sample indeed has a genetic malfunction.

10 Fluorescent in situ hybridisation (FISH) overcomes the problem of non-specific staining and is able to distinguish Her2 amplification from polysomy 17, whilst retaining the morphology of the tissue analysed. FISH is now established as a gold standard in the detection of Her2 amplification in breast cancers for the selection of patients who might benefit from Herceptin treatment. However, this is a
15 cumbersome method that is generally not available, and the need for fluorescent microscopy and photography is disruptive to the workflow of most diagnostic laboratories. In addition, as fluorescence signals fade with time, it is essential to photograph each slide soon after the assay is done to ensure there is a permanent record.

20 As an alternative, Silver enhanced in situ hybridization (SISH) is a fully automated technique that uses bright field microscopy with molecular analysis, allowing pathologists to evaluate tissue within 6 hours from procurement. Researchers have found that HER2/neu gene amplification can be determined as
25 reliably by the SISH method as compared to FISH. The decreased performance time allows for same day results (e.g. within 6 hours), the preformulated reagents remove testing variability, and the use of conventional bright field microscopy makes this technology a consistent, less expensive alternative to direct-labeled manual FISH.

30 Because SISH combines bright field microscopy with molecular analysis and full automation, it appears to be particularly suited for routine application in surgical pathology. However, the cost of performing SISH is still high, mainly due to the price of the commercial probes used in these assays. In addition to the test case, one must also factor in the cost of performing the assay on a positive and negative

control specimen. This means the cost of performing SISH for a single case is considerable high, when the control specimens are factored in. The high cost of this assay has limited the availability of this useful diagnostic test for many patients with breast cancer.

5

Cell-image analysis may provide a more cost-effective supplementary or alternative analysis to the above methods. Conventional cell-image analysis methods are generally directed to segmenting or counting the relevant cells. For example, US Patent Publication No. 2007/0020697 discloses segmenting positive and negative cells which have been immunostained and calculating a population statistic, e.g. percentage of brown stained cells over the total number of cells, using colour channels in RGB and HSI colour spaces.

10

A need therefore exists to provide a method and system for analysing a stained bio-specimen that seek to address at least one of the above problems, or at least to provide an alternative.

15

SUMMARY

20

In accordance with a first aspect of the present invention, there is provided a method for analysing a stained bio-specimen, the method comprising the steps of:

25

converting an image of the stained bio-specimen from a Red-Green-Blue (RGB) colour space to an Intensity-Colours colour space comprising two colour channels defining a colour plane;

determining a collection of points in the colour plane representing respective combinations of values of the colour channels that appear at least once in a pixel in the Intensity-Colours colour space;

30

fitting an ellipse to the collection of points in the colour plane;
determining a major axis length of the fitted ellipse; and
classifying the bio-specimen based on the major axis length.

The determining of the collection of points in the colour plane may comprise:

obtaining a two-dimensional histogram for the two colour channels in the Intensity-Colours colour space; and

converting the two-dimensional histogram into a binary image in the colour plane whereby the collection of points comprises image pixels in the binary image.

5

The Intensity-Colours colour space may comprise an Intensity-Orange-Blue (IOB) colour space.

The converting of the image of the stained bio-specimen from the RGB colour space to the Intensity-Colours colour space may comprise calculating an Orange channel value based on Green and Red channel values in the RGB colour space, and calculating an Blue channel value based on Blue channel value in the RGB space.

15

The converting of the image of the stained bio-specimen from the RGB colour space to the Intensity-Colours colour space may comprise calculating an Intensity channel value based on the Green, the Red and the Blue channel values in the RGB colour space.

20

The major axis length may be determined based on the following formula:

$$major\ axis\ length = 2 \left(\sqrt{\frac{\mu_{2,0} + \mu_{0,2} + ((\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2)}{(0.5\mu_{0,0})}} \right)$$

where $\mu_{p,q} = \sum_{(x,y) \in R} (x - \bar{x})^p (y - \bar{y})^q$ ($p, q = 0, 1, 2$), region R represents the

binary image comprising N pixels, and $\bar{x} = \frac{1}{N} \sum_{(x,y) \in R} x$, $\bar{y} = \frac{1}{N} \sum_{(x,y) \in R} y$

25

The bio-specimen may be classified as containing cancerous cells if the major axis length is greater than a threshold.

