

Abstract:

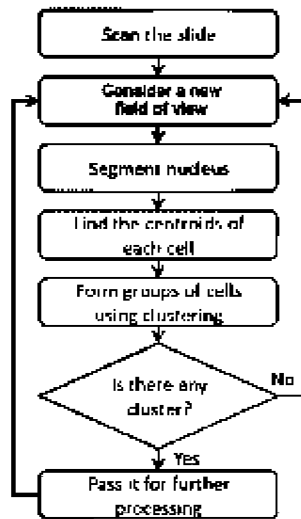


Fig. 1

(EN)The present invention is related to a method or system for the detection of abnormalities in a biological sample comprising cells. Said method or system comprises at least the following steps, or is capable to carry out, and/or comprises means capable to carry out, at least the following steps, respectively: a) acquiring an image of said sample by digital image acquisition, b) optionally, carrying out digital image processing, c) selecting a field of view, d) determining, by digital image processing, whether or not, in said field of view, cell aggregates exist, and e1) selecting a new field of view and carrying on with step c) if, in step d), it turns out that, in said field of view, the determination of cell aggregates is negative, or e2) carrying out further process steps if, in step d), it turns out that, in said field of view, the determination of cell aggregates is affirmative.

1/13

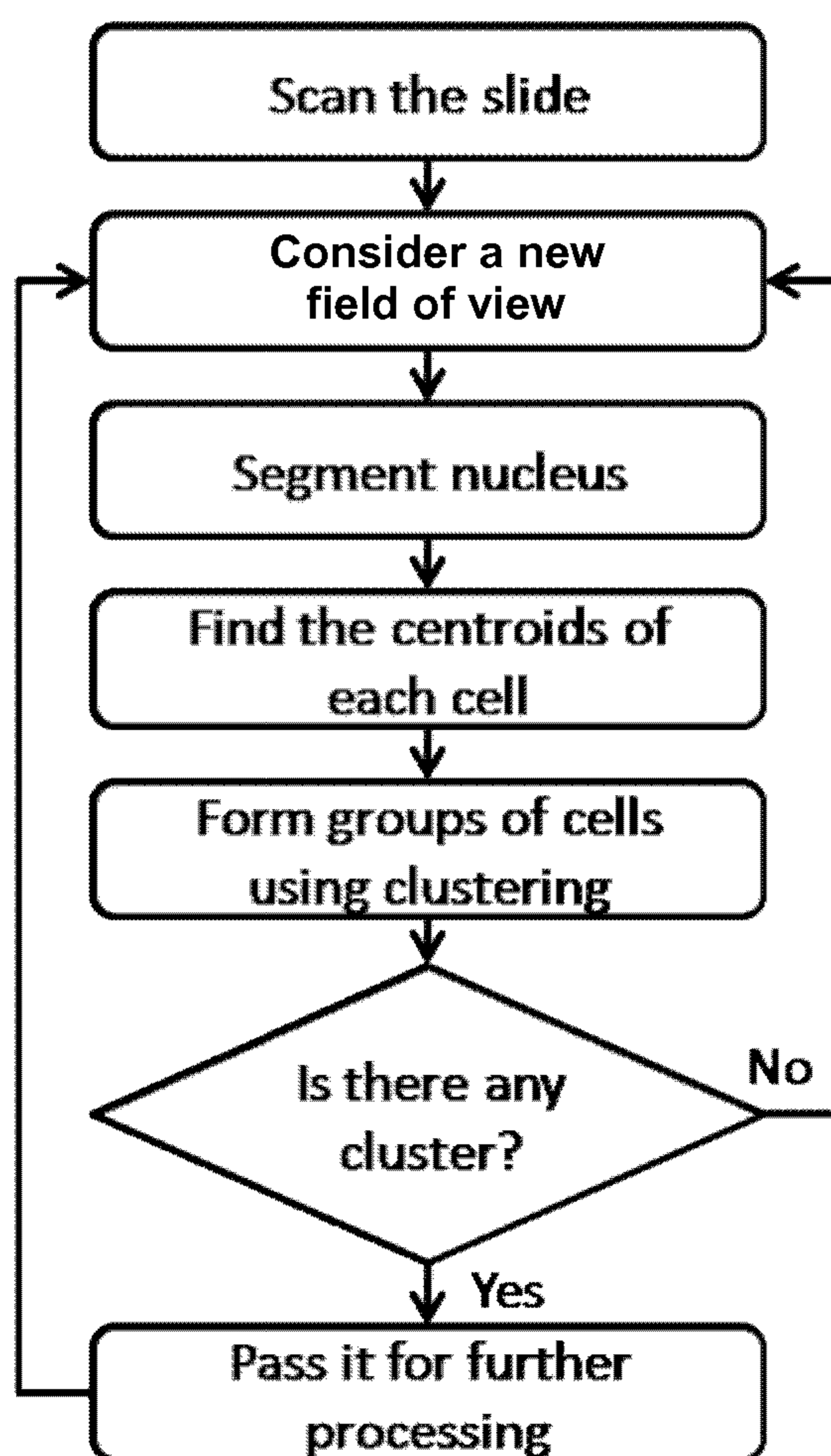


Fig. 1

2/13

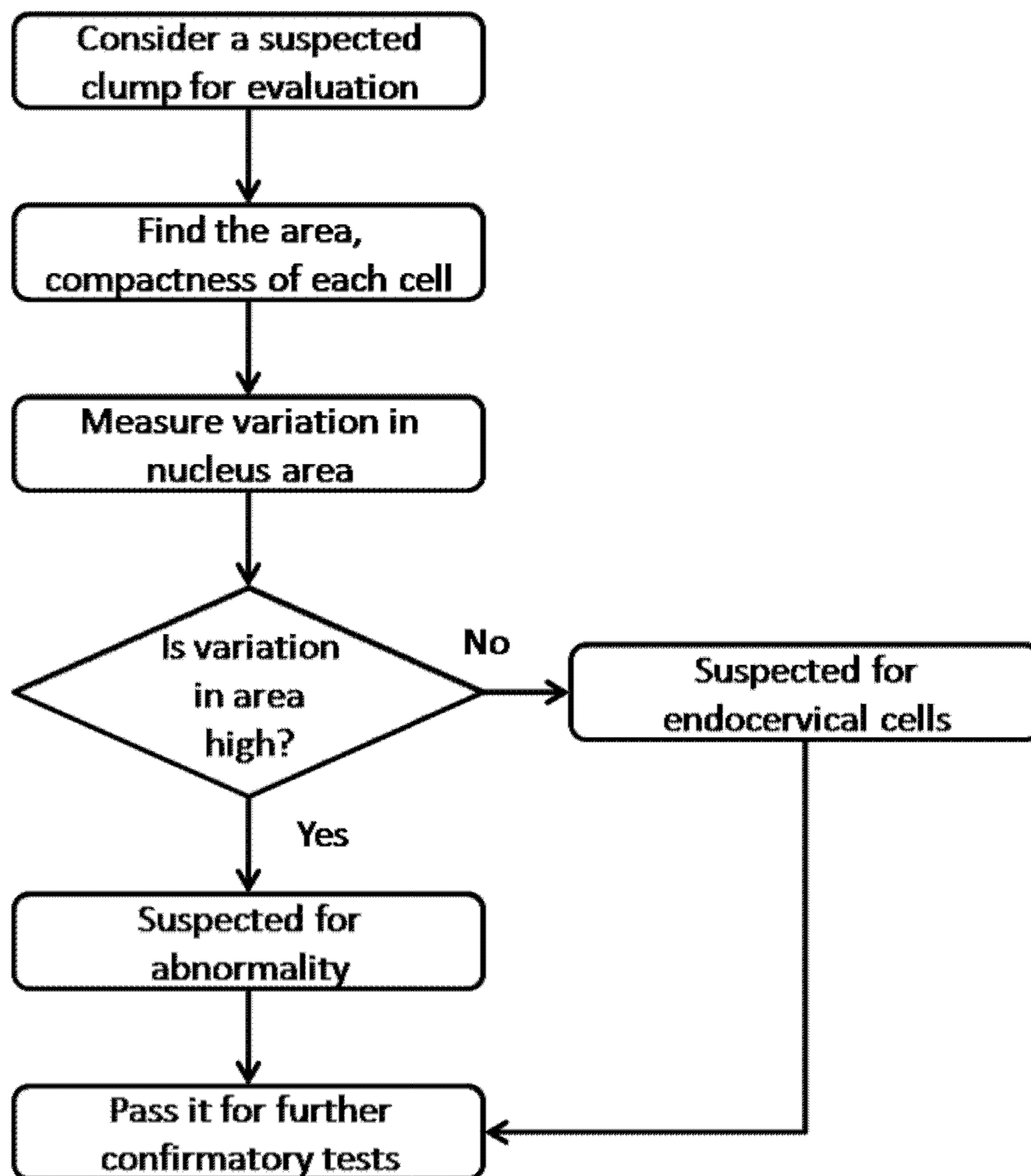


Fig. 2

3/13

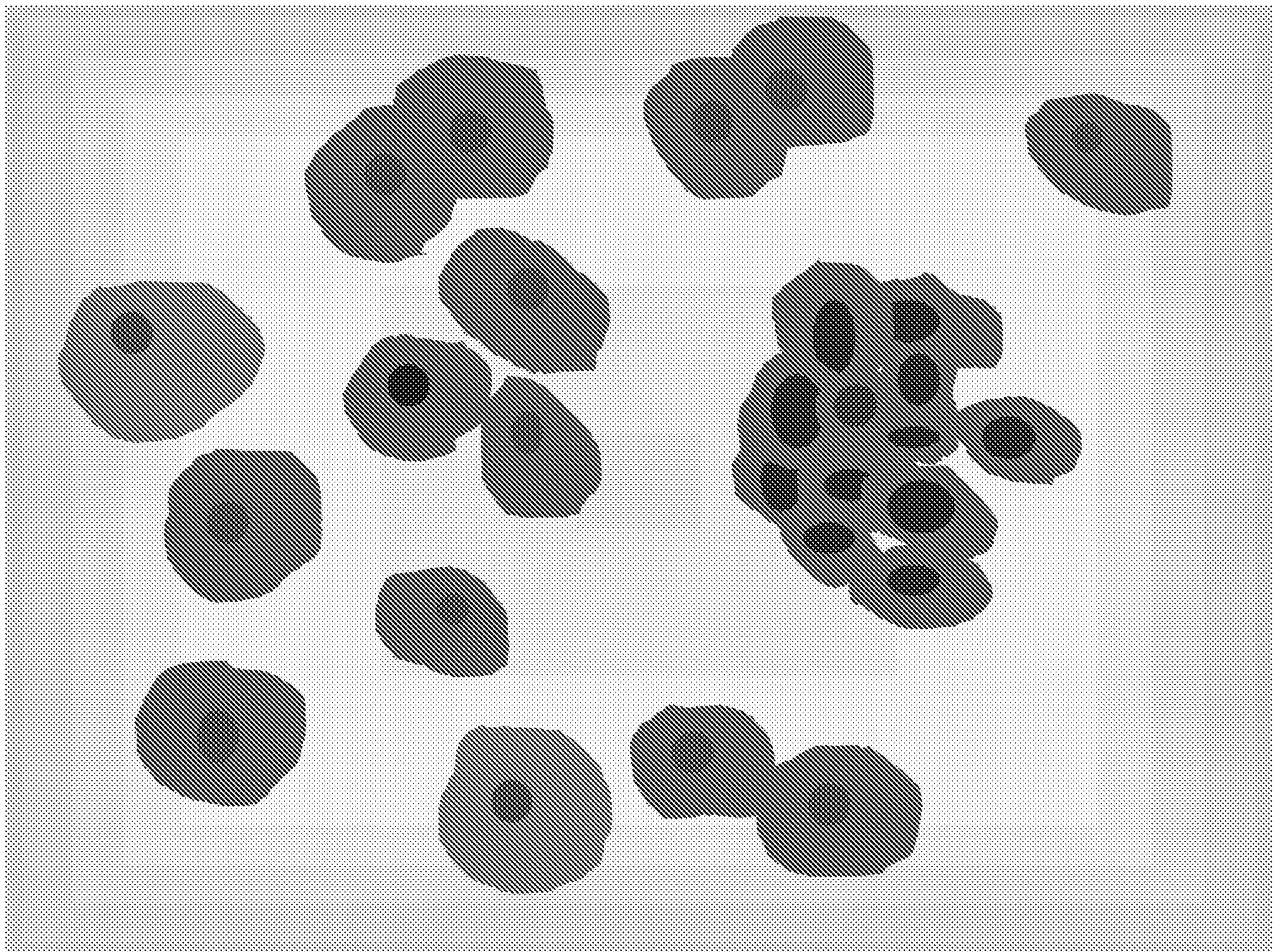


Fig. 3

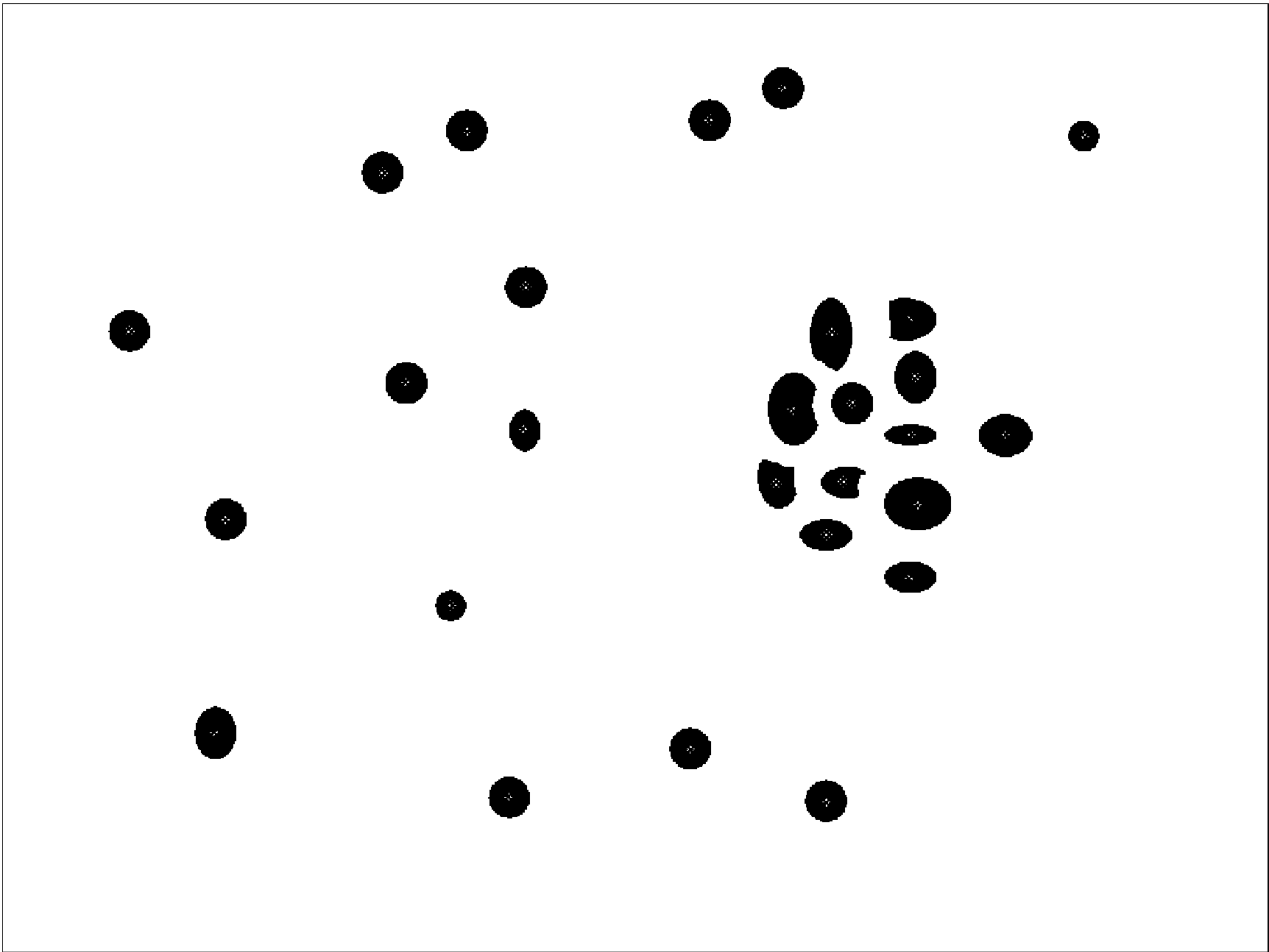


Fig. 4

