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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 April 2018 (12.04.2018)

W!P O | PCT

(10) International Publication Number
WO 2018/067234 Al

(51) International Patent Classification:
G06T 7/00 (2017.01) *G06T 7/136* (2017.01)
G06T 7/11 (20 17.0 1) *G06K 9/00* (2006 .01)

(21) International Application Number:
PCT/US20 17/047004

(22) International Filing Date:
15 August 2017 (15.08.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
15/288,776 07 October 2016 (07.10.2016) US

(71) Applicant: **SONY CORPORATION** [JP/JP]; 1-7-1 Konan, Minato-ku, Tokyo 108-0075 (JP).

(72) Inventor; and
(71) Applicant *(for US only):* **WANG, Su** [US/US]; 3040 Rosato Ct., San Jose, California 95 135 (US).

(72) Inventors: **XU, Xun;** 800 E. Charleston Rd., #7, Palo Alto, California 94303 (US). **NAKAMURA, Akira;** 13 15 Petal Way, San Jose, California 95 129 (US).

(74) Agent: **FUJII, Harold T.;** 16535 Via Esprillo, MZ1029, San Diego, California 92127 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.1 7(H))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.1 7(Hi))*

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

(54) Title: AUTOMATED NUCLEI AREA/NUMBER ESTIMATION FOR IHC IMAGE ANALYSIS

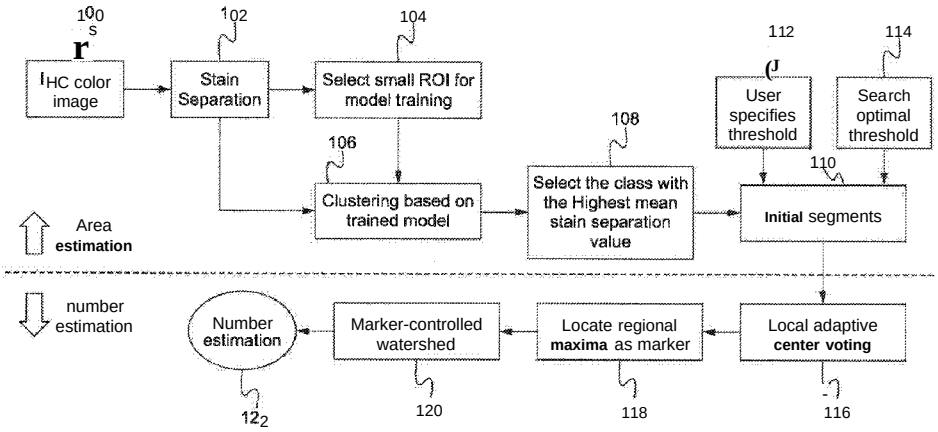

Fig. 1

(57) Abstract: Automated nuclei area/number estimation utilizes a two-stage estimation framework - area estimation first followed by number estimation. After determining area information, each local patch's shape features are able to be extracted to define a local voting rule. The resulting voting score determines the strength of each local voting peak. The number of voting peaks is exactly the number of nuclei.

WO 2018/067234 A1

AUTOMATED NUCLEI **AREA/NUMBER ESTIMATION** FOR IHC IMAGE ANALYSIS

FIELD OF THE INVENTION

The present invention relates to the field of imaging. More specifically, the present invention relates to medical imaging.

BACKGROUND OF THE INVENTION

Immunohistochemistry (IHC) refers to the process of detecting proteins in cells of a tissue section. IHC staining is widely used in the diagnosis of abnormal cells such as those found in cancerous tumors. Common practice in pathology laboratories is to score IHC-stained images. By indicating a tumor is negative or positive, the percentage of positively stained tumor cell nuclei is able to be reported, which is able to assist pathologists for the final scoring purpose.

Some research has been done for the percentage estimation of positively stained tumor cell nuclei. The goal is achieved by using a color de-convolution algorithm for separating the staining components (diammobenzidine and hematoxylin) and adaptive thresholding for nuclear area segmentation. The quantitative results are calibrated using cell counts defined visually as the gold standard.

Most of the nuclei area estimation algorithms require a user to manually specify a cut-off threshold value for defining positive/negative. Although this type of user interaction is allowed, it is able to be improved.

However, most of the existing research work is performing nuclei area estimation, which is fast in speed but does not provide nuclei number estimation. According to pathologists, number information is a plus and is able to be provide extra hints when scoring IHC-stained images.

Nuclei's shapes and image intensities vary significantly. Touching cases (e.g., when nuclei are connected with each other) makes the number estimation even more challenging. Under-estimation and over-estimation are two major issues when developing automated nuclei number estimation.