The threshold may be 170 pixels.

30

In accordance with a second aspect of the present invention, there is provided a system for analysing a stained bio-specimen, the system comprising:

means for converting an image of the stained bio-specimen from a Red-Green-Blue (RGB) colour space to an Intensity-Colours colour space comprising two
5 colour channels defining a colour plane;

means for determining a collection of points in the colour plane representing respective combinations of values of the colour channels that appear at least once in a pixel in the Intensity-Colours colour space;

means for fitting an ellipse to the collection of points in the colour plane;

10 means for determining a major axis length of the fitted ellipse; and

means for classifying the bio-specimen based on the major axis length.

The means for determining the collection of points in the colour plane may be configured to obtain a two-dimensional histogram for the two colour channels in the
15 Intensity-Colours colour space, and convert the two-dimensional histogram into a binary image in the colour plane whereby the collection of points comprises image pixels in the binary image.

The Intensity-Colours colour space may comprise an Intensity-Orange-Blue
20 (IOB) colour space.

The means for converting the image of the stained bio-specimen from the RGB colour space to the Intensity-Colours colour space may be configured to calculate an Orange channel value based on Green and Red channel values in the
25 RGB colour space, and calculate a Blue channel value based on Blue channel value in the RGB space.

The means for converting the image of the stained bio-specimen from the RGB colour space to the Intensity-Colours colour space may be configured to
30 calculate an Intensity channel value based on the Green, the Red and the Blue channel values in the RGB colour space.

The major axis length may be determined based on the following formula:

$$\text{major axis length} = 2 \left(\sqrt{\frac{\mu_{2,0} + \mu_{0,2} + ((\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2)}{(0.5\mu_{0,0})}} \right)$$

where $\mu_{p,q} = \sum_{(x,y) \in R} (x - \bar{x})^p (y - \bar{y})^q$ ($p, q = 0, 1, 2$), region R represents the

binary image comprising N pixels, and $\bar{x} = \frac{1}{N} \sum_{(x,y) \in R} x$, $\bar{y} = \frac{1}{N} \sum_{(x,y) \in R} y$

- 5 The bio-specimen may be classified as containing cancerous cells if the major axis length is greater than a threshold.

The threshold may be 170 pixels.

- 10 In accordance with a third aspect of the present invention, there is provided a computer storage medium having stored thereon computer code means for instructing a computing device to execute a method for analysing a stained bio-specimen, the method comprising the steps of:

- 15 converting an image of the stained bio-specimen from a Red-Green-Blue (RGB) colour space to an Intensity-Colours colour space comprising two colour channels defining a colour plane;

determining a collection of points in the colour plane representing respective combinations of values of the colour channels that appear at least once in a pixel in the Intensity-Colours colour space;

- 20 fitting an ellipse to the collection of points in the colour plane;
determining a major axis length of the fitted ellipse; and
classifying the bio-specimen based on the major axis length.

25 BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention will be better understood and readily apparent to one of ordinary skill in the art from the following written description, by way of example only, and in conjunction with the drawings, in which:

Figure 1 shows a flow chart illustrating a method for analysing a stained bio-specimen according to an example embodiment.

5 Figures 2A and 2B show images of immunohistochemically stained tissues according to an example embodiment.

Figures 3A and 3B shows example ellipses that best fit binary images corresponding to cancerous and non-cancerous samples of Figures 2A and 2B respectively according to an example embodiment.

10

Figure 4 shows a schematic diagram of a computer system for implementing the method and system for analyzing a stained bio-specimen according to an example embodiment.

15 Figure 5 shows a flow chart illustrating a method for analysing a stained bio-specimen according to an example embodiment.

20 DETAILED DESCRIPTION

20

Figure 1 shows a flow chart 100 illustrating a method for analysing a stained bio-specimen according to an example embodiment. At step 102, an image of the stained bio-specimen is converted from a Red-Green-Blue (RGB) colour space to an Intensity-Colours colour space. At step 104, a two-dimensional histogram for the colour channels of the converted image is obtained. At step 106, the two-dimensional histogram is converted into a binary image. At step 108, a major axis length of an ellipse that best fits the binary image is calculated. At step 110, the bio-specimen is classified based on the calculated major axis length.