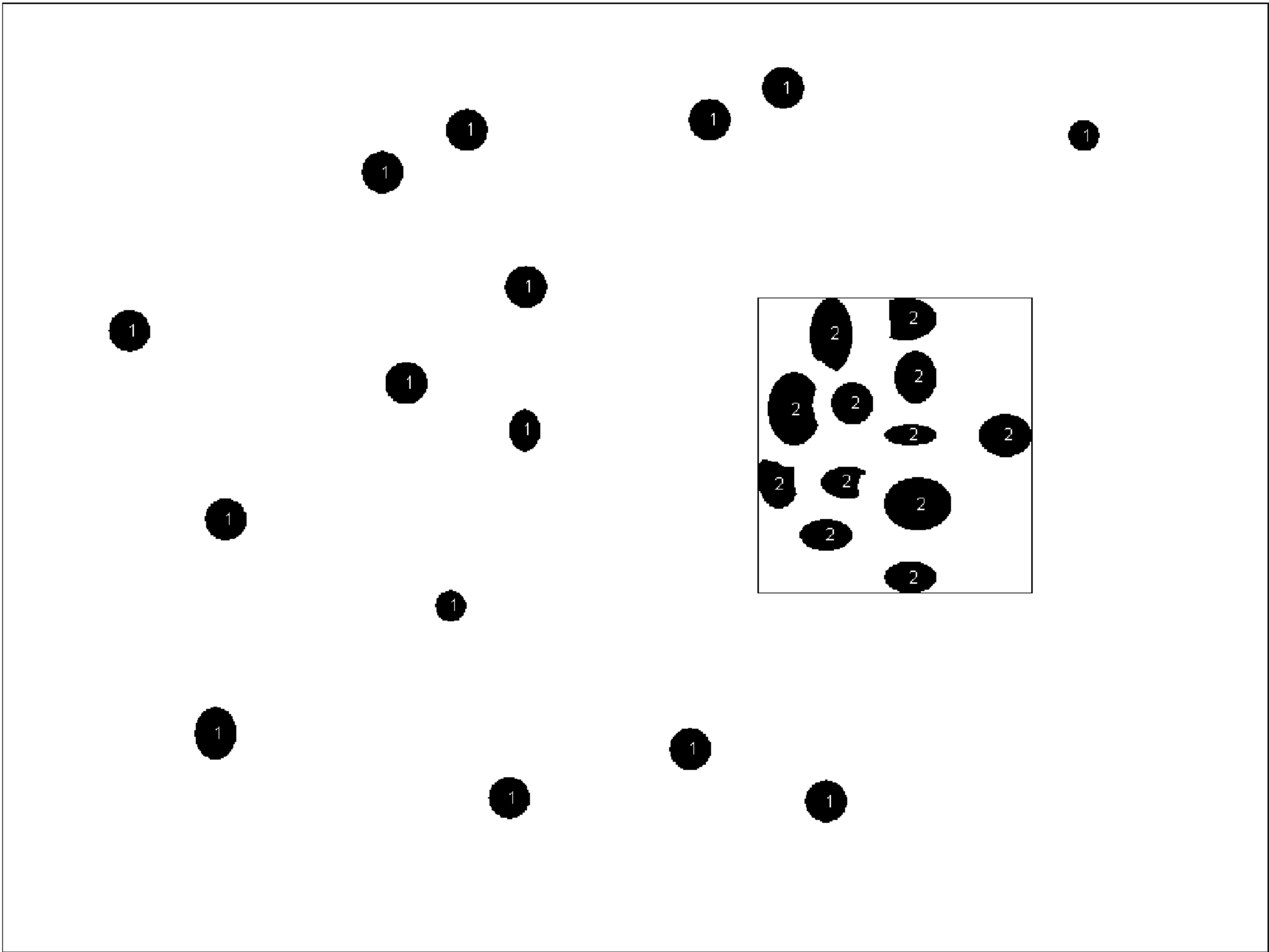


Fig. 5

6/13

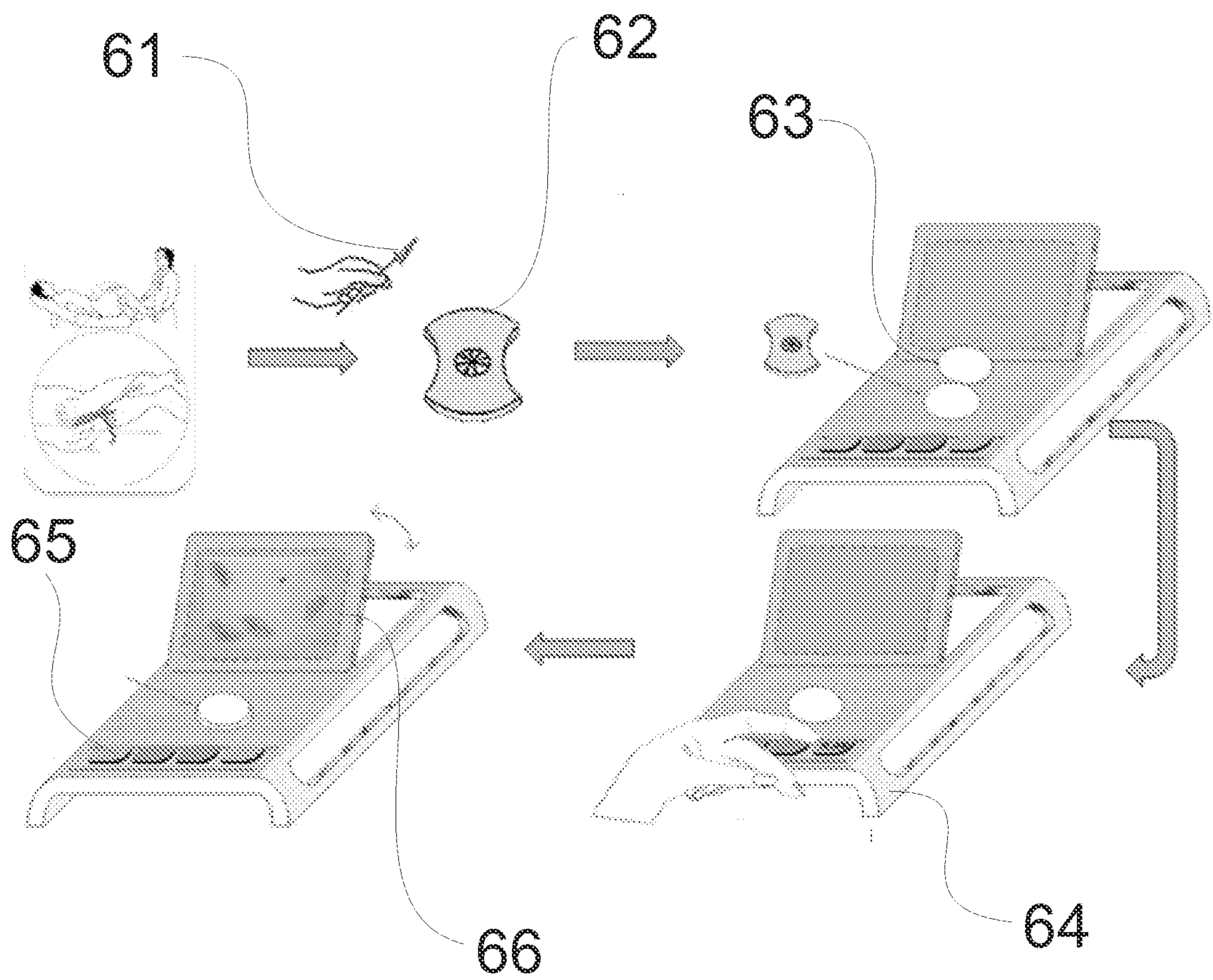


Fig. 6

7/13

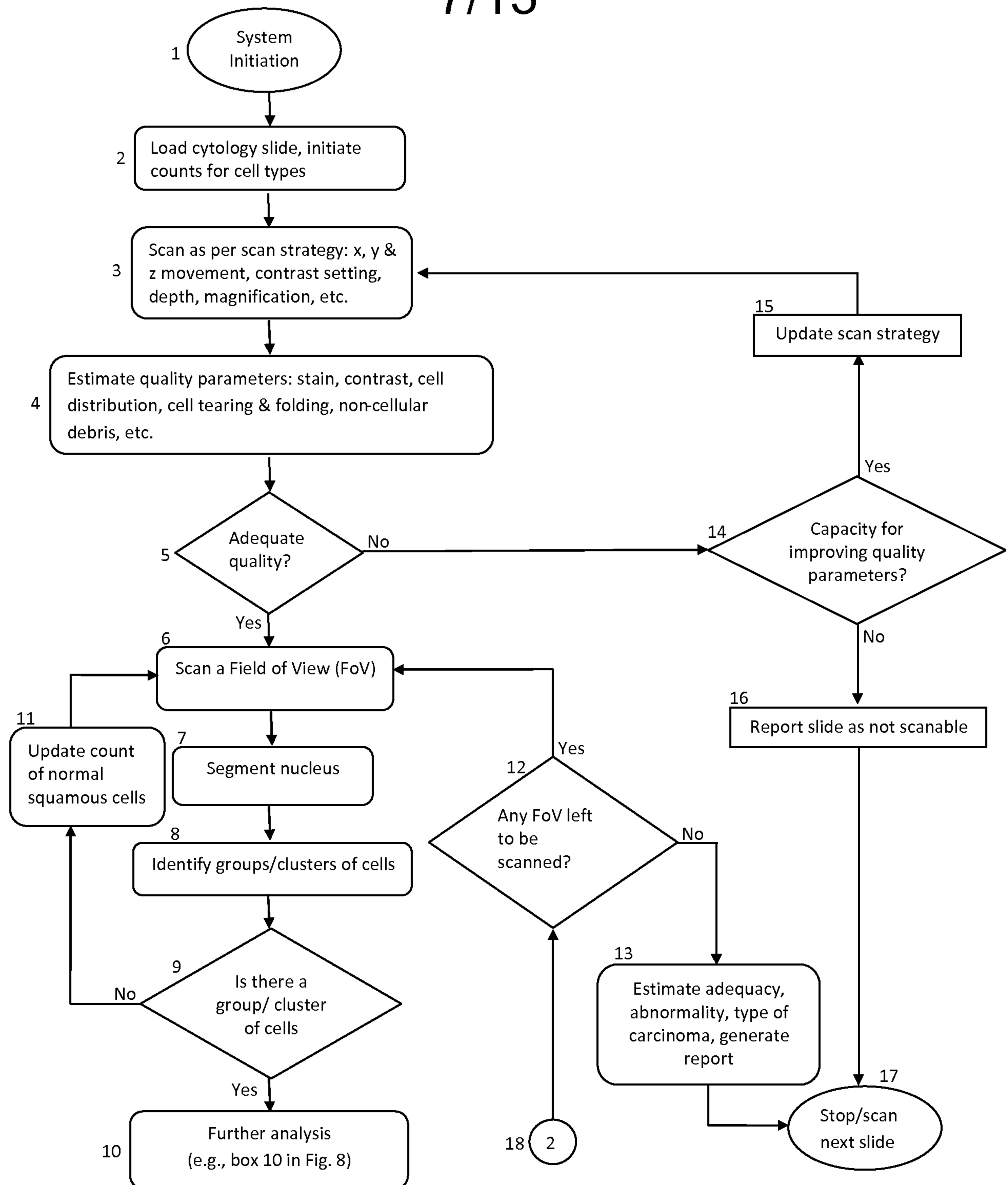


Fig. 7

8/13

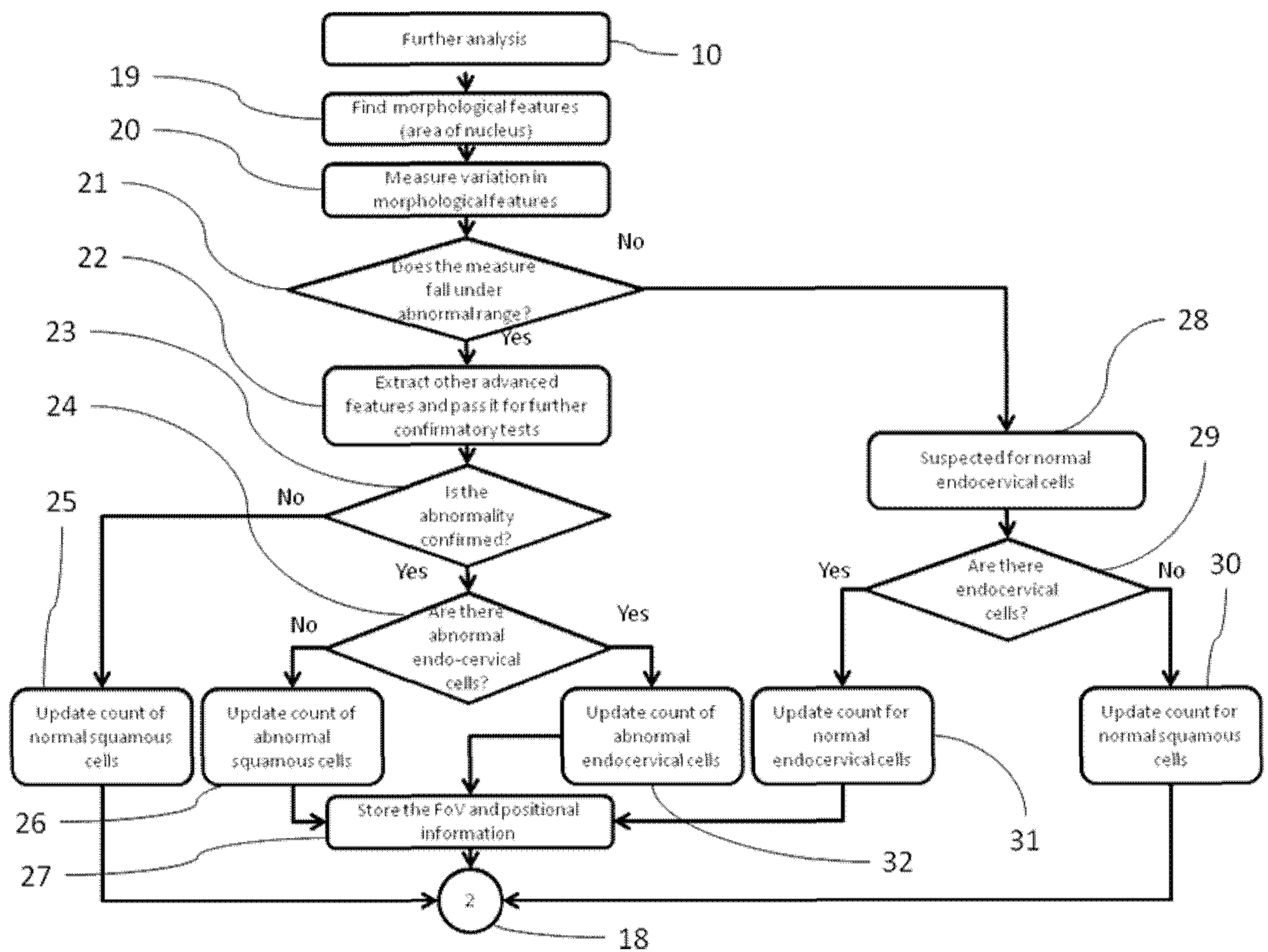


Fig. 8

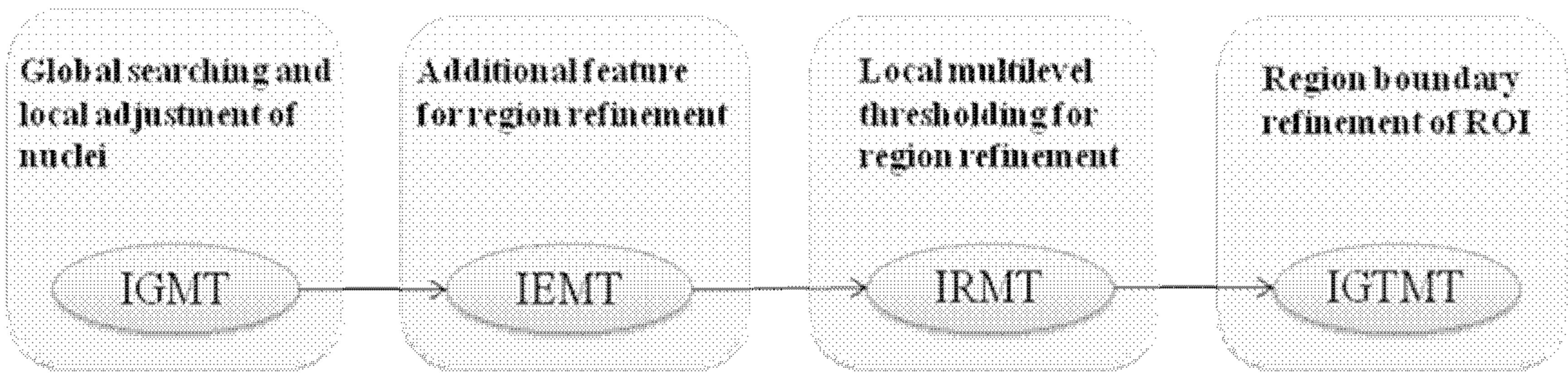


Fig. 9

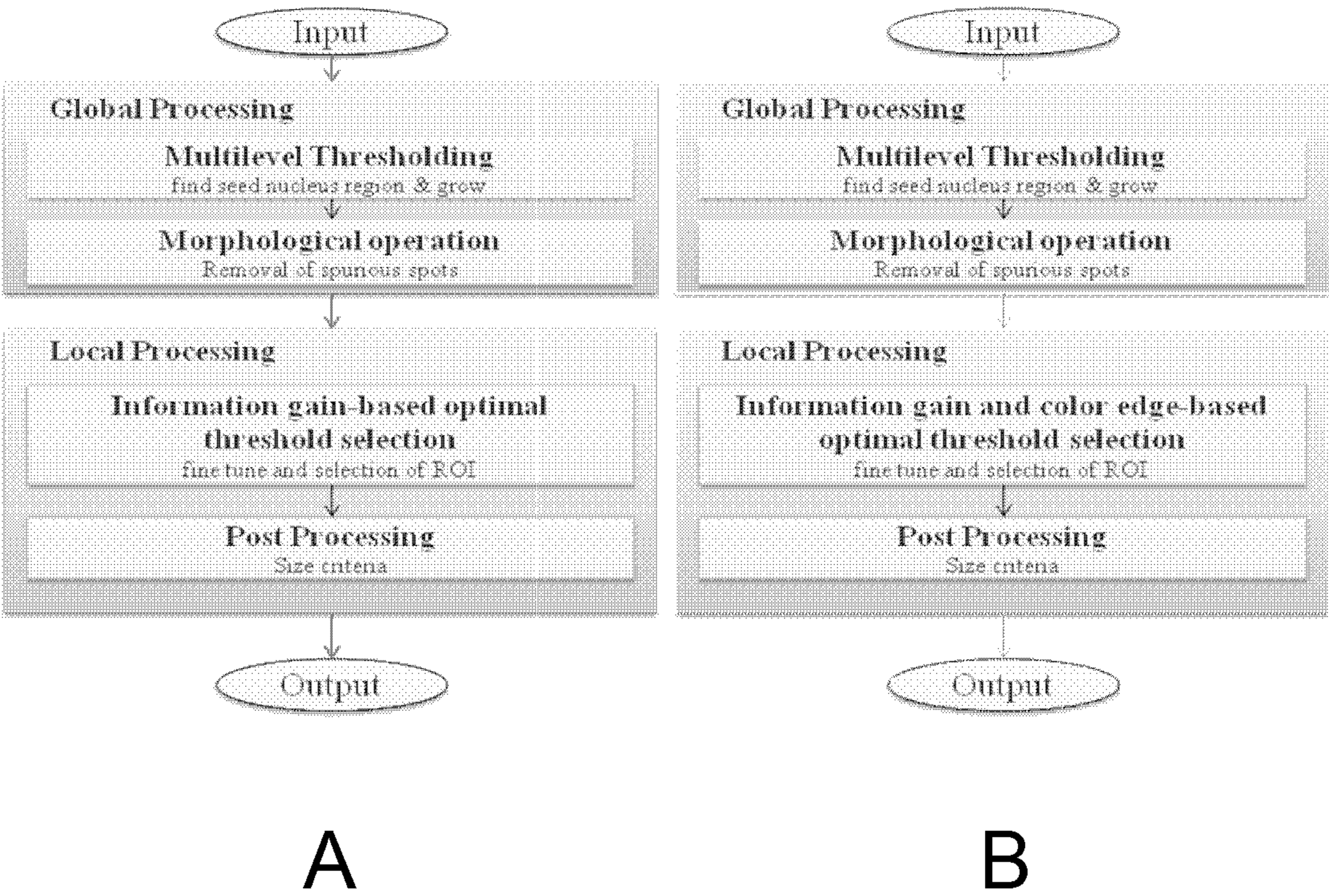
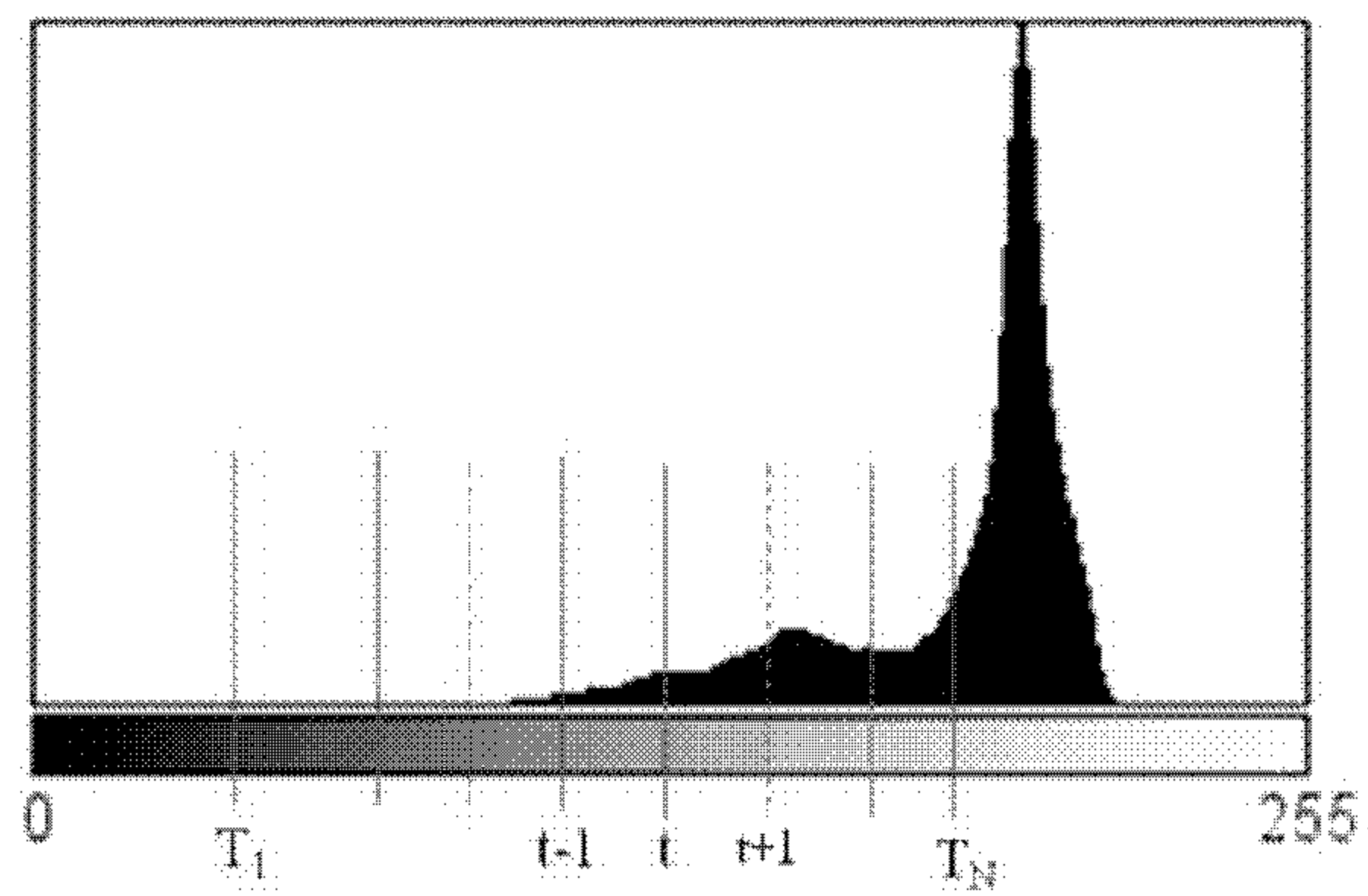
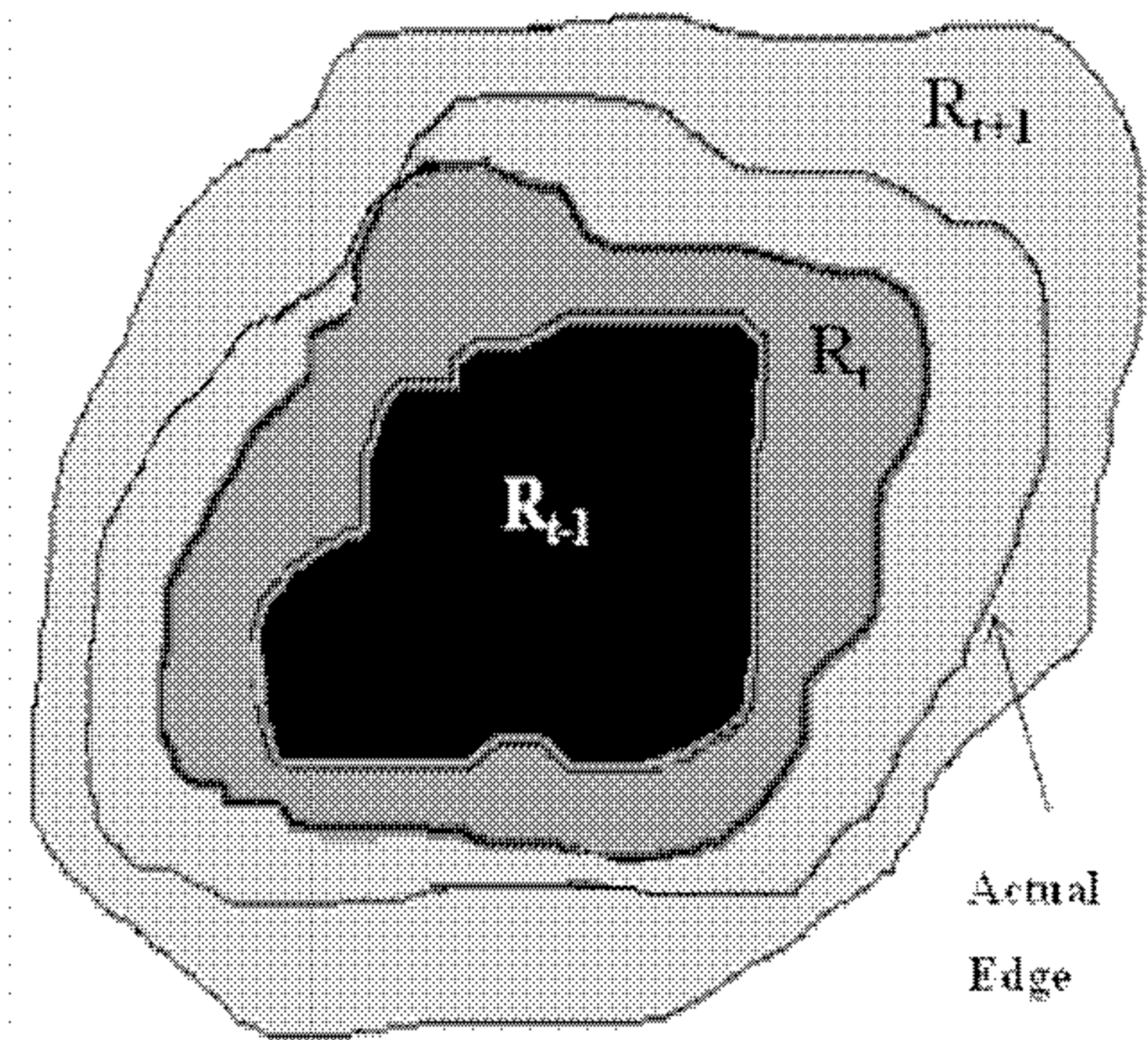


