

SYSTEM FOR CLASSIFYING SLIDES USING SCATTER PLOT DISTRIBUTIONS

FIELD OF THE INVENTION

This invention relates to systems for analyzing biological specimens, and
5 more particularly to systems for classifying biological specimens as normal,
abnormal or suspicious using scatter plot distributions.

BACKGROUND OF THE INVENTION

In the medical industry, there is often a need for a laboratory technician, e.g.,
10 a cytotechnologist, to review a cytological specimen for the presence of specified cell
types. For example, there is presently a need to review a cervico-vaginal
Papanicolaou (Pap) smear slides for the presence of malignant or pre-malignant
cells. Since its introduction over fifty years ago, Pap smears have been a powerful
tool for detecting cancerous and precancerous cervical lesions. During that time, the
15 Pap smear has been credited with reducing mortality from cervical cancer by as
much as 70%. This once precipitous drop in the death rate has slowed however,
and the mortality rate in the United States for this preventable disease has remained
virtually constant, at about 5,000 per year since the mid-eighties. Therefore, about
one-third of the 15,000 women diagnosed with cervical cancer annually still die,
20 because the cancer was detected too late. A further cause for concern is National
Cancer Institute data that shows an annual 3% increase in the incidence of invasive
cervical cancer in white women under 50 years of age since 1986.

A number of factors may be contributing to this current threshold, not the least
of which is the fact that many women, particularly in high risk populations, are still

not participating in routine cervical cancer screening. Another contributing factor that has received much attention is the limitation of the traditional Pap smear method itself.

The reliability and efficacy of a cervical screening method is measured by its ability to diagnose precancerous lesions (sensitivity) while at the same time avoiding false positive diagnosis (specificity). In turn, these criteria are dependent on the accuracy of the cytological interpretation. Traditionally, a pathologist may perform a single cell analysis on a biological specimen by looking at the characteristics of individual cell nuclei, or a contextual analysis on the biological specimen by looking for characteristic patterns in the architecture of the cells as they appear on the slide.

Even when performing both single cell and contextual analyses, the conventional Pap smear has false negative rates ranging between 10-50%. This is due in large part to the vast number of cells and objects (typically as many as 100,000 to 200,000) that must be reviewed by a technician to determine the possible existence of a small number of malignant or pre-malignant cells. Thus, Pap smear tests, as well as other tests requiring detailed review of biological material, have suffered from a high false negative rate due to fatigue imposed on the technician.

To facilitate this review process, automated screening systems have been developed. In a typical system, an imager is operated to provide a series of images of a cytological specimen slide, each depicting a different portion of the slide. A processor then processes these images to furnish quantitative and prognostic information about the specimen. The processor can perform either a single cell analysis or a contextual analysis, or both, in providing this diagnostic information. When performing a single cell analysis, the processor may determine such cell characteristics as cytoplasmic area ratio, nuclear integrated optical density, average

nuclear optical density, nuclear texture, and cytoplasmic hue. When performing a contextual analysis, the processor may determine such architectural characteristics as (1) average internuclear distances within clusters; (ii) standard deviation of the cytoplasmic grey values; (iii) means of the area of bare nuclei; (iv) standard deviation
5 of the area of the bare nuclei; (v) average cytoplasm area of clusters; and (vi) inverse intercell distances. The results of the single cell and contextual analyses are combined to arrive at a diagnosis.

In some automated screening systems, the processor uses the diagnostic information to delineate between "normal" and "suspicious" slides. In the clinical
10 laboratory, only those slides flagged as suspicious would be examined by a cytotechnologist, thus reducing the work load. In other automated screening systems, the processor uses the diagnostic information to delineate between normal and suspicious biological material within each specimen. That is, the processor will focus the cytotechnologist's attention on the most pertinent cells, with the potential to
15 discard the remaining cells from further review. In this case, the screening device uses the diagnostic information to determine the most pertinent biological objects (e.g., the cells and cell clusters mostly likely to have attributes consistent with malignant or pre-malignant cells), and their locations on the slide. This location information is then passed onto a microscope, which automatically proceeds to these
20 locations and centers on the biological objects for review by the cytotechnologist. The cytotechnologist will then electronically mark the biological objects that are the most pertinent (for example, the objects having attributes consistent with malignant or pre-malignant cells) for further review by a pathologist.

In general, the use of automated screening systems has proved to be
25 successful, since the technician's attention is focused on the those slides that are

suspicious or on a limited number of more pertinent objects within each slide. No automated device, however, has thus far eliminated false negative and false positive results at sufficiently low cost to be practical. In addition, there is also a commercial aspect that must be taken into account. The cost borne by laboratories to review
5 cytological specimens, such as Pap smear specimens, is tied, at least in part, to the time taken for a technician to review each slide. That is, the more slides that a cytotechnologist must review, or the more time it takes for a technician to review each slide slide, the more cost in labor the laboratory incurs. Conversely, the less slides that a technician must review, and the less time that it takes for the technician
10 to review a slide, the more money the laboratory can save.