30

Some portions of the description which follows are explicitly or implicitly presented in terms of algorithms and functional or symbolic representations of operations on data within a computer memory. These algorithmic descriptions and functional or symbolic representations are the means used by those skilled in the data processing arts to convey most effectively the substance of their work to others skilled in

the art. An algorithm is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. The steps are those requiring physical manipulations of physical quantities, such as electrical, magnetic or optical signals capable of being stored, transferred, combined, compared, and otherwise manipulated.

5

Unless specifically stated otherwise, and as apparent from the following, it will be appreciated that throughout the present specification, discussions utilizing terms such as "scanning", "calculating", "converting", "determining", "classifying", "generating", "initializing", "outputting", or the like, refer to the action and processes of a computer system, or similar electronic device, that manipulates and transforms data represented as physical quantities within the computer system into other data similarly represented as physical quantities within the computer system or other information storage, transmission or display devices.

10

15

The present specification also discloses apparatus for performing the operations of the methods. Such apparatus may be specially constructed for the required purposes, or may comprise a general purpose computer or other device selectively activated or reconfigured by a computer program stored in the computer. The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general purpose machines may be used with programs in accordance with the teachings herein. Alternatively, the construction of more specialized apparatus to perform the required method steps may be appropriate. The structure of a conventional general purpose computer will appear from the description below.

20

25

In addition, the present specification also implicitly discloses a computer program, in that it would be apparent to the person skilled in the art that the individual steps of the method described herein may be put into effect by computer code. The computer program is not intended to be limited to any particular programming language and implementation thereof. It will be appreciated that a variety of programming languages and coding thereof may be used to implement the teachings of the disclosure contained herein. Moreover, the computer program is not intended to be limited to any particular control flow. There are many other variants of the computer program, which can use different control flows without departing from the spirit or scope of the invention.

30

Furthermore, one or more of the steps of the computer program may be performed in parallel rather than sequentially. Such a computer program may be stored on any computer readable medium. The computer readable medium may include storage devices such as magnetic or optical disks, memory chips, or other storage devices suitable for interfacing with a general purpose computer. The computer readable medium may also include a hard-wired medium such as exemplified in the Internet system, or wireless medium such as exemplified in the GSM mobile telephone system. The computer program when loaded and executed on such a general-purpose computer effectively results in an apparatus that implements the steps of the preferred method.

Figures 2A and 2B show images of immunohistochemically stained tissues according to an example embodiment. The images are acquired through e.g. microscopic imaging of normal breast tissues as well as breast tissues of varying grades of cancer. They are immunohistochemically stained such that cell walls of normal breast tissues (e.g. Figure 2B) are blue while those of abnormal breast tissues (e.g. Figures 2A) are of varying degrees of orange and brown. Typically, the nuclei of the cells in the tissue take up the blue stain. The images, e.g. of dimensions 1300 pixels by 1030 pixels, are then exported and saved in Tagged Image File Format (TIFF) format in the example embodiment. It should be appreciated that these images are manually focused in the example embodiment such that they are optimized for study and classification by the histopathologist. The same tissues may be subjected to SISH for determining the cancer status.

Referring to Figure 1, the steps of the method according to an example embodiment are now described in detail. At step 102, an image of the stained bio-specimen is converted from a Red-Green-Blue (RGB) colour space to an Intensity-Colours colour space comprising two colour channels defining a colour plane. Preferably, the Intensity-Colours colour space comprises an Intensity-Orange-Blue (IOB) colour space. As will be appreciated by a person skilled in the art, the RGB colour space is not frequently used in image processing and analysis due to e.g. lack of control. Further, the IOB colour space according to the example embodiment can advantageously simplify further processing because, as mentioned above, the cancer status of the cells can be distinguished by the relative amounts of blue and orange colours in the image.

In addition, the orange and blue colours can better represent the cancer images than other conventional colour models such as the CIELAB colour space which is developed for other applications. The blue and orange colour channels can directly measure the amount of blue and orange colours in the image. Further, computation of IOB may be advantageously simpler and hence faster than CIELAB.

