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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 October 2024 (24.10.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/220521 A1

(51) International Patent Classification:
G06V 20/69 (2022.01)

(21) International Application Number:
PCT/US2024/024967

(22) International Filing Date:
17 April 2024 (17.04.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/460,879 20 April 2023 (20.04.2023) US

(71) Applicant: **LIFE TECHNOLOGIES CORPORATION** [US/US]; 5823 Newton Drive, Carlsbad, California 92008 (US).

(72) Inventors: **CHANDA, Deepankar**; Life Technologies Corporation, 5823 Newton Drive, Carlsbad, California 92008 (US). **LIU, Sitong**; Life Technologies Corporation, 5823 Newton Drive, Carlsbad, California 92008 (US). **COX, Conor**; Life Technologies Corporation, 5823 Newton Drive, Carlsbad, California 92008 (US). **SARKAR, Saradwata**; Life Technologies Corporation, 5823 Newton Drive, Carlsbad, California 92008 (US).

(74) Agent: **BAI, Rong**; Attn: IP Legal Dept., 5823 NEWTON DRIVE, Carlsbad, California 92008 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,

(54) Title: QUALITY CONTROL AUTOMATION IN POLYMERASE CHAIN REACTION ANALYSIS

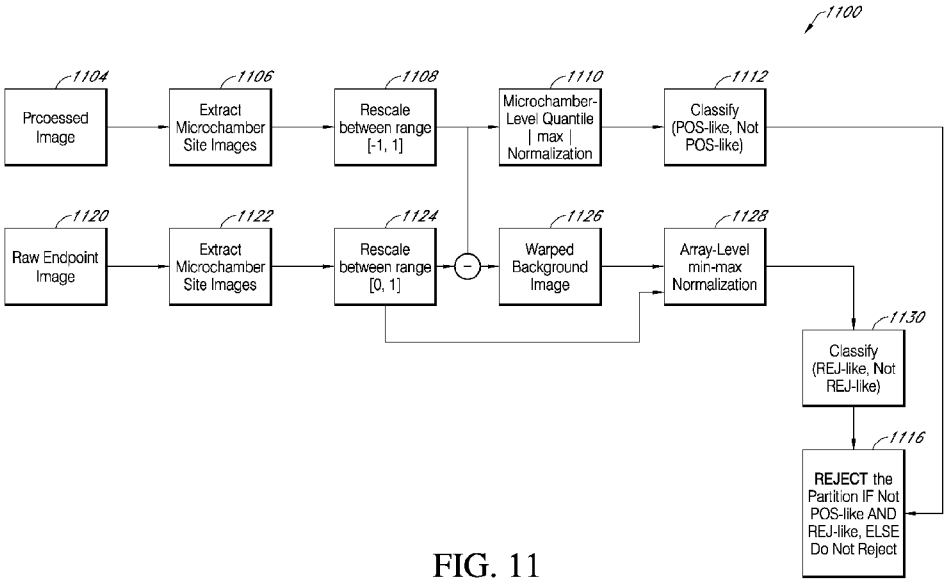

FIG. 11

(57) Abstract: Methods and systems are disclosed for computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a target concentration result. Image data representing a plurality of partitions disposed in a container is obtained, including background image data captured prior to amplification cycles of a dPCR assay, and endpoint image data captured after amplification. Individual partition image data for the plurality of partitions is extracted from the image data. The partition background image data and the partition endpoint image data are pre-processed to obtain processed image data for input into one or more machine learning models, which are used to generate classification results. The one or more classification results are used to determine whether to reject image data corresponding to the partition of the plurality of partitions from use in computing the target concentration result.

WO 2024/220521 A1

[Continued on next page]

MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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:

— *of inventorship (Rule 4.17(iv))*

Published:

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

QUALITY CONTROL AUTOMATION IN POLYMERASE CHAIN REACTION ANALYSIS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of US Provisional Application No. 63/460,879 filed April 20, 2023. This application also has some subject matter relationship to commonly assigned Provisional Application numbers: US63/460,882; US63/460,884; and US63/460,885, all filed on April 20th, 2023. The contents of these applications are incorporated herein by reference in their entirety.

TECHNICAL FIELD

[0002] This disclosure relates generally to monitoring, measuring, and/or analyzing biological and biochemical reactions, and more specifically to inventive methods of automating data quality control during polymerase chain reaction analysis for quantifying target concentrations in a biological sample.

BACKGROUND

[0003] Polymerase chain reaction (PCR) is a technique used to quantify biological molecules in a sample. Generally, PCR amplifies nucleic acids with the DNA polymerase enzyme responsible for forming new copies of DNA. Based on the theory that such amplification is exponential, a specific segment of DNA, e.g., nucleic acid molecule or nucleotide sequence, can be amplified millions or billions of times using PCR, producing enough copies to be analyzed based on optical signals associated with one or more different color channels corresponding to one or more fluorescent label associated with "dyes" whose detectable signal increases with amplification of a target in a sample.

[0004] "Analog" quantification (e.g., in analog PCR) relies on extrapolating measurements based on measured patterns. For example, a target analyte may be quantified by comparing the number of amplification cycles and amount of PCR end-product to those of a reference sample. However, this type of quantification can be complicated by uncertainties and inaccuracies. Detection efficiency in a test sample may be different from that of reference samples. For example, in PCR, initial amplification cycles may not be exponential, and PCR amplification may plateau after an uncertain number of cycles. Particularly, low initial concentrations of target analytes may be missed completely when they do not amplify to detectable levels.

[0005] "Digital" quantification methods (e.g., digital PCR or dPCR) are a biotechnological refinement of analog methods offering more robust absolute quantification of analytes with higher accuracy and precision than analog methods. Digital quantification is more adept at detecting and quantifying concentrations of hard to detect rare targets, providing a more precise quantitation of samples or analysis (e.g., nucleotide sequences), and measuring low fold changes in analyte concentration. Consequently, digital quantification has many applications in basic research, clinical diagnostics, and environmental testing. For example, digital PCR has been applied to pathogen detection and cancer monitoring, copy number variation analysis, single gene expression analysis, rare sequence detection, gene expression profiling and single-cell analysis, detection of DNA contaminants in bioprocessing, validation of gene edits, and detection of specific methylation changes in DNA as biomarkers of cancer.

[0006] In contrast to an analog measurement that relies on extrapolating certain measurements based on measured patterns (e.g., exponential amplification cycles), digital quantification methods can quantitatively and discretely measure a certain analyte. Digital quantification can be performed on biological samples that contain or are suspected to contain a target analyte of interest, such as a cell, tissue, or specimen such as hair, a biological fluid such as blood, urine, saliva, etc., a

cell cluster such as a microbial colony, or an organism, cell, microbe, bacterium, virus, protein, antibody, or nucleic acids such as such as DNA or RNA molecules. Target analytes include "original" analytes that were originally present in the biological sample as well any "synthetic" analytes that are indicative of the presence of original analytes which may be added or generated during detection, including PCR amplicons, antigen-antibody complexes, etc. Digital quantification (e.g., digital PCR) begins with a sample including a relatively small number of a target analyte, e.g., a polynucleotide or nucleotide sequence template DNA (or RNA). The sample is partitioned into a large number of smaller test samples, which will ideally contain either one target analyte or none of the target analytes such that a separate detection reaction can be conducted in each partition individually. Suitable partitions are individual targets that are sufficiently distanced from other individual targets to allow for individual detection or quantification, which may or may not be fluidically isolated from each other. Partitions may or may not include separating barriers such as walls or membranes or liquids that are immiscible with the sample, or semisolid media. Exemplary partitions include individually distanced targets, e.g., deposited on a substrate such as a glass slide, a tube, open or closed well, droplet, vesicle, chamber or bead, or any representation of an individual signal derived from a target that is distinguishable over background or noise, for example a bright spot over a darker background in a digital or analog image. In digital PCR methods, when the samples are thermally cycled using a PCR apparatus, the samples containing the target concentration are amplified and produce a positive detection signal, while the samples that do not contain the target concentration are not amplified and produce no detection signal. After multiple PCR amplification cycles, the samples are imaged and analyzed for fluorescence, which is used to quantify the target concentration in the samples.

[0007] In digital PCR, biological samples may be affected by systemic failures or spurious external artifacts present on the partitions of the array holding the sample. Such failures may result from physical or chemical processes occurring during the digital PCR process. Errors may be introduced when measuring target

concentrations due to spurious external artifacts such as debris or dust present in or on the array partitions, for example. Bubbles or gas may also be present in some partitions of the array which may introduce errors. Array partitions may also be subject to systemic failures in the entire array such as delamination, leaks, or fill issues in the partitions.

[0008] In addition, optical saturation errors may occur when the signal from an imaged channel exceeds the operating range of the detecting sensor on the PCR apparatus. As a consequence, the digital images of the partitions may appear excessively bright or saturated. Bridging phenomenon observed as occurring across an array of partitions may also result in data quality errors for particular biological samples.

SUMMARY

[0009] Particular processes for improving the accuracy of quantifying target concentrations in a biological sample using digital quantification, including processes for automating quality control by determining which partitions to accept for analysis, are disclosed herein. Identifying partitions with systemic failures or spurious external artifacts can present a variety of technical challenges that can adversely affect the goal of obtaining useful test results. Methods of automating quality control to flag such partitions can help identify the level of severity and type of failure and improve the accuracy of results from a dPCR sample.

[0010] Various computer-implemented systems, methods, and articles of manufacture for classifying partitions whether to accept or reject them for use in analyzing one or more target concentrations in a biological sample using an analyte detection (e.g., a dPCR) apparatus, and for training a machine-learning model used for classifying partitions used in analyzing one or more biological samples by an analyte detection (e.g., a dPCR) apparatus, are described herein.

[0011] In one embodiment, a method of automating quality control when quantifying one or more target concentrations in a biological sample using an

analyte detection (e.g., a PCR) apparatus configured to analyze an array of partitions (e.g., an array of about 1000-5000, 10,000-50,000, about 100,000, or even significantly more than 100,000 partitions) of the biological sample is provided.

[0012] In one embodiment, methods and systems are disclosed for computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a target concentration result. Image data representing a plurality of partitions disposed in a container is obtained, including background image data captured prior to amplification cycles of a dPCR assay, and endpoint image data captured after amplification. Individual partition image data for the plurality of partitions is extracted from the image data. The partition background image data and the partition endpoint image data are pre-processed to obtain processed image data for input into one or more machine learning models, which are used to generate classification results. The one or more classification results are used to determine whether to reject image data corresponding to the partition of the plurality of partitions from use in computing the target concentration result.

[0013] In one embodiment, a method of training a machine-learning model used for analyzing one or more biological samples by an analyte detection (e.g., a PCR) apparatus is provided. The method is performed by one or more computing devices and comprises obtaining a first plurality of images identified as reject partition images and obtaining a second plurality of images identified as non-reject partition images. The method further comprises generating one or more datasets using the first plurality of images and the second plurality of images and determining, by the one or more computing devices, a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets. A trained machine-learning model is configured based on the set of parameters to analyze one or more target concentrations in the one or more biological samples.

[0014] Various objects, features, aspects, and advantages of the inventive subject matter will become more apparent from the following specification, along with the accompanying drawings in which like numerals represent like components.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0016] FIG. 1 illustrates a block diagram of a PCR apparatus and an array of partitions disposed in a microfluidic array plate.

[0017] FIG. 2 illustrates a flowchart of an exemplary workflow utilizing microfluidic array plate (MAP) technology.

[0018] FIG. 3 illustrates a flowchart of an exemplary data source analysis pipeline.

[0019] FIG. 4 illustrates a plurality of exemplary pre-PCR images and post-PCR images representing one or more arrays of partitions disposed in a microfluidic array plate.

[0020] FIG. 5A illustrates an image of an array of partitions and template images used to determine corner locations of the array.

[0021] FIG. 5B illustrates exemplary visual indications of a match between a template image and an edge.

[0022] FIG. 6A illustrates an array of partitions represented in an image.

[0023] FIG. 6B illustrates a zoomed-in image representing a corner area of the array in FIG. 6A.

[0024] FIG. 7 illustrates four examples of arrays of partitions, one representing a normal sample, and three representing various quality control issues.

[0025] FIG. 8 illustrates examples of normal partition images and examples of partition images showing data quality issues.

[0026] FIG. 9 illustrates examples of graphs of dPCR fluorescence data across various channels showing normal fluorescence data and quality control-afflicted data.

[0027] FIG. 10A is a flowchart illustrating a method for a quality control workflow for an array of partitions according to various embodiments.

[0028] FIG. 10B is a flowchart illustrating a method for conducting blank image error check according to various embodiments.

[0029] FIG. 11 is a flowchart illustrating a method for classifying individual microchamber site images in an array according to various embodiments.

[0030] FIG. 12 is a flowchart illustrating a method for generating a processed image according to various embodiments.

[0031] FIG. 13 is a flowchart illustrating a method for classifying partition images according to various embodiments.

[0032] FIG. 14 is a flowchart illustrating a method for training a machine learning model according to various embodiments.

[0033] FIG. 15 illustrates a block diagram of a computer system that can be used for implementing one or more aspects of the various embodiments.

[0034] FIG. 16A is a flowchart illustrating a method for generating a synthetic array image according to various embodiments.

[0035] FIG. 16B is a flowchart illustrating a method for computing microchamber or partition color according to various embodiments.

[0036] While the invention is described with reference to the above drawings, the drawings are intended to be illustrative, and other embodiments are consistent with the spirit, and within the scope, of the invention.

DETAILED DESCRIPTION

[0037] To provide a more thorough understanding of the present invention, the following description sets forth numerous specific details, such as specific configurations, parameters, examples, and the like. It should be recognized, however, that such description is not intended as a limitation on the scope of the present invention but is intended to provide a better description of the exemplary embodiments.

[0038] As used herein, the term "biological sample" means a sample or solution containing any type of biological chemical or component and/or any target molecule of interest to a user, manufacturer, or distributor of the various embodiments of the present invention described or implied herein, as well as any sample or solution containing related chemicals or compounds used for the purpose of conducting a biological assay, experiment, or test. These biological chemicals, components, or target molecules may include, but are not limited to, DNA sequences (including cell-free DNA), RNA sequences, genes, oligonucleotides, molecules, proteins, biomarkers, cells (e.g., circulating tumor cells), or any other suitable target biomolecule. A biological sample may comprise one or more of at least one target nucleic acid sequence, at least one primer, at least one buffer, at least one nucleotide, at least one enzyme, at least one detergent, at least one blocking agent, or at least one dye, marker, and/or probe suitable for detecting a target or reference nucleic acid sequence. In various embodiments, such biological components may be used in conjunction with one or more PCR methods and systems in applications such as fetal diagnostics, multiplex dPCR, viral detection, and quantification standards, genotyping, sequencing assays, experiments, or protocols, sequencing validation, mutation detection, detection of genetically modified organisms, rare allele detection, and/or copy number variation.

[0039] In various embodiments, the devices, instruments, systems, and methods described herein may be used to detect one or more types of biological components of interest. These biological components of interest may be any suitable biological target including, but are not limited to, DNA sequences (including cell-free DNA), RNA sequences, genes, oligonucleotides, molecules, proteins, biomarkers, cells (e.g., circulating tumor cells), or any other suitable target biomolecule.

[0040] In various embodiments, such biological components may be used in conjunction with various digital PCR methods and systems in applications such as multiplex digital PCR, viral detection and quantification standards, genotyping, sequencing validation, mutation detection, detection of genetically modified organisms, fetal diagnostics, rare allele detection, and copy number variation.

[0041] Embodiments of the present disclosure are generally directed to devices, instruments, systems, and methods for automating quality control while measuring or quantifying a biological reaction for a large number of partitions of samples. The partitions can include droplets, small sample volumes contained in microfluidic structure including microchambers, microwells, through-holes, or any combinations thereof.

[0042] The term "min-max normalization" as the term is used in the specification and/or claims means the maximum value in the data is mapped to a fixed chosen maximum and the minimum value in the data is mapped to a fixed chosen minimum.