The occurrence of false positives sometimes results from biological specimens with low-quality stains. For example, a technician may fail to follow the proper protocol for applying a particular stain to the biological specimens, or if correctly applied, the shelf-life of the stain may have expired. Oftentimes, specimens
15 are not flagged as having a low-quality stain until after they have been manually reviewed by the cytotechnologist, thereby wasting valuable time.

SUMMARY OF THE INVENTION

In one embodiment of the invention, a method of classifying a biological
20 specimen comprises identifying a number of objects of interest (OOIs) within the biological specimen. This can be accomplished by acquiring images of the biological specimen and processing the images to identify OOIs. The method further comprises measuring a first (e.g., nuclear area) and a second (e.g., nuclear integrated optical density (IOD)) feature of each OOI, and then generating a scatter
25 plot of the measured first features versus the measured second features. The

specimen is then classified (e.g., as normal or suspicious) based at least in part on a distribution of points within the scatter plot. In one embodiment, the specimen is classified as normal based on a spread of the point distribution, which is relatively small in the case of a specimen exhibiting normal cellular attributes, and relatively

5 large in the case of a specimen exhibiting abnormal cellular attributes. In this case, the specimen will be classified as normal if the point distribution spread is below a threshold. The specimen may also be classified as suspicious at least partially based on a distribution of points within the scatter plot. In this case, the specimen will be classified as suspicious if the point distribution spread is above a threshold.

- 10 This classification can be applied in any one of a variety of scenarios, e.g., to remove normal slides from subsequent review or providing a standard for other devices.

In one embodiment, normal specimens are flagged as not requiring further review, in which case, subsequent review of the specimen can be conveniently passed, thereby reducing the number of slides that a cytotechnologist must review.

- 15 In contrast, a suspicious specimen is flagged as requiring further review.

Subsequent review of the specimen can be performed in any one of a variety of manners, but preferably a plurality of fields of interest that include the OOI are generated in order to focus the cytotechnologist's attention on the suspicious objects.

- 20 In one embodiment, the point distribution spread is quantified using any one of a variety of techniques. For example, the quantification step can comprise fitting a geometric shape (e.g., an ellipse) to the points, determining at least one of the dimensions of the geometric shape, and computing a value that is a function of the dimension(s). For example, a composite score (in the case of two dimensions) can
- 25 be computed. The computed value can then be compared to the threshold value.

The method can optionally comprise measuring at least one additional feature of the specimen, wherein the specimen is classified as normal based further on the additional measured feature(s). By way of non-limiting example, this additional information can be used to determine if the scatter plot is reliable. . The additional
5 feature can be any suitable feature including the total number of OOIs in the specimen, total number of artifacts in the specimen, and mean nuclear optical density of the OOIs.

In another embodiment of the invention, a biological screening system for classifying a biological specimen is provided. The system comprises an imaging
10 station for obtaining images of the specimen and generating image data from the images. The system further comprises at least one processor for processing (e.g., filtering) the image data to identify a number of OOIs within the biological specimen, measuring a first feature of each OOI, measuring a second feature of each OOI, generating a scatter plot of the first measured feature versus the second measured
15 feature, and classifying the specimen at least in part based on the spread of the point distribution. These functions can be performed in the same manner as previously described above. If fields of interest (FOIs) are optionally generated, the system comprises an automated or semi-automated reviewing station for scanning a field of view (FOV) relative to each FOI to present the one or more OOIs in each
20 FOI.

BRIEF DESCRIPTION OF THE DRAWINGS

The drawings illustrate the design and utility of embodiments of the invention, in which similar elements are referred to by common reference numerals, and in
25 which:

Fig. 1 is a plan view of a standard microscope slide carrying a biological specimen;

Fig. 2 is a plan view of a biological screening system constructed in accordance with one embodiment of the invention;

5 Fig. 3 is a scatter plot of the nuclear area (y-axis) versus the nuclear integrated optical density (x-axis) of objects of interest found in a normal biological specimen;

Fig. 4 is a scatter plot of the nuclear area (y-axis) versus the nuclear integrated optical density (x-axis) of objects of interest found in a suspicious
10 biological specimen;

Fig. 5 is a plan view of another standard microscope slide carrying a biological specimen;

Fig. 6 is a plan view of a biological screening system constructed in accordance with another embodiment of the invention; and

15 Fig. 7 is a view of a field of interest and marker indicator as shown through a field of view of a microscope used in the system of Fig. 5.

DETAILED DESCRIPTION OF THE ILLUSTRATED EMBODIMENTS

Referring to Fig. 2, a biological screening system 10 constructed in
20 accordance with one embodiment of the invention is described. The system 10 is configured to process a series of microscope slides 12 in order to classify a biological specimen 14 (best shown in Fig. 1) as either "normal," which would not require review by a cytotechnologist, or "suspicious," which would require review by the cytotechnologist.