Fig. 10

10/13



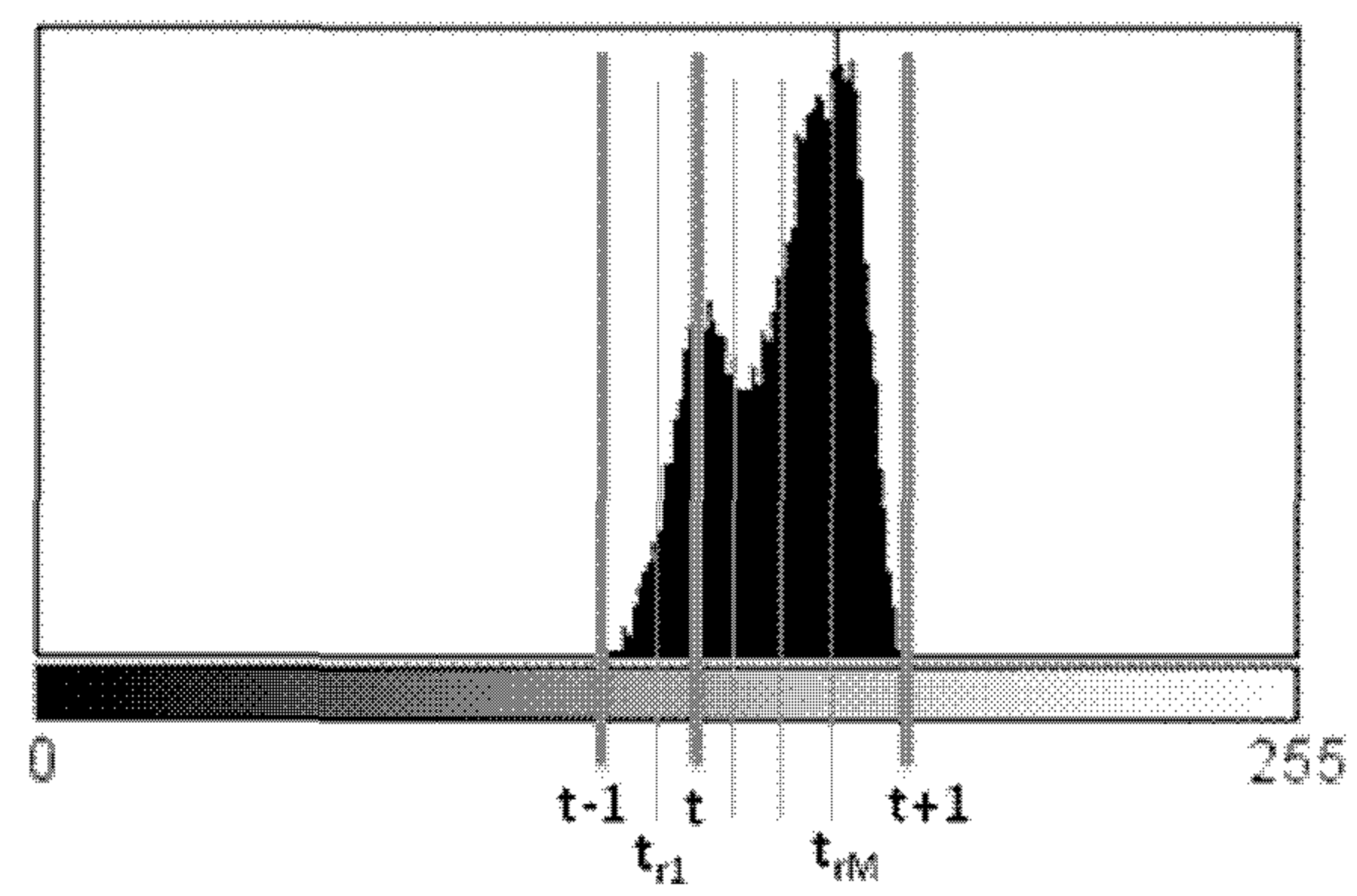
A



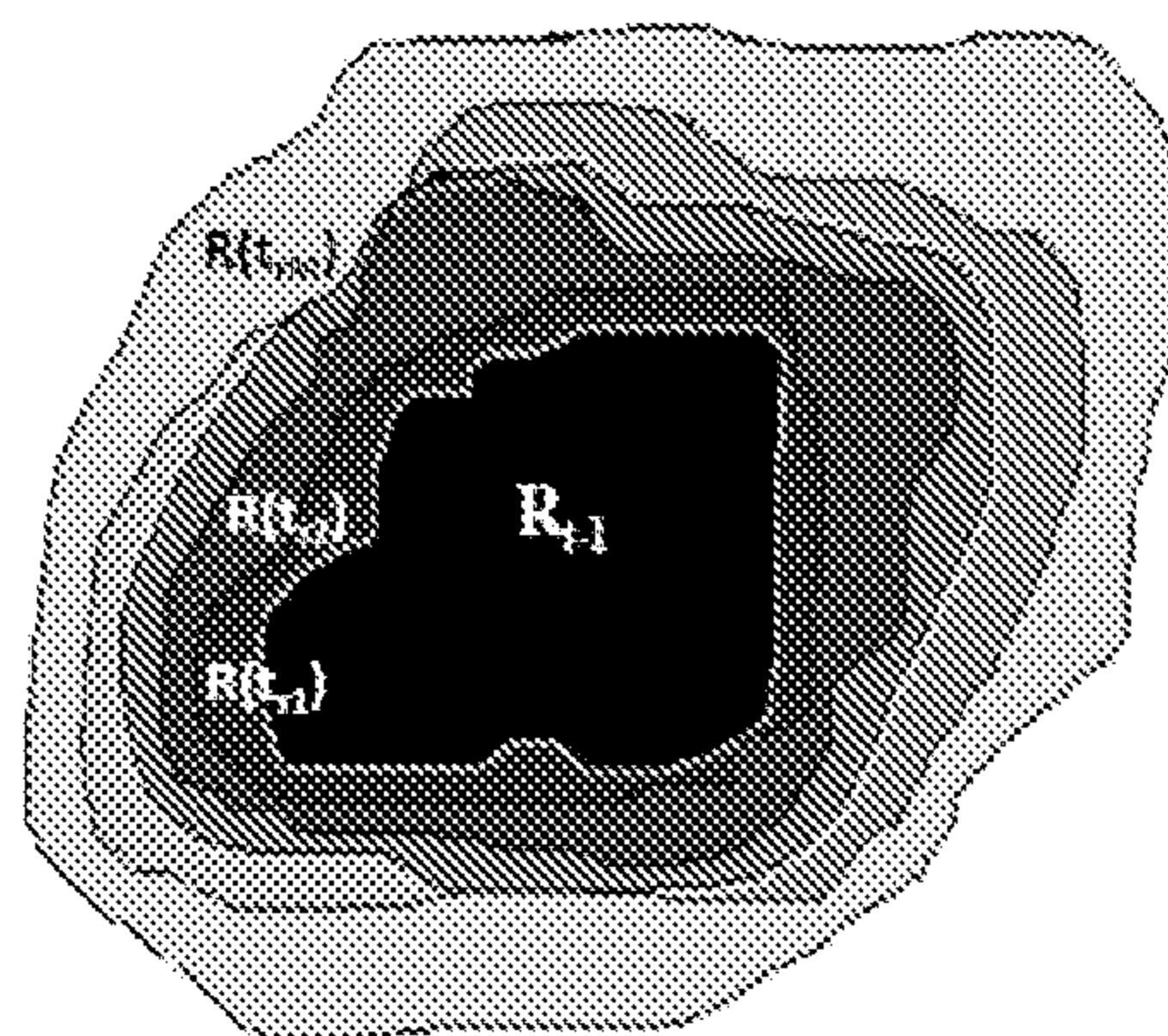
B

Fig. 11

11/13



A



B

Fig. 12

12/13

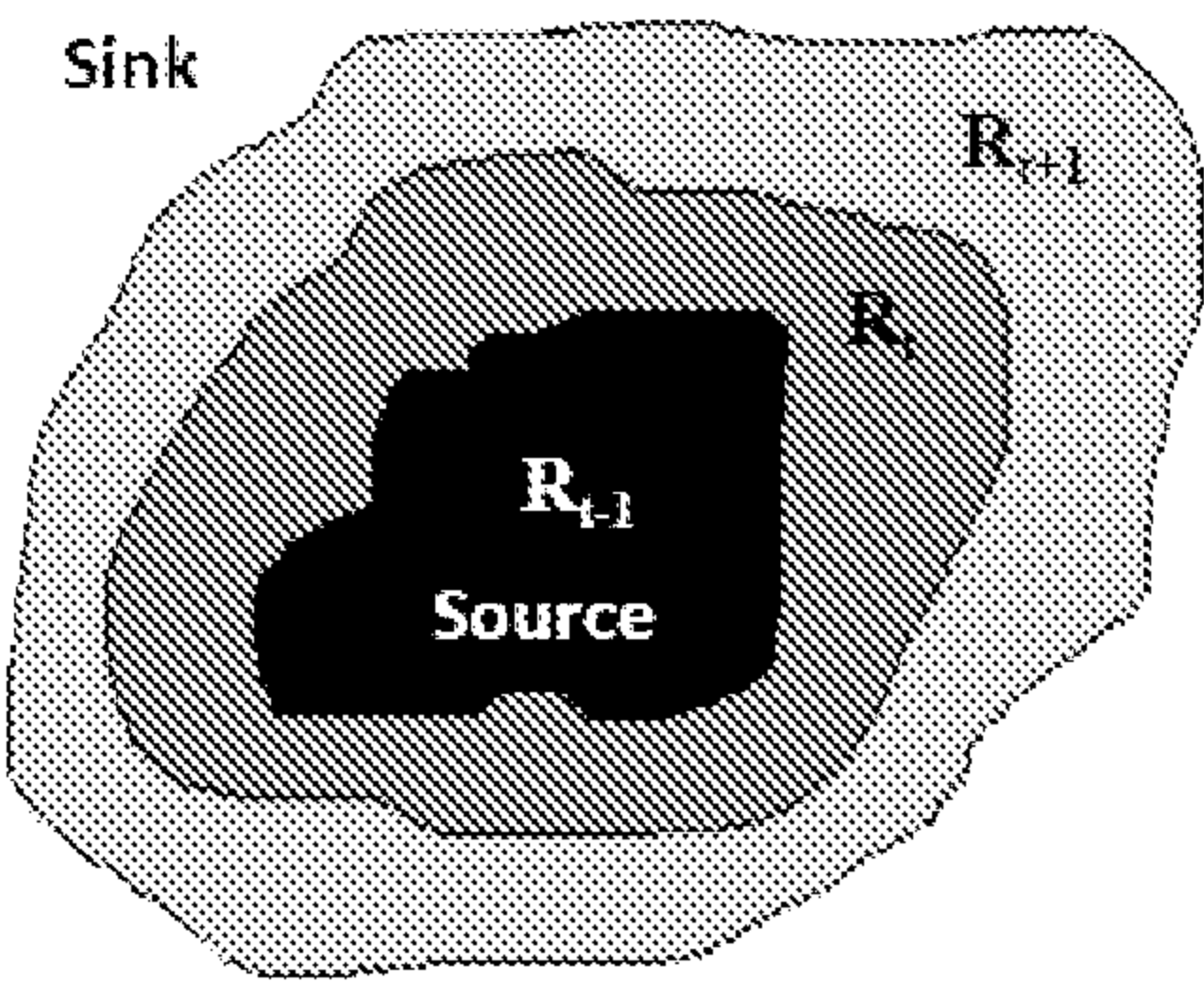


Fig. 13

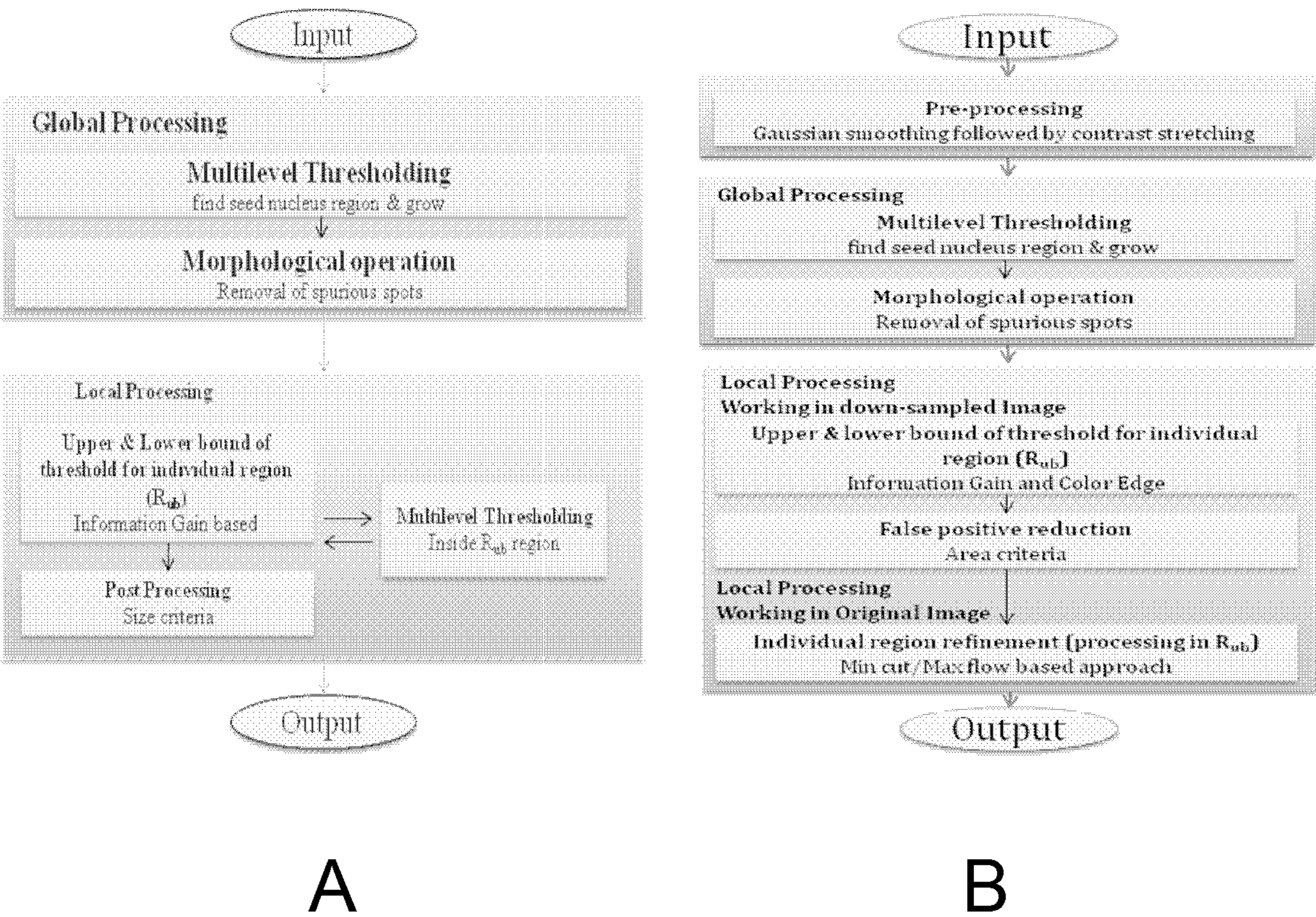


Fig. 14

13/13

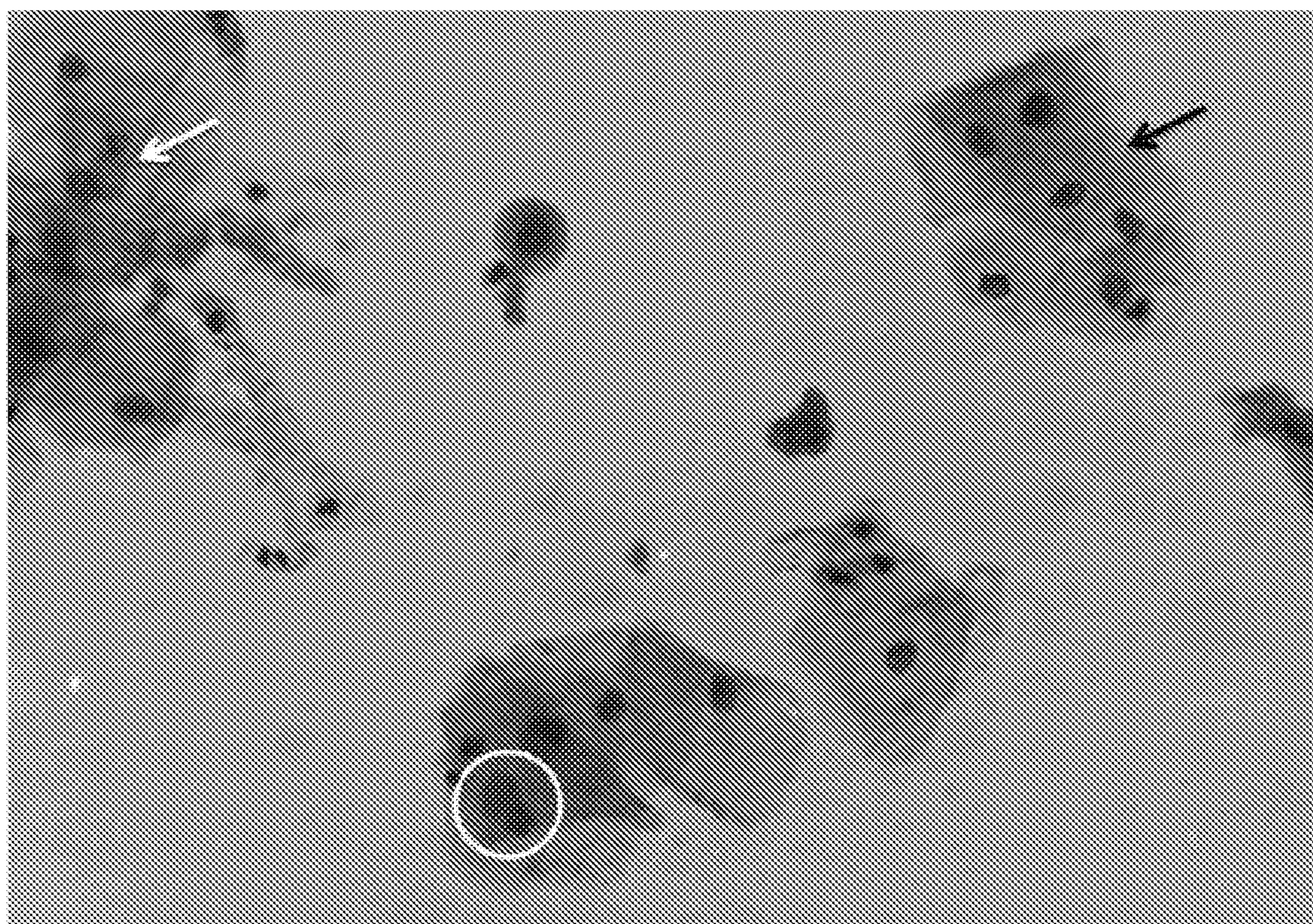


Fig. 15

System and method for the detection of abnormalities in a biological sample

FIELD OF THE INVENTION

The invention relates to the field microscopic detection of abnormalities in a
5 biological sample comprising cells.

BACKGROUND OF THE INVENTION

Cancer arising from cervix is the number one cancer in women in many industrialized countries as well as emerging countries. About 30% of cancers in women are
10 due to cervical cancer with more than 100,000 new cases diagnosed every year, e.g., in India. The estimated compounded annual growth rate (CAGR) for cervical cancer cases is 2.56% and at this growth rate approximately 175,000 new cases of cervical cancer will be detected in the year 2012.

One of the recommended tools for screening of cervical cancer is to detect
15 cytological precursors of cancer in Papanicolaou tests (also called Pap-smear, Pap-test, cervical smear, or smear test), which is a screening test used in gynecology to detect premalignant and malignant processes in the cervical canal especially in the transformation zone.

In taking a Pap smear, a speculum is used to gather cells from the outer
20 opening of the cervix of the uterus and the endocervix. The cells are examined under a microscope to look for abnormalities. The test aims to detect potentially pre-cancerous changes, which are, among others, caused by sexually transmitted human papilloma viruses. The test remains an effective, widely used method for early detection of pre-cancer and cervical cancer. The test may also detect infections and abnormalities in the endocervix and
25 endometrium.

This procedure has been effective in bringing down the incidence of cervical cancer in the developed countries. However, Pap smear has a false negative rate of 10-29%. Reasons for the false-negative results are numerous and include sample collection errors (failure to obtain adequate cells on the slide in terms of cells representing the transformation
30 zone), screening errors (failure to find abnormal cells on the slide), interpretation errors (failure to properly interpret abnormal cells), and miscellaneous laboratory errors related to staining problems, mislabeling, etc.

One of the major factors is that it is challenging for a pathologist to go through each of the cell in the slide. Each Pap smear slide has more than 10,000 cells of different

morphological features. Depending on the stage of the cancer it is not unlikely that only a small fraction of cells (e.g., $\ll 1\%$) in the sample is abnormal. This abnormality is detected by changes in the morphological features of the cell such as nuclear features, nuclear membrane, nuclear cytoplasmic ratio etc. Thus, careful observation of each cell feature is required to prevent false negative impression. This is a challenging task, considering the limited number of qualified pathologist in the field, the tremendous economical pressure under which they work, the limited time budget they have for each patient and the huge workload they are exposed to.

10 SUMMARY OF THE INVENTION

According to a first aspect of the invention, a method for the detection of abnormalities in a biological sample comprising cells is provided. Said method or system comprises at least the following steps, or is capable to carry out, and/or comprises means capable to carry out, at least the following steps, respectively:

- a) acquiring an image of said sample by digital image acquisition
- b) optionally, carrying out digital image processing
- c) selecting a field of view
- 20 d) determining, by digital image processing, whether or not, in said field of view, cell aggregates exist, and
 - e1) selecting a new field of view and carrying on with step c) if, in step d), it turns out that, in said field of view, the determination of cell aggregates is negative, or
 - 25 e2) carrying out further process steps if, in step d), it turns out that, in said field of view, the determination of cell aggregates is affirmative.

The term “cell aggregates”, as used herein, refers to a group of at least two cells being in physical contact with one another and forming a two- or three-dimensional cluster.

The term “the determination of cell aggregates is negative” as used herein, means that no cell aggregates could be found in the actual field of view. The term “the determination of cell aggregates is affirmative” as used herein, means that cell aggregates

have been found in the actual field of view, suggesting that the actual field of view could comprise abnormal cells.

The process according to the invention automatically scans the sample for fields of view which are likely to comprise images of abnormal cells – which is usually the case for cells which appear in clusters, or aggregates. The approach thus carries out a preselection in which only those fields of view are passed over to further analysis which comprise images of cell aggregates, while those fields of view which do not comprise images of cell aggregates are discarded, because it is unlikely that they comprise images of abnormal cells. This again saves time and reduces computational efforts.

Further, this methodological approach comprises a feedback loop, which significantly reduces the computational time for complex image processing algorithms as well as it simplifies the tedious task of examining each of the cells found in the sample, namely by reducing the evaluation time. The approach does not analyse the acquired image as a whole, but selects subunits of said image, which are called “field of view” herein, for image processing. An overview of the described method is given in Fig. 1.

Generally, the cells which have the highest risk to become cancerous are ectocervical cells, while endocervical cells have a smaller risk to become cancerous (in which case an adenocarcinoma is formed). Ectocervical cells are also called squamous cells, while endocervical cells are also called columnar cells.

However, it has been observed that also cells which have only a small risk to become cancerous also tend to form aggregates, which may create some confusion under particular circumstances. In a preferred embodiment, the method according to the invention thus further comprises the steps of:

- f) determining the degree of variation of at least one given morphological feature of at least two cells in the field of view, and
