



US 20050260583A1

(19) **United States**
 (12) **Patent Application Publication** (10) **Pub. No.: US 2005/0260583 A1**
Jackway et al. (43) **Pub. Date: Nov. 24, 2005**

(54) **CHROMATIN SEGMENTATION** (52) **U.S. Cl. 435/6**

(76) Inventors: **Paul Jackway**, Toowong Queensland
 (AU); **Andrew Menhert**, St Lucia
 Queensland (AU)

(57) **ABSTRACT**

Correspondence Address:

WORKMAN NYDEGGER
(F/K/A WORKMAN NYDEGGER & SEELEY)
60 EAST SOUTH TEMPLE
1000 EAGLE GATE TOWER
SALT LAKE CITY, UT 84111 (US)

A method of segmenting chromatin particles in a nucleus of a cell by locating regional minima in an image, computing a zone of influence (ZOI) around each regional minimum, and segmenting a single chromatin blob within each ZOI using a region growing procedure. The method can be used as the basis of a method of qualitatively characterizing the distribution of nuclear chromatin by computing features for individual chromatin particles. Chromatin features can be synthesized from the features of individual particles and particle features can be synthesized into nucleus features and slide features. The method is useful for detecting malignancy associated changes and changes during neoplasia. The method may also be used more generally to assess chromatin patterns in living cells during the cell life cycle. This makes it possible to measure alternations in the evolving patterns that result from pathological or environmental influences.