Although the system 10 can be used to classify any biological specimen as normal or suspicious, the system 10 lends itself particularly well to the presentation of cytological cervical or vaginal material, such as that typically found on a Pap smear slide. In this case, the cells may reflect abnormalities (e.g., cytolysis, atrophy, infection, damage), malignancy or premalignancy, such as Low Grade Squamous Intraepithelial Lesions (LGSIL) or High Grade Squamous Intraepithelial Lesions (HGSIL), as well as all other cytologic categories as defined by The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnosis. The biological specimen 14 will typically be placed on the slide 12 as a thin cytological layer. Preferably, a cover slip (not shown) is adhered to the specimen 14, thereby fixing the specimen 14 in position on the slide 12. The specimen 14 may be stained with any suitable stain, such as a Papanicolaou stain.

During processing, the system 10 obtains images of biological specimens 14 carried by slides 12, which are typically contained within a cassette (not shown).

During the imaging process, the slides 12 are removed from the respective cassettes, digitally imaged, and then returned to the cassettes in a serial fashion. Then, the system 10 analyzes the digital images (either serially as each slide is processed, or offline when processing of all slides is complete) in order to classify the specimens 14 as either being normal or suspicious.

In performing the imaging operation, the system 10 comprises a camera 16, a microscope 18, and a motorized stage 20. The camera 16 captures magnified images of the slide 12 through the microscope 18. The camera 16 may be any one of a variety of conventional cameras, such as a charge coupled device (CCD) camera, which alone or in conjunction with other components, such as an analog-to-digital (A/D) converter, can produce a digital output of sufficient resolution to allow

processing of the captured images, for example a digital image having a resolution of 640 X 480 pixels. Preferably, each pixel is converted into an eight-bit value (0 to 255) depending on its optical transmittance, with "00000000" being the assigned value for least amount of light passing through the pixel, and "11111111" being the
5 assigned value for a greatest amount of light passing through the pixel. The slide 12 is mounted on the motorized stage 20, which scans the slide 12 relative to the viewing region of the microscope 18, while the camera 16 captures images over various regions of the biological specimen 12. The shutter speed of the camera 16 is preferably relatively high, so that the scanning speed and/or number of images
10 taken can be maximized.

In performing the analysis function, the system 10 comprises an image processor 22 that performs a variety of operations on the digital images obtained from the camera 16. For example, the image processor 22 may perform primary segmentation, secondary segmentation, classification, and contextual analysis
15 operations to determine if each slide 12 is normal or suspicious.

In the primary segmentation operation, the image processor 22 removes artifacts from further consideration. The image processor 22 accomplishes this by masking pixels in the digital image from further consideration that, by virtue of their lightness, are unlikely to be cell nuclei. The remaining pixels in the digital image
20 form "blobs" having all manner of shapes and sizes. The image processor 22 then performs an erosion process on the blobs in order to remove from further consideration blobs that are only a few pixels in diameter and narrow strands extending from blobs or connecting adjacent blobs. The image processor 22 then determines whether each blob in the image is an individual object or a clustered
25 object, depending on the number of pixels in the blob. For example, a blob having

more than 590 pixels can be considered a clustered object, whereas a blob having 590 or less pixels can be considered an individual object. For individual objects, blobs that do not meet certain criteria related to total area, perimeter to area ratio, optical density standard deviation, and grayscale mean pixel value are not

5 considered further.

In the secondary segmentation operation, the image processor 22 removes blobs that are unlikely to be individual cells or clustered cells. For individual objects, the image processor 22 performs a series of erosion operations, which remove small objects and eliminates projections from the remaining blobs, and dilation operations,
10 which remove holes from the remaining blobs. For clustered objects, the image processor 22 sharpens the edges of the object to provide a defined border. From the defined clustered object, the image processor 22 then selects an individual object having the highest nuclear integrated optical density (IOD) and that satisfies a particular shape requirement. The individual objects extracted from clustered
15 objects will be flagged as cluster-extracted objects.

In the classification operation, the image processor 22 measures various features for each of the individual objects and clustered objects, and then calculates an object score for each object based on the measured values of these features. Based on this score, the image processor 22 removes individual objects and
20 clustered objects that are likely to be artifacts. The remaining individual objects are considered objects of interest (OOIs). The image processor 22 then evaluates the OOIs for their nuclear IOD (either corrected or uncorrected), and ranks the OOIs in accordance with the IOD values. The image processor 22 also measures the nuclear area of each of the OOIs for the contextual analysis, described below. For
25 each digital image, the image processor 22 then stores the OOIs, along with their

ranking, and measured nuclear areas and nuclear IODs, within the memory 24 as a frame data record (FDR). In the illustrated embodiment, approximately 2000 digital images are obtained for each slide 12, and thus approximately 2000 FDRs will be stored in memory 24 for each slide 12.

5 In the contextual analysis operation, the image processor 22 obtains from the FDRs, the nuclear area and nuclear IOD's of the 100 highest ranked OOI's, and generates a scatter plot of the nuclear area (y-axis) versus the nuclear IOD (x-axis) of each of the OOI's. For example, the scatter plot illustrated in Fig. 3 was generated from a slide carrying a biological specimen with normal cells, whereas the scatter
10 plot illustrated in Fig. 4 was generated from a slide carrying a biological specimen with abnormal cells. Based on this scatter plot, the image processor 22 will either classify the specimen 14 as normal and flag the biological specimen 14 as not requiring further review by a cytotechnologist, or classify the specimen 14 as suspicious and flag the biological specimen 14 as requiring further review by a
15 cytotechnologist.