- g1) classifying the sample as “likely to comprise abnormal cells” if the degree of variation of said morphological feature exceeds a predetermined threshold, or
- g2) classifying the sample as “likely to comprise normal cells” if the degree of variation of said morphological feature falls below a predetermined threshold.

The term “classifying the sample as likely to comprise abnormal cells” is equivalent to “suspecting a sample for abnormality”, as shown in Fig. 2. The term “classifying the sample as likely to comprise normal cells” is equivalent to “suspecting a sample for normal endocervical cells”, as shown in Fig. 8, or, although clinically imprecise, “suspecting a sample for endocervical cells”, as shown in Fig. 2.

As used herein, the term “abnormal cells” relates to cells which are in a process of becoming cancerous, or malignant, or are cancerous, or malignant, already.

Preferably, it is provided that said morphological feature is at least one selected from the group consisting of

- cell nucleus size, or area
- regularity of shape of a cell, and/or a cell nucleus
- size ratio of cytoplasm to nucleus, in a given cell, and/or
- ratio of areas of cytoplasm and nucleus, in a given cell.

In this complimentary approach, the variability of said morphological feature can be used as a further distinguishing feature, because in abnormal cell aggregates the nucleus size varies widely, whereas in normal cell aggregates the nucleus sizes remain uniform.

This complimentary approach thus serves to identify abnormal cells, particularly in those regions of interest which have earlier been identified, by cell aggregation analysis, as suspicious. If the variation said morphological feature in the field of view exceeds, statistically, a given threshold, the sample can be classified as “likely to comprise abnormal cells”.

The cell nuclei sizes in the field of view can for example be determined by calculating the respective image area, e.g., by counting the number of pixels in the respective region. The variation of nuclei sizes in the field of view can for example be expressed as standard deviation of these areas, or by determining the variation in major axis or minor axis of the ellipse encircling these regions, or ratio of these values. While no fixed threshold value exists for the variation of sizes of the nuclei, thresholds can be determined a priori using ground truth data which could also vary on account other factors such as magnification, resolution of the image, etc.

The term “shape”, as used herein, relates to the two-dimensional shape of a cell nucleus image, in the field of view. A cell nucleus shape is considered to have a high degree of regularity in case the shape of its image is circular, or close to circular. The

criterion to determine the cellular shape is by employing properties like form factor, perimeter, major axis, minor axis cell membrane signature.

If the regularity of the cell nucleus shapes is, statistically, below a given threshold, the sample can as well be classified as likely to comprise abnormal cells.

5 The ratio of cytoplasm size, or area and nucleus size, or area, in a given cell, is another indicator which can be used in the context of the present invention. While normal cells have large cytoplasm and small nucleus, abnormal cells tend to have large nuclei and small cytoplasm.

10 The cellular Nucleus harbours the most significant changes in precancerous and cancerous cells. Hence, identifying the nucleus automatically can be a useful approach to detect abnormal cells in cervical smears. However, the segmentation of nucleus is a challenging task due to the varied morphological appearance with clumps and artefacts (see Fig. 15). Moreover, in some cases, the Pap smear images are blurry and highly affected by unwanted noises, e.g., blood, air artefacts (bubbles), or vagina discharge. In some cases, Pap-smear images of that kind may lead to an increase in false alarms when analyzed with a segmentation algorithm. Existing techniques are unable to tackle all these problems. Hence a fast and accurate nucleus segmentation technique is necessary to solve this issue.

20 According to a particularly preferred embodiment, the cell nucleus is therefore detected by an optical technique encompassing multilevel thresholding. Image histograms are usually the basis for thresholding. A histogram is unimodal if there is one hump, bimodal if there are two humps and multimodal if there are many humps. Histograms of Pap-smear images are multimodal in nature (see, e.g., Fig. 11A). According to this embodiment, multilevel thresholding has been introduced to locate initial seeds for nucleus segmentation. This is the first and foremost common step for all the above four methods.

25 In a more preferred embodiment, said optical technique encompassing multilevel thresholding is at least one selected from the group consisting of

- IGMT (Information Gain-based Multilevel Thresholding technique)
- IEMT (Information gain and color Edge-based Multilevel Thresholding technique)
- IRMT (Information gain-based Recursive Multilevel Thresholding technique), and/or
- 30 • IGTMT (Information and Graph Theory-based Multilevel Thresholding technique)

In IGMT, an information gain-based local region refinement is introduced after multiple thresholding to segment out nucleolus region. In IEMT, color edge analysis is introduced along with information gain to improve the accuracy of segmentation. However, region growing based on global multiple threshold levels may lead to chance of under/over segmentation of nuclei.

IRMT may be used to refine the region based on local multilevel thresholding, although it may be unable to solve the boundary leak problem. Hence, local boundary adjustment is necessary to solve this problem. IGTMT may be used to fine-tune the boundary region of Pap-smear nuclei. The different approaches will be discussed in detail in the following:

Information Gain-based Multilevel Threshold (IGMT)

In this approach, information gain is employed along with clustering technique to segment out Pap-smear nuclei. Histograms of Pap-smear images are multimodal in nature. Hence in the first step (global processing) multilevel thresholding has been introduced to locate initial seed for nucleus segmentation. In the following step, information gain-based local region selection and refinement technique has been introduced to fine-tune and isolate the nucleus region (see Fig. 10 (A)). If change in multilevel threshold yields abrupt change in information gain for a particular region of interest (ROI), region growing using multilevel thresholding stops.