(21) Appl. No.: **10/484,530**

(22) PCT Filed: **Jul. 19, 2002**

(86) PCT No.: **PCT/AU02/00969**

(30) **Foreign Application Priority Data**

Jul. 19, 2001 (AU)..... PR 6478

Publication Classification

(51) **Int. Cl.⁷ C12Q 1/68**

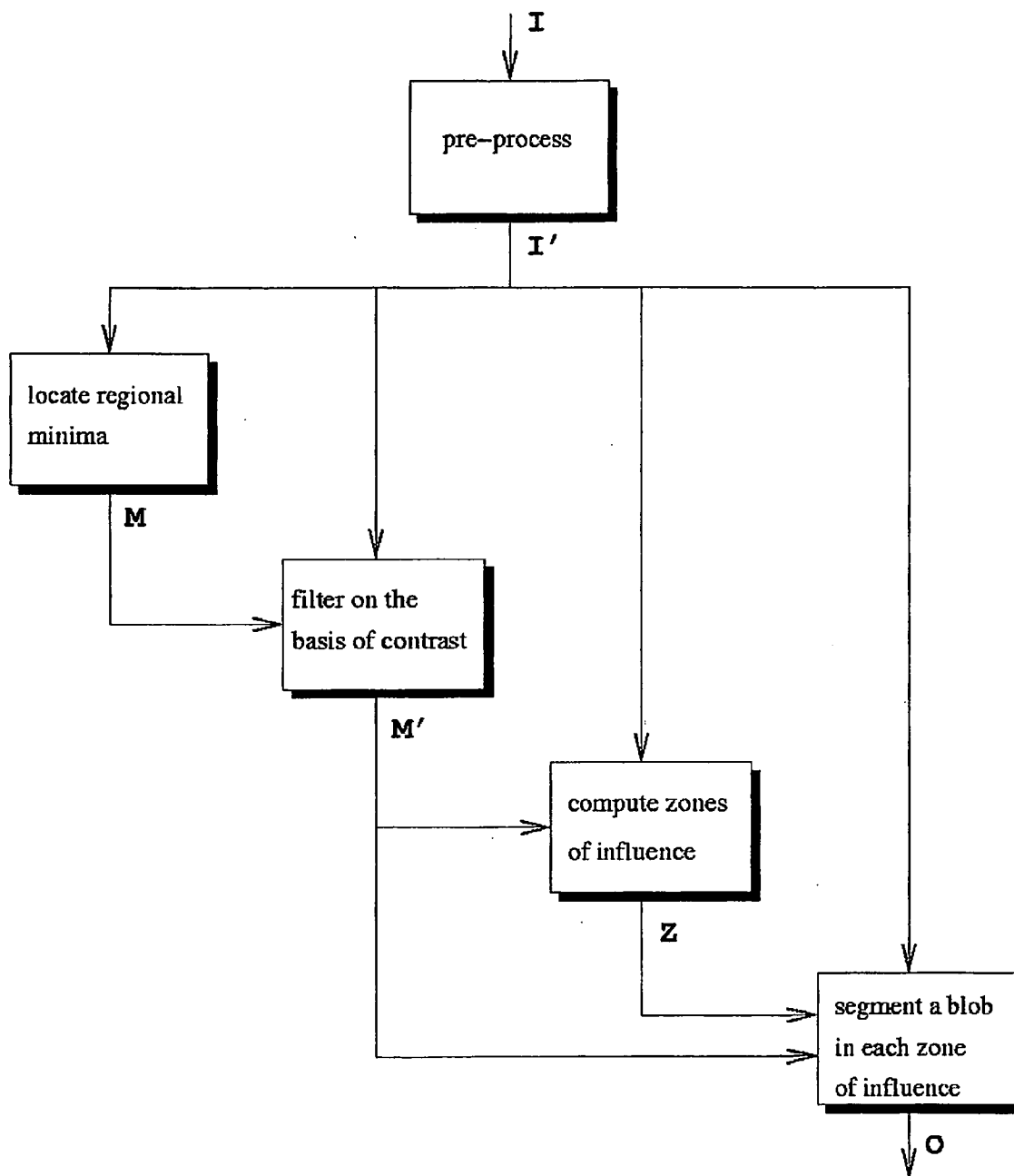


FIG 1

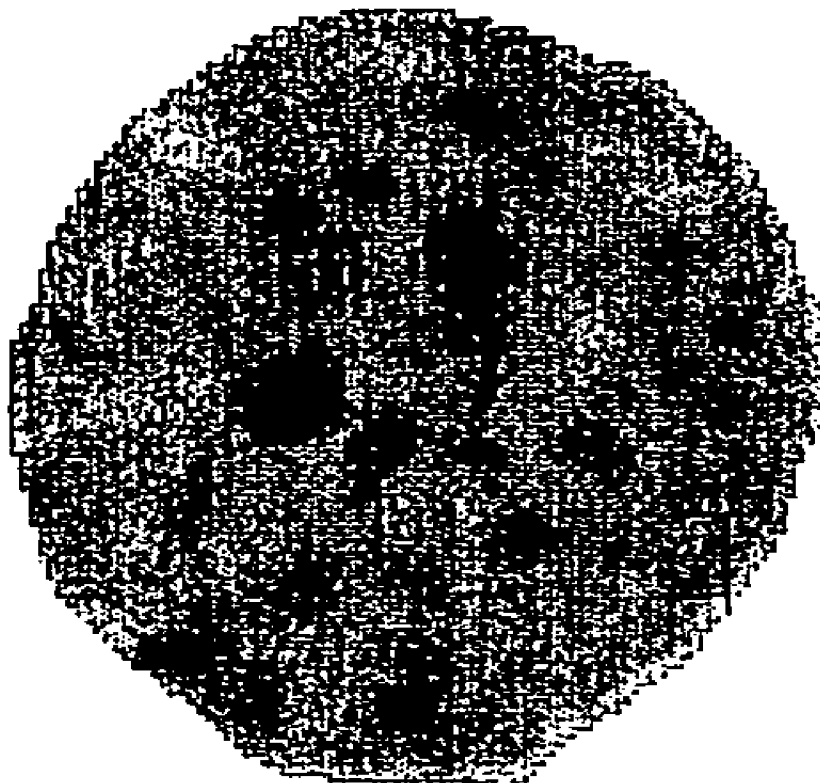


FIG 2

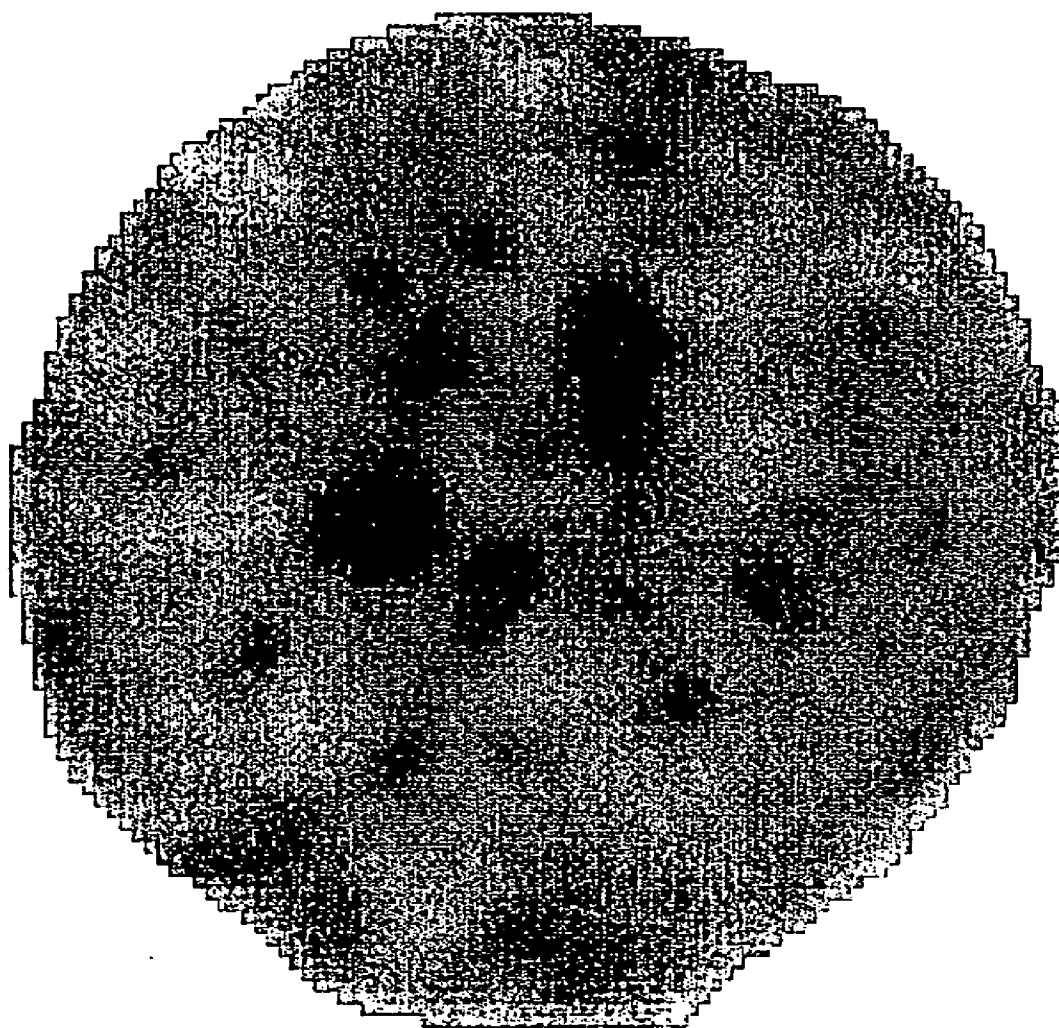


FIG 3

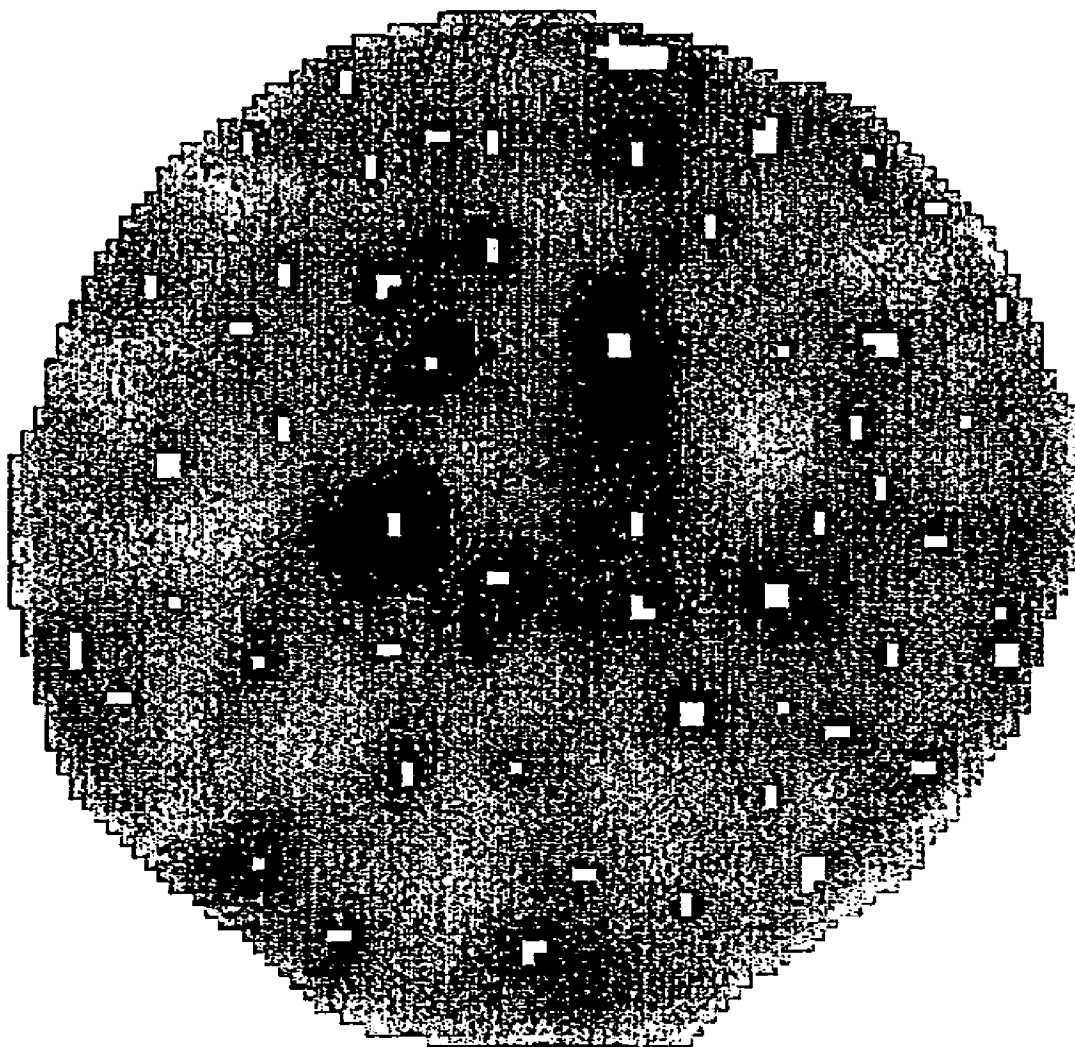


FIG 4

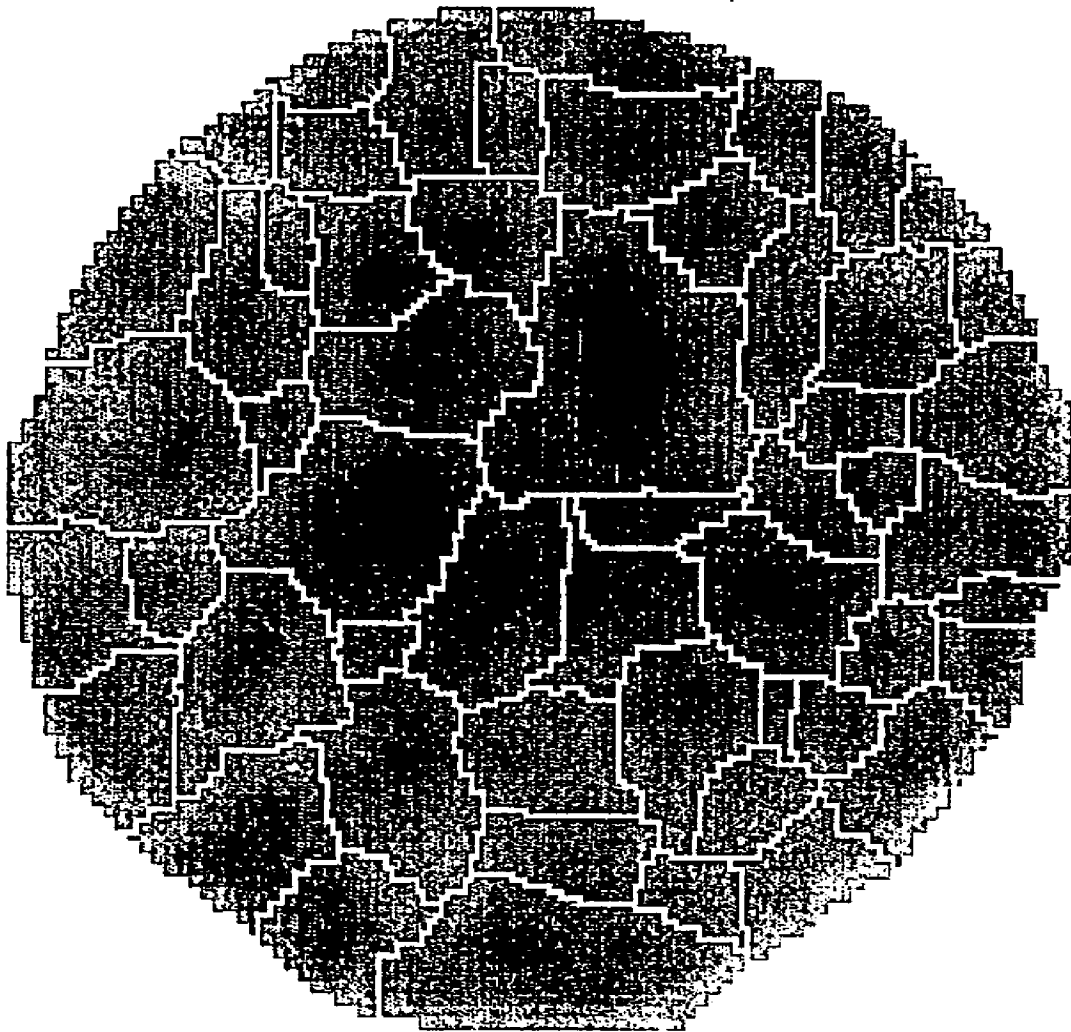