In general, biological specimens that contain cells that reflect abnormalities tend to have a wider distribution in terms of both nuclear area and nuclear IOD (i.e., the scatter points tend to form a loosely bounded region), as illustrated in Fig. 4. In contrast, biological specimens that contain normal cells tend to have a narrow
20 distribution in terms of both nuclear area and nuclear IOD (i.e., the scatter points tend to form a tightly bounded region), as illustrated in Fig. 3. Thus, the spread of the scatter point distribution provides an indication of whether the biological specimen 14 located on the slide 12 is normal or suspicious. The difference in the scatter plots illustrated in Figs. 3 and 4 stems from the fact that abnormal cells will
25 tend to have larger nuclear areas and higher nuclear IODs (by virtue of increased

amount of chromatin) than normal cells, and thus will spread the scatter points away from the otherwise well-defined region defined by the scatter point distribution resulting from normal cells.

The scatter point distribution can be quantified using any one of a variety of ways, but in a preferred embodiment, a geometric fit is performed on the scatter points, as illustrated in Figs. 3 and 4. Notably, the scatter point distribution from a normal slide will tend to be somewhat elliptical, and thus an elliptical fit is preferred. Using the general principle that the biological specimen incrementally becomes more suspicious as the scatter point distribution varies from this tightly bound elliptical shape, the size of the ellipse, as measured by its minor and major axes, will provide an indication of whether the biological specimen is normal or suspicious. The dimensions of the major and minor axes may be weighted and used to calculate a composite score.

Once the score has been computed, it is compared against a threshold value. Assuming that the score is computed in a manner such that the suspicion of a biological specimen increases as the score increases, a score higher than the threshold value will indicate that the biological specimen is suspicious, while a score lower than the threshold value will indicate that the biological specimen is normal. In this case, the image processor 22 will classify the specimen 14 as normal if the score is less than the threshold value, and will classify the specimen 14 as suspicious if the score is greater than the threshold value.

Significantly, selection of the threshold value depends on two competing factors: the generation of false negatives (i.e., slides containing abnormal cells that have been identified as normal) and the generation of false positives (i.e., slide containing normal cells that have been identified as suspicious). In general, the

generation of false negatives is less tolerable than the generation of false positives, since the possibility of misdiagnosing disease weighs heavily against the possibility of requiring unnecessary review by a cytotechnologist. Given this, the threshold value is preferably selected, such that the percentage of false negatives is

5 maintained at a low level, e.g., 5%, while allowing the percentage of false positives to increase to a high level, e.g., 50%. Although, in this case, twice as many slides than necessary are subsequently reviewed, the reduction in the number of slides actually reviewed by the cytotechnologist has been cut in half.

Notably, the limitation of the number of OOs analyzed (i.e., only 100 OOs
10 are used to generate the scatter plot), not only reduces processing requirements, but also helps to safeguard against false negatives, since the scatter plot will presumably be generated from an enriched set of suspicious OOs.

To further minimize false negatives, additional information can be used to ensure that the scatter plot generated from the highest ranked OOs is reliable. For
15 example, a scatter plot based on the nuclear areas and nuclear IODs of all OOs can be generated, and used to generate another score. This score, like the score derived from the scatter plot generated from the highest ranked OOs, can be compared to a threshold value and used to determine whether the specimen 14
should be classified as normal or suspicious. In this case, if either of the scores from
20 both scatter plots are above the respective threshold values, the specimen 14 will be classified as suspicious. Thus, for the specimen 14 to be classified as normal both scores must be below the respective threshold values. Alternatively, the scores derived from both scatter plots can be weighted and combined into a total score, which can then be compared to a single threshold value.

Other information can be used to determine if the scatter plot generated from the highest ranked OOs is reliable. For example, the number of artifacts in the biological specimen, an elevated number of which may result in a deceptive score, can be measured and compared against a threshold value. If the number of artifacts is above this threshold value, the specimen 14 may be classified as suspicious, even though the score derived from the scatter plot may indicate that the slide 12 is normal.

Also, the number of OOs can be measured and compared to a threshold value or a threshold range. For example, a relatively small number of OOs may indicate that the pool of OOs from which the scatter plot was generated is statistically too small to provide accurate results. Or a relative large number of OOs may result in a large number of artifacts that may skew the results obtained from the scatter plot. In this case, if the number of OOs is less than a low threshold value or greater than a high threshold value, i.e., the number of OOs falls outside of a threshold range, the specimen 14 may be classified as suspicious, even though the score derived from the scatter plot may indicate that the slide 12 is normal.