Information gain and color Edge-based Multilevel Thresholding (IEMT)

IGMT is capable of segmenting nucleus regions with higher accuracy than other threshold based techniques. However, due to unfocused Pap-smear images, improper staining and other artifacts discussed above, information gain may sometimes not be sufficient enough for region refinement. Hence, colour edge analysis is introduced along with information gain to fine-tune and isolate the nucleus region (see Fig. 10 (B)). If a change in multilevel threshold yields abrupt change in information gain as well as normalized color edge for a particular region of interest (ROI), region growing using multilevel thresholding stops. This additional information (color edge) obtained with IEMT yields increasing accuracy of the nucleus segmentation than the IGMT method.

IGMT and IEMT methods are capable of finding the nucleus region in Pap-smear images. However, region growing for a ROI based on multiple threshold levels may lead to chance of under/over segmentation of nuclei (see, e.g., Fig. 11(B)).

5 Information gain-based Recursive Multilevel Thresholding (IRMT)

Recursive multilevel-based thresholding is introduced in IRMT to reduce the possibility of over/under segmentation of nuclei in IGMT and IEMT methods (see Fig. 7(A)). IRMT consists of two major steps: (i) global seed selection, followed by (ii) local
10 region refinement.

Global seed selection of Pap-smear images as carried out in the IRMT method is identical to IGMT and IEMT. In case of a local region refinement, the IRMT method provides two major steps:

15 (ii.a.) Selection of upper and lower bound of threshold for individual region:

First of all, optimal threshold t is computed using information gain and color edge for individual ROI. This optimal threshold t sometimes yields over/under segmented nuclei. Therefore, region (R_{ub}) having gray value in-between $t-1$ and $t+1$ and connected with the ROI ($R_t \cup R_{t+1}$) is chosen for farther processing (see Fig. 11(B)). The idea behind this step is that
20 the exact nucleus boundary lies in between this bounding region (R_{ub}).

(ii.b.) Recursive multilevel threshold for region refinement: The same multilevel thresholding technique is applied on the histogram of bounding region (R_{ub}) (see Fig. 12 (A) and (B)). Sub-thresholds thus obtained will increase the chance of getting more
25 refined optimal threshold using information gain and color edge-based region growing method. This will help in fine-tuning the individual ROI obtained by the IGMT or IEMT method.

30 Information and Graph Theory-based Multilevel Thresholding (IGTMT)

The proposed IRMT method may sometimes fails to solve the boundary leak problem. The basic idea behind the IGTMT approach is to utilize the graph cut theory for local boundary refinement (see Fig. 14 (B)).

Like the previous approaches, global multilevel thresholding is the first step in the IGTMT method, to carry out seed initialization of the probable region of the nucleus. This is followed by the selection of upper and lower bound of threshold for individual region which is similar to the IRMT method. Thereafter, to increase the accuracy of segmentation scheme, IGTMT introduces the min cut/max flow based graph theory approach. The IGTMT method uses the similarity measure based on gray level difference of neighbourhood pixels in R_{ub} region. This affinity measures are used as weights in graph where image pixels are represents as nodes with pre-calculated single source and sink (see Fig. 13). Min cut finds a partitioning of graph nodes which minimizes the sum of weights of cut edges. This introduces the granularity concept helping to achieve the exact boundary of the nucleus.

In another preferred embodiment, it is provided that the method further comprises the step of determining at least one feature selected from the group of

- brightness intensity variation within a cell and/or a cell nucleus
- jazziness of a cellular membrane
- texture, and/or
- fractal dimension of the nucleus

As used herein, the term “jazziness” shall mean the variation of distance between cell boundary points from centre of the cell or a fixed reference point preferably inside the cell. High jazziness in texture can be considered as an indication of abnormality.

As used herein, the term “texture” shall mean the spatial arrangements of colors or intensities in a nucleus or cell region. High variations in texture can be considered as an indication of abnormality.

As used herein, the term “fractal dimension of the nucleus” relates to a statistical quantity that gives an indication of how completely a fractal appears to fill space of a given cell, as one zooms down to finer and finer scales.

morphological feature	algorithm/technique for determination
aggregation of cells	determining inter-nuclei distance
cell nucleus size, or area	calculating the respective image area, e.g., by counting the number of pixels in the respective region
inter-nuclei distance	distance between the centre of any two nucleus
regularity of shape of a cell, and/or a cell nucleus	employing properties like form factor, perimeter, major axis, minor axis cell membrane signature.
ratio of cytoplasm to nucleus in terms of size, or ratio of areas of cytoplasm and nucleus size, or area, in a given cell	see cell nucleus size
Brightness/ intensity variation within a cell and/or a cell nucleus	variance of intensity in the cell nucleus
jazziness of a cellular membrane	the variation of distance between cell boundary points from centre of the cell or a some fixed reference point preferably inside the cell
Texture	the spatial arrangements of colors or intensities in a nucleus or cell region
fractal dimension of the nucleus to measure the deviation from the uniformity	Hausdorff dimension, box counting dimension, Renyi dimension

In still another preferred embodiment, the method according to the invention further comprises at least one step selected from the group consisting of:

5

- a) detailed inspection of the cervix by colposcopy;
- b) carrying out a HPV DNA test in said biological sample, or in new sample with comparable properties;

- c) carrying out a Biomarker test in said biological sample, or in new sample with comparable properties; and/or
- d) visual inspection of said biological sample, or of a new sample with comparable properties, by a qualified pathologist.

5

Alternatively, the method according to the invention further comprises the step of recommending further investigation by at least one step selected from the group consisting of

- 10 a) detailed inspection of the cervix by colposcopy;
- b) carrying out a HPV DNA test in said biological sample, or in new sample with comparable properties;
- c) carrying out a Biomarker test in said biological sample, or in new sample with comparable properties; and/or
- 15 d) visual inspection of said biological sample, or of a new sample with comparable properties, by a qualified pathologist.

Colposcopy is a medical diagnostic procedure to examine an illuminated, magnified view of the cervix and the tissues of the vagina and vulva. Primarily in order to
20 detect premalignant lesions and malignant lesions which may result in cancer. Colposcopy is done using a colposcope, which provides an enlarged view of the areas, allowing the colposcopist to visually distinguish normal from abnormal appearing tissue and take directed biopsies for further pathological examination. The main goal of colposcopy is to prevent cervical cancer by detecting precancerous lesions early and treating them.

25 A HPV DNA test detects cervical infection with human papilloma virus (HPV), which is one of the most important infectious causes of cervical cancer. 84% of new cervical cancers were in the developing world, compared with about 50% of all new cancers. HPV DNA test kits are today commercially available. Such test may be carried out during a routine smear test, as described above (in which case part of the smear sample is taken for the
30 HPV DNA test, while another part is taken for the method according to the invention, or with a newly taken sample with comparable properties, and can be used to improve, confirm or falsify the diagnostic significance of the method according to the invention.

Biomarker tests have been developed to investigate whether or not a patient suspected to be predisposed for cervical cancer, or a patient who is suspected for having

cervical cancer, or in which cervical cancer has already been diagnosed, has, in its genome or proteome, an abnormality which coincides with increased or decreased likelihood of getting a given cancer, or which coincides with increased or decreased responsiveness towards a given therapy. Such abnormality is, for example, a mutation in a given gene, an abnormality in an
5 epigenomic feature, like DNA methylation, or an abnormality with respect to expression of a given gene.

In still another preferred embodiment of the method according to the invention, the image acquisition is carried out by means of a scanner. As an alternative, a two dimensional imaging device can be used. In both cases the imaging device is preferably a
10 CCD (linear or two dimensional) or a CMOS (linear or two dimensional).

In another preferred embodiment of the method according to the invention, the image acquisition is carried out by means of an optical magnification device. Said optical magnification device is, for example, a microscope.

In still another preferred embodiment, the method according to the invention,
15 further comprises, prior to step a), a step in which an image of the sample is acquired at lower magnification, as is the case in step a). In this embodiment, an overview image is made first. The low magnification slide overview is processed by the algorithm to identify the regions suspicious for abnormality and those suspected region are further scanned with higher magnification. This will provide an advantage of quick scanning of the slides. In yet another
20 preferred embodiment of the method according to the invention, steps b) and following are carried out while the digital image acquisition and/or the digital image processing is still in process.

This methodological approach, which is also called “on the fly” or “real time”, reduces the computational time for complex image processing algorithms as well as it
25 simplifies the tedious task of examining each of the cells found in the sample by reducing the evaluation time.

Further, this approach allows to forgo an image archive of the raw images. In such embodiment, it is possible to only store only the selected processed images (see, e.g., Fig. 5), or even only the non-image data extracted from such images.

30 Further, such real time approach enables a pathologist/operator to intervene in the process, e.g., to fine tune discriminating variables, or to teach the system in that results of a given method step are accepted, or rejected as false.

In yet another preferred embodiment of the method according to the invention, step d) comprises at least the steps of

- Segmentation of cell nuclei, and
- Determining the centroids of each cell.

Preferably, the latter can be done by determining the inter-nuclei distance of at least two cells.

As used herein, the term “segmentation of cell nuclei” refers to the process of partitioning a digital image comprising the image of at least one cell into multiple segments (sets of pixels in order to identify the cell nuclei. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s). When applied to a stack of images, typical in Medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like marching cubes.

As used herein, the term “centroid of a cell” relates to the geometric center, or barycenter, of a cell's two-dimensional image, as determined with digital image processing methods, e.g., by image moments.

In another preferred embodiment of the present invention, the biological sample comprising cells is a cervical sample. However, because principles of cancer genesis and cell transformation are ubiquitous, the method can also be used with samples from other body tissues which have to be checked for abnormalities, like breast samples, prostate samples, liver samples, lung samples and so forth.

In case the biological sample comprising cells is a cervical sample, it is further preferred the method further comprises at least one step selected from the group consisting of:

- classifying cells comprised in a sample which is likely to comprise normal cells as “normal endocervical” or “normal ectocervical” based on a set of predetermined criteria, and/or
- classifying cells comprised in a sample which is likely to comprise abnormal cells as “abnormal endocervical” or “abnormal ectocervical” based on a set of predetermined criteria.

It is again important to mention that the term “ectocervical cells” is used, interchangeably, with the term “squamous cells” herein.

Furthermore, it is preferred that the method according to the invention further comprises at least one step of counting a given cell type, or updating an existing count thereof.

It is further preferred that the biological sample comprising cells comprises at least one sample selected from the group consisting of

- smear sample
- tissue slice
- liquid sample, and/or
- any other cytology sample.

A smear sample is for example similar or identical to those samples used in the Papanicolaou tests (also called Pap smear, Pap test, cervical smear, or smear test). A tissue slice is for example, sliced by a microtome. A liquid sample can preferably consist of a suspension of cells, e.g., obtained by a smear.

Other suitable samples comprise, but are not restricted to, fine needle aspiration cytology (FNAC) samples, abrasive cytology samples and/or exfoliated samples.

It needs to be said that in the flow charts, the term slide is used. In many cases a sample is indeed placed on a slide to make it available for investigation, e.g. a tissue slice, or a smear. However, other devices can also be used to carry a sample, e.g. a small cuvette in case the sample is a liquid sample or a cartridge in case the sample is a brush sample. The term slice as used in the flow charts is thus by no means construed to limiting the scope of the present invention.

In a particularly preferred embodiment of the present invention, the biological sample comprising cells is stained, preferably prior to step a) of image acquisition. Dyes which are preferably used comprise Pap-stain, ultra fast Pap-stain, Romanowsky-type stain, Haris Haematoxylin stain, fluorescent stains like Achrodyn Orange, and H & E stain.

In another preferred embodiment of the method according to the invention optical and/or digital image enhancement approaches are used.

Optical image enhancement is preferably carried out prior to step a) of image acquisition. Preferred methods, though non-limiting, comprise dark field microscopy, phase contrast, differential interference contrast (DIC) and/or reflected interference contrast (RIC). Digital contrast enhancement is preferably carried out after step a) of image acquisition.

Preferred methods comprise bright field microscopy, for example a typical transmission microscope.

In still another preferred embodiment of the method according to the invention steps b) and following are carried out while the data related to the acquired image, or parts thereof, is still in a volatile memory.

As used herein, the term volatile memory is used interchangeably with the term temporary memory, and shall be understood in such way that the data related to the acquired image are not yet stored on the hard disk or on a flash storage. A preferred form of such volatile memory is a random access memory (RAM) used by the image processor, or by the computer's CPU.