FIG 5

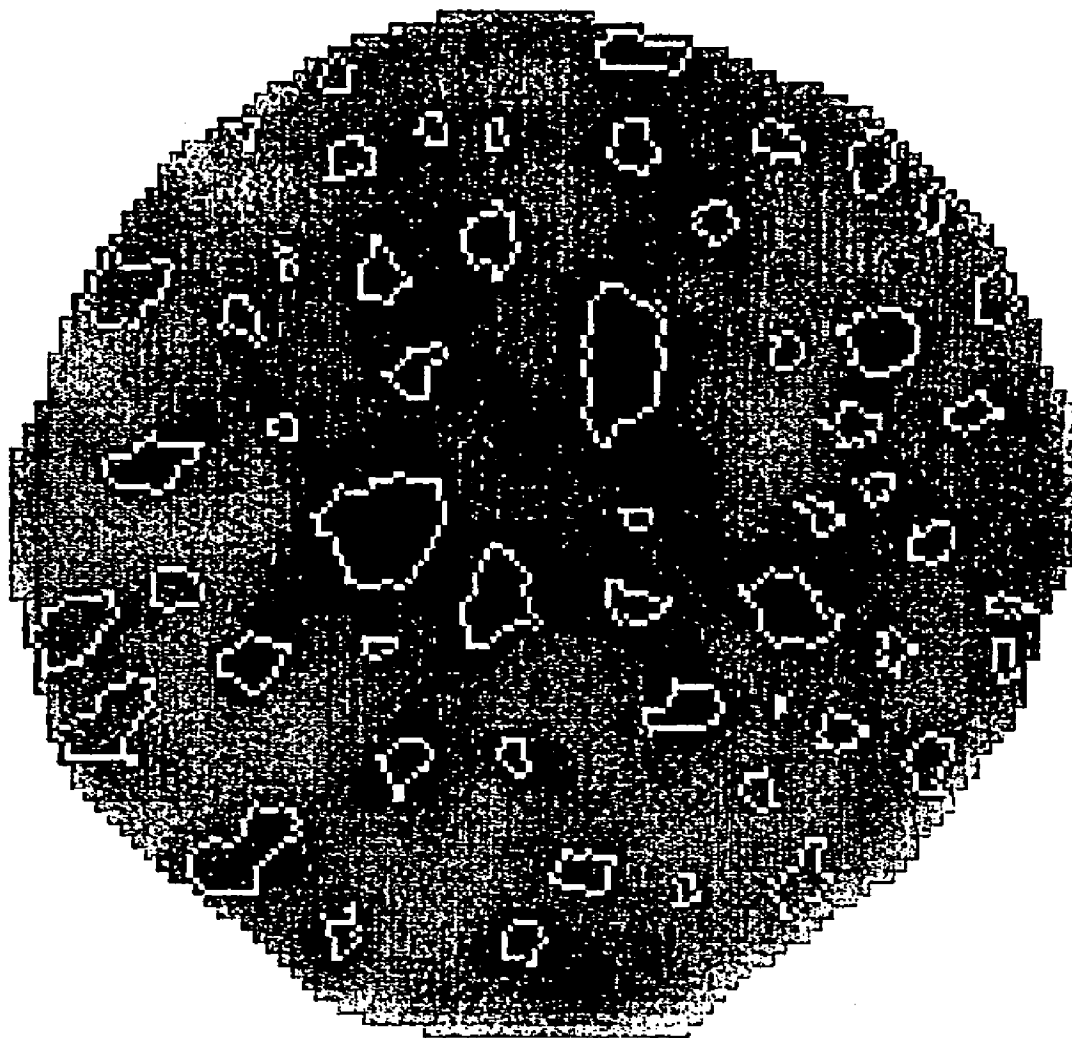


FIG 6

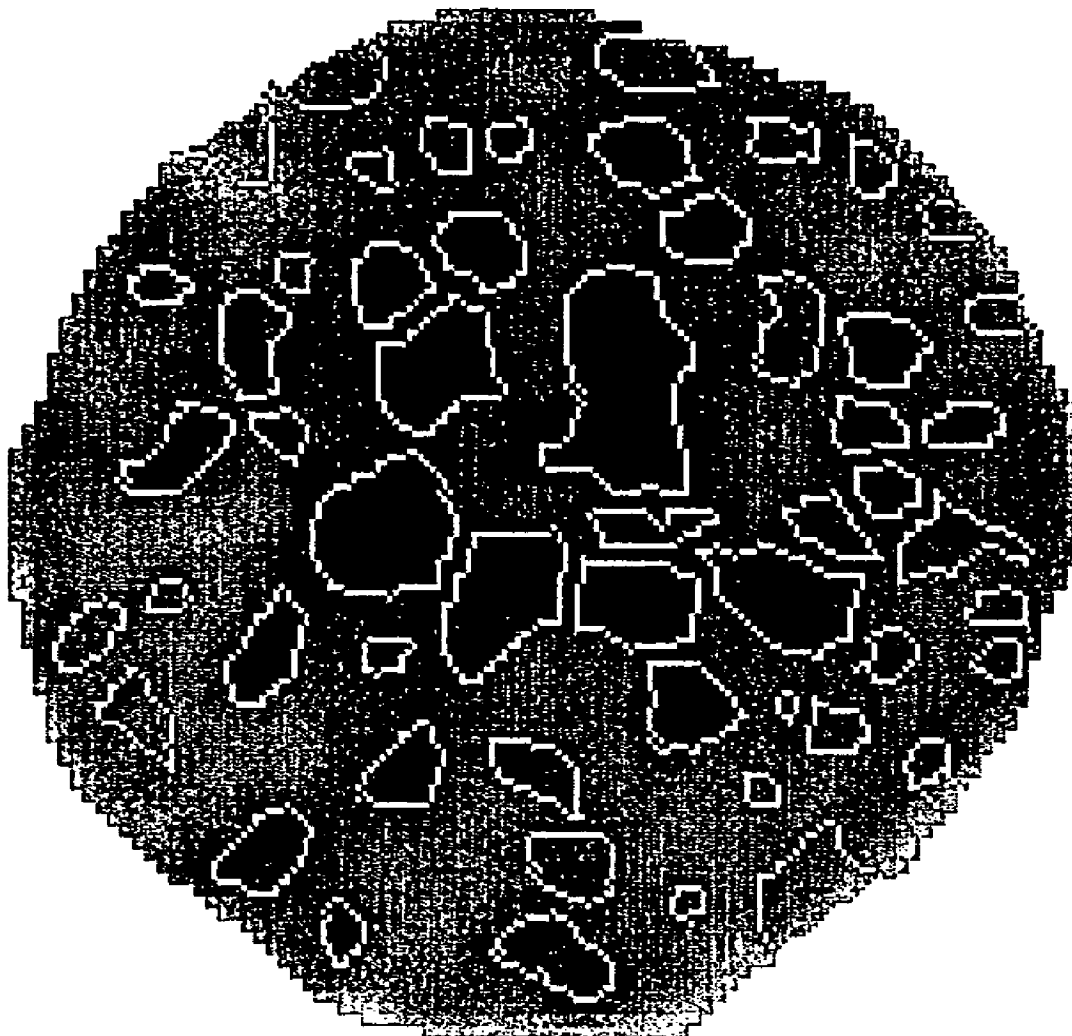


FIG 7

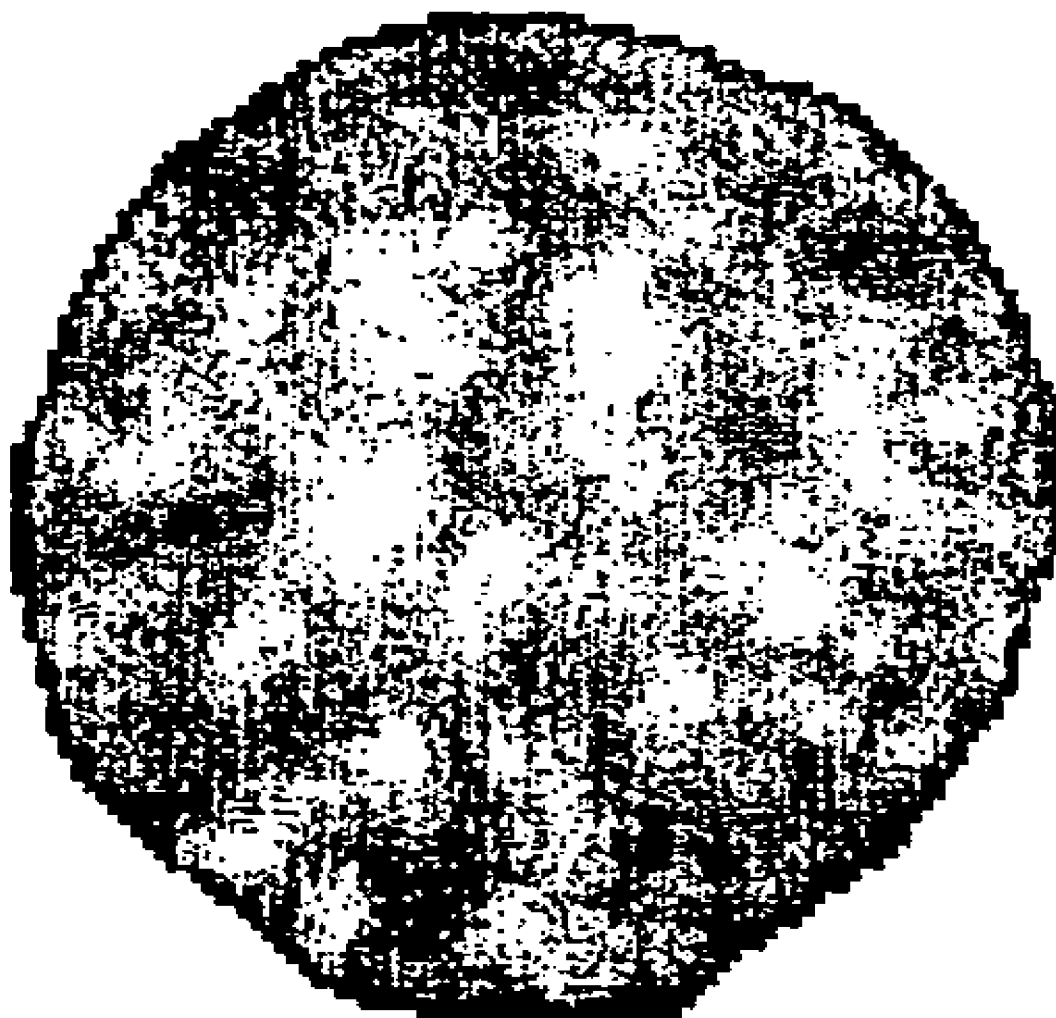


FIG 8

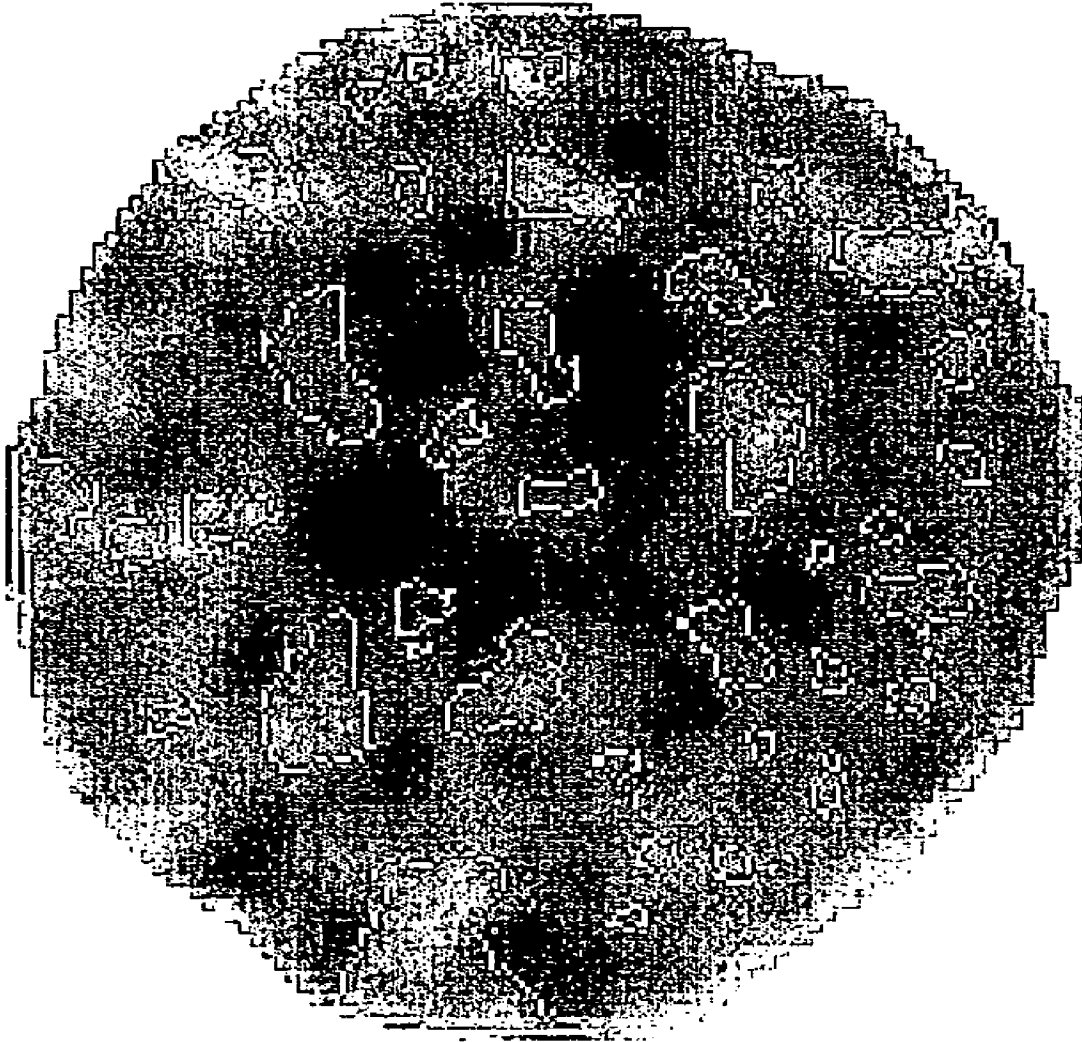


FIG 9

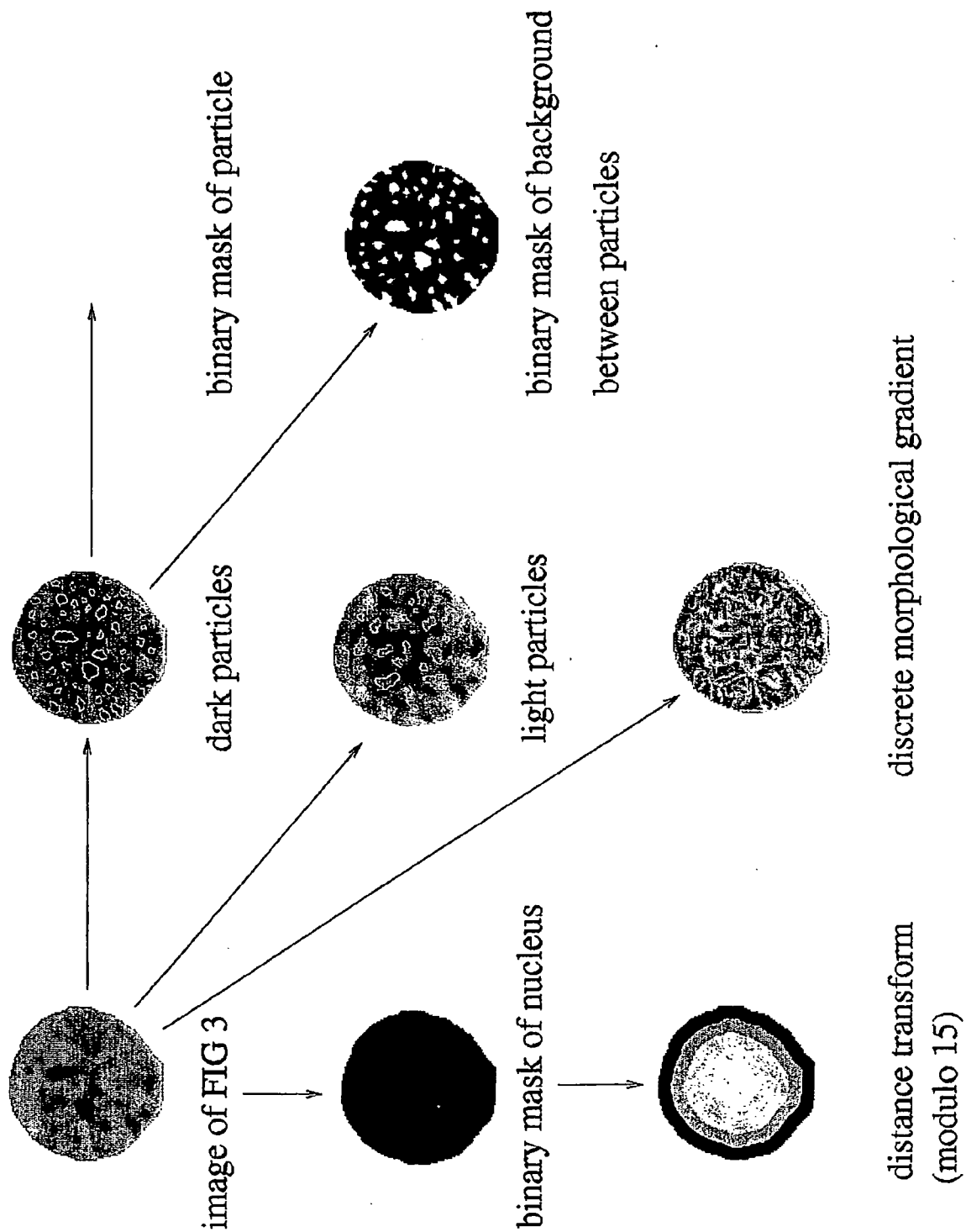


FIG 10

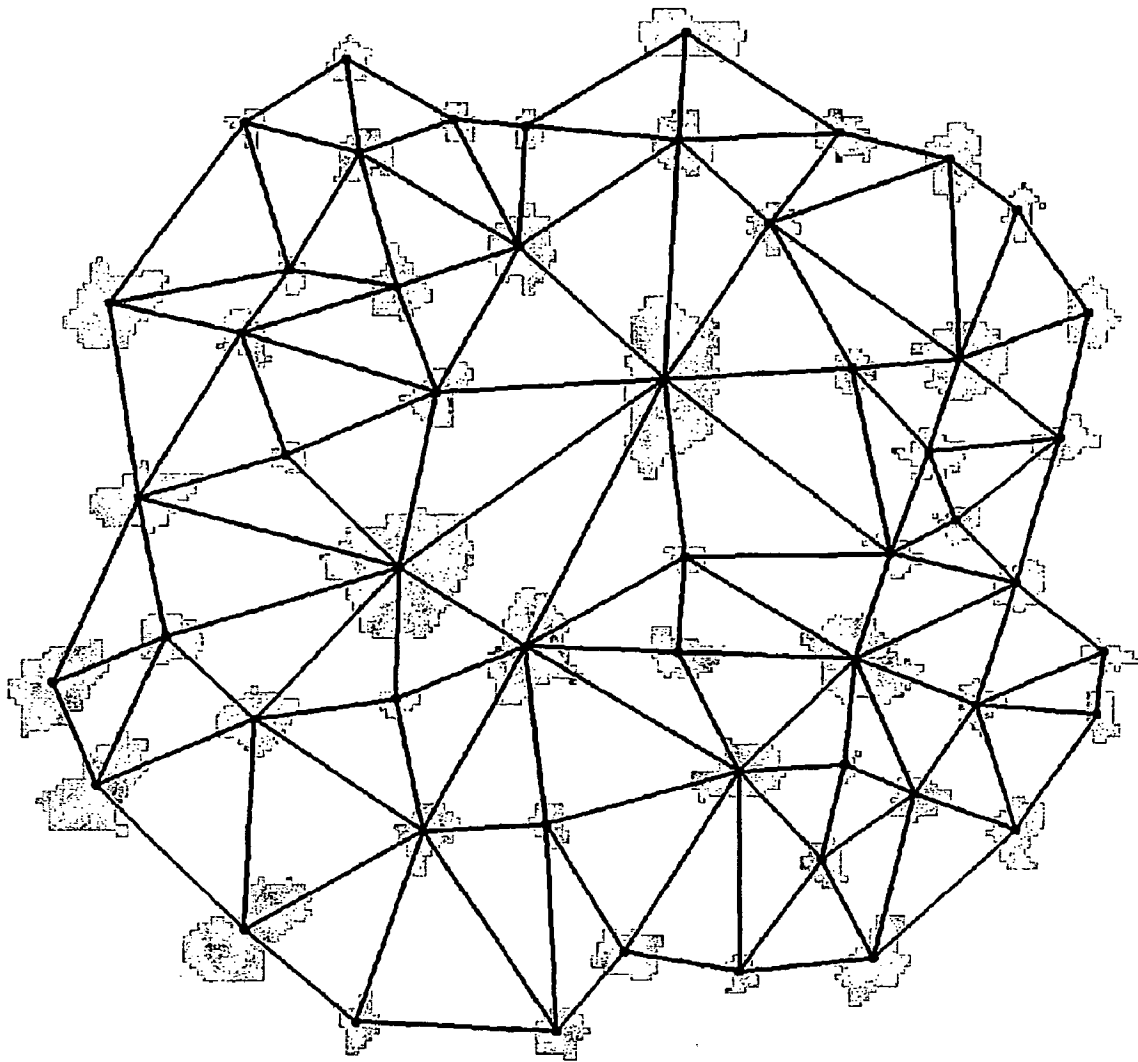


FIG 11

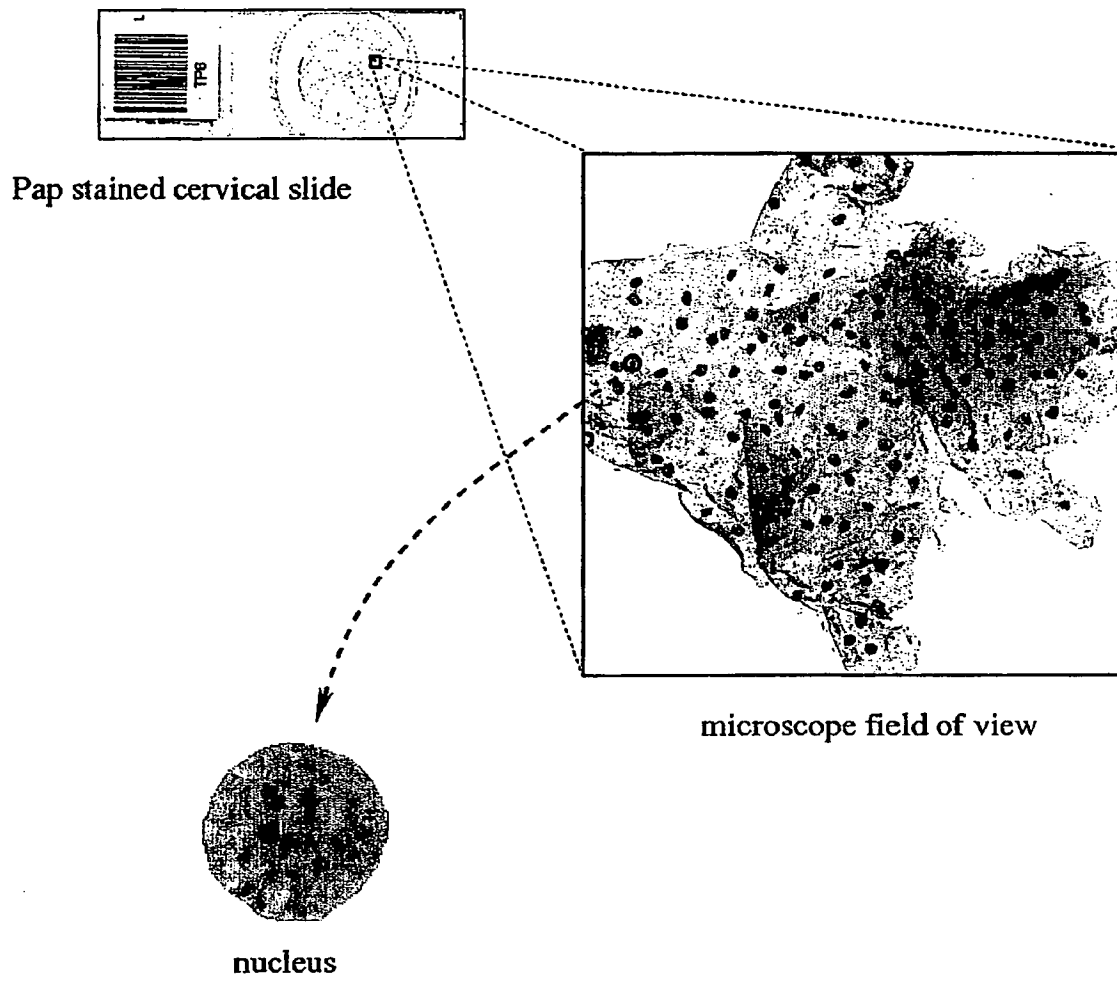


FIG 12

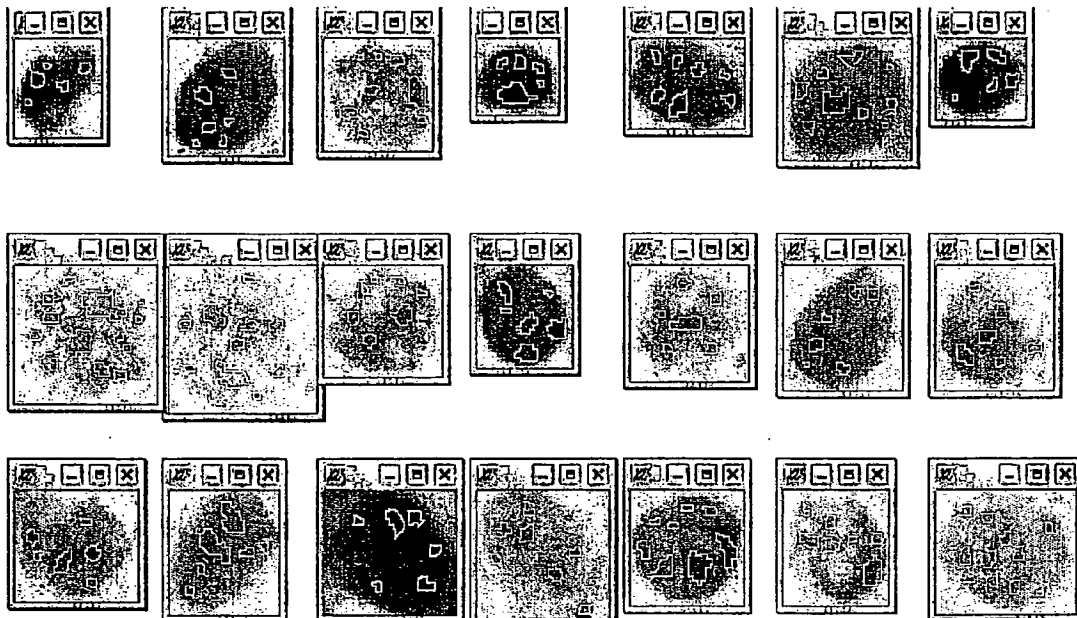


FIG 13

CHROMATIN SEGMENTATION

[0001] This invention is a method for segmenting (delineating) the chromatin in the nucleus of a cell. Chromatin is visualized by light microscopy as a patchwork of light and dark regions. The dark regions correspond to areas of high optical density and are blob-like or particle-like in appearance. The light regions correspond to areas of clearing and can also be interpreted as blobs or particles. In particular, though not exclusively, this invention is a method for segmenting these light and dark particles.