[0043] The term "rescale", "rescaling", or any variation used in the specification and/or claims, refers to the process of modifying the size of range of pixel values in the image.

[0044] The term "warp", "warping", "warpage", or any variation used in the specification and/or claims refer to the process of modifying the image by

applying an affine transformation to match points that are common to the given image and a reference image.

[0045] While generally applicable to digital quantification such as PCR, it should be recognized that any other suitable quantification method may be used in accordance with various embodiments described herein. Suitable PCR methods include, but are not limited to, digital PCR, allele-specific PCR, asymmetric PCR, ligation-mediated PCR, multiplex PCR, nested PCR, qPCR, genome walking, and bridge PCR, for example.

[0046] As used herein, thermal cycling may include using a thermal cycler, isothermal amplification, thermal convention, infrared mediated thermal cycling, or helicase dependent amplification, for example.

[0047] According to various embodiments, detection of a target may be, but is not limited to, fluorescence detection, detection of positive or negative ions, pH detection, voltage detection, or current detection, alone or in combination, for example.

[0048] Various embodiments described herein are particularly suited for digital PCR (dPCR). In digital PCR, a solution containing a relatively small number of a target analyte, e.g., a polynucleotide or nucleotide sequence, may be subdivided into a large number of small test samples, such that each sample generally contains either one molecule of the target analyte, e.g., a nucleotide sequence, or none of the target. When the samples are subsequently thermally cycled in a PCR protocol, procedure, or experiment, the sample containing the target are amplified and produce a positive detection signal, while the samples containing no target are not amplified and produce no detection signal. Using Poisson statistics, the number of targets in the original solution may be correlated to the number of samples producing a positive detection signal.

[0049] One should appreciate that the disclosed techniques provide many advantageous technical effects including automated methods for quality control

during quantification of one or more target concentrations in a biological sample using an analyte detection (e.g., a PCR) apparatus. The techniques described herein employ logic to automate various processes, including processes currently performed using manual human effort. Further, the disclosed techniques have been designed to support data accuracy and allow for processing data algorithms and complex permutations on a scale and speed that cannot be achieved using manual human effort.

[0050] It should also be appreciated that the following specification is not intended as an extensive overview, and as such, concepts may be simplified in the interests of clarity and brevity.

[0051] FIG. 1 illustrates a block diagram of a PCR apparatus 100 and a microfluidic array plate 110 having an array of partitions 120 disposed therein. In one embodiment, PCR apparatus 100 is a digital PCR. As described above, digital PCR (dPCR) uses a solution including a relatively small number of a target analyte, e.g., a polynucleotide or nucleotide sequence template DNA (or RNA), fluorescence-quencher probes, primers, and a PCR master mix comprising DNA polymerase and reaction buffers at optimal concentrations. The solution is partitioned into a large number of small test samples, e.g., tens of thousands of microchambers disposed within a microfluidic array plate 110. As shown in FIG. 1, microfluidic array plate 110 includes an array of microchambers formed by multiple rows and columns of microchambers (e.g., nano-liter sized microchambers). A microchamber may have nucleic acid binding surfaces for better binding with a target in the test sample. A small test sample is disposed in each microchamber of the array, thereby forming an array of partitions 120. A partition 122 thus includes a microchamber and a small test sample disposed therein. Typically, some partitions of array 120 include a test sample that has one or more targets (e.g., a target analyte such as a nucleotide sequence) and some partitions do not. In FIG. 1, for example, partition 122 includes a target but partition 124 does not include one. Thermal cycling is

subsequently performed with respect to array of partitions 120 using PCR apparatus 100.

[0052] FIG. 2 illustrates a flowchart of an exemplary workflow 200 utilizing microfluidic array plate (MAP) technology that may be utilized in conjunction with some embodiments of the present invention. In step 202, a sample is prepared, e.g., using a bulk reaction mix. In step 204, the sample is loaded (e.g., pipetted) onto a microfluidic array plate (MAP), which distributes or digitizes the sample into thousands of small independent reactions each in a micro-chamber, e.g., so that each microchamber contains either one or zero copies of the target. Statistical methods are then used to calculate the original concentrations based on the number of positive and negative microchambers. In one embodiment of the invention, the MAP comprises a MAP16 Digital PCR consumable utilizing microinjection molded plate technology, having 16 samples (16 arrays or chips) per plate, and approximately 20,000 micro-chambers per reaction, with greater than 95% of the loaded volume analyzed. In some embodiments, the MAP comprises a sample plate containing 32 arrays (or chips). The sample plate with 32 Digital PCR arrays (or chips) can comprise approximately 10240 microchambers per sample plate. In step 206, the digital PCR thermal cycling process is run as described above. In one embodiment, up to four optical channels may be multiplexed to enable multiple targets to be measured per sample, possibly saving time and reagents. In one embodiment of the invention, using control software, parameters such as plate layout, optical channels, and thermal conditions can be easily modified to allow users to quickly start generating data. Specific protocols may also be saved and reloaded to streamline the workflow. In step 208, results are displayed to the user, allowing the user to visualize multiplexed data or calculate concentrations across replicates for example. In one embodiment, results can be exported for downstream analysis or compiled into a report.

[0053] In some embodiments, increasing the precision of calculating concentrations of targets includes more accurate identification and rejection of

defective microchambers in a MAP technology array plate. Automated quality control may increase throughput of the dPCR system, is agnostic to dyes, channels and assays run, and also help users to identify and report issues with dPCR runs.

[0054] FIG. 3 illustrates a flowchart of an exemplary data source analysis pipeline 300. The blocks illustrated in FIG. 3 represent processes and/or methods that can be implemented and/or executed by one or more computing elements described below (e.g., by one or more computing devices). The computing elements can be standalone computing elements, networked computing elements, distributed computing elements, and/or embedded computing elements. For example, the computing elements can be integrated with PCR apparatus 100, an onsite standalone computing device, a cloud-based computing device, or a combination thereof. In one embodiment, image 302 of a plurality of partitions in an array is captured following a background cycle, prior to the PCR. Image 302 shows uniform fluorescence, with no observed change in intensities due to the PCR occurring.

[0055] Image 304 is a post-PCR fluorescence image representing the plurality of partition in the array shown in image 302, after the amplification of one or more targets in a biological sample is performed by thermal cycling. Each bright spot shown in image 304 thus represents a partition of the array shown in image 302 after amplification. It is understood that the brightness of the spots in image 304 shown in FIG. 3 are for illustration purposes and can vary in any manner. For example, some partitions in the array shown in image 304 may appear to be less bright or dark (e.g., because the partitions do not include a target). In some embodiments, image 304 is associated with a particular fluorescence channel of a PCR apparatus having multiplexing capabilities.

[0056] FIG. 4 illustrates a plurality of exemplary pre-PCR images 402 and post-PCR images 404 representing array of partitions 120 disposed in microfluidic array plate 110. In FIG. 4, images 402 are pre-PCR images representing array of partitions 120 before the PCR amplification of one or more targets in a biological sample is performed. Images 404 are post-PCR images representing array of

partitions 120 after the amplification of one or more targets is performed. In one embodiment, images 402 and 404 are fluorescence images provided by multiple fluorescence channels of PCR apparatus 100.

[0057] PCR apparatus 100 can have multiplexing capabilities and is thus able to quantify multiple targets simultaneously. Multiplexing capabilities are obtained by having multiple fluorescence channels in PCR apparatus 100. In some embodiments, multiple fluorescence channels use different types of dyes for generating signals having different spectral wavelengths. Different dyes may each bind with a different target and produce signals with a different fluorescence color or spectrum. In FIG. 4, five exemplary types of dyes are used in different fluorescence channels in a PCR apparatus. The five different dyes may include a Carboxyfluorescein (FAM) based dye that produces signals having a blue fluorescence color, a Hexachloro-fluorescein (HEX) based dye that produces signals having a green fluorescence color, a 6-carboxy-X-rhodamine (ROX) based dye that produces signals having a red fluorescence color, a Tetramethyl rhodamine (TAMRA) based dye that produces signals having a red or yellow fluorescence color, and a TYE™ based dye that produces a dark red fluorescence color. In some embodiments, the ROX based dye is used as a control reference dye.

[0058] In some embodiments, pre-processing step 306 in analysis pipeline 300 may include a process for determining corner locations of an array of partitions. With reference to FIG. 5A, an image 500 represents an array of partitions 502. For illustration purposes, image 500 only shows a portion of array 502. While the below description uses a post-PCR image as an example, it is understood that a pre-PCR image can also be used to determine corner locations of an array of partitions.

[0059] As a part of pre-processing step 306, a computing device uses image 500 to determine locations associated with corners of array 502. Corner locations of an array of partitions are used subsequently for determining expected locations of partitions in array 502. Observed locations are then determined based on the expected locations and are used to obtain images of individual partitions. Next, the

images of individual partitions are classified, the results of which are used for performing the automated quality control classification analysis described herein for the individual partitions, and calculating target concentrations. The partition location determination process and the classification process are described in more detail below.

[0060] With reference to FIG. 5A, array 502 is a two-dimensional array having four corners. A corner location determination process may begin with any of the four corners. FIG. 5A uses the top left corner of array 502 as an illustration. To determine the location of the top left corner of array 502, corner area 504 is first selected to indicate that the top left corner is under analysis. Corner area 504 may be annotated using, for example, a square or a rectangle shaped box having a predetermined dimension.

[0061] In one embodiment, dimensions of corner area 504 are predetermined such that corner area 504 is not too small or too large. An overly small corner area may not include enough partitions for correlating with template images to correctly find an edge. For example, as shown in FIG. 5A, template image 506 represents seven partitions disposed along a horizontal edge of a template array of partitions. Correspondingly, the dimensions of corner area 504 may be predetermined such that 20-40 partitions disposed along a horizontal edge are included in corner area 504 for analysis. Template image 506 is described below in more detail.

[0062] An overly large corner area may also affect the correlation using template images due to, for example, warpage of the image of the array of partitions. The warpage of the image may cause poor correlation between template images with the image of the array of partitions, making it difficult to find edges of the array of partitions. Thus, a properly dimensioned corner area should include sufficient number of partitions based on the template images but also not be overly large to cause correlation difficulty. It is understood that proper dimensions of a corner area can be predetermined, by a computing device and/or by a user input,

based on the dimensions of the array of partitions and the dimensions of the template images.

[0063] In one embodiment shown in FIG. 5A, corner area 504 is annotated and displayed. The annotation may be, for example, a colored box overlaying an image of the array of partitions to indicate the selected corner area. In FIG. 5A, corner area 504 is annotated by a colored box (e.g., a green box) overlying image 500 to indicate that the top left corner is selected to be analyzed.

[0064] As described above, template images are used for determining corner locations. A template image includes several portions that can be used to find an edge of an array. FIG. 5A illustrates a template image 506 used to correlate with a top edge 512 of array 502 and a template image 508 used to correlate with a left edge 514 of array 502. Template image 506 includes two portions. A first portion of template image 506 comprises several bright spots representing a plurality of partitions that are predetermined to be partitions disposed along a top edge. These predetermined partitions of template image 506 form, for example, a line pattern as shown in FIG. 5A. A second portion of template image 506 represents an area predetermined to have no partitions. In template image 506, this second portion is immediately adjacent to the first portion and is dark. Therefore, template image 506 is an image representing a known pattern of partitions disposed along a known top edge.

[0065] FIG. 5A further illustrates another template image 508, which is an image representing a known pattern of partitions disposed along a known left edge. In FIG. 5A, template image 508 also includes two portions. A first portion of template image 508 comprises several bright spots representing a plurality of partitions that are predetermined to be partitions disposed along a left edge. These predetermined partitions form, for example, a staggered pattern as shown in FIG. 5A. The staggered pattern may also be viewed as comprising two staggered lines. A part of the partitions represented in template image 508 form a first vertical line and another part of the partition form a second vertical line. The first and second

vertical lines are offset from each other in the horizontal direction. For example, as shown in FIG. 5A, the partitions forming the first line may be disposed more towards the left than the partition forming the second line. And the partitions of the first line and the second line are staggered in alternating rows.

[0066] Like template image 506, template image 508 also has a second portion that represents an area predetermined to have no partitions. In FIG. 5A, this portion is shown as a dark area immediately adjacent to the first portion with the partitions forming the staggered line pattern. Unlike template image 506, the second portion of which is oriented in a horizontal direction, the second portion of template image 508 is oriented in a vertical direction. The two portions of the template image 508 thus represent a known pattern of partitions disposed along a known left edge. While FIG. 5A illustrates template images used for the top edge and the left edge, it is understood that similar template images can be configured and obtained for the bottom edge and the right edge.

[0067] A corner of a two-dimensional array is formed by two edges. For instance, in FIG. 5A, the top left corner of array 502 is formed by top edge 512 and left edge 514. As such, if the locations of the two edges are determined, the location of the corresponding corner can be determined. In one embodiment, two template images are used to determine locations of two edges, and in turn, a corner location formed by the two edges. For example, in FIG. 5A, template image 506 and template image 508 are obtained and used for determining locations of top edge 512 and left edge 514, respectively, that form the top left corner. Using template image 506 as an example, to determine the location of top edge 512, template image 506 is moved and/or rotated to correlate with corner area 504 to find top edge 512. If at least a part of the template image 506 is matched with top edge 512, the location of top edge 512 may be determined. For instance, by moving and/or rotating template image 506, if at least four or five out of the total seven partitions shown in template image 506 match with at least four or five partitions shown in corner area 504 of image 500, and if the dark portion shown in template image 506 also matches with

the dark portion above top edge 512 in corner area 504, then top edge 512 is found. In some embodiments, if an initial match is found, template image 506 is further moved along a horizontal direction to see if additional or continued matches can be found. If so, there is a high probability that top edge 512 is found. Once top edge 512 is found, its location can be readily determined by, for example, measuring the number of pixels from top edge 512 to the top edge of image 500.

[0068] In a similar manner, to determine the location of left edge 514 of array 502, template image 508 is moved and/or rotated to correlate with left edge 514. If at least a part of template image 508 is matched with left edge 514, the location of left edge 514 may be determined. For instance, by moving and/or rotating template image 508, if at least four or five out of the total seven partitions shown in template image 508 match with four or five partitions in corner area 504 and if the dark portion shown in template image 508 also matches with the dark portion to the left of left edge 514 in corner area 504, then left edge 454 is found. In some embodiments, if an initial match is found, template image 508 is further moved along a vertical direction to see if additional or continued matches can be found. If so, there is a high probability that left edge 514 is found. Once left edge 514 is found, its location can be readily determined by, for example, measuring the number of pixels from left edge 514 to the left edge of image 500.

[0069] When a match between a template image and an edge is found, visual indications may be displayed on image 500 and in other manners. FIG. 5B illustrates such exemplary visual indications using the top edge as an example. In FIG. 5B, the brightness of certain partitions disposed along top edge 512 is increased to indicate a likely match between template image 506 and top edge 512.

[0070] With reference to FIG. 5A, after the location of top edge 512 and the location of left edge 514 are determined, the location of the top left corner of array 502 can be determined. In one embodiment, the location of the top left corner is represented by horizontal and vertical coordinates in number of pixels from the left and top edges, respectively. The location of the top left corner of array 502 shown in

FIG. 5A may thus be represented by a coordinate pair such as (135, 133). While FIG. 5A uses the top left corner as an example, it is understood that locations of other corners of array 502 can be determined in a similar manner. Moreover, when additional images of the same array 502 or different arrays are obtained, for example, from different fluorescence channels, corner locations of the same array or different arrays can be similarly determined using the above-described techniques.

[0071] With reference back to FIG. 3, the corner locations determined in pre-processing step 306 are provided to extract microchamber site images 308 in which images of individual partitions in an array of partitions are determined and extracted. As described above, an array of partitions may be arranged in multiple rows and columns. FIG. 6A illustrates such an array 602 represented in image 600. FIG. 6B shows a zoom-in image representing the top left corner area of array 502. As illustrated in FIGs. 5A and 5B, array 502 includes many partitions (e.g., about 20,000) arranged in rows and columns. These partitions are illustrated by the bright spots in FIGs. 5A and 5B. Ideally, these partitions are equally spaced from one another in the underlying microfluidic array plate, such that the partitions form a regular repeating pattern. Further, the images of the individual partitions would also ideally show that the partitions are arranged in rows and columns with equal spacing. If the partitions themselves and the images of them are both ideal, then locations of the partitions can be readily determined from the corner locations of the array of partitions by a simple calculation.