SUMMARY OF THE INVENTION

Automated nuclei area/number estimation utilizes a two-stage estimation framework - area estimation first followed by number estimation. After determining area information, each local patch's shape features are able to be extracted to define a local voting rule. The resulting voting score detemiines the strength of each local voting peak. The number of voting peaks is exactly the number of nuclei.

In one aspect, a method programmed in a non-transitory memory of a device comprises performing nuclei area estimation and performing nuclei number estimation for detecting abnormal cells. Performing nuclei area estimation comprises: receiving a color image. Performing nuclei area estimation comprises: utilizing stain separation to separate two dominating color components, a first color corresponding to positive stains and a second color corresponding to negative stains. Performing nuclei area estimation comprises: adaptive thresholding based on each color channel. Performing nuclei area estimation comprises: wherein a small region of interest is selected for model training. Performing nuclei area estimation comprises: utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via model training and selection. Perfonning nuclei area estimation comprises: applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution, and clustering is performed for a remaining area. Performing nuclei area estimation comprises: selecting a class with the highest mean stain separation value. Performing nuclei area estimation comprises: hard thresholds are applied to adaptively-enhanced stain separation images to determine a nuclear area. The thresholds are user-specified values. The thresholds are searched optimum values. Performing nuclei number estimation comprises: after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting. Performing nuclei number estimation comprises: voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center. Perfonning nuclei number estimation comprises: filtering peaks caused by artifacts using local shape-determined rules. Performing nuclei number estimation comprises:

determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei. The device comprises a personal computer, a laptop computer, a computer workstation, a server, a mainframe computer, a handheld computer, a personal digital assistant, a cellular/mobile telephone, a smart appliance, a gaming console, a digital camera, a digital camcorder, a camera phone, a smart phone, a portable music player, a tablet computer, a mobile device, a video player, a video disc writer/player, a high definition disc writer/player, an ultra high definition disc writer/player), a television, a home entertainment system, or a smart watch.

In another aspect, a method programmed in a non-transitory memory of a device comprises performing nuclei area estimation including: receiving a color image, utilizing stain separation to separate two dominating color components of the color image, a first color corresponding to positive stains and a second color corresponding to negative stains, performing adaptive thresholding based on each color channel, selecting a small region of interest for model training, utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via the model training and selection, applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution and performing clustering for a remaining area, selecting a class with the highest mean stain separation value, applying hard thresholds to the adaptively-enhanced stain separation image to determine a nuclear area and performing nuclei number estimation for detecting abnormal cells including: after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting, voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center, filtering peaks caused by artifacts using local shape-determined rules and determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei. The hard thresholds are user-specified values. The hard thresholds are searched optimum values. The device comprises a personal computer, a laptop computer, a computer workstation, a server, a mainframe computer, a handheld computer, a personal digital assistant, a cellular/mobile telephone, a smart appliance, a gaming console, a digital camera, a digital camcorder, a camera

phone, a smart phone, a portable music player, a tablet computer, a mobile device, a video player, a video disc writer/player, a high definition disc writer/player, an ultra high definition disc writer/player), a television, a home entertainment system, or a smart watch.

In another aspect, an apparatus comprises a non-transitory memory for storing an application, the application for: performing nuclei area estimation including: receiving a color image, utilizing stain separation to separate two dominating color components of the color image, a first color corresponding to positive stains and a second color corresponding to negative stains, performing adaptive thresholding based on each color channel, selecting a small region of interest for model training, utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via the model training and selection, applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution and performing clustering for a remaining area, selecting a class with the highest mean stain separation value and applying hard thresholds to the adaptively-enhanced stain separation image to determine a nuclear area and performing nuclei number estimation for detecting abnormal cells including: after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define mles for local center voting, voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center, filtering peaks caused by artifacts using local shape-determined rules and determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei and a processing component coupled to the memory, the processing component configured for processing the application. The hard thresholds are user-specified values. The hard thresholds are searched optimum values.

In yet another aspect, an apparatus comprises a non-transitory memory for storing an application, the application for: performing nuclei area estimation and performing nuclei number estimation for detecting abnormal cells and a processing component coupled to the memory, the processing component configured for processing the application. Performing nuclei area estimation includes: receiving a color image, utilizing stain separation to separate two dominating color components of the color image, a first color corresponding to positive stains and a second color corresponding to negative stains, performing adaptive thresholding based on each color channel, selecting a small region of interest for model training, utilizing a user-selected region of interest as training data to enliance the image quality of stain separation via the model training and selection, applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution and performing clustering for a remaining area, selecting a class with the highest mean stain separation value and applying hard thresholds to the adaptively-enhanced stain separation image to determine a nuclear area. Performing nuclei number estimation for detecting abnormal cells further includes: after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting, voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center, filtering peaks caused by artifacts using local shape-determined rales and determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei.