BACKGROUND TO THE INVENTION

[0002] The distribution of the chromosomes or DNA in the nuclei of cells can be quantitatively measured using a computer and image analysis techniques. Moreover, these measurements or features can be used to detect both malignancy associated changes (MACs), and changes during neoplasia. The features that appear to have the most discriminatory power are texture features. Such features quantitatively describe the intensity variation of the chromatin (the substance that constitutes the chromosomes or DNA and is readily visualized by staining) in the nucleus of a cell. The most widely used chromatin texture features are based on a statistical or probabilistic assessment of the gray-levels (intensity levels or optical density levels) in the cell nucleus. Unfortunately these features are difficult to relate to the terms and adjectives (e.g. granularity, compaction, margination, clumping, chromatin particles, condensation) used by cytologists to describe chromatin texture. Moreover they are usually defined at the pixel level and therefore fail to take into account the structural aspect of the chromatin distribution; e.g. this is true of all but the discrete texture features described in the United States Patent of Palcic et al. [System and method for automatically detecting malignant cells and cells having malignancy-associated changes; U.S. Pat. No. 6,026,174 dated Feb. 15, 2000]. Therefore their efficacy in relation to quantifying chromatin distribution is questionable.

[0003] An alternative approach to computing chromatin texture features is to segment the chromatin into aggregates and then to synthesize chromatin features from quantitative features computed for these aggregates. This approach has two advantages: (i) the segmentation step introduces structural information, and (ii) the synthesized features can be related to qualitative descriptions of chromatin texture made by cytologists. The key to this approach is the segmentation step. Several different methods of chromatin segmentation have been published in the literature. A characteristic they have in common is that they require the a priori specification of one or more operational parameters such as threshold values and region merging criteria. Moreover these parameters need to be tuned to the particular application. As a consequence these methods are not robust to changes in, or non-uniformity of, illumination and staining. The quality of the segmentations produced is arguably poor. This in turn affects the quality of the chromatin features that are computed from such segmentations. This is likely one of the major reasons that such features, with the exception of those of Young, Verbeek, and Mayall [Characterization of chromatin distribution in cell nuclei; Cytometry; vol. 7; 1986; pp. 467-474], have not found widespread use. Another possible reason is that the software implementation of the segmentation step is complicated.

[0004] Three existing methods of chromatin segmentation deserve special attention. The first is the method of Young, Verbeek, and Mayall because it is the basis of the discrete texture features detailed in the aforementioned United States Patent of Palcic et al. The second is the method of Wolf, Beil, and Guski [Chromatin structure analysis based on a hierarchic texture model; Analytical and Quantitative Cytology and Histology; vol. 17; no. 1; 1995; pp. 25-34] because it utilizes the watershed transform as does the preferred embodiment of the present invention. The third is the method of Kondo and Taniguchi [Evaluation of the chromatin for cell images; Systems and Computers in Japan; vol. 17; no. 9; 1986; pp. 11-19] because it represents the closest known prior art to the present invention.

[0005] The Young, Verbeek, and Mayall (YVM) method of chromatin segmentation takes as input a digitized image of a cell nucleus visualized by light microscopy. Young, Verbeek, and Mayall illustrate their method on images obtained from foam cells in human nipple aspirate fluid, and rat urothelial cells. The method of staining is unspecified. The YVM chromatin segmentation method involves nothing more than partitioning the gray-level histogram of the nucleus image into three parts (i.e. choosing two gray levels), and then using this division to label each nucleus pixel. The result is a segmentation comprising regions of low, medium, and high optical density. The manner in which the partition points (threshold values) are determined must be a priori specified. Although the YVM segmentation method is simple (hence its relative popularity) the quality of the segmentation is questionable for the following two reasons. Firstly, the method of segmentation utilizes only the intensity histogram and does not take into account any spatial information. Secondly, the method requires the specification of two threshold values. The manner in which these are chosen must be a priori specified. Moreover they must be tuned to the particular application.

[0006] The Wolf, Beil, and Guski (WBG) method of chromatin segmentation takes as input a digitized image of a cell nucleus, visualized by light microscopy, from a cervical tissue section obtained by colposcopic biopsy and stained by the Feulgen method. The WBG method consists of two steps. The first step involves determining the watershed of the gradient of the input image. This is done using a modification of the classic watershed algorithm of Vincent and Soille [Watersheds in digital spaces: an efficient algorithm based on immersion simulations; IEEE Transactions on Pattern Analysis and Machine Intelligence; vol. 13; no. 6; 1991; pp. 583-598]. The result is an oversegmentation; i.e. too many regions are delineated and as a consequence the result does not correspond very well to the chromatin patches in the original image. The second step involves selectively merging the regions segmented in the first step. Specifically, this step involves fitting a plane to each segmented region using standard least-squares techniques and then iteratively merging neighboring regions based on merging criteria related to the standard deviation of gray-levels in the regions. The decision to merge two regions is based on the evaluation of a single parameter that is then compared to a threshold value. This is a drawback of the method because the manner in which this threshold value is determined must be a priori specified.

[0007] The Kondo and Taniguchi (KT) method of chromatin segmentation represents the closest known prior art to

the present invention. The method takes as input a digitized image of a cell nucleus, visualized by light microscopy, from a Pap smear. The method comprises three steps: (i) local maxima (with respect to optical density) are located in the input image (these correspond to local minima of intensity), (ii) the input image is partitioned into sub-images (regions), each containing a single maximum, and (iii) a chromatin granule (densely stained blob of chromatin) is segmented from each region in turn using local adaptive thresholding. Kondo and Taniguchi propose three different methods for the partitioning step: (i) partitioning using a Voronoi neighborhood, (ii) region partitioning by directed tree, and (iii) area expansion by difference direction. A drawback of the Voronoi neighborhood method is that it does not use the topography of the input image to determine a region around each minimum. Consequently it is possible that the region determined around a minimum cuts through one or more adjacent chromatin particles. A drawback of the directed tree method is that it is necessary to a priori select a sensitivity parameter to control growth. A drawback of the density difference method is that the growth is not prescribed by geodesic distance (i.e. if the image is viewed as a landscape then the growth is not prescribed by the topography of the landscape). The local adaptive thresholding method of segmenting a granule from each region has several potential drawbacks including sensitivity to noise and non-uniform illumination, and the need to prescribe the manner in which the threshold value is determined.

[0008] The present invention is specifically designed for the purpose of segmenting chromatin particles in the nucleus of a cell. The method takes as input an image of the nucleus of a cell. Consequently the task of segmenting a cell from a field of cells and the task of segmenting the nucleus from a single cell, are not the subjects of this invention. Indeed details of these tasks are described in International Patent Application number PCT/AU01/00787 (WO 02/03331) and co-pending International Patent Application number PCT/AU99/00231 (WO 99/52074) respectively.

OBJECT OF THE INVENTION

[0009] It is an object of the present invention to provide a means of segmenting chromatin in a nucleus of a cell.

SUMMARY OF THE INVENTION

[0010] In one form, although it need not be the only, or indeed the broadest form, the invention resides in a method of segmenting chromatin particles in the nucleus of a cell including the following steps of:

- [0011] (i) locating regional minima in the image;
- [0012] (ii) computing a zone of influence (ZOI) around each regional minimum; and
- [0013] (iii) segmenting a single chromatin blob within each ZOI using a region growing procedure.

[0014] The input image is preferably a two-dimensional gray-scale image comprising only those pixels that define the nucleus of a cell. It will be appreciated that the method is not limited to two-dimensional gray-scale images. The method can be applied to three-dimensional images in which case regional minima will be sets of voxels rather than pixels. The method can also be applied to multi-valued (including multispectral) images.

[0015] The method may further include the step of evaluating the contrast of each regional minimum and discarding those regional minima that do not satisfy a priori specified contrast criteria.

[0016] The method may also include a preliminary step of pre-processing the input image to correct for degradations. The pre-processing step suitably removes degradations such as noise and blurring. The pre-processed image may optionally be up-sampled.

[0017] The regional minima are suitably regions of constant gray-value that are surrounded by pixels of strictly higher (lighter) gray-value. Each regional minimum identifies the location of a dark blob.

[0018] The method may also be applied to the photographic negative of an image to identify the locations of light blobs. This is equivalent to identifying regional maxima in the original (positive) image.

[0019] In preference, the step of computing a zone of influence is performed by means of seeded region growing, or catchment basins determined by the watershed transform, or influence zones (IZs) with respect to an a priori specified metric.

[0020] The segmentation step is suitably performed by means of either a watershed transform or a seeded region growing algorithm.

[0021] In a further form, the invention resides in a method of quantitatively characterizing the structure of nuclear chromatin including the steps of:

- [0022] (i) obtaining an image of a cell nucleus showing chromatin texture;
- [0023] (ii) locating regional minima in the image;
- [0024] (iii) computing a zone of influence (ZOI) around each regional minimum;
- [0025] (iv) segmenting a single chromatin blob within each ZOI using a region growing procedure; and
- [0026] (v) computing features for individual chromatin blobs.

[0027] The method may further include the step of:

- [0028] (vi) synthesizing chromatin features from the features computed for individual chromatin blobs.

[0029] The method may suitably include the further step of repeating steps (ii) to (v) for the negative of the image of step (i) and using the computed features in step (vi).

[0030] The method may further include the step of evaluating the contrast of each regional minimum in the image of the cell nucleus and discarding those regional minima that do not satisfy a priori specified contrast criteria.

[0031] The image of a cell nucleus may be a suitable gray-scale image.