[0072] With reference to FIGs. 6A and 6B, partitions in array 602 are arranged in rows and columns. Therefore, array 602 has four corners. The locations of the four corners of array 602 can be determined using techniques described above. The locations of the corners can be represented in number of pixels in a pair of X and Y coordinates (i.e., number of pixels measured from the relevant edges in the horizontal and vertical directions). For example, the location of the top left corner of array 602 may be represented as a coordinate pair measured in pixels such as (135, 132), where "135" is the distance of the top left corner from

the left edge of image 500 in numbers of pixels and "132" is the distance of the top left corner from the top edge of image 500 in numbers of pixels. The locations of the top right corner, the bottom left corner, and the bottom right corner of array 502 may be similarly represented as (2073, 127), (130, 2049), and (2085, 2044), respectively. In the above examples, the location of top left corner of image 500 is set to be the base coordinate point (0, 0), and the locations of the corners are measured with respect to the base coordinate point. It is understood that any other base coordinate point or system can be used.

[0073] After corner locations of an array of partitions are determined, expected locations of partitions in the array can be calculated. Expected locations of partitions are locations calculated according to design and/or manufacturing specifications of the microfluidic array plate. For example, the design and/or manufacturing specifications may provide a row spacing between any two immediately neighboring partitions in the same row. Similarly, a column spacing between any two immediately neighboring rows may also be provided by the specifications. Further, in some embodiments, two immediately neighboring rows of partitions may be offset in the horizontal direction and the offset is also known according to the design and/or manufacturing specifications. An example of such an offset is illustrated in FIG. 6B, where two immediately neighboring rows may be offset by about the width of a partition. Using one or more of the known row spacing, column spacing, and offset values provided by the design and/or manufacturing specifications, expected locations of the partitions can be calculated based on the determined corner locations.

[0074] With reference back to FIG. 3, in process 310, fluorescence signals represented in the partition images are summarized. In some embodiments, the dimension of a partition is about 16 pixels by 16 pixels. In other embodiments, the dimension of a partition is 9 pixels by 11 pixels.

[0075] In some embodiments, to summarize the fluorescence signals represented in a partition image of a particular partition, a 16x16 grid or frame is

applied at the observed location of the partition. Next, several brightest pixels in the corresponding 16x16 grid are ignored. Fluorescence signals of the next group of brightest pixels are then averaged to obtain the summarized value of the signals in a particular partition represented by the partition image. In some embodiments, several darkest pixels are also ignored in calculating the summarized value of the signals. Fluorescence signals in a partition represented by a non-positive partition (e.g., a negative partition) image may also be summarized in a similar manner or in a different manner (e.g., simply taking the average of all the pixels). It is understood that other methods of summarizing signals of a partition may also be used. For instance, summarizing signals may include, but not limited to, integrating, summing, averaging, weighted averaging, etc. of the signals.

[0076] With reference back to FIG. 3, in some embodiments, a background image subtraction process during the summarize fluorescence data step 310 can also be performed to improve the quality of images of partitions before they are provided to a machine-learning model-based classification process, described below with respect to FIGs. 11-14, which may occur during the determine user thresholds step 312.

[0077] With reference back to FIG. 3, using the observed locations of partitions determined in step 308, images of individual partitions can be obtained (e.g., by image cropping at or near the observed locations or reproducing using the original image of the array of partitions). These images of individual partitions are provided to a classification process in step 312 with or without the background image subtraction process during the summarize fluorescence data step 310. In some embodiments, classification process in step 312 is a machine-learning model-based classification process that can identify reject-like partition images and non-reject-like partition images, and positive-like partition images and non-positive-like partition images.

[0078] With reference still to FIG. 3, in estimate target concentration step 314, target presence is detected, or concentrations are quantified. In some

embodiments, to detect target presence or quantify the target concentrations, the number of total partitions, the number of positive partition images (less the ones rejected in the previous various processes), and the total number of accepted partition images (including both positive and negative less the ones rejected in the previous various processes) are obtained. Using Poisson statistics, the target presence can be detected, or concentrations can be quantified within a defined confidence interval. For example, the number of targets in the original solution may be correlated to the number of samples producing a positive detection signal represented by the positive partition images. By performing one or more of the above-described processes (e.g., machine-learning model-based classification, observed location determination, spectral compensation, partition rejection by thresholding, rejecting long consecutive positive partitions), the target concentration can be quantified in a more accurate manner than traditional analog-type PCR apparatuses.

[0079] Also, in or following step 314, visual reports and charts of the results of target detection or concentration quantification can be generated and displayed to a user.

[0080] In some embodiments, quality control is a feature that helps identify sample failures or spurious external artifacts. To increase the throughput of a system, automated quality control becomes an important aspect to help identify and report issues in timely fashion, and inform customers of failures that may occur during the course of a dPCR run. It is critical to warn the user of any quality control issues so that the user receives adequate warning that the reliability of the dPCR results may be adversely impacted.

[0081] Occasionally, a sample is lost due to a systemic failure or some other error. FIG. 7 illustrates four examples of arrays of partitions, one representing a normal sample 702, and three representing various quality control issues. Array 702 representing the normal sample shows uniform fluorescence throughout. The quality control issues illustrated in FIG. 7 include an array of partitions 704

manifesting failure due to excess gas present in the partitions, which shows some nonuniformity in fluorescence in portions of array 704. Array of partitions 706 illustrates a quality control issue in the array, which is manifesting failure due to fill issues, e.g., incompletely or unevenly filled partitions or microchambers, which shows some nonuniformity in fluorescence particularly on the left side of the array, which may be due to incompletely filled or empty partitions on that side of the array where the fluid dPCR solution did not reach. Array 708 shows a more catastrophic failure, a leak in the array, which may be caused by delamination, where there may be a defect or separation in an injection molded microfluidic device comprising an array of partitions or s separation of a thermoplastic thin film that caps and seals (e.g., hermetically seals) the partitions or microchambers, and connecting siphon apertures and microchannels, from the microfluidic device. The thermoplastic thin film may be at least partially gas permeable when a pressure differential is applied across the thermoplastic thin film. Other types of failures that may occur include bubbles, debris, corner finding failures, saturation, bridging, debris, and dust. Such quality control issues are identified and flagged in embodiments described herein, including the level of severity and type of failure.

[0082] FIG. 8 illustrates examples of normal (e.g., positive) partition images 802 extracted from raw sample images of microchamber sites. Positive partition images represent positive partitions, which correspond to an existence of one or more target concentrations in the partitions. A positive partition includes a target (e.g., a DNA molecule) that has been amplified by thermal cycling so that it is detectable via fluorescence signals. A positive partition thus includes a large number of target PCR amplicons of the original DNA. The positive partition images are subsequently used for quantifying the target PCR amplicon concentration in a biological sample. The images show uniform fluorescence with an area of bright intensity near the center of the partition. A negative partition image (not shown) corresponds to a non-existence of a target presence in a partition. If a particular partition does not have a target (e.g., a DNA target molecule such as a PCR amplicon), then thermal cycling in PCR did not produce target amplicons. As a

result, the fluorescence image of the negative partition appears mostly or completely dark due to the lack of fluorescence signals. Negative partition images may be classified as non-positive partition images.

[0083] FIG. 8 also illustrates examples 804 of partition images showing data quality issues. Non-positive partition images also include images associated with defective microchambers of partitions, images associated with contaminated partitions, images associated with defective fillings of microchambers of partitions, defective images of partitions, and any post-PCR partition images other than positive partition images.

[0084] Images associated with defective partitions are also examples of non-positive partition images. For example, a defective partition can include design and/or manufacturing defects in a microchamber of the partition, defects in microfluidic array plate such as a reflective surface of the plate, delamination, leaks, gas, a defective filling of a microchamber of the partition (e.g., fill issues), and any other defects of a partition or the microfluidic array plate. Such defective partitions may or may not include a target (e.g., a DNA target molecule such as nucleic acids originally present in the biological sample or PCR amplicons generated from such nucleic acids). While thermal cycling in PCR may produce target amplicons of a target, defective partitions may affect the fluorescence signal detection and image generation. Images generated based on defective partitions may not be used to accurately quantify the target concentration in the biological sample under analysis. Some of these images can generate unique signatures, and such examples are shown in 804 of FIG. 8. For example, gas, trapped air, or bubbles may exhibit as a ring across one section in the partition image. Fill issues may exhibit as non-uniformly filled or illuminated partition images. These images may be classified as non-positive partition images. Partition images exhibiting signatures of quality control issues may be used to train machine learning classifiers which may be employed to quickly identify and reject defective partition images, and therefore improve the quality of dPCR target concentration estimates.

[0085] Images associated with contaminated partitions are additional examples of non-positive partition images. Partitions may be contaminated with, for example, dusts, fibers, particles, parasite DNAs, or the like. Contaminated partitions may or may not include a target. While thermal cycling in PCR may produce amplicons of a target, contaminated partitions may affect the fluorescence signal detection and image generation. Images generated based on contaminated partitions may not be used to accurately quantify the target concentration in the biological sample under analysis.

[0086] Images themselves can also be defective. Defective images may also be classified as non-positive partition images. Imaging defects can include, for example, defects associated with the imaging system such as dust on a camera lens, a corrupted image, image distortion, or the like. These images may not be used to accurately quantify the target concentration in the biological sample under analysis. These images may also be classified as non-positive partition images.

[0087] However, many factors may affect the partitions in a microfluidic array plate and the images of the partitions, rendering deviations from the ideal arrangement of the partitions and/or the ideal image thereof. For example, the microfluidic array plate and/or some microchambers of partitions may have defects such that the spacings between some partitions may be smaller or greater than a designed or desired spacing. In addition, contamination of the partitions and/or the microfluidic array plate (e.g., dust, fiber, surface contamination) may affect the images of the partitions, such that they are not a precise representation of the partition arrangement. The imaging system may introduce errors as well. As an example, when capturing the images of the partitions, the camera position may have some error, so the microfluidic array plate and/or the partitions may not be always centered in the image. The microfluidic array plate and/or the partitions may have some rotations with respect to the camera such that the rows and columns of the partitions are not perfectly horizontal and vertical in the image. Also, the camera lenses may introduce some distortion such that the rows and

columns of the partitions are not perfectly straight, but rather they may have some small curvatures that are dependent on locations in the image.

[0088] While negative partition images, images of defective partitions, images of contaminated partitions, and defective images are described above as examples of reject-like or non-positive-like partition images, it is understood that any other partition images that are not positive partition images may also be classified as non-positive-like partition images.

[0089] FIG. 9 illustrates examples of graphs of dPCR fluorescence data across various channels showing normal fluorescence data 902, and graphs of dPCR fluorescence data across various channels showing quality control-afflicted (e.g., spurious) data 904. In the graphs of each fluorescence channel in 902 and 904, the horizontal axis represents the partition numbers (e.g., total of 20,000 partitions) and the vertical axis represents the fluorescence signal intensity. In the graphs of 902 and 904, the fluorescence signals are plotted according to the detected signals. The graphs of 902 depicting normal data show a fairly uniform pattern of fluorescence signal intensity for target concentrations and for negative partitions that do not contain any target molecules. As described above, many of the fluorescence signals in 904 representing reject-like or non-positive-like may be inaccurate due to a variety of factors such as partition defects, image artifacts, contaminations, or the like. The inaccuracies are manifested in more nonuniform, random fluorescence values in the graphs of 904.

[0090] FIG. 10A is a flowchart illustrating a method for a quality control workflow 1000 for an array of partitions according to various embodiments. For this sample level data quality check, three levels may be considered for data quality validation: (1) PASS: No quality control issue encountered in the current sample; (2) WARN: The current sample is impacted by a quality control issue that can (possibly) be corrected by adjusting user-set thresholds. The decision to use the results is up to the customer; and (3) FAIL: The current sample has suffered a failure due to a severe quality control issue that may impact the results, e.g.,

concentration estimates. If a sample quality control status is FAIL, no results are reported to the user. In one embodiment of the workflow, each unit (array of partitions, or chip containing partitions) operates independently of each other. Thus, this quality control workflow assumes inter-array independence. This implies that a quality control issue in a given array does not affect other arrays. For, e.g., currently if a specific array has leaks, the information from that array is not used to make quality control assessments about adjacent arrays.

[0091] In the embodiment shown in FIG. 10A, framework 1000 will report the type of failure and warning as follows. Specifically, for warnings the cause is delineated by one or more of five categories including saturation 1010, bridging 1012, distribution of positive 1013, excessive quality control rejects 1014, or overloaded array 1015. If any of warning categories 1010, 1012, 1013, 1014, or 1015 is met, the system will issue a warning at steps 1020, 1018, 1015, 1016, and 1019 respectively and continue the workflow until the sample array passes at step 1022. For a failure, the cause can include one or more of (a) blank image error 1003, which includes images with no signal or out of focus images and requires user to turn off or remove channel at issue from analysis at 1005; (b) image registration error 1004, which can cause sample to fail and exit the quality control check at 1006; and (c) sample failure 1006, which will cause the sample to fail and exit the quality control check at 1008.

[0092] In some embodiments, in the category of sample failures, the root cause may be one of a number of possible systemic failures including leaks, delamination, gas or other systemic issues, but a root-cause level of granularity to disambiguate between the root causes of a sample failure may not be provided. In other embodiments, a root cause of a sample failure could potentially be provided based on information from the machine learning classifier disclosed herein or other information.

[0093] A data quality check may follow the order of precedence illustrated in FIG. 10A in some embodiments, which implies that quality control may be triggered

first by one issue, and not triggered simultaneously or at various times by different issues. The first quality control issue that is encountered raises the corresponding quality control flag. For saturation, if a sample is failed and it is saturated, then the sample is reported as a FAIL for quality control purposes, though the cause may be reported as saturation. The workflow of the analysis pipeline follows the logical order in which issues may occur. For example, blank image error check 1003 and/or the image registration error check 1004 always precede every other step in workflow 1000 because if something else fails with image registration or receives blank image, the user should be informed about the first point of failure. If a sample fails under step 1006, further warnings about the sample are similarly not needed. For warnings, saturation is detected first at step 1010, then a bridging check at step 1012, and finally an excessive quality control channel reject check at step 1014. The saturation check is performed before the bridging check, as saturation can lead to bridging, but not the other way around. Excessive quality control threshold reject check 1014 is independent of the other warnings.

[0094] Further example implementation details of the data quality control framework of some embodiments are set forth below. Blank image error during digital quantification methods, including digital PCR, can occur under several conditions. In some embodiments, the conditions include failure to load any sample, running controls intentionally without an assay or no template, or incorrect software configuration by a customer or on the system side. Blank image errors can trigger image failure metrics, which require the image to be handled differently. According to some embodiments, as shown in FIG. 10B at 1031, blank image check 1003 can include determining whether an image have any location error. A location error, in some embodiments, measures goodness of fit along rows and columns of the estimated sample partition (e.g., microchamber site) locations on the sample image. A location error can be determined when the pipeline is unable to accurately estimate the locations of the sample partition (e.g., microchamber) sites. For example, a distance can be determined between two adjacent microchambers along rows and columns on the sample image and a difference can be determined to

compare with the median variation of this distance on the same sample image, and a location error can be raised if the differences are too great and/or too numerous. If the image does not have location error, as shown in 1032, the image is considered normal in terms of blank image error check. In response to an image having location error, as shown in 1033, a blank score is compared with a predetermined thresh hold. According to embodiments, the blank score is obtained by comparing intensity of a raw image with intensity of a transformed image. The transformed image includes an image that has been transformed via amplifying signal in the shape of the sample partition (e.g., microchamber) to emphasize high values (i.e., maxpool (which is max of small subsets of pixels) of an image convolved with sample partition shape/mean of raw image). In response to the blank score not greater than the threshold, blank image error and/or out of focus error are flagged out at block 1035. The channel with the blank image error and/or out of focus error, as shown in 1037, is then turned off to eliminate influence on analysis from this channel. In response to the blank score greater than the threshold, image registration error check is conducted at 1039.