The mean optical density (either corrected or not) of small OOs or all OOs can be measured and compared against a threshold value or threshold range. For example, a relatively low optical density may indicate that the specimen was not sufficiently stained, whereas a relatively high optical density may indicate that the specimen was overly stained. Either case may skew the nuclear IOD measurements, thereby creating a deceptive score. In this case, if the mean optical density is less than a low threshold value or greater than a high threshold value, i.e., the mean optical density falls outside of a threshold range, the specimen 14 may be

classified as suspicious, even though the score derived from the scatter plot may indicate that the slide 12 is normal.

It should be noted that the system 10 can classify specimens for purposes other than determining whether a slide should be reviewed by a cytotechnologist.

5 For example, normal specimens, as classified by the system 10, can be used to standardize equipment, such as screening systems that require the cytotechnologist to review all slides. For example, the normal specimens can be processed through the screening system to provide a standardized mean nuclear optical density.

Subsequent unprocessed specimens can then be processed through the screening
10 system. If the subsequent specimens have measured mean nuclear optical densities that vary from the standardized mean nuclear optical densities a given amount (e.g., if the subsequent specimens have been stained too much or too little), the results will be rejected.

The use of scatter plots can also be used by screening systems that delineate
15 between normal and suspicious cellular material within each specimen. Specifically, with reference to Fig. 6, a biological screening system 110 constructed in accordance with another embodiment of the invention is illustrated. The screening system 110 is generally the same as the previously described system 10, with the exception that the system 110 requires the cytotechnologist to subsequently review
20 all slides 112 (one shown in Fig. 5), focusing the cytotechnologist's attention only on those regions of the specimen 114 (referred to as fields of interest (FOIs)) that are suspicious. That is, no slides 112 (even if suspicious) are discarded prior to review by a cytotechnologist. The system 110 uses a scatter plot to indicate which slides 112 are suspicious, thereby cautioning the cytotechnologist to pay particular
25 attention to the suspicious slides.

The system 110 generally comprises (1) an imaging station 118 for obtaining scanned images of the biological material contained on the slide 112 and generating electronic image data from the images; (2) a server 120 for filtering the image data to obtain the OOI and for assigning one or more of the OOIs to each FOI; and (3) a plurality of reviewing stations 122 (3 shown), each of which provides a field of view (FOV) (illustrated in Fig. 7) that is scanned relative to each FOI in order to present the OOIs to the cytotechnologist. The system 110 may also comprise a user interface (not shown), including a monitor, keyboard, and mouse (all not shown), so that the cytotechnologist can interact with the system 110.

The imaging station 118 is configured to image the slide 112 in a manner similar to that described above with respect to the screening system 10. To this end, the imaging station 118 comprises a camera 124, microscope 126, and motorized stage 128 that are similar to the camera 16, microscope 18, and motorized stage 20 described above. The motorized stage 128, however, additionally keeps track of the x-y coordinates of the images as they are captured by the camera 124. For example, encoders (not shown) can be coupled to the respective motors of the motorized stage 128 in order to track the net distance traveled in the x- and y- directions during imaging. These coordinates are measured relative to fiducial marks 116 affixed to the slide 112 (shown in Fig. 5). As will be described in further detail below, these fiducial marks 116 will also be used by each reviewing station 122 to ensure that the x-y coordinates of the slide 112 during the review process can be correlated with the x-y coordinates of the slide 112 obtained during the imaging process.

The server 120 comprises (1) an image processor 130 that is configured to obtain the OOIs from the image data acquired from the camera 124; (2) a FOI

processor 132, which is configured to assign OOs to each FOI; (3) a routing processor 134, which is configured to map a routing path that the reviewing station 122 will use to scan from one FOI to the next; and (4) memory 136 configured for storing the OOs and FOIs, the ranking and x-y coordinates of the OOs, and the routing path for the FOIs. It should be appreciated that the functions performed by the respective processors 132, 134, and 136 can be performed by a single processor, or alternatively, performed by more than three processors. Likewise, it can be appreciated that the memory 136 can be divided into several memories.

The image processor 130 is similar to the previously described image processor 22, with the exception that it ranks the ABCs (i.e., the abnormal clusters) and stores the ABCs, along with their rankings, with the FDRs. The image processor 130 additionally stores the coordinates of the OOs and ABCs within the FDRs. In the same manner as the image processor 22, the image processor 130 obtains from the FDRs, the nuclear area and nuclear IOD's of the 100 highest ranked OOs, and generates a scatter plot of the nuclear area (y-axis) versus the nuclear IOD (x-axis) of each of the OOs. The image processor 130 then classifies the slide 112 as normal or suspicious based on the scatter plot, and optionally any additional information. If the slide 112 is classified as suspicious, the image processor 130 will store a flag in memory 112 indicating so.

The FOI processor 132 defines FOIs based on the x-y coordinates of the OOs. In the illustrated embodiment, 22 FOIs are defined, 20 of which are reserved for OOs, and 2 of which are reserved for ABCs, i.e., abnormal clusters. In the simplest case, 20 of the FOIs are centered over the 20 highest ranked OOs, and the remaining 2 FOIs are centered over the 2 highest ranked ABCs. Alternatively, the FOIs are defined in a manner that avoids inclusion of OOs and ABCs in an FOI that

are already included within another FOI. In this manner, the OOI and ABCs can be grouped within the FOIs in a coordinated manner, so that the number of OOI and ABCs that are included within FOIs can be maximized. Alternatively, if the number of FOIs is not fixed, the number of FOIs required to include all of the OOI and ABCs
5 can be minimized.