BRIEF DESCRIPTION OF THE **DRAWINGS**

FIG. 1 illustrates a flowchart of a method of automated nuclei area and number estimation according to some embodiments.

FIG. 2 illustrates a block diagram of an exemplary computing device configured to implement the automated nuclei area/number estimation method according to some embodiments.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

An automated nuclei area and number estimation method and system enable improved Immunohistochemistry (IHC) image analysis.

The automated nuclei area and number estimation system uses a two-stage estimation framework: nuclear area estimation (e.g., number of nuclear pixels) followed by nuclear number estimation. Nuclear area is estimated from a binarized patch or patches, and these segmented patches provide local shape features which are able to facilitate number estimation.

To better distinguish a nuclear target and artifacts, the image quality of stain separation is enhanced by performing adaptive clustering based on a user-selected Region of Interest (ROI) via model training/selection.

To estimate nuclei area, stain separation is applied to separate two dominating color components, one corresponding to positive stains and the other one corresponding to negative stains. Area estimation is performed based on each color channel by adaptive thresholding. The system intelligently utilizes a user-selected ROI as the training data to enhance the image quality of stain separation via model training and selection. Regarding model training, a Gaussian Mixture Model (GMM) is applied to parameterize user-selected data distribution, and clustering (model selection) is performed for the remaining area. To determine a nuclear area, adaptively-enhanced stain separation images are hard thresholded, either by user-specified values or searched optimum values.

Nuclei number estimation is based on the aforementioned nuclear area estimation. More specially, after detennining segmented patches, connected component analysis (CCA) is applied to analyze each local patch's shape. These shape features help define the rules for local center voting. Local center voting is an important algorithm to determine nuclei numbers. This gradient-based algorithm votes the center of each nuclear, such that the higher the voting score, the more likely to be a real nuclear center. And those local shape-determined rules help to filter out those peaks caused by artifacts. Finally, the number of voting peaks is exactly the number of nuclei. The following shape features are utilized (although others are able to be used):

- Convex ratio

- Major axis length/minor axis length

Figure 1 illustrates a flowchart of a method of automated nuclei area and number estimation according to some embodiments. In the step 100, an IHC color image is acquired. For example, the IHC color image is retrieved from a data storage. In the step 102, stain separation is utilized to separate two dominating color components, one corresponding to positive stains and the other one corresponding to negative stains. Stain separation is able to he performed in any maimer; for example, automatically detecting two different colors. Area estimation is performed based on each color channel by adaptive thresholding such that the threshold changes dynamically over the image. In the step 104, a small ROI (e.g., less than a specified percent of an image such as 2%) is seleeted for model training. The system intelligently utilizes a user-selected (or computer-selected) ROI as the training data to enhance the image quality of stain separation via model training and selection. Regarding model training, a Gaussian Mixture Model (GMM) is applied to parameterize user-selected data distribution, and clustering (model selection) is performed for the remaining area, in the step 106. The clustering generates clusters or classes based on mean stain separation values. in the step 108, the class with the highest mean stain separation value is selected. In the step 110, to determine the nuclear area, adaptively-enhanced stain separation images are hard ihresholded, either by user-specified values, in the step 112, or searched optimum values, in the step 114. For example, a specified threshold is utilized to estimate the nuclear area.

Nuclei number estimation is based on the aforementioned nuclear area estimation. More specially, after determining segmented patches, connected component analysis (CCA) is applied to analyze each local patch's shape. These shape features help define the rules for local center voting, in the step 116. Local center voting is an important algorithm to determine nuclei numbers. Local center voting is a gradient-based algorithm which votes on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center, in the step 118. The local shape-determined rules help to filter out the peaks caused by artifacts, in the step 120. Finally, the number of voting peaks is exactly the number of nuclei, in the step 122.