[0095] According to embodiments, image registration error check 1004 measures the goodness of fit along the rows and columns of the estimated partition (microchamber site) locations on the sample image. The error is flagged when the pipeline is unable to accurately estimate the locations of the microchamber sites. For example, a distance can be determined between two adjacent microchambers along rows and columns on the sample image and a difference can be determined to compare with the median variation of this distance on the same sample image, and an image registration error can be raised if the differences are too great and/or too numerous. Example reasons for image registration errors include corner finding failures, or unloaded samples.

[0096] In some embodiments, sample failure check 1008 is performed by a site classifier process as set forth in FIGs. 11 and 12 discussed below. The unit of measure is the number of sites rejected by the site classifier. When the number of

sites rejected by the site classifier exceeds an analytical threshold (e.g., 6.1% of the total microchambers in the sample), a sample failure may be raised. The site classifier as discussed in FIG. 11 below is a sophisticated algorithm that captures most kinds of severe failures discussed above, including delamination, bubbles, gas, leaks, and fill issues.

[0097] Saturation check 1010 is performed in some embodiments when sites are rejected by an optical saturation detection algorithm. Optical saturation is observed when the signal from an imaged channel exceeds the operating range of the detecting sensor on the dPCR instrument, for example, as a consequence, the digital images appear excessively bright or saturated. In one embodiment, it is detected on the raw microchamber site images at the endpoint cycle, and manifests on a digital site image as extremely high intensity values on the pixels that belong to the microchamber's area. A small region belonging to the microchamber's signal is used for detection of saturation in individual microchambers. In one embodiment, if 8% of total microchamber sites in the sample are saturated, the channel is saturated. Saturation often occurs as spuriously bright luminosities typically in the FAM channel, although it can also manifest in the ROX channel due to JUN2ROX spectral crosstalk.

[0098] Bridging check 1012 manifests as a length of a stream of contiguous positives (e.g., bright microchambers) in an area of negatives (e.g., a sparsely lit (bright) area). In one embodiment, if 2.5% of the total microchambers in the sample are flagged for bridging errors, the bridging check warning is raised for the user to check if the thresholds for the bridging warning are incorrectly set. Distribution of positive error 1013 may occur when the positive partitions (reaction sites) are not well distributed across a sample plate, which leads to inconsistencies in data interpretation. The image of partitioned samples may exhibit variations and/or discrepancies in distribution of a target analyte/molecule. For example, uneven distribution of target analytes/molecules within a sample, leading to some partitions containing many target molecules and others containing very few. These

variations and/or discrepancies can introduce uncertainty in the data analysis process, potentially impacting the accuracy and reliability of experimental findings. In digital PCR, samples can be run at concentrations outside the range of Poisson statistical computation. For instance, this can happen due to suboptimal dilution.

[0099] In some embodiments, distribution of positive error at 1013 is determined by assessing uniformity of the distribution of positive sample partitions over nine regions of a chip or an array of a sample plate. The nine regions can include a top left region, a top center region, a top right region, a middle-left region, a middle-center region, a middle-right region, a bottom left region, a bottom center region, and a bottom right region. In some embodiments, the nine regions contain substantially the same amount of sample partitions (e.g., microchambers or droplets). According to embodiments, a coefficient of variation (CV) for positive count across the nine regions being greater than 0.3 indicates distribution of positive error. The coefficient of variation is calculated by $CV=\sigma/\mu$, where σ is the standard deviation of positive counts for the nine regions, and μ is mean of the positive counts in the nine regions. In some embodiments, distribution of positive error check at 1013 is protected against unreliable coefficient of variation values caused by a low count of positives. For instance, any target with a mean of less than 50 positives will be assigned "NAN" rather than distribution of positive error and/or unreliable coefficient of variation value. In embodiments, a sample with the number of positive partitions (e.g., microchambers) greater than 90% of total number of positive partitions (e.g., microchambers) and less than 450 is designated with unreliable coefficient of variation. According to embodiments, QC (quality control) warning for a sample is triggered when the number of positive sample partitions is less than 90% of total partition counts for a sample and greater than 450 and the coefficient of variation on all the regions is greater than 0.3. In embodiments, unreliable coefficient of variation can be caused by partial bridging in an array background issues, and/or incorrect thresholding.

[00100] According to embodiments, excessive quality control threshold rejects check 1014 is measured by the number of sites that lie outside the quality control threshold range. In one embodiment, if 5% of the total microchambers in the sample lie outside the quality control threshold range, the warning is triggered, and the user may check to see if the lower quality control threshold is set too high or the upper quality control threshold is set too low, resulting in an excessive number of rejects.

[00101] According to embodiments, overloaded array error 1017 is triggered when an analysis detects substantially no negative sample partitions or no negative sample partitions. Per Poisson statistics, the average number of molecules per sample partition is obtained by $\lambda = -ln(z/n)$ where z is the number of negative partitions and n is the total number of partitions. Hence, overloaded arrays with no negative partitions can impact computation and data analysis based on Poisson statistics. In response to overloaded array error at 1017, concentration for a sample run needs to be corrected, and/or threshold for the overloaded dye(s) needs to be adjusted.

[00102] If the sample image of the array of partitions passes all the checks of quality control framework 1000 as shown in FIG. 10A, no quality control issue is detected, and the image passes at step 1022. In some embodiments, the sample image may exit the quality control framework check 1000 at the first raised warning and the framework does not check for subsequent warnings if a warning was raised already.

[00103] Using a trained machine-learning model, many of the reject-like and non-positive signals can be removed or filtered out by correctly classifying the partition images as non-positive and reject-like (and hence rejected) partition images. Accordingly, the machine-learning model-based classification process for sample failure check 1008 as described with respect to FIG. 10A and FIG. 11 can perform defect detection at a microchamber level, by evaluating microchamber site images. In some embodiments, machine learning and data analysis may be

leveraged to significantly improve the accuracy of determining positive partition images.

[00104] A machine-learning model can use images that are pre-identified as positive-like partition images and non-positive-like or reject-like or non-reject-like partition images for training the model. For example, both the normal partition images and quality control-afflicted partition images in FIG. 8 can be used as training images for training the machine-learning model, e.g., as positive-like or non-reject-like, and non-positive-like and reject-like partition images, respectively.

[00105] Positive-like partition images and non-positive-like partition images can be obtained from past PCR tests. For example, a group of post-PCR images can be classified and annotated (manually or automatically) as non-reject-like partition images and reject-like partition images. In some embodiments, training of a machine-learned model may need many annotated images to provide enough quantity and enough variety of both positive-like and non-positive-like, and reject-like and non-reject-like partition images. Therefore, images obtained only from past PCR analyses may or may not be sufficient for training a machine-learning model.

[00106] If images obtained from past PCR analyses are insufficient, additional training images may be necessary. To obtain additional images for training a machine-learning model, in some embodiments, a set of non-positive-like partition images can be generated based on another set of non-positive-like partition images. For example, a first set of images is obtained from past PCR tests. The images in the first set are pre-identified and annotated as non-positive-like partition images. The images in the first set can be modified to obtain a second set of images. Such image modification can include, but not limited to, one or more of rotating, editing, cropping, distorting, mirroring, brightening, darkening, changing a contrast of, changing a color of, and changing a pattern of, the first set of images. The modified images in the second set can also be pre-identified and annotated as non-positive-like partition images. The second set can thus be included in the group of non-

positive-like partition images used for training the machine-learning model, thereby increasing the quantity and variety of the available training images.

[00107] In a similar manner, in some embodiments, a set of non-positive-like partition images can be generated based on a set of positive-like partition images. For example, a third set of images is obtained from past PCR tests. The images in the third set are pre-identified and annotated as positive-like partition images. The positive-like partition images in the third set can be modified to obtain a fourth set of images, which can be non-positive-like partition images. Such image modification can include, but not limited to, one or more of rotating, editing, cropping, distorting, mirroring, brightening, darkening, changing a contrast of, changing a color of, and changing a pattern of, the first set. The modified images in the fourth set can be pre-identified and annotated as non-positive-like partition images. The fourth set of images can thus be included in the group of non-positive-like partition images used for training the machine-learning model, thereby increasing the quantity and variety of the available training images.

[00108] In the above examples, a set of non-positive-like partition images and/or a set of positive-like partition images are modified and used to generate another set of non-positive-like partition images. In a similar manner, a set of non-positive-like partition images and/or a set of positive-like partition images can be modified and used to generate another set of positive-like partition images. In some embodiments, the non-positive-like partition images and/or positive-like partition images being modified are also referred to as seed images. By modifying seed images, many other images can be generated to expand the quantity of the training images. Modification of the seed images can also be configured to generate different types of non-positive-like partition images and/or positive-like partition images, thereby greatly improving the variety of the training images for the machine-learning model.

[00109] In some embodiments, a set of positive-like partition images and a set of non-positive-like partition images, and/or a set of reject-like images and non-

reject-like images, are combined to generate one or more datasets for training the machine-learning model. The datasets may include a training dataset, a validation dataset, and a testing dataset. For example, 1/3 of the images are used as a training dataset; 1/3 of the images are used as a validation dataset; and 1/3 of the images are used as a testing dataset. The datasets may include image files and/or features extracted from the image files.

[00110] Using the one or more datasets, one or more computing devices (e.g., a server) iteratively train the machine-learning model to determine a set of parameters. The set of parameters can include, for example, weights and features of the machine-learning model. The training can be performed using, for example, a stochastic gradient descent method and/or its extensions and variants such as the Adaptive Moment Estimation (Adam) method, implicit updates method (ISGD), the momentum method, the averaged stochastic gradient descent method, the Adaptive Gradient (AdaGrad) method, and the root mean square propagation (RMSProp) method.

[00111] According to various embodiments, the machine-learning model used for classification of images can be, for example, a convolutional neural network (CNN). A CNN includes an input layer, one or more hidden layers, and an output layer. The hidden layers include layers that perform convolutions such as one or more convolutional layers. A CNN may also include local and/or global pooling layers along with the convolutional layers. The pooling layers are for reducing the dimensions of data. While CNN is used as an example of the machine-learning model for classification of images, it is understood that extensions and variants of CNN and/or other types of neural networks for image classification may also be used. In one embodiment, the CNN may comprise a VGG19 image classification convolutional neural network that is 19 layers deep, that may be modified or optimized to classify the smaller pixel dimensions of the microchamber site images.

[0112] After training, the trained machine-learning model is configured with the set of parameters determined from the training. The trained machine-learning

model can then be used for classifying images to positive-like partition images or non-positive-like partition images, or reject-like or non-reject like partition images. As described above, observed locations of partitions represented in a post-PCR image are determined. Based on the observed locations, images of individual partitions can be obtained (e.g., by image cropping or reproducing at or near the observed locations). These individual partition images are then provided to the trained machine-learning model. Optionally, before the partition images are provided to trained machine-learning model, image subtraction is performed to improve the post-PCR image quality. The trained machine-learning model then classifies each individual partition images as a positive-like partition image or a non-positive-like partition image, or alternatively, as a reject-like partition image or a non-reject-like partition image. In some embodiments, the trained machine-learning model determines the probability that a particular partition image is a positive-like partition image, or a probability that a particular partition image is a reject-like partition image. Based on the probability, the image is classified as a positive-like or non-positive-like partition image in the machine learning model, or in another machine learning model, the image is classified as a reject-like or non-reject-like partition image.

[0113] Some embodiments use a combination of two machine learning classifiers that may be similarly structured in architecture but are differently trained and are executed on differently processed input derived from the same background and endpoint images to perform separate classification tasks, the results of which are then logically mapped to final determinations of whether to reject data from a particular sample partition. Specifically, one trained machine-learning model classifies partition images as positive-like versus non-positive-like, and another trained machine-learning model classifies reject-like versus non-reject-like partition images. The inputs to each model, in a specific example, undergo different pre-processing using the relevant source images as further described below.

[0114] FIG. 11 is a flowchart illustrating method 1100 for classifying individual microchamber site images in an array according to various embodiments. At step 1104 a processed image is obtained, and at step 1120 a raw endpoint image is obtained as described below with respect to FIG. 12. At step 1106, microchamber site images are extracted from the processed image, and at step 1122, corresponding microchamber site images are extracted from the raw endpoint image.

[0115] For discriminating {POS-like, Not-POS-like, where POS=positive} sites each microchamber image is normalized individually, which is distinct from the {REJ-like, Not-REJ-like, where REJ=reject} normalization. At step 1106, a microchamber site image is extracted from the processed image ($C_{processed} = C_{endpoint} - C_{warped\_background}$, eg. $C_{processed}$ = C40 – C15). The dimensions of this image are 16 pixels x 16 pixels. After the microchamber image is extracted from a processed image, the values in the processed images lie in the range [-65535, 65535] for 16 bits. This range is mapped non-linearly to [-1, 1] in step 1108.

[0116] At step 1110, the 90%-ile value of all pixels in that microchamber image is computed. In some embodiments, there are 256 pixels in the image of an individual microchamber. The quantile value is computed by ordering both the positive and the negative values. The absolute value of the 90%-ile value is then computed, and then every pixel in this image is divided by this value. Finally, values lying outside the range bounds of [-1, 1] are clipped to these range limits. At step 1112, these values are run through the classifier and a positive-like or non-positive-like classification is issued. In one embodiment, if the classifier issues a positive classification, the reject-like, non-reject-like classification is not issued to increase computational efficiency.

[0117] At step 1122, the microchamber image is extracted from the raw endpoint image ($C_{endpoint}$, eg. C40). The values in the raw images lie in the range [0, 65535]. In step 1124, this range is mapped linearly to [0, 1]. This mapping is done as follows. For a given channel (FAM, VIC, etc.), find the maximum and minimum pixel values in all microchamber images taken from $C_{warped\_background}$ and $C_{endpoint}$.

Linearly scale every pixel value in every ($C_{warped\_background}$, $C_{endpoint}$) image pair for that channel to the range [0, 1] using min-max normalization in one embodiment.

[0118] At step 1126, a warped background image is generated by subtracting the processed image from the raw endpoint image as discussed below with respect to FIG. 12. The warped background image is scaled to [0, 1] due to the inputs being scaled.

[0119] At step 1128, the scaled warped background image and the scaled endpoint image for each microchamber for each channel (e.g., FAM, VIC) is fed into the array-level min-max normalization step 1128. The minimum and maximum values across the whole array of partitions for both the scaled warped background image and the scaled endpoint images for a given channel is determined, and these values are used to normalize (e.g., linearly scale) the pixel values in both in the scaled warped background image and the scaled endpoint image using min-max normalization in one embodiment. Both of these images are concatenated together and fed into the machine learning-based classifier at step 1130 (e.g., each partition is represented by an e.g., 16x16x2 array). Note that in step 1130 and step 1122, other types of normalization generally known in the art may be used.

[0120] At step 1116, the output from positive-like and not-positive-like classifier 1112 and reject-like and not-reject-like classifier 1116 are evaluated logically and the partition is rejected if the output of classifier 1112 is positive-like and the output of classifier 1130 is reject-like, otherwise the partition is not rejected.

[0121] FIG. 12 is a flowchart illustrating method 1200 for generating a processed image according to various embodiments. A raw background image is obtained in 1210 from the background cycle before the PCR takes place. A raw endpoint image is also obtained in 1220 after PCR is completed. A warping operation is performed in 1230 to align the partitions of the background image and the partitions of the raw endpoint image to obtain a warped background image 1240. In one embodiment, the warping may be performed using a homography

transformation. The warped background image 1240 is subtracted from the raw endpoint image 1220 in subtraction operation 1250 to form processed image 1260, which is stored along with the raw endpoint image for analysis. In one embodiment, the site locations are not computed for the background cycle image, and additionally, the raw background cycle images may be discarded after image warping. The warped background image may be used as a substitute background image for later analysis. The warped background image may be obtained by subtracting the processed image from the raw endpoint cycle image (on a pixel-by-pixel basis).