In the example embodiment, the intensity channel value of the IOB colour space is calculated as a weighted average of the Red, Green and Blue channel values, while the orange channel value is calculated as an average of the Red and Green channel values. Example pseudocodes for converting the image from the RGB colour space to the IOB colour space are as follows:

```

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

      for x = 1 to m {
        for y = 1 to n {
          IOB_intensity(x,y) = 0.299Red(x,y) + 0.587Green(x,y) + 0.114Blue(x,y);
20          IOB_orange(x,y) = (Red(x,y) + Green(x,y))/2;
          IOB_blue(x,y) = Blue(x,y);
        };
      };

```

From the above step, sets of (i,o,b) values corresponding to IOB_intensity, IOB_orange and IOB_blue respectively are obtained in the example embodiment. It will be appreciated that different weights can be used for the respective Red, Green, Blue channels for calculating the intensity channel values. It is noted that while the intensity channel is used to define a colour model with 3 dimensions, in the subsequent steps, only the 2 colour channels are used in the analysis according to the example embodiment.

At step 104, a two-dimensional (2D) histogram for the colour channels of the converted image is obtained. Conceptually, for each pixel in the IOB image, the O channel value and the B channel value are plotted on a graph to count the number of coordinates with the same sets of (O,B) values. In the example embodiment, the graph comprises the O channel value on the x-axis, the B channel value on the y-axis and the count as the z-axis. After plotting all the (O,B) values on the graph, a

2D histogram (a 2D histogram is an XYZ graph) is obtained in the example embodiment. Example pseudocodes for obtaining the 2D histogram are as follows

```

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

      for x = 1 to m {
        for y = 1 to n {
10      (i,o,b) = get pixel value at (x,y);
        histogram(o,b) = histogram(o,b) + 1;
        };
      };

```

15 In the above step, for each particular coordinate (O,B), if another point in the IOB colour space with the same set of (O,B) values is encountered, the count at that coordinate is increased by 1. In the example embodiment, the above scanning is repeated for all points, thereby forming a 2D histogram.

At step 106, the two-dimensional histogram is converted into a binary image. 20 In the example embodiment, for every cell (or pixel) in the histogram, if the pixel 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. Example pseudocodes for converting the 2D histogram into the binary image are as follows:

```

25      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;
30      };
        };

```

The binary image obtained after the above step may contain noise, which can be eliminated or minimized e.g. by removing of small spots and filling up holes. 35 In the example embodiment, the removal of small spots in the image is done by erosion operation followed by dilation operation. Conversely, the filling up of holes in the image is done by dilation operation followed by erosion operation. Further, in the example embodiment, disks of diameter of about 5 pixels are used for the erosion and dilation operations described above. However, it will be appreciated that disks 40 of different sizes can be used.

Figures 3A and 3B shows example ellipses that best fit binary images corresponding to cancerous and non-cancerous samples of Figures 2A and 2B respectively according to an example embodiment. At step 108, from the cleaned binary image, a major axis length of the ellipse that best fits the binary image is calculated in the example embodiment using equations (1) and (2) below:

$$major\ axis\ length = 2 \left(\sqrt{\frac{\mu_{2,0} + \mu_{0,2} + ((\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2)}{(0.5\mu_{0,0})}} \right) \quad (1)$$

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

where $p, q = 0, 1, 2$, region R represents the binary image comprising N pixels, and

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

$$\bar{y} = \frac{1}{N} \sum_{(x,y) \in R} y \quad (5)$$

At step 110, the bio-specimen is classified based on the calculated major axis length (MAL). In the example embodiment, if the major axis length is greater than a threshold, e.g. about 170 pixels, the image is classified as having positive SISH output, i.e. the bio-specimen contains cancerous cells. On the other hand, if the major axis length is less than about 170 pixels, the image is classified as having negative SISH output, i.e. the bio-specimen does not contain cancerous cells.

Thus, the method according to the example embodiment can advantageously determine the relative amounts of blue and brown/orange colours in the image of the stained bio-specimen into a quantitative value.