BRIEF DESCRIPTION OF THE DRAWINGS

[0032] The preferred embodiment of the invention is described with reference to the following figures in which:

[0033] **FIG. 1** is a flowchart showing the steps of the invention;

[0034] FIG. 2 is an example input to the present invention;

[0035] FIG. 3 shows the result after the application of a 3×3 median filter to FIG. 2 and then up-sampling by factor 2;

[0036] FIG. 4 shows the location of the regional minima of FIG. 3 (depicted as small white regions or connected components);

[0037] FIG. 5 shows the watershed lines (white lines) produced by the application of the watershed transform to the image in FIG. 3 using the white regions of FIG. 4 as markers;

[0038] FIG. 6 shows the watershed lines (white lines) produced by the application of the watershed transform to the morphological gradient of the image in FIG. 3 using the white regions of FIG. 4 and the white lines of FIG. 5 as markers. The white lines in this figure delineate the chromatin particles defined by the topography of the gradient image;

[0039] FIG. 7 shows the boundary lines (white lines) produced by the application of a seeded region growing algorithm to the image in FIG. 3 using the white regions of FIG. 4 and the white lines of FIG. 5 as seeds. The white lines in this figure delineate chromatin particles defined by the homogeneity criteria of the seeded region growing algorithm (in this case the homogeneity of gray values);

[0040] FIG. 8 shows the negative of the image in FIG. 2;

[0041] FIG. 9 shows the segmentation produced when FIG. 8, rather than FIG. 2, is used as the input to the preferred embodiment of the present invention;

[0042] FIG. 10 shows examples of the intermediate images used to compute the preferred features of a particle;

[0043] FIG. 11 shows a Delaunay graph for the dark particles of FIG. 10;

[0044] FIG. 12 shows a cytology slide, a magnified image of a particular field on the slide, and a magnified image of one of the nuclei from this field of view; and

[0045] FIG. 13 shows a small gallery of nuclei for which the dark particles have been segmented.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0046] The preferred embodiment of the present invention is similar to the KT method in the sense that it involves the steps of locating minima in the intensity image, partitioning the image into regions each containing a single minimum, and then segmenting a single blob in each region. In contrast to the KT method the present invention locates regional minima (with respect to intensity) rather than local minima. A regional minimum is a region or connected component of constant gray-value that is surrounded by pixels of strictly higher (lighter) gray-value, whereas a local minimum is a pixel whose neighbors are of strictly higher gray-value. In contrast to the KT method the present invention, in its preferred embodiment, grows these minima using the watershed transform. If the input image is viewed as a topographic landscape (with the light pixels representing high points and the dark pixels -representing low points), then the boundaries of the regions determined by the watershed transform

are guaranteed to lie between the minima as determined by the topography of the surface. In contrast to the KT method the present invention, in its preferred embodiment, segments a single blob in each region using the watershed transform.

[0047] FIG. 1 is a flowchart of the steps in the preferred embodiment of the present invention. The input image is denoted by the letter I on the flowchart. It is a gray-scale image in which the chromatin is visualized as a patchwork of light and dark regions. The darker the region, the more optically dense is the chromatin. FIG. 2 is an example of such an image. The white area surrounding the gray area is not considered to be part of the image. The first optional step is pre-processing. Depending on the quality of I, this step may include filtering to attenuate noise, deconvolution to correct for blur, and up-sampling to facilitate line rendering in subsequent steps. The preferred method of pre-processing involves the application of a 3×3 median filter [Gonzalez & Woods; Digital image processing; Addison-Wesley; Reading, Mass.; ISBN: 0-201-50803-6; 1992; pp. 191-195] followed by up-sampling by factor 2. The preferred method of up-sampling by factor 2 involves nothing more than replacing each pixel with a 2×2 block of pixels of the same gray-value. FIG. 3 shows the result after applying the preferred method of pre-processing to the image I of FIG. 2. The pre-processed image is denoted I' on the flowchart.

[0048] The next step is to locate the regional minima in the pre-processed image I'. A regional minimum is a region or connected component, comprising one or more pixels of constant gray value, that is surrounded by pixels of strictly higher (lighter) gray value. The output from this step is a binary image M containing connected components, each of which marks the location of a regional minimum. Several methods for detecting regional minima exist in the literature; e.g. the grayscale reconstruction method of Vincent [Morphological grayscale reconstruction in image analysis: applications and efficient algorithms; IEEE Transactions on Image Processing; vol. 2; no. 2; April 1993; pp. 184-185]. The present invention does not require that any particular method be employed. FIG. 4 shows the location of the regional minima of the pre-processed image I' of FIG. 3 (the image M consists of the small white connected components).

[0049] As mentioned earlier, the preferred embodiment is described by reference to two-dimensional gray-scale images. It will be appreciated that the process may be generalized to locate sets of voxels in a three-dimensional image rather than sets of pixels in a two-dimensional image.

[0050] The next optional step is to filter the image M according to a priori specified contrast criteria. This involves computing a contrast valuation, with respect to the pre-processed image I', for each regional minimum, and then discarding those minima that do not satisfy the contrast criteria. The filtered image is denoted M' on the flowchart. Two particularly useful contrast measures are dynamics devised by Grimaud [A new measure of contrast: the dynamics; Proceedings of the SPIE—The International Society for Optical Engineering; vol. 1769; 1992; pp. 292-305] and symmetrical dynamics devised by Vachier and Vincent [Valuation of image extrema using alternating filters by reconstruction; Proceedings of the SPIE—The International Society for Optical Engineering; vol. 2568; 1995; pp. 94-103]. The preferred measure of contrast is dynamics. The

preferred contrast criterion is to select all regional minima with a dynamics value of 1 or greater. Effectively this preferred criterion leads to all of the connected components of M being retained; i.e. $M'=M$. Whilst from a segmentation point of view the preferred contrast criterion has no effect on the final segmentation, the computed dynamics values do provide a set of valuable features from which chromatin features can subsequently be computed (an intended application of the present invention).

[0051] The next step is to compute a zone of influence (ZOI) around each of the connected components in M' . The image containing the ZOIs is denoted by Z on the flowchart. Depending on the method of implementation, Z may be either a binary image of lines delineating the zones of influence or it may be a gray-scale image in which each ZOI has its own unique numerical label. The preferred method for computing Z is to apply the watershed transform [Beucher & Meyer; *The morphological approach to segmentation: the watershed transformation*; in *Mathematical morphology in image processing*; E. R. Dougherty (Ed.); chapter 12; Marcel Dekker, Inc., New York; ISBN: 0-8247-8724-2; 1993; pp. 433-481] to the pre-processed image I' using the connected components of M' as markers. Another possibility is to apply a scan-order-independent seeded region growing algorithm—such as that of the inventors, Mehnert and Jackway [An improved seed region growing algorithm; *Pattern Recognition Letters*; vol. 18; 1997; pp. 1065-1071] or that of Beare and Talbot [Exact seeded region growing for image segmentation; *Proceedings of the Fifth International/National Biennial Conference on Digital Image Computing, Techniques, and Applications (DICTA'99)*; 7-8 December, Perth, Western Australia; 1999; pp. 132-137]—the pre-processed image I' using the connected components of M' as seeds. Yet another possibility is to determine the influence zone (IZ) around each connected component of M' using an a priori specified metric. The IZ of a connected component is the set of all pixels that are closer to it than to any other connected component. FIG. 5 shows the result (white lines) of the preferred method; i.e. the application of the watershed transform to image I' of FIG. 3 using the connected components M' of FIG. 4 as markers.

[0052] The next step is to segment a single blob (chromatin particle) within each ZOI in Z . This is done using a region growing procedure. The preferred method for doing this is to apply the watershed transform to the modulus of the gradient of the pre-processed image I' using both the connected components of M' and the boundary lines of Z as markers (these boundary lines may or may not include the boundary of the entire nucleus itself; the preferred method is not to include this). Numerous methods for computing the modulus of the gradient exist in the literature. The preferred method is the discrete morphological gradient [Rivest, Soille, & Beucher; *Morphological gradients*; *Journal of Electronic Imaging*; vol. 2; no. 4; 1993; pp. 326-336]. Another possibility for segmenting the chromatin blobs is to apply a scan-order-independent seeded region growing algorithm to the pre-processed image I' using both the connected components of M' and the boundary lines of Z as seeds (once again these boundary lines may or may not include the boundary of the entire nucleus itself). FIG. 6 shows the result of the preferred method; i.e. the application of the watershed transform to the morphological gradient of the image I' of FIG. 3 using the connected components of M'

shown in FIG. 4 and the watershed lines (boundary lines of Z) shown in FIG. 5 as markers. FIG. 7 shows the result of the alternative method; i.e. the application of seeded region growing to the image I' of FIG. 3 using the connected components of M' shown in FIG. 4 and the watershed lines (boundary lines of Z) shown in FIG. 5 as seeds.

[0053] If the input image I is replaced with its photographic negative (for an 8-bit gray-scale image with gray-values ranging from 0 to 255 this amounts to subtracting each pixel's value from 255) then the (dark) particles segmented by the present invention will correspond to light particles in the original positive image. FIG. 8 shows the photographic negative of the image in FIG. 2. FIG. 9 shows the segmentation produced by the preferred embodiment of the present invention when FIG. 8 is used as the input.

[0054] Although the preceding discussion has been based upon two-dimensional gray-scale images, the inventors envisage that the invention can be applied to higher dimensional images, and to multispectral and other multi-valued images. For example, the invention can be applied to color images for which each pixel has multiple components. The nature of these components is dictated by the underlying color model such as HLS (hue, luminance, saturation), YIQ (luminance, in-phase, quadrature), CMY (cyan, magenta, yellow), and RGB (red, green, Blue). HLS, YIQ, CMY, and RGB are all multi-value images. The seeded region growing algorithms aforementioned readily generalize to three-dimensional gray-scale images and to color images. A three-dimensional version of the watershed transform also exists [Cotsaces and Pitas; *Computing the watersheds of large three-dimensional images using limited random access memory*; in *Mathematical Morphology and its Applications to Image and Signal Processing*; Dordrecht, The Netherlands; Kluwer Academic Publishers; ISBN 0-7923-5133-9; 1998; pp. 239-246].

[0055] A characteristic of all of the known chromatin segmentation methods is that they require the a priori specification of one or more operational parameters such as threshold values and region merging criteria. The present invention differs in that it is, in its preferred embodiment, a parameter-free method of segmenting chromatin particles. In addition, in comparison to existing methods, the present invention produces a discernibly better segmentation of nuclear chromatin; i.e. for any given nucleus image the method yields a segmentation of chromatin particles that corresponds well with what a human observer might intuitively perceive to be blobs or particles.

[0056] Persons skilled in the relevant art may realize variations from the preferred embodiment that will nonetheless fall within the scope of this invention. One such variation, for example, is to determine the ZOIs without rendering boundary lines, and then to segment a blob in each ZOI in turn (sequentially and independently).

[0057] The inventors believe that the most important application of the present invention is to the development and computation of features that quantitatively characterize the structure of nuclear chromatin. In particular these features can be synthesized from features computed for the individual chromatin particles including area, perimeter, volume, surface area, average gray-level, circularity, dynamics, number of neighbors, average distance to neighbors, and distance to nuclear boundary. In turn the inventors

believe that such features have two very important applications. The first is that such features can be used to detect changes during neoplasia, and malignancy associated changes. The second is that such features can be used to select cell nuclei with sufficient texture for subsequent analysis using standard texture features (such as those described in the aforementioned patent of Palcic et al.) and/or features computed from the current invention.

[0058] The following discussion describes the computation of particle features and the synthesis of nucleus features and slide features. The discussion follows the gray-scale example used above.