[0122] FIG. 13 illustrates various methods 1300 for automating quality control by determining whether to reject partition image data for use in computing dPCR-based target concentration results. Further details about these methods 1300 are discussed throughout the specification and particularly with respect to FIGs. 11 and 12. At step 1320, image data representing a plurality of partition is obtained, including background image data and endpoint image data. At step 1340, partition background image data and partition endpoint image data for the plurality of partitions are extracted. At step 1360 the partition background and endpoint image data are pre-processed to obtain processed image data for input into a machine learning model. In step 1380, the plurality of partition images is classified by one or more trained machine-learning models as reject-like and non-reject-like partition images, and as positive-like partition images or non-positive-like partition images. The trained machine-learning model is trained by using one or more images identified as reject-like, non-reject-like, positive-like and non-positive like, as described with respect to FIG. 14. In one embodiment, classifying the plurality of partition images includes, for each partition image of the plurality of partition images, determining a probability that the partition image is a reject-like partition image and a non-positive-like partition image, as described with respect to FIG. 11. Based on the probability, the partition image data is rejected at step 1390 from use in computing the dPCR-based target concentration result.

[0123] With reference to FIG. 14, method 1400 is provided. Method 1400 is for training a first machine-learning model used for analyzing one or more biological samples to determine whether a partition image is considered reject-like or not-reject-like. A method analogous to method 1400 may also be employed for training a second machine-learning model used for analyzing one or more biological samples to determine whether a partition image is considered positive-like or not-positive-like. Method 1400 is performed by one or more computing devices.

[0124] In some embodiments, method 1400 begins with step 1420, during which a first plurality of images (e.g., images shown in 804 of FIG. 8) identified as reject-like partition images is obtained. The reject-like partition images are associated with reject-like partitions, which may have one or more of the unique signatures of partition images having data quality issues. The positive-like partition images are associated with positive-like partitions, which may exhibit characteristics of fluorescence corresponding to the existence of one or more target concentrations in the partitions. A positive partition includes a target (e.g., a DNA molecule) that has been amplified by thermal cycling so that it is detectable via fluorescence signals. A positive partition thus includes a large number of target PCR amplicons of the original DNA. The positive partition images are subsequently used for quantifying the target PCR amplicon concentration in a biological sample.

[0125] In step 1440, a second plurality of images (e.g., the normal images shown in 802 of FIG. 8, or negative partition images) identified as non-reject-like partition images is obtained. The second plurality of images may include one or more images modified from one or more other images identified as non-positive-like partition images. The non-positive-like partition images may also include at least one negative partition image associated with a negative partition, which corresponds to a non-existence of a target concentration in the negative partition. The non-positive-like partition images may also include at least one image associated with a defective microchamber of a partition. The non-positive-like partition images may also include at least one image associated with a

contaminated partition. The non-positive partition images may also include at least one image associated with a defective filling of a microchamber of a partition. The non-positive partition images may also include at least one defective image of a partition.

[0126] In some embodiments, obtaining the second plurality of images (step 1440) includes obtaining a first subset of the second plurality of images identified as non-positive partition images. The first subset of the second plurality of images is modified to obtain a second subset of the second plurality of images, which are identified as non-positive partition images. Step 1440 also includes the second subset in the second plurality of images. The modifying of the first subset of the second plurality of images includes one or more of rotating, editing, cropping, distorting, mirroring, brightening, darkening, changing a contrast of, changing a color of, and changing a pattern of, the first subset of the second plurality of images.

[0127] In some embodiments, obtaining the second plurality of images (step 1440) includes obtaining a first subset of the first plurality of images identified as positive partition images. The first subset of the first plurality of images is modified to obtain a third subset of the second plurality of images, which are identified as non-positive partition images. Step 1440 also includes the third subset in the second plurality of images. The modifying of the first subset of the first plurality of images includes one or more of rotating, editing, cropping, distorting, mirroring, brightening, darkening, changing a contrast of, changing a color of, and changing a pattern of, the first subset of the first plurality of images.

[0128] In step 1460, one or more datasets are generated using the first plurality of images (e.g., images shown in 804 of FIG. 8) and the second plurality of images (e.g., image shown in 802 of FIG. 8, and other images as discussed above). In one embodiment, the one or more datasets includes a training dataset, a validation dataset, and a testing dataset.

[0129] In step 1480, a set of parameters of the machine-learning model is determined. The set of parameters is determined by iteratively training the

machine-learning model using the one or more datasets. Based on a result of the iterative training, the set of parameters of the machine-learning model is determined. In one embodiment, the machine-learning model includes a convolutional neural network (CNN) as described above. In one embodiment, the convolutional neural network may be a variant of a VGG19 image classifier. In another embodiment, the CNN comprises 19 layers. In yet another embodiment, the CNN comprises 19 layers, including 16 convolutional layers and 3 fully connected layers.

[0130] According to embodiments, a synthetic image is generated as the visual representation of one or more chips of a sample plate after the QC process as shown in FIGS. 10-14. The synthetic image can be generated based on QC results and QC process as shown in FIGS. 10-14. As shown in FIG. 16A, process 1600 of generating the synthetic image includes determining whether channels exist at 1602. If yes, at 1604, whether images exist is determined. In response to images exist, process 1600 includes obtaining unit level rejection information including sample failure-based rejection and sample quality based rejection information at 1606. As shown in FIG. 16A, for each channel in a unit, process 1600 can include computing locations of sample partitions (e.g., microchambers) in the image coordinate space at 1608. Computing at 1608 includes the computation of individual sample partition (e.g., microchamber) locations that follows once corners of the array. The process tiles a known numbers of sample partitions across the image. These tiles then can be adjusted in a large rolling chunk to avoid minor variations in their positions using a template matching method to detect peaks corresponding to sample partition (e.g., microchamber) locations on the image. According to embodiments, process 1600 includes performing min-max normalization of a processed image using positive and negative sites to generate normalized image data at 1610. Process 1600 can further include rescaling the normalized image data to a range of [0.5 to 1.0] at 1612 to generate rescaled normalized image data. Rescaling at 1612 is configured to prevent color de-saturation of dim intensities. In embodiments, as shown at 1614, process 1600 can include computing partition (e.g., microchamber) color based on

type (positive or negative) and intensity of each partition (e.g., a microchamber or site) using the rescaled normalized image data. According to some embodiments, as shown at 1616, process 1600 includes generating image from computed attributes. The computed attributes can include the color of the site, the label of the site (whether the partition is positive, negative or rejected), location of the site, or any combination thereof. As shown at 1618, process 1600 includes adjusting image parameters of the generated image to produce an adjusted image. The image parameters can include brightness, contrast, image resolution, overall appearance of the image, or any combination thereof. At 1618, adjusting can include increasing the brightness and/or contrast of the generated image. Process 1600 can include returning the adjusted image as the synthetic image for the one or more chips of the sample plate, as shown at 1620.

[0131] According embodiments, as shown in FIG. 16B, step 1614 of computing color for partition of the one or more chips of the sample plate can include, at 1632, obtaining all positive sites. The information can include a label that indicates whether the site is positive (i.e. it is above the threshold used to discriminate between positive and negative sites). Step 1614 can include obtaining site summaries for each partition (e.g., microchamber) as shown at 1634. In some embodiments, site summaries at 1634 can include the median or mean of the pixels belonging to the sample partition, or any combination thereof. Step 1614 can include rescaling the site summaries as an alpha value as shown at 1636. The alpha value, in embodiment, can be obtained by performing a min-max normalization of the site summaries of the positive and negative sample partition so that the minimum site summary maps to 0.5 and the maximum site summary maps to 1.

[0132] Step 1614, as shown in FIG.. 16B at 1638, can include generating colors for partitions (e.g., microchambers) of the image based on the types of the sites (partition), which includes positive, negative, and rejected. As shown at 1640, for positive sites, the color is generated by blending the alpha value with the color of

the corresponding dye channel. As shown at 1642, for negative sites, the color is generated by blending the alpha value with the negative color. The negative color, in some embodiments, is chosen based on some user-interface testing. The blending at 1640 and 1642 includes mapping the RGB color values to the HSL (Hue, Saturation and Lightness) color space. In the HLS color space, the hue and saturation values are preserved while the lightness value is modified using the calculation: min(1.0, alpha * lightness of microchamber base color). The base color can correspond to a fixed color value for the positive sample partition. It is different for each filter channel supported by the instrument. The exemplary hexadecimal color values used for positives include: dark red for JUN dye "0xC00000", yellow for ABY dye "0xFFC000", green for VIC dye "0x18CF2F", blue for FAM dye "0x4472C4", light red for ROX dye"0xFF8080", other dye color "0xC0C0C0". As shown at 1644, for rejected sites, a fixed color is applied thereto. The fixed color can include reject-on color "0x595959" and reject-off color "0x000000".

[0133] Systems, apparatus, and methods described herein may be implemented using digital circuitry, or using one or more computers using well-known computer processors, memory units, storage devices, computer software, and other components. Typically, a computer includes a processor for executing instructions and one or more memories for storing instructions and data. A computer may also include, or be coupled to, one or more mass storage devices, such as one or more magnetic disks, internal hard disks and removable disks, magneto-optical disks, optical disks, etc.

[0134] Systems, apparatus, and methods described herein may be implemented using computers operating in a client-server relationship. Typically, in such a system, the client computers are located remotely from the server computers and interact via a network. The client-server relationship may be defined and controlled by computer programs running on the respective client and server computers. Examples of client computers can include desktop computers,

workstations, portable computers, cellular smartphones, tablets, or other types of computing devices.

[0135] Systems, apparatus, and methods described herein may be implemented using a computer program product tangibly embodied in an information carrier, e.g., in a non-transitory machine-readable storage device, for execution by a programmable processor; and the method processes and steps described herein, including one or more of the steps of FIGS. 2, 3, and 10-14, may be implemented using one or more computer programs that are executable by such a processor. A computer program is a set of computer program instructions that can be used, directly or indirectly, in a computer to perform a certain activity or bring about a certain result. A computer program can be written in any form of programming language, including compiled or interpreted languages, and it can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment.

[0136] A high-level block diagram of an exemplary apparatus that may be used to implement systems, apparatus and methods described herein is illustrated in FIG. 15. Apparatus 1500 comprises a processor 1510 operatively coupled to a persistent storage device 1520 and a main memory device 1530. Processor 1510 controls the overall operation of apparatus 1500 by executing computer program instructions that define such operations. The computer program instructions may be stored in persistent storage device 1520, or other computer-readable medium, and loaded into main memory device 1530 when execution of the computer program instructions is desired. For example, processor 1510 may comprise one or more components of PCR apparatus 100. Thus, the method steps of FIGS. 2, 3, and 10--14 can be defined by the computer program instructions stored in main memory device 1530 and/or persistent storage device 1520 and controlled by processor 1510 executing the computer program instructions. For example, the computer program instructions can be implemented as computer executable code programmed by one skilled in the art to perform an algorithm defined by the method steps of FIGS. 2, 3,

and 10-14. Accordingly, by executing the computer program instructions, processor 1510 executes an algorithm defined by the method steps of FIGS. 2, 3, and 10-14. Apparatus 1500 also includes one or more network interfaces 1580 for communicating with other devices via a network. Apparatus 1500 may also include one or more input/output devices 1590 that enable user interaction with apparatus 1500 (e.g., display, keyboard, mouse, speakers, buttons, etc.).

[0137] Processor 1510 may include both general and special purpose microprocessors and may be the sole processor or one of multiple processors of apparatus 1500. Processor 1510 may comprise one or more central processing units (CPUs), and one or more graphics processing units (GPUs), which, for example, may work separately from and/or multi-task with one or more CPUs to accelerate processing, e.g., for various image processing applications described herein. Processor 1510, persistent storage device 1520, and/or main memory device 1530 may include, be supplemented by, or incorporated in, one or more application-specific integrated circuits (ASICs) and/or one or more field programmable gate arrays (FPGAs).

[0138] Persistent storage device 1520 and main memory device 1530 each comprise a tangible non-transitory computer readable storage medium. Persistent storage device 1520, and main memory device 1530, may each include high-speed random access memory, such as dynamic random access memory (DRAM), static random access memory (SRAM), double data rate synchronous dynamic random access memory (DDR RAM), or other random access solid state memory devices, and may include non-volatile memory, such as one or more magnetic disk storage devices such as internal hard disks and removable disks, magneto-optical disk storage devices, optical disk storage devices, flash memory devices, semiconductor memory devices, such as erasable programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), compact disc read-only memory (CD-ROM), digital versatile disc read-only memory (DVD-ROM) disks, or other non-volatile solid state storage devices.

[0139] Input/output devices 1590 may include peripherals, such as a printer, scanner, display screen, etc. For example, input/output devices 1590 may include a display device such as a cathode ray tube (CRT), plasma or liquid crystal display (LCD) monitor for displaying information to a user, a keyboard, and a pointing device such as a mouse or a trackball by which the user can provide input to apparatus 1500.

[0140] Any or all of the functions of the systems and apparatuses discussed herein may be performed by processor 1510, and/or incorporated in, an apparatus such as PCR apparatus 100. Further, PCR apparatus 100 and/or apparatus 1500 may utilize one or more neural networks or other deep-learning techniques performed by processor 1510 or other systems or apparatuses discussed herein.

[0141] One skilled in the art will recognize that an implementation of an actual computer or computer system may have other structures and may contain other components as well, and that FIG. 15 is a high-level representation of some of the components of such a computer for illustrative purposes.

[0142] Throughout the specification and claims, the following terms take the meanings explicitly associated herein, unless the context clearly dictates otherwise:

[0143] The phrase "in one embodiment" as used herein does not necessarily refer to the same embodiment, though it may. Thus, as described below, various embodiments of the invention may be readily combined, without departing from the scope or spirit of the invention.

[0144] As used herein, the term "or" is an inclusive "or" operator and is equivalent to the term "and/or," unless the context clearly dictates otherwise.

[0145] The term "based on" is not exclusive and allows for being based on additional factors not described unless the context clearly dictates otherwise.

[0146] As used herein, and unless the context dictates otherwise, the term "coupled to" is intended to include both direct coupling (in which two elements that are coupled to each other contact each other) and indirect coupling (in which at least

one additional element is located between the two elements). Therefore, the terms "coupled to" and "coupled with" are used synonymously. Within the context of a networked environment where two or more components or devices are able to exchange data, the terms "coupled to" and "coupled with" are also used to mean "communicatively coupled with," possibly via one or more intermediary devices.

[0147] In addition, throughout the specification, the meaning of "a," "an," and "the" includes plural references, and the meaning of "in" includes "in" and "on".

[0148] Although some of the various embodiments presented herein constitute a single combination of inventive elements, it should be appreciated that the inventive subject matter is considered to include all possible combinations of the disclosed elements. As such, if one embodiment comprises elements A, B, and C, and another embodiment comprises elements B and D, then the inventive subject matter is also considered to include other remaining combinations of A, B, C, or D, even if not explicitly discussed herein. Further, the transitional term "comprising" means to have as parts or members, or to be those parts or members. As used herein, the transitional term "comprising" is inclusive or open-ended and does not exclude additional, unrecited elements or method steps.

[0149] Throughout the following disclosure, numerous references may be made regarding servers, services, interfaces, engines, modules, clients, peers, portals, platforms, or other systems formed from computing devices. It should be appreciated that the use of such terms is deemed to represent one or more computing devices having at least one processor (e.g., ASIC, FPGA, DSP, x86, ARM, ColdFire, GPU, multi-core processors, etc.) configured to execute software instructions stored on a computer readable tangible, non-transitory medium (e.g., hard drive, solid state drive, RAM, flash, ROM, etc.). For example, a server can include one or more computers operating as a web server, database server, or other type of computer server in a manner to fulfill described roles, responsibilities, or functions. One should further appreciate the disclosed computer-based algorithms, processes, methods, or other types of instruction sets can be embodied as a

computer program product comprising a non-transitory, tangible computer readable medium storing the instructions that cause a processor to execute the disclosed steps. The various servers, systems, databases, or interfaces can exchange data using standardized protocols or algorithms, possibly based on HTTP, HTTPS, AES, public-private key exchanges, web service APIs, or other electronic information exchanging methods. Data exchanges can be conducted over a packet-switched network, a circuit-switched network, the Internet, LAN, WAN, VPN, or other type of network.