After the FOIs have been generated, the FOI processor 134 stores the x-y coordinates of all of the FOIs in memory 138 for later use by the routing processor 136. Specifically, the routing processor 136 maps the x-y coordinates of the FOIs using a suitable routing algorithm, such as a "traveling salesman" algorithm, which
10 determines the most efficient viewing route for presenting the FOIs in the reviewing station 122. The routing processor 136 then stores the x-y coordinates of the FOIs, along with the routing plan (which in the illustrated embodiment, is accomplished by simply placing the FOIs in a list in the order that they will be reviewed), in memory 134 for subsequent access by the reviewing station 122.

15 Referring still to Fig. 6, in one embodiment, a total of three reviewing stations 122 are shown coupled to the server 120, so that up to three cytotechnologists have simultaneous access to the pertinent information stored in the server 120. Notably, the system 10 can typically process the slides 112 much quicker than a cytotechnologist can review them. Even if the specimen processing speed of the
20 system 110 is slower than the specimen review speed of a cytotechnologist, the system 110 can generally be operated 24 hours a day, whereas the typical cytotechnologist will only work 8 hours a day. Thus, the bottleneck in the screening process occurs at the human level, i.e., the detailed review of the biological material contained on the slides 112. Thus, it can be appreciated that the use of multiple

reviewing stations 122 alleviates this bottleneck, thereby providing for a much more efficient process.

Before discussing the details of the reviewing stations 122, reference is made to Fig. 7, which illustrates an exemplary FOV that each reviewing station 122 centers over a FOI. In the illustrated embodiment, the FOV has a diameter of 2.2 mm, and the FOI is defined by a 0.4 mm x 0.4 mm square circumscribed by the FOV. In the actual embodiment, the borders of the FOI are imaginary and cannot be seen, so that the cytotechnologist's view of any OOI is not obstructed. In order to more quickly direct the cytotechnologist's attention to the FOI and to provide a reference that generally indicates the exact region bound by the imaginary borders of the FOI, an L-shaped mark indicator 142 is provided. The mark indicator 142 captures the FOI (i.e., an open square portion 144 of the mark indicator 142 borders the left and bottom sides of the FOI). A 0.05 mm margin is provided between the mark indicator 142 borders and the imaginary borders of the FOI, so that the portions of OOI extending outside of the left and bottom borders of the FOI (resulting from an OOI that is included within the FOI, but centered near the left or bottom border of the FOI) will not be obstructed by the mark indicator 142. The mark indicator 142 also serves to provide a means for the cytotechnologist to electronically mark the FOI (e.g., by pushing a button that electronically colors the mark indicator 52) as requiring further review by a pathologist (e.g., if an OOI has malignant or pre-malignant attributes).

Referring back to Fig. 6, each reviewing station 122 comprises a microscope 138 and a motorized stage 140. The slide 112 (after image processing) is mounted on the motorized stage 140, which moves the slide 112 relative to the viewing region of the microscope 138 based on the routing plan and x-y coordinates of the FOIs obtained from memory 136. The slide 112 (after image processing) is mounted on

the motorized stage 140, which moves the slide 112 relative to the viewing region of the microscope 138 based on the routing plan and a transformation of the x-y coordinates of the FOIs obtained from memory 138. Specifically, these x-y coordinates, which were acquired relative to the x-y coordinate system of the imaging station 118, will be transformed into the x-y coordinate system of the reviewing station 122 using the fiducial marks 116 affixed to the slide 112 (shown in Fig. 5). Thus, it is ensured that the x-y coordinates of the slide 112 during the reviewing process are correlated to the x-y coordinates of the slide 112 during the imaging process. The motorized stage 140 will then move in accordance with the transformed x-y coordinates of the FOIs, as dictated by the routing plan. In the illustrated embodiment, to advance from one FOI to another, the cytotechnologist presses an activation switch (not shown). In this sense, the reviewing station 122 is semi-automatic. Alternatively, the FOIs are automatically advanced from one to the next. In this case, the motorized stage 140 may optionally pause for a predetermined amount of time for each FOI. In this sense, the reviewing station 122 is fully automatic.

As the selected FOIs are presented in the FOV of the microscope 138, the cytotechnologist reviews the FOIs and makes decisions about the level of cell abnormality, if any. The cytotechnologist will then electronically mark FOIs that are suspect. The cytotechnologist is capable of returning to a previously viewed FOI, and manually moving to (and viewing) locations on the slide not encompassed by FOIs. Following review of the slide 112, if any of the FOIs have been marked by the cytotechnologist, the reviewing station 122 automatically scans the entire biological specimen 112, so that 100% viewing coverage is ensured. The cytotechnologist is

able to pause the autoscan and to move the stage 140 in order to reposition and access locations on the slide 112, as desired.