Particle Features

[0059] In the preferred embodiment of the invention, features are computed for both the dark and light particles of a nucleus image. The former particles are found from a segmentation (using the procedure already described) of the nucleus image. The latter are found from a segmentation of the photographic negative of the nucleus image. FIG. 10 shows examples of the intermediate images used to compute the preferred features of a particle. FIG. 10 is not to scale. The reduction in scale can be gauged by comparing the top left image of FIG. 10 with FIG. 3. Each particle (an image object or region) is a distinct set of gray-level pixels in the nucleus image. The cardinality of each particle can be represented by a single connected component (binary mask). Numerous features can be found in the literature to characterize the size, shape, boundary and texture of image objects. Compendia of such features can be found in [Grohs & Husain (Eds.); Automated cervical cancer screening; IGAKU-SHOIN Medical Publishers; New York; ISBN: 0-89640-255-X; 1994; pp. 24-38] and in the aforementioned United States Patent of Palcic et al. Although these compendia describe features for the purpose of characterizing cell nuclei, the features can be applied more generally to any image object, including nuclear particles segmented by the present invention. In the preferred embodiment of the present invention, the following features are computed for each particle:

[0060] Morphometric Features

[0061] These features characterize the size, shape, and boundary of a connected component. In the preferred embodiment of the present invention, the following morphometric features are computed from the binary mask of a particle.

[0062] 1. Area

[0063] This is defined as the number of pixels belonging to the particle; i.e. the number of pixels in the binary mask of the particle.

[0064] 2. Perimeter

[0065] This is computed using the digital version of Crofton's formula [Serra; Image Analysis and Mathematical Morphology; Volume 1; Academic Press; London; ISBN 0-12-637240-3; 1982; pp. 220-223].

[0066] 3. G factor

[0067] This is computed using the method described by Danielsson [A new shape factor; Computer Graphics and Image Processing; vol. 7; pp. 292-299; 1978].

[0068] Densitometric Features

[0069] These features characterize the gray-level (optical density or intensity) variation of the pixels in an image object. These features characterize gray-level variation only (they do not take into account the positions of the pixels). In the preferred embodiment of the present invention, the image I' together with the binary mask of a particle are used to compute the following densitometric features.

[0070] 1. Volume

[0071] This is computed from the set of gray pixels in I' that correspond to the binary mask of the particle. It is defined as the sum of these gray values.

[0072] 2. Mean Gray Value

[0073] This is computed from the set of gray pixels in I' that correspond to the binary mask of the particle. It is defined as the mean of these gray values.

[0074] 3. Dynamics

[0075] This is defined to be the dynamics value of the regional minimum associated with the particle (the value determined during segmentation).

[0076] Texture Features

[0077] These features characterize the gray-level (optical density or intensity) variation of the pixels in a gray-level image object, taking into account the position of the pixels. In the preferred embodiment of the present invention, the image I' together with the binary mask of a particle are used to compute the following texture feature:

[0078] 1. Surface Area

[0079] This is computed from the set of gray pixels in the discrete morphological gradient of I' that correspond to the binary mask of the particle. It is defined as the sum of these gray values [Rivest & Soille; Physical significance of image measurements; IEEE Transactions on Instrumentation and Measurement; vol. 44; no. 3; June 1995; pp. 751-754].

[0080] Contextual Features

[0081] These features characterize relationships between one or more image objects. In the preferred embodiment of the present invention, relationships between a nucleus and its particles, and relationships between the particles themselves are characterized. With respect to the first case the following feature is computed:

[0082] 1. Distance to Nucleus Boundary

[0083] This is computed from the set of gray pixels in the Euclidean distance transform [Mehner & Jackway; On computing the exact Euclidean distance transform on rectangular and hexagonal grids; Journal of Mathematical Imaging and Vision; vol. 11; 1999; pp. 223-230] of the binary mask of the nucleus that correspond to the binary mask of the particle. It is defined as the mean of these gray values.

[0084] With respect to the second case, contextual features are computed from (i) a neighborhood graph defined on the dark particles; (ii) a neighborhood graph defined on the light particles; and (iii) a neighborhood graph defined on both the light and dark particles. Vincent [Graphs and mathematical

morphology; Signal Processing; vol. 16; 1989; pp. 365-388] describes and presents algorithms for a range of such neighborhood graphs. The preferred type of neighborhood graph is the Delaunay graph. The Delaunay graph for the dark particles of FIG. 10 is shown in FIG. 11. In the preferred embodiment of the invention, the following graph-based features are computed for particles:

[0085] 2. Number of Dark Particle Neighbors a Dark Particle Has

[0086] This is determined from the neighborhood graph defined on the dark particles; e.g. it can be seen in FIG. 11 that the dark particle shown as a binary mask in FIG. 10 has 9 dark particle neighbors.

[0087] 3. Mean Distance Between a Dark Particle and its Dark Particle Neighbors

[0088] This is determined from the neighborhood graph defined on the dark particles. The distance is the centroid-to-centroid Euclidean distance; e.g. this distance is the length of a graph edge connecting 2 particles in FIG. 11.

[0089] 4. Number of Light Particle Neighbors a Light Particle Has

[0090] This is determined from the neighborhood graph defined on the light particles.

[0091] 5. Mean Distance Between a Light Particle and its Light Particle Neighbors

[0092] This is determined from the neighborhood graph defined on the light particles. The distance is the centroid-to-centroid Euclidean distance.

[0093] 6. Number of Dark Particle Neighbors

[0094] This is determined from the neighborhood graph defined on both the light and dark particles.

[0095] 7. Mean Distance to Dark Particles

[0096] This is determined from the neighborhood graph defined on both the light and dark particles. The distance is the centroid-to-centroid Euclidean distance.

[0097] 8. Number of Light Particle Neighbors

[0098] This is determined from the neighborhood graph defined on both the light and dark particles.

[0099] 9. Mean Distance to Light Particles

[0100] This is determined from the neighborhood graph defined on both the light and dark particles. The distance is the centroid-to-centroid Euclidean distance.

[0101] 10. Number of Particle Neighbors

[0102] This is determined from the neighborhood graph defined on both the light and dark particles.

[0103] 11. Mean Distance to Particle Neighbors

[0104] This is determined from the neighborhood graph defined on both the light and dark particles. The distance is the centroid-to-centroid Euclidean distance.

Nucleus Features

[0105] The particles segmented from a nucleus can be used to compute a variety of nucleus features. One set of such features is obtained from summary statistics (including moments and order statistics) computed for each particle feature; e.g. mean, median, mode, variance, interquartile range, skewness, and kurtosis of particle area. These summary statistics can be computed for the dark particles only, the light particles only, and for both types of particle. In the preferred embodiment of the present invention, the following features are computed:

[0106] 1. mean of each particle feature except dynamics;

[0107] 2. standard deviation of each particle feature except dynamics;

[0108] 3. median dynamics;

[0109] 4. interquartile range of dynamics;

[0110] 5. number of dark particles; and

[0111] 6. number of light particles.

[0112] Another set of nucleus features can be computed from (i) the histogram of gray-values in the distance transform of the binary mask representing the nucleus background between the dark particles (see FIG. 10); (ii) the histogram of gray-values in the distance transform of the binary mask representing the nucleus background between the light particles; and (iii) the histogram of gray-values in the distance transform of the binary mask representing the nucleus background between both the light and dark particles. Summary statistics of these histograms yield nucleus features. Alternatively, the method of Russ [The image processing handbook; Florida; CRC Press; ISBN 0849342333; 1992; p. 336] can be used to compute a feature for each histogram that characterizes the particle clustering. In the preferred embodiment of the invention, the following features are computed:

[0113] 7. mean of the histogram obtained from the Euclidean distance transform of the nucleus background between the dark particles;

[0114] 8. standard deviation of the histogram obtained from the Euclidean distance transform of the nucleus background between the dark particles;

[0115] 9. mean of the histogram obtained from the Euclidean distance transform of the nucleus background between the light particles;

[0116] 10. standard deviation of the histogram obtained from the Euclidean distance transform of the nucleus background between the light particles;

[0117] 11. mean of the histogram obtained from the Euclidean distance transform of the nucleus background between both the light and dark particles; and

[0118] 12. standard deviation of the histogram obtained from the Euclidean distance transform of the nucleus background between both the light and dark particles.

[0119] Several additional sets of nucleus features can be computed from the 3 neighborhood graphs aforementioned (one defined on the dark particles only, one defined on the light particles only, and one defined on both the light and

dark particles). For a given graph, a co-occurrence matrix can be defined for each particle feature as described in [Toriwaki & Yokoi; Voronoi and related neighbors on digitized two-dimensional space with applications to texture analysis; in Computational morphology: A computational geometric approach to the analysis of form; Toussaint (Ed.); Elsevier Science Publishers; Amsterdam; ISBN: 0444704671; 1988; pp. 207-228]. For example, from the histogram of dark particle areas and the neighborhood graph defined on these particles, it is possible to construct a matrix such that the entry in the *i*-th row and the *j*-th column represents the number of times a particle of area *i* is adjacent to a particle of area *j*. To keep the matrix size manageable or to avoid having a sparse matrix the number of bins in the histogram of the feature under consideration can be reduced. For each co-occurrence matrix, co-occurrence matrix features can be computed and used as nucleus features. A compendium of co-occurrence matrix features can be found in [Walker; Adaptive multi-scale texture analysis: with application to automated cytology; PhD Thesis; School of Computer Science and Electrical Engineering; The University of Queensland; Australia; 1997; Chapter 2]. In the preferred embodiment of the invention the following features are computed:

[0120] 13. contrast and

[0121] 14. energy computed from the co-occurrence matrix obtained from the neighborhood graph defined on the dark particles, using the particle feature area. To keep the matrix manageable, the feature area is binned into two classes: (i) small containing particles with an area less than $\frac{1}{6}$ of the nucleus area and (ii) large containing all other particles.

[0122] Slide Features Derived From Nucleus Features

[0123] FIG. 12 (not to scale) shows a cytology slide, an image of a single field of view from the slide taken by a CCD camera mounted on a microscope, and one of the nuclei from the field of view. Features can be synthesized for a slide (be it a cytology slide or a histology slide) by computing summary statistics (including moments and order statistics) for each feature for all or a subset of nuclei obtained from the slide. In the preferred embodiment of the invention the mean and standard deviation are computed for each nucleus feature from a statistically representative number of nuclei extracted from a slide.

[0124] Slide Features Derived From particle Features

[0125] Slide features can also be synthesized by computing summary statistics (including moments and order statistics) directly from particle features computed for all or a subset of nuclei extracted from a slide. In the preferred embodiment of the invention the mean and standard deviation are computed for each particle feature from a statistically representative number of nuclei extracted from a slide.

[0126] A person skilled in the relevant art may realize possible variations to the described process that will nonetheless fall within the scope of this invention. One such variation is to use different estimators than those described to compute features such as perimeter, e.g. particle perimeter could be computed using the method described in the aforementioned patent of Palcic et al. for computing the perimeter of a nucleus. Another variation is to employ a different type of neighborhood graph (e.g. a Gabriel graph)

for the purpose of computing features. Another variation is to compute graph-based features using *k*-adjacency rather than just 1-adjacency; i.e. 2 particles are considered to be *k*-neighbors if a sequence of *k* edges connects them in the neighborhood graph. Another variation is that a region adjacency graph can be defined on the zones of influence image *Z*, computed during segmentation, and that this be used in place of the neighborhood graph for the purpose of computing features. Another variation is that nucleus features can be computed from the set of lengths of shared borders between pairs of zones of influence. Another variation is that nucleus features can be computed from the histogram of centroid-to-centroid distances for pairs of neighboring particles (as defined by a neighborhood graph). Another variation is that the segmentation of the nucleus into dark particles and light particles creates a partition of the nucleus pixels into 3 sets and that these 3 sets can be used to compute some of the features that Young, Verbeek, and Mayall compute for their method of chromatin segmentation described earlier (note that the exact meaning and value of these will be completely different for the present invention). Another variation is that nucleus features can be computed from histograms of particle features for particles that are adjacent to the nucleus boundary. Another variation is to compute particle features directly from the nucleus image *I* rather than *I'* (in which case the particle masks must be down-sampled or the image *I* up-sampled). Another variation is that features can be normalized; e.g. particle area can be expressed as a percentage of the nucleus area. Another variation is that distances other than Euclidean distance can be used to compute features; e.g. chessboard distance, city-block distance, 5-7-chamfer distance [Heijmans; Morphological image operators; Academic Press; San Diego; ISBN: 0-12-014599-5; 1994; pp. 325-331].