[0150] As used in the description herein and throughout the claims that follow, when a system, engine, server, device, module, or other computing element is described as being configured to perform or execute functions on data in a memory, the meaning of "configured to" or "programmed to" is defined as one or more processors or cores of the computing element being programmed by a set of software instructions stored in the memory of the computing element to execute the set of functions on target data or data objects stored in the memory.

[0151] It should be noted that any language directed to a computer should be read to include any suitable combination of computing devices or network platforms, including servers, interfaces, systems, databases, agents, peers, engines, controllers, modules, or other types of computing devices operating individually or collectively. One should appreciate the computing devices comprise a processor configured to execute software instructions stored on a tangible, non-transitory computer readable storage medium (e.g., hard drive, FPGA, PLA, solid state drive, RAM, flash, ROM, etc.). The software instructions configure or program the computing device to provide the roles, responsibilities, or other functionality as discussed below with respect to the disclosed apparatus. Further, the disclosed technologies can be embodied as a computer program product that includes a non-transitory computer readable medium storing the software instructions that causes a processor to execute the disclosed steps associated with implementations of computer-based algorithms, processes, methods, or other instructions. In some

embodiments, the various servers, systems, databases, or interfaces exchange data using standardized protocols or algorithms, possibly based on HTTP, HTTPS, AES, public-private key exchanges, web service APIs, known financial transaction protocols, or other electronic information exchanging methods. Data exchanges among devices can be conducted over a packet-switched network, the Internet, LAN, WAN, VPN, or other type of packet switched network; a circuit switched network; cell switched network; or other type of network.

[0152] In the context of the specification, at least the following embodiments are described. Embodiment 1 is a A method of computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the method comprising obtaining image data representing a plurality of partitions disposed in a container, the image data including background image data corresponding to a background image captured prior to amplification cycles of a dPCR assay of the biological sample, and endpoint image data corresponding to an endpoint image captured after an endpoint cycle of the dPCR assay of the biological sample; extracting, from the obtained image data, individual partition image data for the plurality of partitions, including partition background image data and partition endpoint image data; pre-processing the partition background image data and the partition endpoint image data to obtain processed image data for input into one or more machine learning models; generating for a partition of the plurality of partitions, one or more classification results using the one or more machine learning models; and determining, using the one or more classification results, whether to reject image data corresponding to the partition of the plurality of partitions from use in computing the dPCR-based target concentration result.

[0153] Embodiment 2 is the method of embodiment 1, wherein the one or more classification results comprises a positive-like classification and a not-positive-like classification. Embodiment 3 is the method of any of embodiments 1 and 2, wherein the one or more classification results further comprises a reject-like classification and a not-reject-like classification. Embodiment 4 is the method of embodiment 3, wherein the image data corresponding to the partition is rejected if the partition is classified as not-positive-like and

reject-like. Embodiment 5 is the method of any of embodiments 1-4, wherein extracting individual partition image data for the plurality of partitions further comprises determining one or more corner locations for the plurality of partitions. Embodiment 6 is the method of embodiment 5, wherein extracting individual partition image data for the plurality of partitions further comprises determining individual partition locations in an array of partitions. Embodiment 7 is the method of any of embodiments 1-6, wherein extracting individual partition image data for the plurality of partitions further comprises determining a microchamber site summary from the individual partition image data for each partition in the plurality of partitions. Embodiment 8 is the method of embodiment 7, further comprising performing a quality control process on the image data. Embodiment 9 is the method of embodiment 8, wherein the quality control process comprises flagging an image registration error prior to generating the one or more classification results in response to one or more corner locations, individual partition locations, or a microchamber site summary cannot be determined.

[0154] Embodiment 10 is the method of embodiment 8, further comprising conducting blank image error check prior to checking image registration error. Embodiment 11 is the method of embodiment 10, wherein the blank image error check includes determining whether the image has any location error; comparing a blank score of the image with a predetermined threshold in response to the image having any location error; and designating the image as having a blank image error or an out of focus error in response to the blank score being greater than the pre-determined threshold. Embodiment 12 is the method of any of embodiments 10 and 11, further comprising turning off a channel in which the image is obtained, in response to the blank image error or out of focus error.

[0155] Embodiment 13 is the method of any of embodiments 8 to 12, wherein the quality control process further comprises generating one or more warnings in response to one or more non-failure quality control errors. Embodiment 14 is the method of embodiment 13, wherein one or more the non-failure quality control errors include a saturation error, a bridging error, a distribution of positive error, an excessive quality control rejects error, overloaded array error, or any combination thereof. Embodiment 15 is the method of embodiment 14, wherein the distribution of positive is determined by assessing uniformity of the distribution of positive over two or more regions of a chip or an array of a sample plate. Embodiment 16 is the method of any

of embodiments 1-15, wherein pre-processing the partition background image data and the partition endpoint image data comprises warping a background image corresponding to the background image data with an endpoint image corresponding to the endpoint image data such that individual partition image data corresponding to one or more of the plurality of partitions are aligned to form a warped background image. Embodiment 17 is the method of embodiment 16, wherein the warping is performed using a homography transformation. Embodiment 18 is the method of any of embodiments 16 and 17, wherein pre-processing the partition background image data and the partition endpoint image data further comprises subtracting individual partition image data corresponding to one or more of the plurality of partitions in the endpoint image from individual partition image data corresponding to the one or more of the plurality of partitions in the warped background image to form processed image data. Embodiment 19 is the method of embodiment 18, wherein a first machine learning model is used on the processed image data to generate one or more classification results corresponding to one or more of the partitions in the processed image. Embodiment 20 is the method of embodiment 19, wherein the one or more classification results of the first machine learning model comprises a positive-like classification and a not-positive-like classification.

[0156] Embodiment 21 is the method of embodiment 20, further comprising performing maximum absolute rescaling on the processed image data to the range [-1, 1] prior to generating the one or more classification results. Embodiment 22 is the method of embodiment 21, further comprising performing quantile-based normalization on the rescaled processed image data prior to classification. Embodiment 23 is the method of any of embodiments 16-22, wherein a second machine learning model is used on the warped background image data and the endpoint image data to generate one or more classification results corresponding to one or more of the partitions in the warped background image. Embodiment 24 is the method of embodiment 23, wherein the one or more classification results of the second machine learning model comprises a reject-like classification and a not-reject-like classification. Embodiment 25 is the method of embodiment 24, further comprising rescaling the warped background image data and the endpoint image data to the range [0, 1] prior to generating the one or more classification results. Embodiment 26 is the method of embodiment 25, further comprising performing min-max normalization on the warped background image data and the endpoint image data prior to generating the one or more classification results. Embodiment 27 is the method of any of embodiments 19-26, wherein one

or more of the first or second machine learning models comprises a convolutional neural network. Embodiment 28 is the method of any of embodiments 1-27, wherein the image data is associated with a first fluorescence channel of a dPCR apparatus comprising a plurality of fluorescence channels including the first fluorescence channel, the plurality of fluorescence channels being associated with different spectral wavelengths. Embodiment 29 is the method of embodiment 28, wherein one of the plurality of fluorescence channels comprises a channel using a 6-carboxy-X-rhodamine (ROX) based dye. Embodiment 30 is the method of any of embodiments 1-29, wherein the container comprises a microfluidic array plate. Embodiment 31 is the method of any of embodiments 1-30, further comprising generating a synthetic image for an array comprising a plurality of partitions. Embodiment 32 is the method of embodiment 31, wherein the synthetic image is generated based on the one or more classification results.

[0157] Embodiment 33 is a non-transitory computer readable medium comprising a memory storing one or more instructions which, when executed by one or more processors of at least one computing device, perform computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, comprising obtaining image data representing a plurality of partitions disposed in a container, the image data including background image data corresponding to a background image captured prior to amplification cycles of a dPCR assay of the biological sample, and endpoint image data corresponding to an endpoint image captured after an endpoint cycle of the dPCR assay of the biological sample; extracting, from the obtained image data, individual partition image data for the plurality of partitions, including partition background image data and partition endpoint image data; pre-processing the partition background image data and the partition endpoint image data to obtain processed image data for input into one or more machine learning models; generating for a partition of the plurality of partitions, one or more classification results using the one or more machine learning models; and determining, using the one or more classification results, whether to reject image data corresponding to the partition of the plurality of partitions from use in computing the dPCR-based target concentration result.

[0158] Embodiment 34 is a system for computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the system comprising one or more processors of at least one computing device; and a memory storing one or more instructions, when executed by the one or more processors, cause the one or more processors to perform processing comprising obtaining image data representing a plurality of partitions disposed in a container, the image data including background image data corresponding to a background image captured prior to amplification cycles of a dPCR assay of the biological sample, and endpoint image data corresponding to an endpoint image captured after an endpoint cycle of the dPCR assay of the biological sample; extracting, from the obtained image data, individual partition image data for the plurality of partitions, including partition background image data and partition endpoint image data; pre-processing the partition background image data and the partition endpoint image data to obtain processed image data for input into one or more machine learning models; generating for a partition of the plurality of partitions, one or more classification results using the one or more machine learning models; and determining, using the one or more classification results, whether to reject image data corresponding to the partition of the plurality of partitions from use in computing the dPCR-based target concentration result.

[0159] Embodiment 35 is a method of training of training a machine-learning model to perform computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the method being performed by one or more computing devices and comprising obtaining a first plurality of images identified as positive-like partition images; obtaining a second plurality of images identified as non-positive-like partition images; generating one or more datasets using the first plurality of images and the second plurality of images; and determining, by the one or more computing devices, a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets, wherein a trained machine-learning model is configured based on the set of parameters to classify individual partitions as non-

positive-like partitions based on one or more quality control issues comprising one or more of bubbles, gas, debris, delamination, leaks, fill issues, saturation, bridging, and dust.

[0160] Embodiment 36 is the method of embodiment 35, wherein the positive partition images are associated with positive-like partitions, wherein the positive-like partitions correspond to an existence of one or more target concentrations in the positive partitions. Embodiment 37 is the method of any of embodiments 35-36, wherein the non-positive-like partition images comprise at least one negative partition image associated with a negative partition, wherein the negative partition corresponds to a non-existence of a target concentration in the negative partition. Embodiment 38 is the method of any of embodiments 35-37, wherein the non-positive-like partition images comprise at least one image associated with a defective microchamber of a partition. Embodiment 39 is the method of any of embodiments 35-38, wherein the non-positive-like partition images comprise at least one image associated with a contaminated partition. Embodiment 40 is the method of any of embodiments 35-39, wherein the non-positive-like partition images comprise at least one image associated with a defective filling of a microchamber of a partition. Embodiment 41 is the method of any of embodiments 35-40, wherein the non-positive-like partition images comprise at least one defective image of a partition. Embodiment 42 is the method of any of embodiments 35-41, wherein the one or more datasets comprise a training dataset, a validation dataset, and a testing dataset. Embodiment 43 is the method of any of embodiments 35-42, wherein determining the set of parameters of the machine-learning model comprises iteratively training the machine-learning model using the one or more datasets; and determining the set of parameters of the machine-learning model based on a result of the iterative training. Embodiment 44 is the method of any of embodiments 35-43, wherein the machine-learning model comprises a convolutional neural network (CNN).

[0161] Embodiment 45 is a method for training a machine-learning model to perform computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the method being performed by one or more computing devices and comprising obtaining a first plurality of images identified as reject-like partition images; obtaining a second plurality of images identified as non-reject-like partition images; generating one or more datasets using the first plurality of

images and the second plurality of images; and determining, by the one or more computing devices, a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets, wherein a trained machine-learning model is configured based on the set of parameters to reject individual partitions based on one or more quality control issues comprising one or more of bubbles, gas, debris, delamination, leaks, fill issues, saturation, bridging, and dust.

[0162] Embodiment 46 is the method of embodiment 45, wherein the one or more datasets comprise a training dataset, a validation dataset, and a testing dataset. Embodiment 47 is the method of any of embodiments 45 and 46, wherein determining the set of parameters of the machine learning model comprises iteratively training the machine learning model using the one or more datasets; and determining the set of parameters of the machine learning model based on a result of the iterative training. Embodiment 48 is the method of any of embodiments 45-47, wherein the machine learning model comprises a convolutional neural network. Embodiment 49 is the method of embodiment 45-48, further comprising obtaining a pre-dPCR image representing the plurality of partitions before amplification of one or more targets in the biological sample; obtaining a post-dPCR image representing the plurality of partitions after amplification of the one or more targets in the biological sample; and performing image subtraction using the pre-dPCR image and the post-dPCR image to obtain the plurality of partition images provided to the trained machine-learning model.

[0163] Embodiment 50 is a non-transitory computer readable medium comprising a memory storing one or more instructions which, when executed by one or more processors of at least one computing device, perform training of a machine-learning model used for computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, by performing processing comprising obtaining a first plurality of images identified as reject-like partition images; obtaining a second plurality of images identified as non-reject-like partition images; generating one or more datasets using the first plurality of images and the second plurality of images; and determining, by the one or more computing devices, a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets, wherein a trained machine-learning model

is configured based on the set of parameters to reject individual partitions based on one or more quality control issues comprising one or more of bubbles, gas, debris, delamination, leaks, fill issues, saturation, bridging, and dust.

[0164] Embodiment 51 is a system for training a machine-learning model used for computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the system comprising one or more processors of at least one computing device; and a memory storing one or more instructions, which, when executed by the one or more processors, cause the one or more processors to perform processing comprising obtaining a first plurality of images identified as reject-like partition images; obtaining a second plurality of images identified as non-reject-like partition images; generating one or more datasets using the first plurality of images and the second plurality of images; and determining, by the one or more computing devices, a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets, wherein a trained machine-learning model is configured based on the set of parameters to reject individual partitions based on one or more quality control issues comprising one or more of bubbles, gas, debris, delamination, leaks, fill issues, saturation, bridging, and dust.

[0165] The foregoing specification is to be understood as being in every respect illustrative and exemplary, but not restrictive, and the scope of the invention disclosed herein is not to be determined from the specification, but rather from the claims as interpreted according to the full breadth permitted by the patent laws. It is to be understood that the embodiments shown and described herein are only illustrative of the principles of the present invention and that various modifications may be implemented by those skilled in the art without departing from the scope and spirit of the invention. Those skilled in the art could implement various other feature combinations without departing from the scope and spirit of the invention.

CLAIMS

What is claimed is:

1. A method of computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the method comprising:

obtaining image data representing a plurality of partitions disposed in a container, the image data including background image data corresponding to a background image captured prior to amplification cycles of a dPCR assay of the biological sample, and endpoint image data corresponding to an endpoint image captured after an endpoint cycle of the dPCR assay of the biological sample;

extracting, from the obtained image data, individual partition image data for the plurality of partitions, including partition background image data and partition endpoint image data;

pre-processing the partition background image data and the partition endpoint image data to obtain processed image data for input into one or more machine learning models;

generating for a partition of the plurality of partitions, one or more classification results using the one or more machine learning models; and

determining, using the one or more classification results, whether to reject image data corresponding to the partition of the plurality of partitions from use in computing the dPCR-based target concentration result.

2. The method of claim 1, wherein the one or more classification results comprises a positive-like classification and a not-positive-like classification.

3. The method of any of claims 1 and 2, wherein the one or more classification results further comprises a reject-like classification and a not-reject-like classification.

4. The method of claim 3, wherein the image data corresponding to the partition is rejected if the partition is classified as not-positive-like and reject-like.