Figure 2 illustrates a block diagram of an exemplar}' computing device configured to implement the automated nuclei area/number estimation method according to some embodiments. The computing device 200 is able to be used to acquire, store, compute, process, communicate and/or display information such as images and videos. In general, a hardware structure suitable for implementing the computing device 200 includes a network interface 202, a memory 204, a processor 206, I/O device(s) 208, a bus 210 and a storage device 212. The choice of processor is not critical as long as a suitable processor with sufficient speed is chosen. The memory 204 is able to be any conventional computer memory known in the art. The storage device 212 is able to include a hard drive, CDROM, CDRW, DVD, **DVDRW,** High Definition disc/drive, ultra-HD drive, flash memory card or any other storage device. The computing device 200 is able to include one or more network interfaces 202. An example of a network interface includes a network card connected to an Ethernet or other type of LAN. The I/O device(s) 208 are able to include one or more of the following: keyboard, mouse, monitor, screen, printer, modem, touchscreen, button interface and other devices. Automated nuclei area'number estimation application(s) 230 used to perform the automated nuclei area'number estimation method are likely to be stored in the storage device 212 and memory 204 and processed as applications are typically processed. More or fewer components shown in Figure 2 are able to be included in the computing device 200. In some embodiments, automated nuclei area-'number estimation hardware 220 is included. Although the computing device 200 in Figure 2 includes applications 230 and hardware 220 for the automated nuclei area'number estimation method, the automated nuclei area/number estimation method is able to be implemented on a computing device in hardware, firmware, software or any combination thereof. For example, in some embodiments, the automated nuclei area/number estimation applications 230 are programmed in a memory and executed using a processor. In another example, in some embodiments, the automated nuclei area'number estimation hardware 220 is programmed hardware logic including gates specifically designed to implement the automated nuclei area/number estimation method.

In some embodiments, the automated nuclei area'number estimation applieation(s) 230

include several applications and/or modules. In some embodiments, modules include one or more sub-modules as well. In some embodiments, fewer or additional modules are able to be included.

Examples of suitable computing devices include a personal computer, a laptop computer, a computer workstation, a server, a mainframe computer, a handheld computer, a personal digital assistant, a cellular/mobile telephone, a smart appliance, a gaming console, a digital camera, a digital camcorder, a camera phone, a smart phone, a portable music player, a tablet computer, a mobile device, a video player, a video disc writer/player (e.g., DVD writer/player, high definition disc writer/player, ultra high definition disc writer/player), a television, a home entertainment system, smart jewelry (e.g., smart watch) or any other suitable computing device.

To utilize automated nuclei area/number estimation, an 1HC color image is analyzed using automated nuclei area/number estimation. Based on the results of automated nuclei area/number estimation, a pathologist is able to further determine if a tumor is cancerous or not.

In operation, automated nuclei area/number estimation provides a benefit based on its two-stage estimation framework - area estimation first followed by number estimation. After determining area information, each iocal patch's shape features are able to be extracted to define a local voting rule. The resulting voting score determines the strength of each local voting peak. The number of voting peaks is exactly the number of nuclei.

SOME EMBODIMENTS OF AUTOMATED NUCLEI **AREA/NUMBER ESTIMATION** FOR IHC MAGE ANALYSIS

1. A method programmed in a non-transitory memory of a device comprising:
   a. performing nuclei area estimation; and
   b. performing nuclei number estimation for detecting abnormal cells.

2. The method of clause 1 wherein performing nuclei area estimation comprises: receiving a color image.

3. The method of clause 2 wherein performing nuclei area estimation comprises: utilizing stain separation to separate two dominating color components, a first color corresponding to positive stains and a second color corresponding to negative stains.

4. The method of clause 3 wherein performing nuclei area estimation comprises: adaptive thresholding based on each color channel.

5. The method of clause 4 wherein performing nuclei area estimation comprises: wherein a small region of interest is selected for model training.

6. The method of clause 5 wherein performing nuclei area estimation comprises: utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via model training and selection.

7. The method of clause 6 wherein performing nuclei area estimation comprises: applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution, and clustering is performed for a remaining area.

8. The method of clause 7 wherein performing nuclei area estimation comprises: selecting a class with the highest mean stain separation value.

9. The method of clause 8 wherein performing nuclei area estimation comprises: hard thresholds are applied to adaptively-enhanced stain separation images to determine a nuclear area.

10. The method of clause 9 wherein the thresholds are user-specified values.

11. The method of clause 9 wherein the thresholds are searched optimum values.

12. The method of clause 1 wherein performing nuclei number estimation comprises: after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting.

13. The method of clause 12 wherein performing nuclei number estimation comprises: voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center.

14. The method of clause 13 wherein performing nuclei number estimation comprises: filtering peaks caused by artifacts using local shape-determined rules.

15. The method of clause 14 wherein performing nuclei number estimation comprises: determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei.

16. The method of clause 1 wherein the device comprises a personal computer, a laptop computer, a computer workstation, a server, a mainframe computer, a handheld computer, a personal digital assistant, a cellular/mobile telephone, a smart appliance, a gaming console, a digital camera, a digital camcorder, a camera phone, a smart phone, a portable music player, a tablet computer, a mobile device, a video player, a video disc writer/player, a high definition disc writer/player, an ultra high definition disc writer/player), a television, a home entertainment system, or a smart watch.