Notably, the reviewing station 122 will apprise the cytotechnologist if the specimen 114 has been classified as suspicious (e.g., by providing an audible or visible signal) in which case, the cytotechnologist will be prompted to perform a more thorough review of the slide 112 as compared to a review that the cytotechnologist would perform on a slide carrying a specimen that has been classified as normal.

The use of scatter plots can also be used to determine the quality of the stain on a batch of specimens. It has been discovered that biological specimens with low quality stains (e.g., the stain protocol has not been followed or the age of the stain has surpassed its shelf life and has deteriorated) tend to have a wider distribution in terms of both nuclear area and nuclear IOD (i.e., the scatter points tend to form a loosely bounded region). In contrast, biological specimens with high quality stains tend to have a narrow distribution in terms of both nuclear area and nuclear IOD (i.e., the scatter points tend to form a tightly bounded region). Thus, the spread of the scatter point distribution provides an indication of the quality of the stain on biological specimens 14 located on the slides 12. The difference between scatter plots generated from high quality and low quality stains stems from the fact that a low-quality stain is non-uniform, and thus will spread the scatter points away from the well-defined region otherwise resulting from a uniform and well-delineated nuclear to cytoplasmic contrast.

The scatter point distribution for each specimen can be quantified using any one of a variety of approaches. In one embodiment, a geometric fit is performed on the scatter points and a value is computed in the same manner described above in determining whether a specimen is normal or suspicious. A combined value is then

calculated from the values computed for a batch of specimens (e.g., 100 specimens), so that the stain quality of the specimen batch, as a whole, can be quantified. The median of the point distribution values for the specimens can be calculated. Alternatively, the median of point distribution values for a percentage of
5 the specimens (e.g., the lowest 10 point distribution values) can be calculated. Rather than using the median, the mode or average of the point distribution values can be used.

Once the median value has been computed, it can be used to determine whether the current specimens have been properly stained. This can be
10 accomplished in a variety of manners. For example, the median value can be compared with a baseline value computed (e.g., by taking the mean of point distribution values) from specimens that are known to be properly stained. Preferably, the number of specimens used to compute the baseline median value is large enough to provide a normal sampling of the specimens processed through the
15 site at which the imaging system is located.

Assuming that the point distribution values are computed in a manner such that the values increase as the stain quality of the specimens decrease, a median value higher than the baseline value by a given amount will indicate that the current batch of specimens may have a low-quality stain. In this case, the image processor
20 22 will generate an error or a flag indicating that the recently analyzed specimens may not be properly stained. A subset of these suspect specimens may then be reviewed manually to visually confirm the quality of the stain.

Alternatively, a Chi-Square statistical method can be used for evaluating the median value of the current batch of specimens compared to the median values of
25 previous batches of specimens. For example, given the equation:

$c^2 = \sum (f_o - f_e)^2 / f_e$, where c^2 equals the Chi-Square, f_o equals the obtained mean values, and f_e equals the expected mean value, it can be determined if the latest batch of specimens has a low-quality stain. For example, if the mean values of the last five batches of specimens are 130, 150, 110, 100, and 170, respectively, f_e

5 equals $\frac{130+150+110+100+170}{5} = 132$, and $c^2 =$

$$\frac{(130-132)^2}{132} + \frac{(150-132)^2}{132} + \frac{(110-132)^2}{132} + \frac{(100-132)^2}{132} + \frac{(170-132)^2}{132} = 0.0303 + 2.4545 + 3.6667 + 7.7575 + 10.9393 = 24.8483$$

Using a standard statistical handbook, the Chi-Square c^2 can then be compared to a cutoff value at a given confidence level to determine whether the mean values of the specimen batches fall within a normal distribution. If not, it is
10 determined that the latest batch of specimens has a low-quality stain.

Alternatively, rather than determining a mean from the computed point distribution values, and then comparing that mean to a baseline value, the point distribution values can be directly compared with the baseline value. In this case, a number of the point distribution values that surpass the baseline value by a given
15 number can be calculated, and compared to a threshold number. For example, if twenty percent of the point distribution values surpass the threshold value by the given number, it may be determined that the quality of the stain is suspect, thereby prompting the generation of an error or a flag.

CLAIMS

1. A biological screening system for classifying a biological specimen, comprising:

an imaging station for obtaining scanned images of the specimen, and

5 generating image data from the scanned images; and

at least one processor for filtering the image data to obtain OOIs, identifying a number of objects of interest (OOIs) within the biological specimen, measuring a nuclear area of each OOI, measuring a nuclear integrated optical density of each OOI, generating a scatter plot of the measured nuclear areas versus the nuclear
10 integrated optical densities, and classifying the specimen as normal at least partially based on a distribution of points within the scatter plot.

2. The biological screening system of claim 1, wherein the at least one processor is further for flagging a normal specimen as not requiring further review.

3. The biological screening system of claim 1, wherein the at least one
15 processor is further for a classifying the specimen as either normal or suspicious based on the point distribution.