5. The method of any of claims 1-4, wherein extracting individual partition image data for the plurality of partitions further comprises determining one or more corner locations for the plurality of partitions.

6. The method of claim 5, wherein extracting individual partition image data for the plurality of partitions further comprises determining individual partition locations in an array of partitions.

7. The method of any of claims 1-6, wherein extracting individual partition image data for the plurality of partitions further comprises determining a microchamber site summary from the individual partition image data for each partition in the plurality of partitions.

8. The method of claim 7, further comprising performing a quality control process on the image data.

9. The method of claim 8, wherein the quality control process comprises flagging an image registration error prior to generating the one or more classification results in response to one or more corner locations, individual partition locations, or a microchamber site summary cannot be determined.

10. The method of claim 8, further comprising conducting blank image error check prior to checking image registration error.

11. The method of claim 10, wherein the blank image error check includes:

determining whether the image has any location error;

comparing a blank score of the image with a predetermined threshold in response to the image having any location error; and

designating the image as having a blank image error or an out of focus error in response to the blank score being greater than the pre-determined threshold.

12. The method of any of claims 10 and 11, further comprising turning off a channel in which the image is obtained, in response to the blank image error or out of focus error.

13. The method of any of claims 8 to 12, wherein the quality control process further comprises generating one or more warnings in response to one or more non-failure quality control errors.

14. The method of claim 13, wherein one or more the non-failure quality control errors include a saturation error, a bridging error, a distribution of positive error, an excessive quality control rejects error, overloaded array error, or any combination thereof.

15. The method of claim 14, wherein the distribution of positive is determined by assessing uniformity of the distribution of positive over two or more regions of a chip or an array of a sample plate.

16. The method of any of claims 1-15, wherein pre-processing the partition background image data and the partition endpoint image data comprises warping a background image corresponding to the background image data with an endpoint

image corresponding to the endpoint image data such that individual partition image data corresponding to one or more of the plurality of partitions are aligned to form a warped background image.

17. The method of claim 16, wherein the warping is performed using a homography transformation.

18. The method of any of claims 16 and 17, wherein pre-processing the partition background image data and the partition endpoint image data further comprises subtracting individual partition image data corresponding to one or more of the plurality of partitions in the endpoint image from individual partition image data corresponding to the one or more of the plurality of partitions in the warped background image to form processed image data.

19. The method of claim 18, wherein a first machine learning model is used on the processed image data to generate one or more classification results corresponding to one or more of the partitions in the processed image.

20. The method of claim 19, wherein the one or more classification results of the first machine learning model comprises a positive-like classification and a not-positive-like classification.

21. The method of claim 20, further comprising performing maximum absolute rescaling on the processed image data to the range [-1, 1] prior to generating the one or more classification results.

22. The method of claim 21, further comprising performing quantile-based normalization on the rescaled processed image data prior to classification.

23. The method of any of claims 16-22, wherein a second machine learning model is used on the warped background image data and the endpoint image data to generate one or more classification results corresponding to one or more of the partitions in the warped background image.

24. The method of claim 23, wherein the one or more classification results of the second machine learning model comprises a reject-like classification and a not-reject-like classification.

25. The method of claim 24, further comprising rescaling the warped background image data and the endpoint image data to the range [0, 1] prior to generating the one or more classification results.

26. The method of claim 25, further comprising performing min-max normalization on the warped background image data and the endpoint image data prior to generating the one or more classification results.

27. The method of any of claims 19-26, wherein one or more of the first or second machine learning models comprises a convolutional neural network.

28. The method of any of claims 1-27, wherein the image data is associated with a first fluorescence channel of a dPCR apparatus comprising a plurality of

fluorescence channels including the first fluorescence channel, the plurality of fluorescence channels being associated with different spectral wavelengths.

29. The method of claim 28, wherein one of the plurality of fluorescence channels comprises a channel using a 6-carboxy-X-rhodamine (ROX) based dye.

30. The method of any of claims 1-29, wherein the container comprises a microfluidic array plate.

31. The method of any of claims 1-30, further comprising generating a synthetic image for an array comprising a plurality of partitions.

32. The method of claim 31, wherein the synthetic image is generated based on the one or more classification results.

33. A non-transitory computer readable medium comprising a memory storing one or more instructions which, when executed by one or more processors of at least one computing device, perform computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, comprising:

obtaining image data representing a plurality of partitions disposed in a container, the image data including background image data corresponding to a background image captured prior to amplification cycles of a dPCR assay of the biological sample, and endpoint image data corresponding to an endpoint image captured after an endpoint cycle of the dPCR assay of the biological sample;

extracting, from the obtained image data, individual partition image data for the plurality of partitions, including partition background image data and partition endpoint image data;

pre-processing the partition background image data and the partition endpoint image data to obtain processed image data for input into one or more machine learning models;

generating for a partition of the plurality of partitions, one or more classification results using the one or more machine learning models; and

determining, using the one or more classification results, whether to reject image data corresponding to the partition of the plurality of partitions from use in computing the dPCR-based target concentration result.

34. A system for computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the system comprising:

one or more processors of at least one computing device; and

a memory storing one or more instructions, when executed by the one or more processors, cause the one or more processors to perform processing comprising:

obtaining image data representing a plurality of partitions disposed in a container, the image data including background image data corresponding to a background image captured prior to amplification cycles of a dPCR assay of the biological sample, and endpoint image data corresponding to an endpoint image captured after an endpoint cycle of the dPCR assay of the biological sample;

extracting, from the obtained image data, individual partition image data for the plurality of partitions, including partition background image data and partition endpoint image data;

pre-processing the partition background image data and the partition endpoint image data to obtain processed image data for input into one or more machine learning models;

generating for a partition of the plurality of partitions, one or more classification results using the one or more machine learning models; and

determining, using the one or more classification results, whether to reject image data corresponding to the partition of the plurality of partitions from use in computing the dPCR-based target concentration result.

35. A method of training of training a machine-learning model to perform computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the method being performed by one or more computing devices and comprising:

obtaining a first plurality of images identified as positive-like partition images;

obtaining a second plurality of images identified as non-positive-like partition images;

generating one or more datasets using the first plurality of images and the second plurality of images; and

determining, by the one or more computing devices, a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets,

wherein a trained machine-learning model is configured based on the set of parameters to classify individual partitions as non-positive-like partitions based on one or more quality control issues comprising one or more of bubbles, gas, debris, delamination, leaks, fill issues, saturation, bridging, and dust.

36. The method of claim 35, wherein the positive partition images are associated with positive-like partitions, wherein the positive-like partitions

correspond to an existence of one or more target concentrations in the positive partitions.

37. The method of any of claims 35-36, wherein the non-positive-like partition images comprise at least one negative partition image associated with a negative partition, wherein the negative partition corresponds to a non-existence of a target concentration in the negative partition.

38. The method of any of claims 35-37, wherein the non-positive-like partition images comprise at least one image associated with a defective microchamber of a partition.

39. The method of any of claims 35-38, wherein the non-positive-like partition images comprise at least one image associated with a contaminated partition.

40. The method of any of claims 35-39, wherein the non-positive-like partition images comprise at least one image associated with a defective filling of a microchamber of a partition.

41. The method of any of claims 35-40, wherein the non-positive-like partition images comprise at least one defective image of a partition.

42. The method of any of claims 35-41, wherein the one or more datasets comprise a training dataset, a validation dataset, and a testing dataset.

43. The method of any of claims 35-42, wherein determining the set of parameters of the machine-learning model comprises:

iteratively training the machine-learning model using the one or more datasets; and

determining the set of parameters of the machine-learning model based on a result of the iterative training.

44. The method of any of claims 35-43, wherein the machine-learning model comprises a convolutional neural network (CNN).

45. A method for training a machine-learning model to perform computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the method being performed by one or more computing devices and comprising:

obtaining a first plurality of images identified as reject-like partition images;

obtaining a second plurality of images identified as non-reject-like partition images;

generating one or more datasets using the first plurality of images and the second plurality of images; and

determining, by the one or more computing devices, a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets,

wherein a trained machine-learning model is configured based on the set of parameters to reject individual partitions based on one or more quality control issues comprising one or more of bubbles, gas, debris, delamination, leaks, fill issues, saturation, bridging, and dust.

46. The method of claim 45, wherein the one or more datasets comprise a training dataset, a validation dataset, and a testing dataset.

47. The method of any of claims 45 and 46, wherein determining the set of parameters of the machine learning model comprises:

iteratively training the machine learning model using the one or more datasets; and

determining the set of parameters of the machine learning model based on a result of the iterative training.

48. The method of any of claims 45-47, wherein the machine learning model comprises a convolutional neural network.

49. The method of claim 45-48, further comprising:

obtaining a pre-dPCR image representing the plurality of partitions before amplification of one or more targets in the biological sample;

obtaining a post-dPCR image representing the plurality of partitions after amplification of the one or more targets in the biological sample; and

performing image subtraction using the pre-dPCR image and the post-dPCR image to obtain the plurality of partition images provided to the trained machine-learning model.

50. A non-transitory computer readable medium comprising a memory storing one or more instructions which, when executed by one or more processors of at least one computing device, perform training of a machine-learning model used for computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions

from use in computing a dPCR-based target concentration result, by performing processing comprising:

obtaining a first plurality of images identified as reject-like partition images;

obtaining a second plurality of images identified as non-reject-like partition images;

generating one or more datasets using the first plurality of images and the second plurality of images; and

determining, by the one or more computing devices, a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets,

wherein a trained machine-learning model is configured based on the set of parameters to reject individual partitions based on one or more quality control issues comprising one or more of bubbles, gas, debris, delamination, leaks, fill issues, saturation, bridging, and dust.

51. A system for training a machine-learning model used for computerized image processing for digital polymerase chain reaction (dPCR) analysis of a biological sample to automatically reject data of individual partitions from use in computing a dPCR-based target concentration result, the system comprising:

one or more processors of at least one computing device; and

a memory storing one or more instructions, which, when executed by the one or more processors, cause the one or more processors to perform processing comprising:

obtaining a first plurality of images identified as reject-like partition images;

obtaining a second plurality of images identified as non-reject-like partition images;

generating one or more datasets using the first plurality of images and the second plurality of images; and

determining, by the one or more computing devices, a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets,

wherein a trained machine-learning model is configured based on the set of parameters to reject individual partitions based on one or more quality control issues comprising one or more of bubbles, gas, debris, delamination, leaks, fill issues, saturation, bridging, and dust.

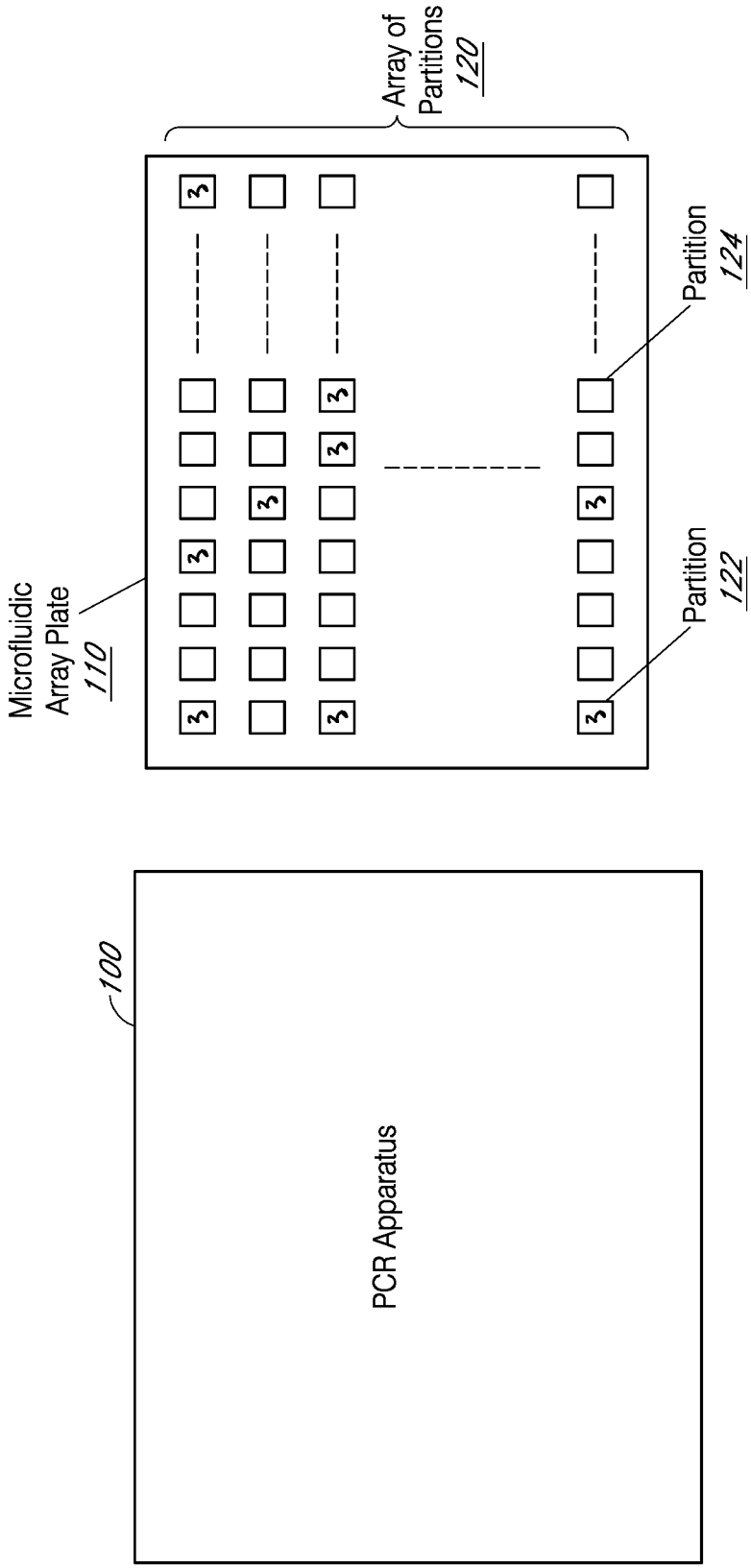
Array of Partitions 120
Microfluidic Array Plate 110
Partition 124
Partition 122
100
PCR Apparatus

FIG. 1

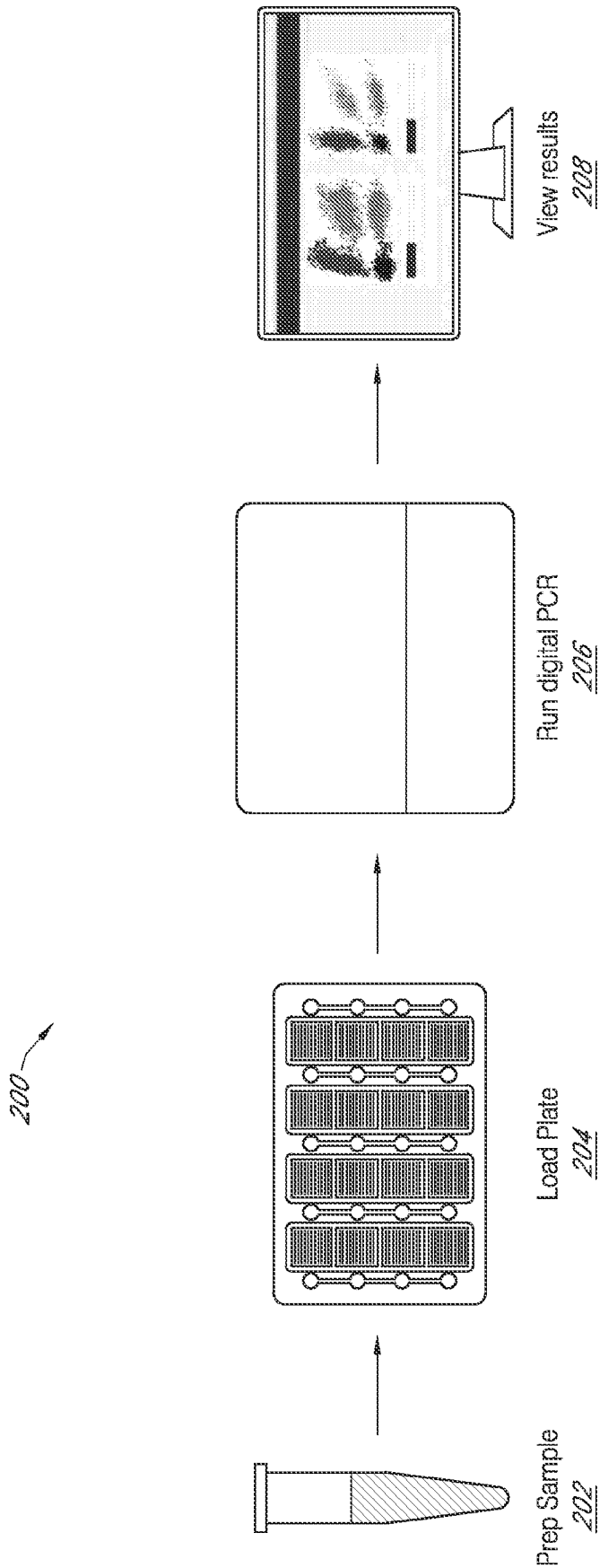
200
Prep Sample
202
Load Plate
204
Run digital PCR
206
View results
208