17. A method programmed in a non-transitory memory of a device comprising:
   a. performing nuclei area estimation including:
      i. receiving a color image;
      ii. utilizing stain separation to separate two dominating color components of

the color image, a first color corresponding to positive stains and a second color corresponding to negative stains;

iii. performing adaptive thresholding based on each color channel;

iv. selecting a small region of interest for model training;

v. utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via the model training and selection;

vi. applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution and performing clustering for a remaining area;

vii. selecting a class with the highest mean stain separation value;

viii. applying hard thresholds to the adaptiveiy-enhanced stain separation image to determine a nuclear area; and

h. performing nuclei number estimation for detecting abnormal cells including:

i. after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting;

ii. voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center;

iii. filtering peaks caused by artifacts using local shape-determined rules; and

iv. determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei.

18. The method of clause 17 wherein the hard thresholds are user-specified values.

19. The method of clause 17 wherein the hard tliresliolds are searched optimum values,

20. The method of clause 17 wherein the device comprises a personal computer, a laptop computer, a computer workstation, a server, a mainframe compuier, a handheld computer, a personal digital assistant, a cellular/mobile telephone, a smart appliance, a

gaming console, a digital camera, a digital camcorder, a camera phone, a smart phone, a portable music player, a tablet computer, a mobile device, a video player, a video disc writer/player, a high definition disc writer/player, an ultra high definition disc writer/player), a television, a home entertainment system, or a smart watch.

21. An apparatus comprising:

a. a non-transitory memory for storing an application, the application for:

i. performing nuclei area estimation including:

(1) receiving a color image;

(2) utilizing stain separation to separate two dominating color components of the color image, a first color corresponding to positive stains and a second color corresponding to negative stains;

(3) performing adaptive thresholding based on each color channel;

(4) selecting a small region of interest for model training;

(5) utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via the model training and selection;

(6) applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution and performing clustering for a remaining area;

(7) selecting a class with the highest mean stain separation value; and

(8) applying hard thresholds to the adaptively-enhanced stain separation image to determine a nuclear area; and

ii. performing nuclei number estimation for detecting abnormal cells including:

(1) after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting;

(2) voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center;

(3) filtering peaks caused by artifacts using local shape-determined rules; and

(4) determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei; and

b. a processing component coupled to the memory, the processing component configured for processing the application.

22. The apparatus of clause 21 wherein the hard thresholds are user-specified values.

23. The apparatus of clause 21 wherein the hard thresholds are searched optimum values.

24. An apparatus comprising:

a. a non-transitory memory for storing an application, the application for:

i. performing nuclei area estimation; and

ii. performing nuclei number estimation for detecting abnormal cells; and

b. a processing component coupled to the memory, the processing component configured for processing the application.

25. The apparatus of clause 24 wherein performing nuclei area estimation includes:

(1) receiving a color image;

(2) utilizing stain separation to separate two dominating color components of the color image, a first color corresponding to positive stains and a second color corresponding to negative stains;

(3) performing adaptive thresholding based on each color channel;

(4) selecting a small region of interest for model training;

(5) utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via the model training and selection;

(6) applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution and performing clustering for a remaining area;

(7) selecting a class with the highest mean stain separation value; and

(8) applying hard thresholds to the adaptively-enhanced stain separation image to determine a nuclear area.

26. The apparatus of clause 24, wherein performing nuclei number estimation for detecting abnormal cells further includes;

(1) after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting;

(2) voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center;

(3) filtering peaks caused by artifacts using local shape-determined rules; and

(4) determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei.

The present invention has been described in terms of specific embodiments incorporating details to facilitate the understanding of principles of construction and operation of the invention. Such reference herein to specific embodiments and details thereof is not intended to limit the scope of the claims appended hereto. It will be readily apparent to one skilled in the art that other various modifications may be made in the embodiment chosen for illustration without departing from the spirit and scope of the invention as defined by the claims.

CLAIMS

What is claimed is:

1. A method programmed in a non-transitory memory of a device comprising:
   a. performing nuclei area esiimation; and
   b. performing nuclei number estimation for detecting abnormal cells.

2. The method of claim 1 wherein performing nuclei area estimation comprises: receiving a color image.

3. The method of claim 2 wherein performing nuclei area estimation comprises: utilizing stain separation to separate two dominating color components, a first color corresponding to positive stains and a second color corresponding to negative stains.

4. The method of claim 3 wherein performing nuclei area estimation comprises: adaptive thresholding based on each color channel.

5. The method of claim 4 wherein performing nuclei area estimation comprises: wherein a small region of interest is selected for model training.

6. The method of claim 5 wherem performing nuclei area esiimation comprises: utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via model training and selection.

7. The method of claim 6 wherem performing nuclei area estimation comprises: applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution, and clustering is performed for a remaining area.