According to still another aspect of the invention a system for the detection of abnormalities in a biological sample comprising cells is provided. The system is capable to carry out, and/or comprises means capable to carry out, at least the following steps:

- a) acquiring an image of said sample by digital image acquisition
- b) optionally, carrying out digital image processing
- c) selecting a field of view
- d) determining, by digital image processing, whether or not, in said field of view, cell aggregates exist, and
- e1) selecting a new field of view and carrying on with step c) if, in step d), it turns out that, in said field of view, the determination of cell aggregates is negative, or
- e2) carrying out further process steps if, in step d), it turns out that, in said field of view, the determination of cell aggregates is affirmative.

It is particularly preferred that said system is further capable to carry out, and/or comprises further means capable to carry out said the steps of:

- f) determining the degree of variation of at least one given morphological feature of at least two cells in the field of view, and
- g1) classifying the sample as "likely to comprise abnormal cells" if the degree of variation of said morphological feature exceeds a predetermined threshold, or
- g2) classifying the sample as "likely to comprise normal cells" if the degree of variation of said morphological feature falls below a predetermined threshold.

Further, said system is preferably capable to carry out, and/or comprises further means capable to carry out the other method steps discussed above.

According to still another aspect of the invention a device for the detection of abnormalities in a biological sample comprising cells, is provided. The device comprises at

5 least the following items:

- a sample receiving unit for receiving the biological sample
- a digital image acquisition unit for acquiring a digital image of the sample
- a digital image processing unit for carrying out at least one of the method steps according to the invention, and
- 10 • a user interface comprising at least one output means and one input means.

Said output means is preferably a display, or a touch screen, while said input means is preferably an array of keys, or buttons, or a touchscreen.

It is important to mention that the method according to the invention is not restricted to the use in such device. It can also be used “stand-alone” for the detection of
15 locations/regions in a slide or set of images being suspected to contain images of abnormal cells.

In a preferred embodiment, the device according to the invention further comprises at least one optical magnification unit.

In another preferred embodiment, the device according to the invention further
20 comprises at least one interface for connecting the device with other equipment. Such interface is, preferably, a GSM interface, a 3G interface, a USB interface, a Bluetooth interface, a Firewire interface and/or a WiFi interface, hypertext terminal, etc.

In another preferred embodiment, the device according to the invention further comprises at least one sample collector and/or at least one cartridge in which the sample is
25 transferred, said cartridge being disposed for placement in the sample receiving unit.

The system are device discussed above is preferably in the form of a point of are device (POC). Preferably, it is provided as a handheld or desktop unit. Even more preferably, it is battery driven and/or portable.

According to another aspect of the invention, use of a method, a system and/or
30 a device according to the invention for at least one purpose selected from the group of

- cancer screening
- cancer diagnosis
- prediction with respect to a given therapy, and/or
- concomitant monitoring of a given cancer therapy

is provided.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments described hereinafter. In the drawings:

Fig. 1 shows a flow diagram identifying regions of interest for the presence or absence of cell aggregates.

Fig. 2 shows a flow diagram for distinguishing whether or not identified cell aggregates comprise abnormal cells or normal cells (termed “endocervical cells” here, which is, strictly speaking, clinically imprecise, because endocervical cells can also become malignant).

Fig. 3 shows an artificially created example of a Pap smear image created to explain the envisaged concept.

Fig. 4 shows the segmentation of the cells and the corresponding nuclei.

Fig. 5 shows results after clustering, i.e., the cluster number and an inner rectangle showing the region of interest for possible abnormality.

Fig. 6 shows the envisaged system according to the invention, and its work flow.

Figs. 7 and 8 show flow diagrams comprising an alternative approach to identify regions of interest.

Fig. 9 shows the Evolution of the Pap-smear nucleus segmentation techniques, from IGMT and IEMT to IGTMT and IRMT.

Fig. 10 shows flow diagrams of (A) IGMT and (B) IEMT technique for Pap-smear nucleus segmentation.

Fig. 11 shows multiple threshold levels in a Papsmear image histogram, and the resulting segmentation. In Fig. 11 (B), R_{t-1} is a single image region below $t-1^{\text{th}}$ threshold as seen in Fig. 11 (A), and R_t is the region having gray value in-between $t-1^{\text{th}}$ and t^{th} threshold as seen in Fig. 11 (A), and R_{t+1} is region having gray value in-between t^{th} and $t+1^{\text{th}}$ threshold as seen in Fig. 11 (A). The actual region boundary of the nucleus is marked by an arrow.

Fig. 12 shows again multiple threshold levels in a Papsmear image histogram, and the resulting segmentation. Fig. 12 (A) shows recursive multiple threshold levels T_{r1} to T_{rM} in ascending order between $t-1^{\text{th}}$ and $t+1^{\text{th}}$ threshold of Fig. 11 (A). Fig. 12 (B) shows regions thresholded by levels T_{r1} to T_{rM} .

Fig. 13 shows R_{t-1} and outside the region R_{t+1} presented as source and sink respectively indicating the flow of a directed graph.

Fig. 14 shows flow diagrams of (A) IRMT and (A) IGTMT for Pap-smear nucleus segmentation.

Fig. 15 shows a Pap-smear image showing overlapping nucleus along with their intensity variation (in white circle) and focused region (marked in black arrow) and unfocused region (marked in white arrow).

10 DETAILED DESCRIPTION OF EMBODIMENTS

While the invention has been illustrated and described in detail in the drawings and foregoing description, such illustration and description are to be considered illustrative or exemplary and not restrictive; the invention is not limited to the disclosed embodiments. Other variations to the disclosed embodiments can be understood and effected by those skilled in the art in practicing the claimed invention, from a study of the drawings, the disclosure, and the appended claims. In the claims, the word “comprising” does not exclude other elements or steps, and the indefinite article “a” or “an” does not exclude a plurality. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage. Any reference signs in the claims should not be construed as limiting the scope.

Figs. 1 and 2 show flow charts which have been described in the text already.

Fig. 3 shows an artificially created example of a Pap smear image to explain the concept of the present invention. In the case of abnormal cells the cytoplasm parts of the cells stick together, and the nuclei of the cells are very close to one another (small inter-nuclei distance).

In contrast thereto, normal cell nuclei are not close to one another even if the cytoplasm parts of these cells contact each other. This is a unique feature that has been observed in many images from different slides imaged as a part of the research project which stands behind the present invention.

Inter-nuclei distance mapping is thus used in the algorithmic identification process to identify abnormal cells, particularly in those regions of interest which have earlier been identified, by cell aggregation analysis, as suspicious.

It has been observed that cells which have only a small risk to become cancerous also tend to form aggregates. In this case, the variability in nucleus size between endocervical cells and abnormal cells from the ectocervix can be used as a distinguishing

feature, because in abnormal cell aggregates the nucleus size varies widely whereas in normal cell aggregates the nucleus sizes remain uniform.

Fig. 4 shows the result of the segmentation of the cells and the corresponding nuclei. Based on the subsequent algorithms, regions of interest can be determined which are highly suspicious to comprise abnormal cells.

Fig. 5 shows the results of such process clustering. Two different clusters have been identified, wherein cluster No 2 – marked by a rectangle – is likely to comprise abnormal cells, due to (i) extensive cell clustering, (ii) high variation of nuclei sizes, and (iii) small inter-nuclei distances.

Fig. 6 shows a system and work flow according to the present invention. The sample is collected by means of a sample collector 61, which may adopt the shape of a brush, similar to a Q-tip, and is transferred to a cartridge 62 which is disposed for placement in the sample receiving unit 63 of a device 64 according to the invention. The device is shown in form of a portable point of care device. Optionally, the sample can be stained in between.

The sample is then scanned by the optical magnification unit and the image acquisition device unit present in the device. Upon actuation of an input means 65, which is embodied, herein, as an array of keys, the device acquires an image from the sample in the cartridge and passes it to the algorithm embedded into the system. Then, the method steps according to the invention are accomplished, and results of the analysis are shown on output means 66, which is embodied, herein, as a display screen.

Figs. 7 and 8 show flow diagrams comprising a more elaborated approach to identify regions of interest which may be likely to comprise abnormal cells. In Fig. 7 the main focus is to identify the cellular regions and find cell aggregates. Once the latter are identified the algorithm passes the respective field of view to further analysis, as explained in Fig. 8. In Box 1 of Fig. 7, the system initiation is done. Here, essentially, the system resets and arrives at a reference position, clears all the buffers, etc. In Box 2 a cytology sample is loaded, in most cases as a slide (see above), and the initiation for cell count of all type of cells, such as normal/abnormal squamous cells and their types, and normal/abnormal endo-cervical cells, is initiated, and the respective result is stored. In Box 3, a scan strategy is chosen. The scan strategy involves the selection of step size of the movement of the sample in x and y direction and the focus depth (z direction) relative to the imaging devices (i.e., either the sample or the imaging device is moved), and the selection of magnification, contrast, etc. Further, the system starts reading the sample/slide, and the images are acquired. In Box 4, the quality parameter of the slide/images are estimated. In Box 5, the quality

parameters are checked for their adequacy (by comparing with a priori data). If the quality is not adequate, the system goes in the loop to check the scope for up-gradation in Box 13. If there is scope to upgrade the scan strategy (i.e, if the scan parameters are within the defined range of the system), then the system provides an on-the-fly feedback to change the scan strategy (Box 12). If there is no scope for upgrading the scan parameters, a report is generated (Box 14) and the systems stops or goes for the next slide (Box 15).

If in Box 5 the quality is found adequate, then the field of view in Box 6 is passed on for segmentation (Box 7), identification of clusters (Box 9) and verification of clusters (Box 10).

If no clusters are found then it is considered that the actual field of view comprises only normal squamous cell, and their number is counted. Here it should be noted that the cells which are not comprised in clusters are assumed to be normal, given the fact that significant abnormalities are rather found in clustered cells than in isolated cells. If in Box 10 clusters are found, they are suspected for abnormality and the system passes the image of the actual field of view, or other data related to the said clusters, for detailed analysis to link 1 in Fig. 8. If in the respective field of view isolated cells as well as clusters of cells have been determined, the image of the actual field of view, or other data related to the said clusters, are passed for further processing in Fig. 8, while isolated cells are just counted.

After detailed analysis in Fig. 8 (see below) the control is returned back to Fig. 7 at link 2. Here, the system checks if there is any other field of view left to be scanned in Box 11. If yes, the system scans the next field of view and repeats the above described process. If all the fields of view are completed the system generates a report on the actual sample (estimation of abnormality, its severity, type of carcinoma, number of cells etc.) and stops, or moves to the next sample in Box 16.

Images of a field of view comprising clusters, or other data related to the said clusters, are then passed on to the algorithm in Fig. 8. Here a more detailed analysis on the clusters is made. The morphological features in each cell of a given cluster are extracted to measure the variability of these features (Box 2 and Box 3). If the measurements are not in an abnormal range (Box 4), then the cell cluster is suspected to comprise normal endocervical cells (Box 11), hence a confirmatory test is done (Box 12). If the assumption is true then the count of normal endocervical cells is done in Box 14. If not, the sample is considered to comprise squamous cells, and a count for squamous cells is done in Box 13. Further, for endocervical cells, their geometric co-ordinates in the sample and the corresponding field of

view are stored because of their importance in interpretation according to the Bethesda system, which is a system for reporting cervical or vaginal cytologic diagnoses used for reporting Pap smear results. This will also help in reviewing the sample, or its image, by a pathologist/cytologist at a later time.

- 5 If, in Box 4, it is decided that the measurements fall under an abnormal range, additional advanced features (such as intensity variation, texture, fractal dimension of the nuclear boundary, jazziness of the membrane, etc.) are extracted in Box 5 and a confirmatory test is done in Box 6. If in Box 6 the abnormality of the clusters is not confirmed, then the cells are considered to be normal squamous cells, hence, their count is updated in Box 8 and
- 10 the control returns back to link 2 in Fig. 7. If, in Box 6, the abnormality of the cluster under consideration is confirmed it is further tested whether these cells are abnormal squamous cells or abnormal endocervical cells (Box 7). If they are not endocervical cells, the count of abnormal squamous cells is updated in Box 9. Otherwise, the count of abnormal endocervical cells is updated (Box 15). The geometric co-ordinates in the slide and the corresponding field
- 15 of view is stored for these categories of cells and the control returns back to link 2 in Fig. 7.