[0127] It is an intention of the inventors that the features computed using the present invention be used in a computerized system (an automated cytometer and classifier) for the purpose of detecting neoplasia and malignancy associated changes in cells and tissue.

EXAMPLE

[0128] The present invention has been applied to hundreds of thousands of images of cell nuclei, or objects that resemble nuclei, obtained from 148 cytology slides using an automated cytometer. FIG. 13 shows a small gallery of nuclei for which the dark particles have been segmented by the invention. Each slide was prepared from a cervical swab taken from a patient. The set of slides include only one slide per patient. The slides were prepared using a liquid-based preparation technique, and stained using the Papanicolaou staining protocol. Each slide has been classified by a pathologist as either negative (showing no signs of cervical intraepithelial neoplasia (CIN)), or positive (showing signs of CIN). In the latter case the pathologist has indicated the relative degree of abnormality: CIN 1, CIN 2, or CIN 3. Of the 148 slides, 101 are negative. The methods used by the cytometer to obtain the images of nuclei, or objects resembling nuclei, are documented in PCT Patent Application number PCT/AU01/00787 and PCT Patent Application number PCT/AU99/00231.

[0129] The cytometer was programmed to extract images of at most 10000 objects resembling nuclei (in terms of size, shape, and intensity) from each slide. A large number of

objects found by the cytometer are artifacts such as blood, mucous, dust, poorly segmented nuclei, and out-of-focus images of nuclei. The invention was used to perform artifact rejection as follows. All objects that contained fewer than 5 dark particles, or that contained very large dark particles (>1/3 of the area of the nucleus) were considered to be artifacts and rejected (too few dark particles, or extremely large dark particles, is indicative of poorly segmented or poorly focused nuclei, or an image artifact such as dust, blood, or mucous).

[0130] The following features were computed and recorded for each nucleus:

[0131] (Note: sdev means standard deviation in the following tables)

Derived directly from the binary mask of the nucleus

nucleus_area
nucleus_perimeter

[0132]

Derived from the invention

nucleus_mean_dark_particle_area
nucleus_sdev_dark_particle_area
nucleus_mean_dark_particle_perimeter
nucleus_sdev_dark_particle_perimeter
nucleus_median_dark_particle_dynamics
nucleus_IQR_dark_particle_dynamics
nucleus_mean_dark_particle_distance_to_nucleus_boundary
nucleus_sdev_dark_particle_distance_to_nucleus_boundary
nucleus_total_dark_particle_area
nucleus_number_of_dark_particles
ratio_of_total_dark_particle_area_to_nucleus_area

[0133] Means and standard deviations of these nucleus features yielded the following slide features:

Slide features derived from nucleus features

slide_mean_of_nucleus_area
slide_sdev_of_nucleus_area
slide_mean_of_nucleus_perimeter
slide_sdev_of_nucleus_perimeter
slide_mean_of_nucleus_mean_dark_particle_area
slide_sdev_of_nucleus_mean_dark_particle_area
slide_mean_of_nucleus_sdev_dark_particle_area
slide_sdev_of_nucleus_sdev_dark_particle_area
slide_mean_of_nucleus_mean_dark_particle_perimeter
slide_sdev_of_nucleus_mean_dark_particle_perimeter
slide_mean_of_nucleus_sdev_dark_particle_perimeter
slide_sdev_of_nucleus_sdev_dark_particle_perimeter
slide_mean_of_nucleus_median_dark_particle_dynamics
slide_sdev_of_nucleus_median_dark_particle_dynamics
slide_mean_of_nucleus_IQR_dark_particle_dynamics
slide_sdev_of_nucleus_IQR_dark_particle_dynamics
slide_mean_of_nucleus_mean_dark_particle_distance_to_nucleus_boundary
slide_sdev_of_nucleus_mean_dark_particle_distance_to_nucleus_boundary
slide_mean_of_nucleus_sdev_dark_particle_distance_to_nucleus_boundary
slide_sdev_of_nucleus_sdev_dark_particle_distance_to_nucleus_boundary

-continued

Slide features derived from nucleus features

slide_sdev_of_nucleus_sdev_dark_particle_distance_to_nucleus_boundary
slide_mean_of_nucleus_total_dark_particle_area
slide_sdev_of_nucleus_total_dark_particle_area
slide_mean_of_nucleus_number_of_dark_particles
slide_sdev_of_nucleus_number_of_dark_particles
slide_mean_of_ratio_of_total_dark_particle_area_to_nucleus_area
slide_sdev_of_ratio_of_total_dark_particle_area_to_nucleus_area

[0134] These 26 slide features, along with the pathologist classifications of negative and positive, were used to train and test a 2-class statistical classifier with the R statistical analysis software [R is free software from the GNU project www.gnu.org; Homepage is http://www.R-project.org/]. The training phase included feature selection (i.e. determining which features are useful for discriminating between the 2 classes). The training was done on a subset of the feature data made up of 70% of the negatives and 70% of the positives chosen at random. The remaining positives and negatives were used to test the classifier. The performance of the classifier was recorded in terms of the area under the receiver operating characteristic (ROC) curve. A classifier for which the area under the ROC curve (AUC) is 0.5 classifies randomly. A classifier for which the AUC is 1 is a perfect classifier. The train/test procedure was independently repeated 100 times (each time randomly splitting the positives and the negatives in the ratio 7:3 for the training and testing data respectively). From the 100 different trials, the following 7 features were selected more than 70% of the time:

- [0135] 1. slide_mean_of_nucleus_number_of_dark_particles
- [0136] 2. slide_mean_of_nucleus_perimeter
- [0137] 3. slide_mean_of_nucleus_mean_dark_particle_area
- [0138] 4. slide_sdev_of_nucleus_mean_dark_particle_area
- [0139] 5. slide_sdev_of_nucleus_sdev_dark_particle_area
- [0140] 6. slide_sdev_of_nucleus_mean_dark_particle_distance_to_nucleus_boundary
- [0141] 7. slide_sdev_of_nucleus_sdev_dark_particle_distance_to_nucleus_boundary

[0142] From another 100 trials with only the 7 features above, the following 4 features were selected more than 75% of the time:

- [0143] 1. slide_mean_of_nucleus_number_of_dark_particles
- [0144] 2. slide_mean_of_nucleus_perimeter
- [0145] 3. slide_mean_of_nucleus_mean_dark_particle_area
- [0146] 4. slide_sdev_of_nucleus_mean_dark_particle_distance_to_nucleus_boundary

[0147] With these 4 features the following estimate of the area under the ROC curve was obtained:

78%±0.06 (mean±standard deviation)

[0148] This indicates that features that summarize the size and distance to the nuclear boundary of the dark particles (condensed chromatin) in the nucleus are good features for detecting MACs. Moreover these features correspond well to descriptive terms like clumping and margination used by cytologists to describe chromatin texture.

[0149] Throughout the preceding specification the aim has been to describe the preferred embodiment of the present invention without limiting the invention to any one embodiment.

1. A method of segmenting chromatin particles in the nucleus of a cell including the following steps of:

- (i) locating regional minima in an image;
- (ii) computing a zone of influence (ZOI) around each regional minimum; and
- (iii) segmenting a single chromatin blob within each ZOI using a region growing procedure.

2. The method of claim 1 wherein the image is a two-dimensional gray-scale image comprising only those pixels that define the nucleus of a cell.

3. The method of claim 1 wherein the image is a three-dimensional image.

4. The method of claim 1 wherein the image is a multi-valued image.

5. The method of claim 1 further including the step of evaluating the contrast of each regional minimum and discarding those regional minima that do not satisfy a priori specified contrast criteria.

6. The method of claim 1 further including a preliminary step of pre-processing the image to correct for degradations.

7. The method of claim 6 wherein the pre-processing step removes degradations such as noise and blurring.

8. The method of claim 6 wherein the pre-processed image is up-sampled.

9. The method of claim 1 wherein the regional minima are regions of constant gray-value that are surrounded by pixels of strictly higher (lighter) gray-value, to identify dark blobs.

10. The method of claim 1 wherein the step of computing a zone of influence is performed by means of region growing or the computation of influence zones (IZs).

11. The method of claim 1 wherein the segmentation step is performed by means of one of a watershed transform or a seeded region growing algorithm.

12. A method of quantitatively characterizing the distribution of nuclear chromatin including the steps of:

- (i) obtaining an image of a cell nucleus showing chromatin texture;
- (ii) locating regional minima in the image;
- (iii) computing a zone of influence (ZOI) around each regional minimum;
- (iv) segmenting a single chromatin blob within each ZOI using a region growing procedure; and
- (v) computing features for individual chromatin blobs.

13. The method of claim 12 further including the step of:

- (vi) synthesizing chromatin texture features from the features computed for individual chromatin blob features.

14. The method of claim 13 including the further step of repeating steps (ii) to (v) for the negative of the image of step (i) and using the computed features in step (vi).

15. The method of claim 12 further including the step of evaluating the contrast of each regional minimum in the image of the cell nucleus and discarding those regional minima that do not satisfy a priori specified contrast criteria.

16. The method of claim 12 wherein the image of the cell nucleus is a gray-scale image.

17. The method of claim 12 wherein the computed features are selected from morphometric features, densitometric features, texture features, and contextual features.

18. The method of claim 17 wherein the morphometric features include one or more of: area; perimeter; and G factor.

19. The method of claim 17 wherein the densitometric features include one or more of: volume; mean gray value; and dynamics.

20. The method of claim 17 wherein the contextual features include one or more of: distance to nucleus boundary; number of dark particle neighbors a dark particle has; mean distance between a dark particle and its dark particle neighbors; number of light particle neighbors a light particle has; mean distance between a light particle and its light particle neighbors; number of dark particle neighbors; mean distance to dark particles; number of light particle neighbors; mean distance to light particles; number of particle neighbors; and mean distance to particle neighbors.

21. The method of claim 17 wherein the synthesized features are selected from nucleus features, slide features derived from nucleus features, and slide features derived from particle features.

22. The method of claim 21 wherein the nucleus features include one or more of: mean of each particle feature except dynamics; standard deviation of each particle feature except dynamics; median dynamics; interquartile range of dynamics; number of dark particles; number of light particles; mean of the histogram obtained from the Euclidean distance transform of the nucleus background between the dark particles; standard deviation of the histogram obtained from the Euclidean distance transform of the nucleus background between the dark particles; mean of the histogram obtained from the Euclidean distance transform of the nucleus background between the light particles; standard deviation of the histogram obtained from the Euclidean distance transform of the nucleus background between the light particles; mean of the histogram obtained from the Euclidean distance transform of the nucleus background between both the light and dark particles; standard deviation of the histogram obtained from the Euclidean distance transform of the nucleus background between both the light and dark particles; contrast; and energy computed from the co-occurrence matrix obtained from the neighborhood graph defined on the dark particles, using the particle feature area.

* * * * *