8. The method of claim 7 wherein performing nuclei area estimation comprises: selecting a class with the highest mean stain separation value.

9. The method of claim 8 wherein performing nuclei area estimation comprises: hard thresholds are applied to adaptively-enhanced stain separation images to determine a nuclear area.

10. The method of claim 9 wherein the thresholds are user-specified values.

11. The method of claim 9 wherein the thresholds are searched optimum values.

12. The method of claim 1 wherein performing nuclei number estimation comprises: after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting.

13. The method of claim 12 wherein performing nuclei number estimation comprises: voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center.

14. The method of claim 13 wherein performing nuclei number estimation comprises: filtering peaks caused by artifacts using local shape-determined rules.

15. The method of claim 14 wherein performing nuclei number estimation comprises: determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei.

16. The method of claim 1 wherein the device comprises a personal computer, a laptop computer, a computer workstation, a server, a mainframe computer, a handheld computer, a personal digital assistant, a cellular/mobile telephone, a smart appliance, a gaming console, a digital camera, a digital camcorder, a camera phone, a smart phone, a portable music player, a tablet computer, a mobile device, a video player, a video disc writer/player, a high definition disc writer/player, an ultra high definition disc writer/player), a television, a home entertainment system, or a smart watch.

17. A method programmed in a non-transitory memory of a device comprising:

a. performing nuclei area estimation including:

i. receiving a color image;

ii. utilizing stain separation to separate two dominating color components of the color image, a first color corresponding to positive stains and a second color corresponding to negative stains;

iii. performing adaptive thresholding based on each color channel;

iv. selecting a small region of interest for model training;

v. utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via the model training and selection;

vi. applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution and performing clustering for a remaining area;

vii. selecting a class with the highest mean stain separation value;

viii. applying hard thresholds to the adaptively-enhanced stain separation image to determine a nuclear area; and

b. performing nuclei number estimation for detecting abnormal cells including:

i. after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting;

ii. voting on the center of each nuclear area, such that the higher the voting

score, the more likely to be a real nuclear center;

iii. filtering peaks caused by artifacts using local shape-determined rules; and

iv. determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei.

18. The method of claim 17 wherein the hard thresholds are user-specified values.

19. The method of claim 17 wherein the hard thresholds are searched optimum values.

20. The method of claim 17 wherein the device comprises a personal computer, a laptop computer, a computer workstation, a server, a mainframe computer, a handheld computer, a personal digital assistant, a cellular/mobile telephone, a smart appliance, a gaming console, a digital camera, a digital camcorder, a camera phone, a smart phone, a portable music player, a tablet computer, a mobile device, a video player, a video disc writer/player, a high definition disc writer/player, an ultra high definition disc writer/player), a television, a home entertainment system, or a smart watch.

21. An apparatus comprising:

a. a non-transitory memory for storing an application, the application for:

i. performing nuclei area estimation including:

(1) receiving a color image;

(2) utilizing stain separation to separate two dominating color components of the color image, a first color corresponding to positive stains and a second color corresponding to negative stains;

(3) performing adaptive thresholding based on each color channel;

(4) selecting a small region of interest for model training;

(5) utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via the model

training and selection;

(6) applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution and performing clustering for a remaining area;

(7) selecting a class with the highest mean stain separation value; and

(8) applying hard thresholds to the adaptively-enhanced stain separation image to determine a nuclear area; and

ii. performing nuclei number estimation for detecting abnormal cells including:

(1) after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting;

(2) voting on the center of each nuclear area, such that the higher the voting score, the more likely to be a real nuclear center;

(3) filtering peaks caused by artifacts using local shape-determined rules; and

(4) determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei; and

b. a processing component coupled to the memory, the processing component configured for processing the application.

22. The apparatus of claim 21 wherein the hard thresholds are user-specified values.

23. The apparatus of claim 21 wherein the hard thresholds are searched optimum values.

24. An apparatus comprising:

a. a non-transitory memory for storing an application, the application for:

i. performing nuclei area estimation; and

si. performing nuclei number estimation for detecting abnormal cells; and

b. a processing component coupled to the memory, the processing component configured for processing the application.

25. The apparatus of claim 24 wherein performing nuclei area estimation includes:

(1) receiving a color image;

(2) utilizing stain separation to separate two dominating color components of the color image, a first color corresponding to positive stains and a second color corresponding to negative stains;

(3) performing adaptive thresholding based on each color channel;

(4) selecting a small region of interest for model training;

(5) utilizing a user-selected region of interest as training data to enhance the image quality of stain separation via the model training and selection;

(6) applying a Gaussian Mixture Model (GMM) to parameterize user-selected data distribution and performing clustering for a remaining area;

(7) selecting a class with the highest mean stain separation value; and

(8) applying hard thresholds to the adaptively-enhanced stain separation image to determine a nuclear area.