The inventors have tested the system according to the example embodiment using 60 samples which have been subjected to SISH. Among the 48 samples that are classified as negative using SISH, 47 are correctly analysed based on the system of the example embodiment. In addition, among the 12 samples that are
5 classified as positive using SISH, 11 are correctly analysed using the system of the example embodiment. Overall, the correct analysis rate is about 96.67% (58/60).

The method and system of the example embodiment can be implemented on a computer system 400, schematically shown in Figure 4. It may be implemented as
10 software, such as a computer program being executed within the computer system 400, and instructing the computer system 400 to conduct the method of the example embodiment.

The computer system 400 comprises a computer module 402, input modules
15 such as a keyboard 404 and mouse 406 and a plurality of output devices such as a display 408, and printer 410.

The computer module 402 is connected to a computer network 412 via a suitable transceiver device 414, to enable access to e.g. the Internet or other
20 network systems such as Local Area Network (LAN) or Wide Area Network (WAN).

The computer module 402 in the example includes a processor 418, a Random Access Memory (RAM) 420 and a Read Only Memory (ROM) 422. The computer module 402 also includes a number of Input/Output (I/O) interfaces, for
25 example I/O interface 424 to the display 408, and I/O interface 426 to the keyboard 404.

The components of the computer module 402 typically communicate via an interconnected bus 428 and in a manner known to the person skilled in the relevant
30 art.

The application program is typically supplied to the user of the computer system 400 encoded on a data storage medium such as a CD-ROM or flash memory carrier and read utilising a corresponding data storage medium drive of a

data storage device 430. The application program is read and controlled in its execution by the processor 418. Intermediate storage of program data maybe accomplished using RAM 420.

5 Figure 5 shows a flow chart 500 illustrating a method for analyzing a stained bio-specimen according to an example embodiment. At step 502, an image of the stained bio-specimen is converted from a Red-Green-Blue (RGB) colour space to an Intensity-Colours colour space comprising two colour channels defining a colour plane. At step 504, a collection of points in the colour plane representing respective
10 combinations of values of the colour channels that appear at least once in a pixel in the Intensity-Colours colour space is determined. At step 506, an ellipse is fitted to the collection of points in the colour plane. At step 508, a major axis length of the fitted ellipse is determined. At step 510, the bio-specimen is classified based on the major axis length.

15 It will be appreciated by a person skilled in the art that numerous variations and/or modifications may be made to the present invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects to
20 be illustrative and not restrictive.

For example, while orange and blue are described as defining the colour plane of the colour space into which the RGB values are converted, other colours may be used in different example embodiments, such that that a collection of points in the colour plane
25 representing respective combinations of values of the colour channels that appear at least once in a pixel in the colour space can be fitted with an ellipse, and the bio-specimen can be classified based on the major axis length of the fitted ellipse. The selection of colours may e.g. depend on the staining methodology/composition used.

CLAIMS

1. A method for analysing a stained bio-specimen, the method comprising the steps of:

5 converting an image of the stained bio-specimen from a Red-Green-Blue (RGB) colour space to an Intensity-Colours colour space comprising two colour channels defining a colour plane;

determining a collection of points in the colour plane representing respective combinations of values of the colour channels that appear at least once in a pixel in
10 the Intensity-Colours colour space;

fitting an ellipse to the collection of points in the colour plane;

determining a major axis length of the fitted ellipse; and

classifying the bio-specimen based on the major axis length.

15 2. The method as claimed in claim 1, wherein determining the collection of points in the colour plane comprises:

obtaining a two-dimensional histogram for the two colour channels in the Intensity-Colours colour space; and

20 converting the two-dimensional histogram into a binary image in the colour plane whereby the collection of points comprises image pixels in the binary image.

3. The method as claimed in claims 1 or 2, wherein the Intensity-Colours colour space comprises an Intensity-Orange-Blue (IOB) colour space.

25 4. The method as claimed in claim 3, wherein converting the image of the stained bio-specimen from the RGB colour space to the Intensity-Colours colour space comprises calculating an Orange channel value based on Green and Red channel values in the RGB colour space, and calculating a Blue channel value based on Blue channel value in the RGB space.