4. The biological screening system of claim 3, wherein the at least one processor is further for flagging a normal specimen as not requiring further review, and flagging a suspicious specimen as requiring further review.

20 5. The biological screening system of claim 1, wherein the specimen is classified as normal based on a spread of the point distribution.

6. The biological screening system of claim 5, wherein the specimen is classified as normal if the point distribution spread is below a threshold.

7. The biological screening system of claim 5, wherein the at least one
25 processor is further for quantifying the point distribution spread, defining a threshold

value, and comparing the quantified point distribution spread to the threshold value, wherein the specimen is classified as normal based on the comparison.

8. The biological screening system of claim 7, wherein the at least one processor quantifies the point distribution spread by fitting a geometric shape to the points, determining at least one dimension of the geometric shape, and computing a value that is a function of the at least one dimension, and wherein the at least one processor compares the quantified point distribution spread to the threshold value by comparing the computed value to the threshold value.

9. The biological screening system of claim 8, wherein the geometric shape is an ellipse.

10. The biological screening system of claim 7, wherein the at least one processor quantifies the point distribution spread by computing a value, wherein the computed value is compared to the threshold value.

11. The biological screening system of claim 1, wherein the at least one processor is further for measuring at least one additional feature of the biological specimen, wherein the specimen is classified as normal based further on the at least one additional measured feature.

12. The biological screening system of claim 11, wherein the at least one processor is further for determining that the scatter plot is reliable based on the at least one additional measured feature.

13. The biological screening system of claim 11, wherein the at least one additional measured feature is a total number of OOIs in the biological specimen.

14. The biological screening system of claim 11, wherein the at least one additional measured feature is a total number of artifacts in the biological specimen.

15. The biological screening system of claim 11, wherein the at least one additional measured feature is a mean nuclear optical density of the OOIs.

16. The biological screening system of claim 1, wherein the number of OOIs is predetermined.

5 17. The biological screening system of claim 3, wherein the at least one processor is further for generating a plurality of fields of interest to include the OOIs if the specimen is classified as suspicious, the system further comprising an automated or semi-automated reviewing station for scanning a field of view relative to each field of interest to present the OOIs.

10 18. A biological screening system for classifying a biological specimen, comprising:

an imaging station for obtaining scanned images of the specimen, and generating image data from the scanned images; and

15 at least one processor for filtering the image data to obtain OOIs, identifying a number of objects of interest (OOIs) within the biological specimen, measuring a first feature of each OOI, measuring a second feature of each OOI, generating a scatter plot of the first measured feature versus the second measured feature, the scatter plot having a distribution of points having a spread, and classifying the specimen as normal at least partially based on the spread of the point distribution.

20 19. The biological screening system of claim 18, wherein the at least one processor is further for flagging a normal specimen as not requiring further review.

20. The biological screening system of claim 18, wherein the at least one processor is further for a classifying the specimen as either normal or suspicious based on the point distribution.

21. The biological screening system of claim 20, wherein the at least one processor is further for flagging a normal specimen as not requiring further review, and flagging a suspicious specimen as requiring further review.

22. The biological screening system of claim 18, wherein the specimen is
5 classified as normal if the point distribution spread is below a threshold.

23. The biological screening system of claim 18, wherein the at least one processor is further for quantifying the point distribution spread, defining a threshold value, and comparing the quantified point distribution spread to the threshold value, wherein the specimen is classified as normal based on the comparison.

10 24. The biological screening system of claim 23, wherein the at least one processor quantifies the point distribution spread by fitting a geometric shape to the points, determining at least one dimension of the geometric shape, and computing a value that is a function of the at least one dimension, and wherein the at least one processor and compares the quantified point distribution spread to the threshold
15 value by comparing the computed value to the threshold value.

25. The biological screening system of claim 24, wherein the geometric shape is an ellipse.

26. The biological screening system of claim 23, wherein the at least one processor quantifies the point distribution spread by computing a value, wherein the
20 computed value is compared to the threshold value.

27. The biological screening system of claim 18, wherein the at least one processor is further for measuring at least one additional feature of the biological specimen, wherein the specimen is classified as normal based further on the at least one additional measured feature.

28. The biological screening system of claim 27, wherein the at least one processor is further for determining that the scatter plot is reliable based on the at least one additional measured feature.

29. The biological screening system of claim 27, wherein the at least one
5 additional measured feature is a total number of OOIs in the biological specimen.

30. The biological screening system of claim 27, wherein the at least one additional measured feature is a total number of artifacts in the biological specimen.

31. The biological screening system of claim 27, wherein the at least one additional measured feature is a mean nuclear optical density of the OOIs.

10 32. The biological screening system of claim 18, wherein the number of OOIs is predetermined.

33. The biological screening system of claim 18, wherein the at least one processor is further for generating a plurality of fields of interest to include the OOIs if the specimen is classified as suspicious, the system further comprising an
15 automated or semi-automated reviewing station for scanning a field of view relative to each field of interest to present the OOIs.