26. The apparatus of claim 24, wherein performing nuclei number estimation for detecting abnormal cells further includes:

(1) after receiving segmented patches, connected component analysis is applied to analyze each local patch's shape which define rules for local center voting;

(2) voting on the center of each nuclear area, such that the higher the

voting score, the more likely to be a real nuclear center;

(3) filtering peaks caused by artifacts using local shape-determined rules; and

(4) determining the number nuclei based on the number of voting peaks, wherein the number of nuclei is the same as the number of nuclei.

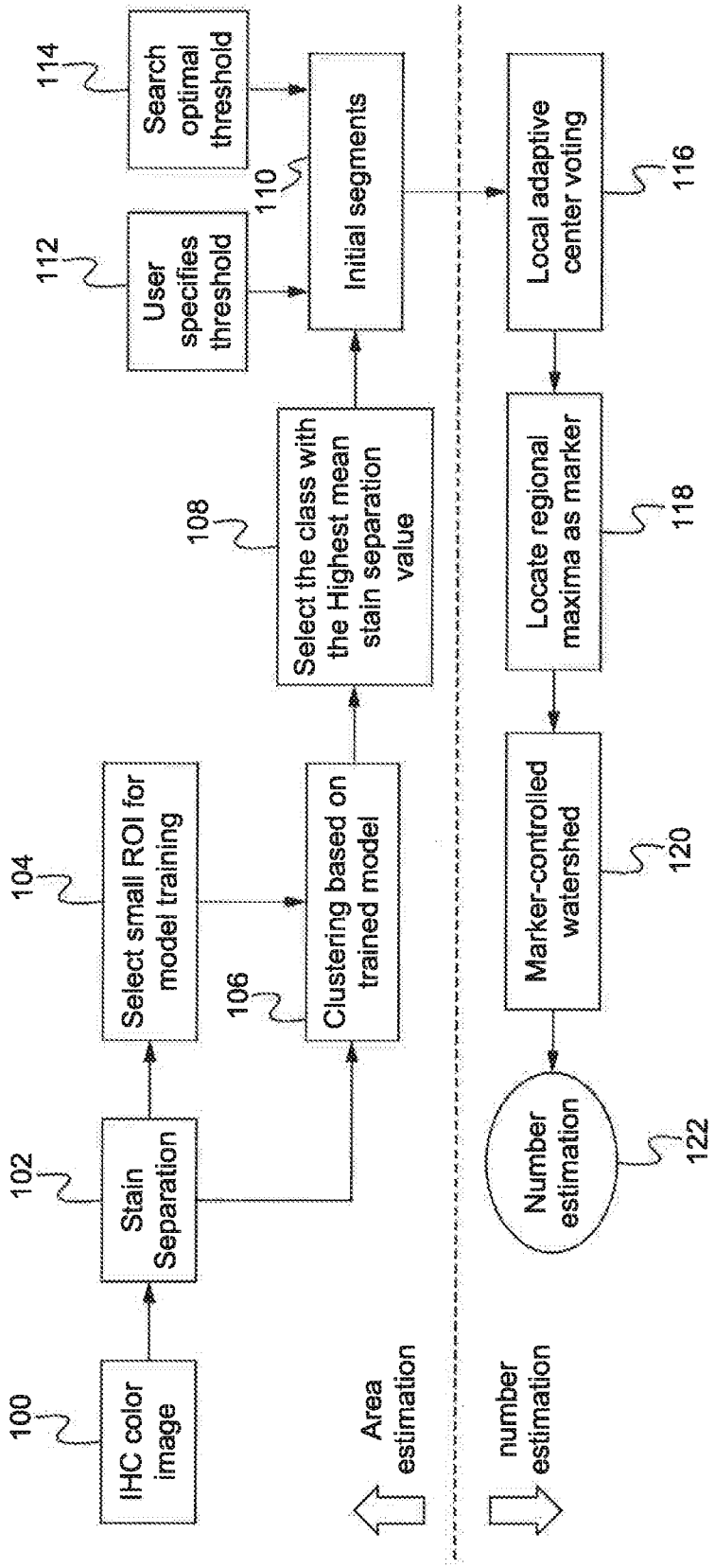
100
IHC color image
102
Stain Separation
104
Select small ROI for model training
106
Clustering based on trained model
108
Select the class with the Highest mean stain separation value
110
Initial segments
112
User specifies threshold
114
Search optimal threshold
Area estimation
number estimation
116
Local adaptive center voting
118
Locate regional maxima as marker
120
Marker-controlled watershed
122
Number estimation