30 5. The method as claimed in claims 3 or 4, wherein converting the image of the stained bio-specimen from the RGB colour space to the Intensity-Colours

colour space comprises calculating an Intensity channel value based on the Green, the Red and the Blue channel values in the RGB colour space.

6. The method as claimed in any one of the preceding claims, wherein
5 the major axis length is determined based on the following formula:

$$\text{major axis length} = 2 \left(\sqrt{\frac{\mu_{2,0} + \mu_{0,2} + ((\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2)}{(0.5\mu_{0,0})}} \right)$$

where $\mu_{p,q} = \sum_{(x,y) \in R} (x - \bar{x})^p (y - \bar{y})^q$ ($p, q = 0, 1, 2$), region R represents the
binary image comprising N pixels, and $\bar{x} = \frac{1}{N} \sum_{(x,y) \in R} x$, $\bar{y} = \frac{1}{N} \sum_{(x,y) \in R} y$

- 10 7. The method as claimed in any one of the preceding claims, wherein the bio-specimen is classified as containing cancerous cells if the major axis length is greater than a threshold.

- 15 8. The method as claimed in claim 7, wherein the threshold is 170 pixels.

9. A system for analysing a stained bio-specimen, the system comprising:

means for converting an image of the stained bio-specimen from a Red-Green-Blue (RGB) colour space to an Intensity-Colours colour space comprising two
20 colour channels defining a colour plane;

means for determining a collection of points in the colour plane representing respective combinations of values of the colour channels that appear at least once in a pixel in the Intensity-Colours colour space;

means for fitting an ellipse to the collection of points in the colour plane;

25 means for determining a major axis length of the fitted ellipse; and

means for classifying the bio-specimen based on the major axis length.

- 10 10. The system as claimed in claim 9, wherein the means for determining the collection of points in the colour plane is configured to obtain a two-dimensional histogram for the two colour channels in the Intensity-Colours colour space, and
30

convert the two-dimensional histogram into a binary image in the colour plane whereby the collection of points comprises image pixels in the binary image.

11. The system as claimed in claims 9 or 10, wherein the Intensity-
5 Colours colour space comprises an Intensity-Orange-Blue (IOB) colour space.

12. The system as claimed in claim 11, wherein the means for converting
the image of the stained bio-specimen from the RGB colour space to the Intensity-
Colours colour space is configured to calculate an Orange channel value based on
10 Green and Red channel values in the RGB colour space, and calculate a Blue
channel value based on Blue channel value in the RGB space.

13. The system as claimed in claims 11 or 12, wherein the means for
converting the image of the stained bio-specimen from the RGB colour space to the
15 Intensity-Colours colour space is configured to calculate an Intensity channel value
based on the Green, the Red and the Blue channel values in the RGB colour space.

14. The system as claimed in any one of claims 9 to 13, wherein the
major axis length is determined based on the following formula:

$$20 \quad \text{major axis length} = 2 \left(\frac{\sqrt{\mu_{2,0} + \mu_{0,2} + ((\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2)}}{(0.5\mu_{0,0})} \right)$$

where $\mu_{p,q} = \sum_{(x,y) \in R} (x - \bar{x})^p (y - \bar{y})^q$ ($p, q = 0, 1, 2$), region R represents the

binary image comprising N pixels, and $\bar{x} = \frac{1}{N} \sum_{(x,y) \in R} x$, $\bar{y} = \frac{1}{N} \sum_{(x,y) \in R} y$

15. The system as claimed in any one of claims 9 to 14, wherein the bio-
25 specimen is classified as containing cancerous cells if the major axis length is
greater than a threshold.

16. The system as claimed in claim 15, wherein the threshold is 170
pixels.

17. A computer storage medium having stored thereon computer code means for instructing a computing device to execute a method for analysing a stained bio-specimen, the method comprising the steps of:

5 converting an image of the stained bio-specimen from a Red-Green-Blue (RGB) colour space to an Intensity-Colours colour space comprising two colour channels defining a colour plane;

determining a collection of points in the colour plane representing respective combinations of values of the colour channels that appear at least once in a pixel in the Intensity-Colours colour space;

10 fitting an ellipse to the collection of points in the colour plane;
determining a major axis length of the fitted ellipse; and
classifying the bio-specimen based on the major axis length.