Fig. 1

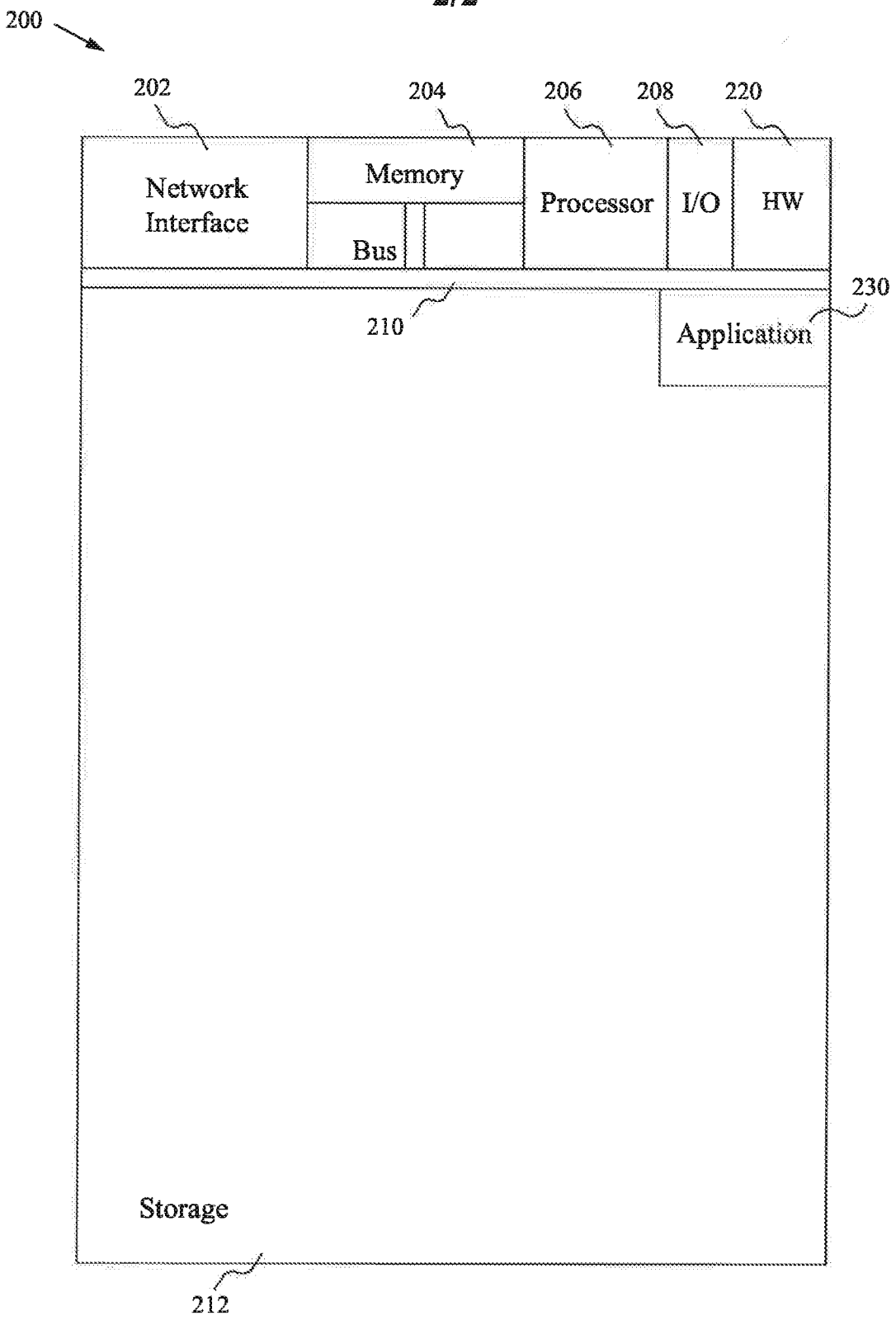
200
202
204
206
208
220
Network Interface
Memory
Bus
Processor
I/O
HW
210
230
Application
Storage
212

Fig. 2

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/047004

A. CLASSIFICATION OF SUBJECT MATTER
INV. G96T7/0O G06T7/11 G06T7/136 G06K9/09
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06T G06K

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, COMPENDEX, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2016/120442 A1 (VENTANA MED SYST INC [US]; HOFFMANN LA ROCHE [CH]] 4 August 2016 (2016-08-04) paragraphs [0009], [0010], [0050], [0063], [0088], [0090], [0101] figure 10 -----	1-26

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

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

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

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

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

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

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 October 2017	14/11/2017
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Winkler, Gregor

INTERNATIONAL SEARCH REPORT

Information on patent Tamlly members

International application No
PCT/US2017/047004

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2016120442 Al	04-08-2016	NONE	