FIG. 2

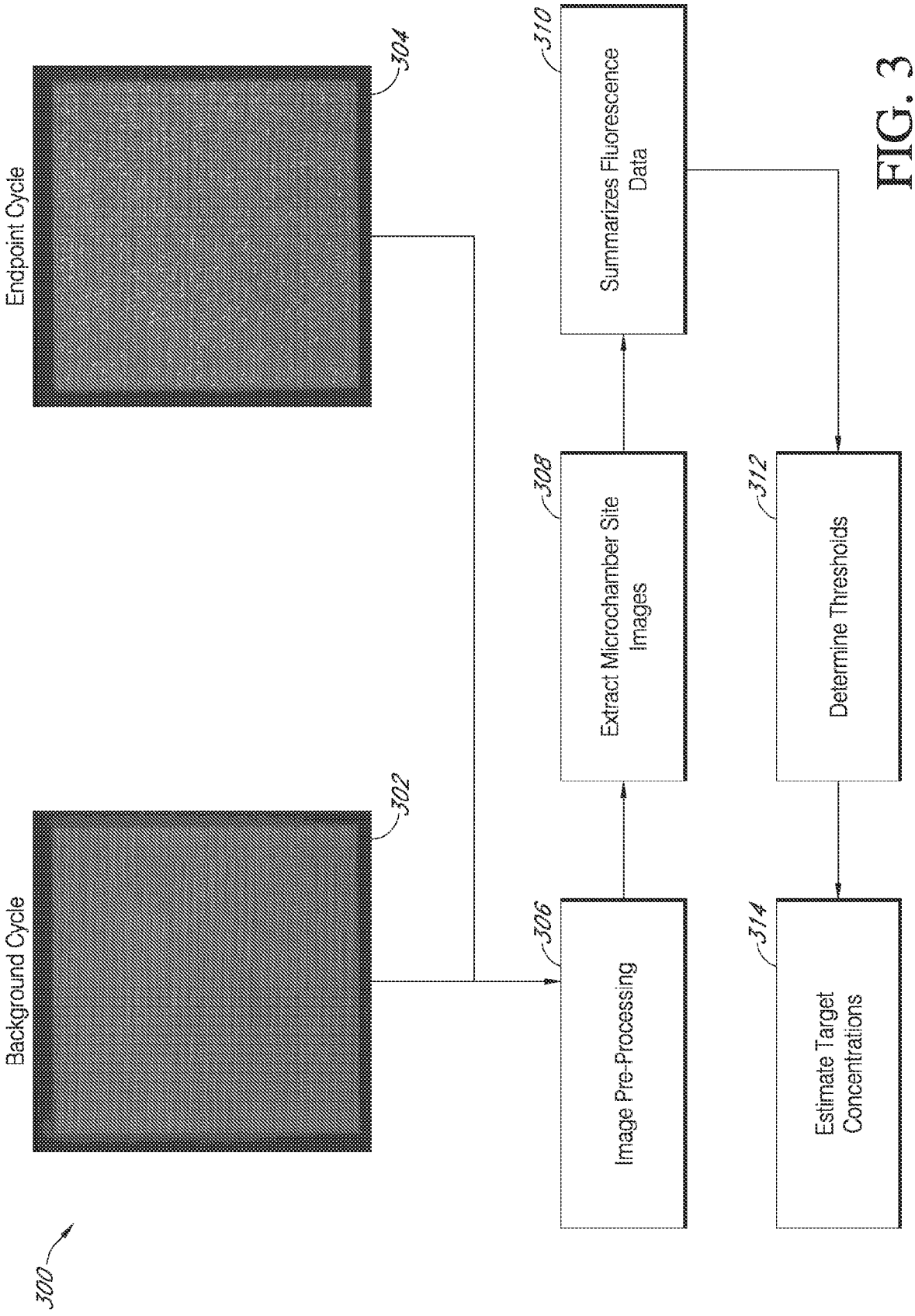
300
Endpoint Cycle
304
Background Cycle
302
Image Pre-Processing
306
Extract Microchamber Site Images
308
Summarizes Fluorescence Data
310
Determine Thresholds
312
Estimate Target Concentrations
314

FIG. 3

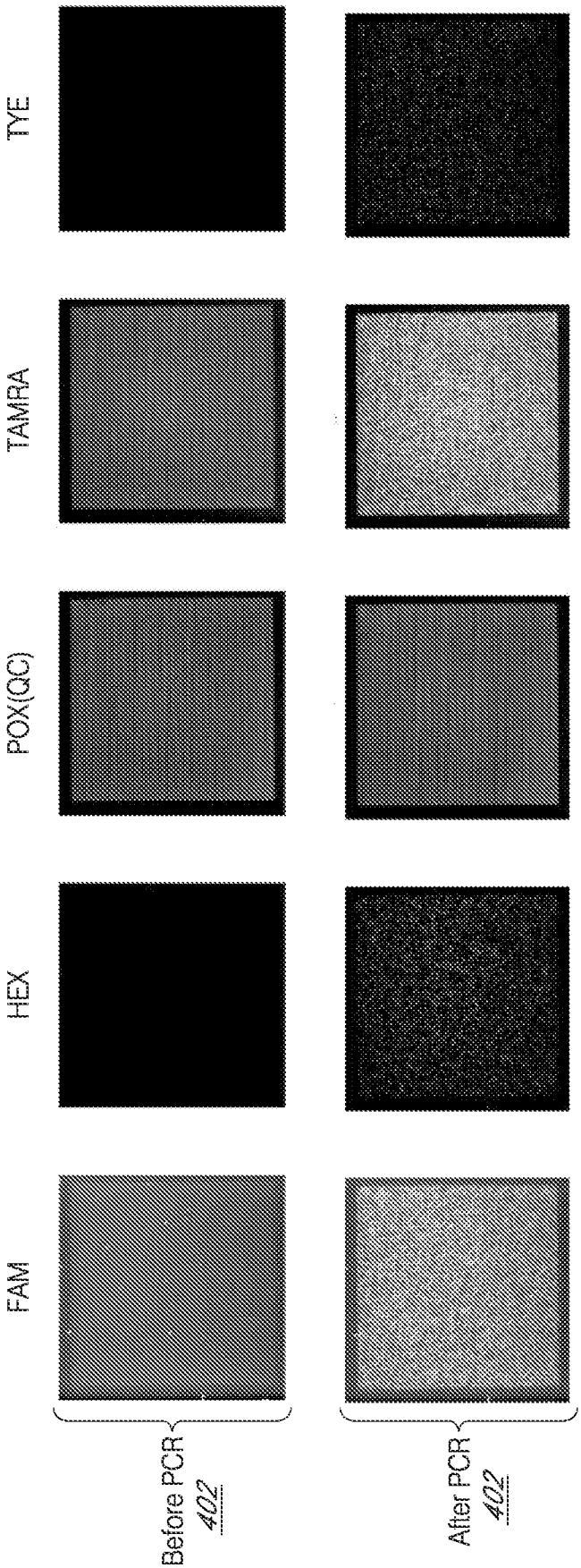

FIG. 4

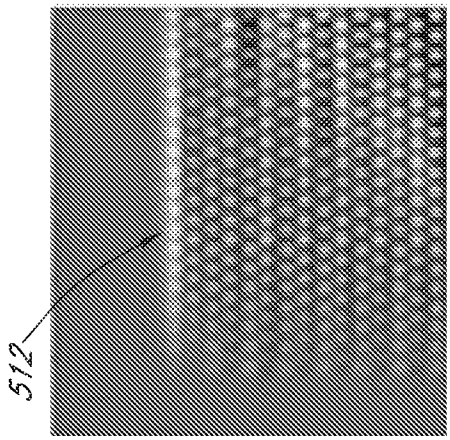

FIG. 5B

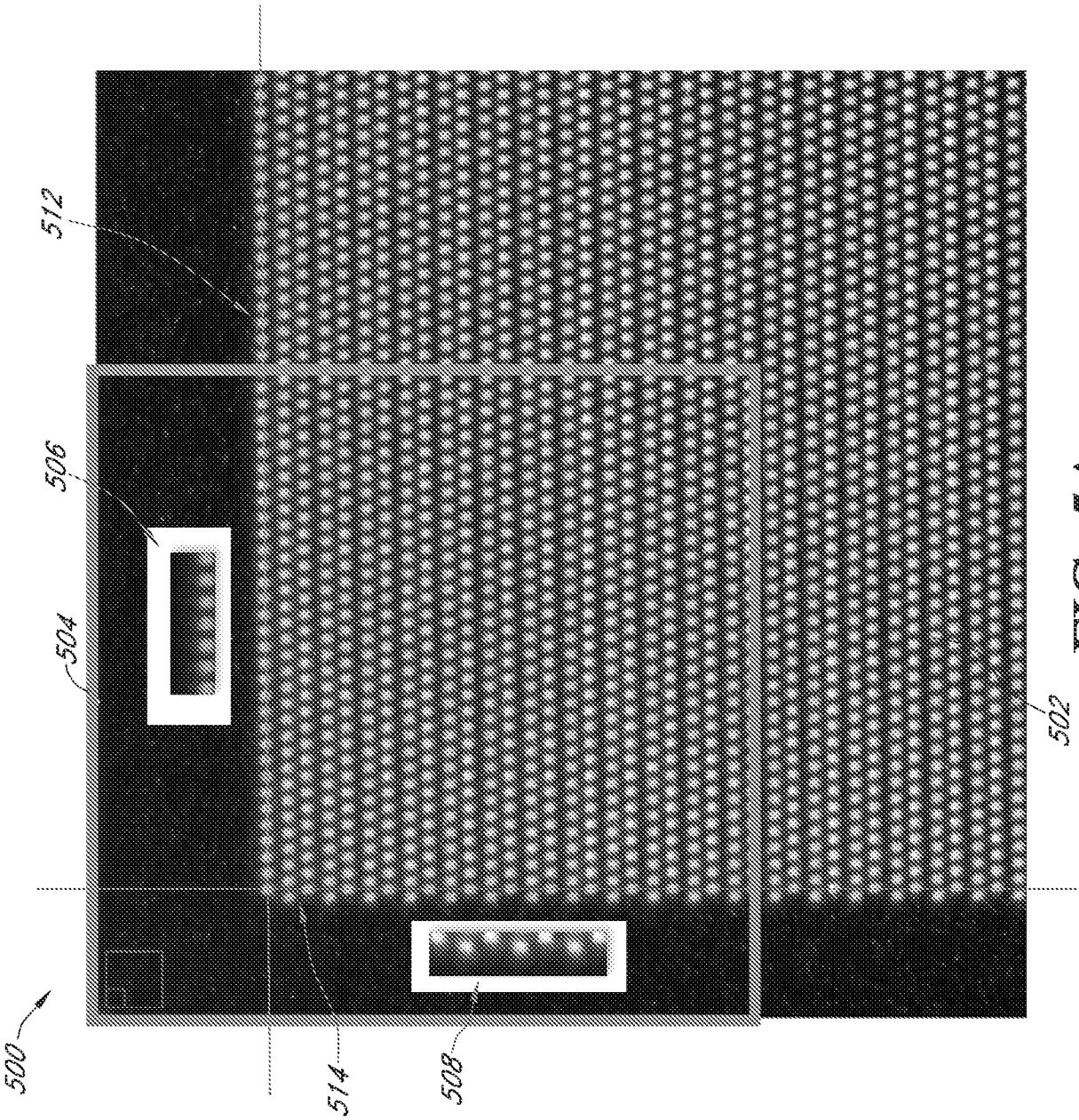

FIG. 5A

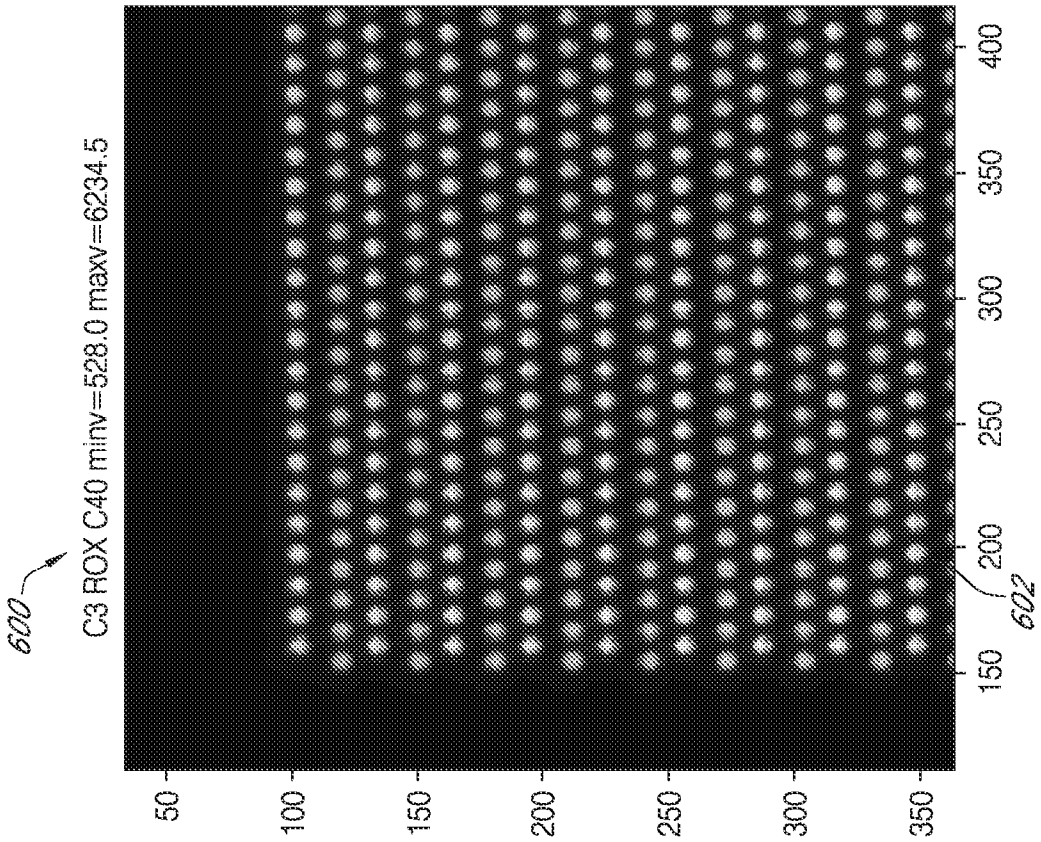
600
C3 ROX C40 minv=528.0 maxv=6234.5
602
0
500
1000
1500
2000
0
250
500
750
1000
1250
1500
1750
2000

FIG. 6A

600
C3 ROX C40 minv=528.0 maxv=6234.5
602
150
200
250
300
350
400
50
100
150
200
250
300
350

FIG. 6B

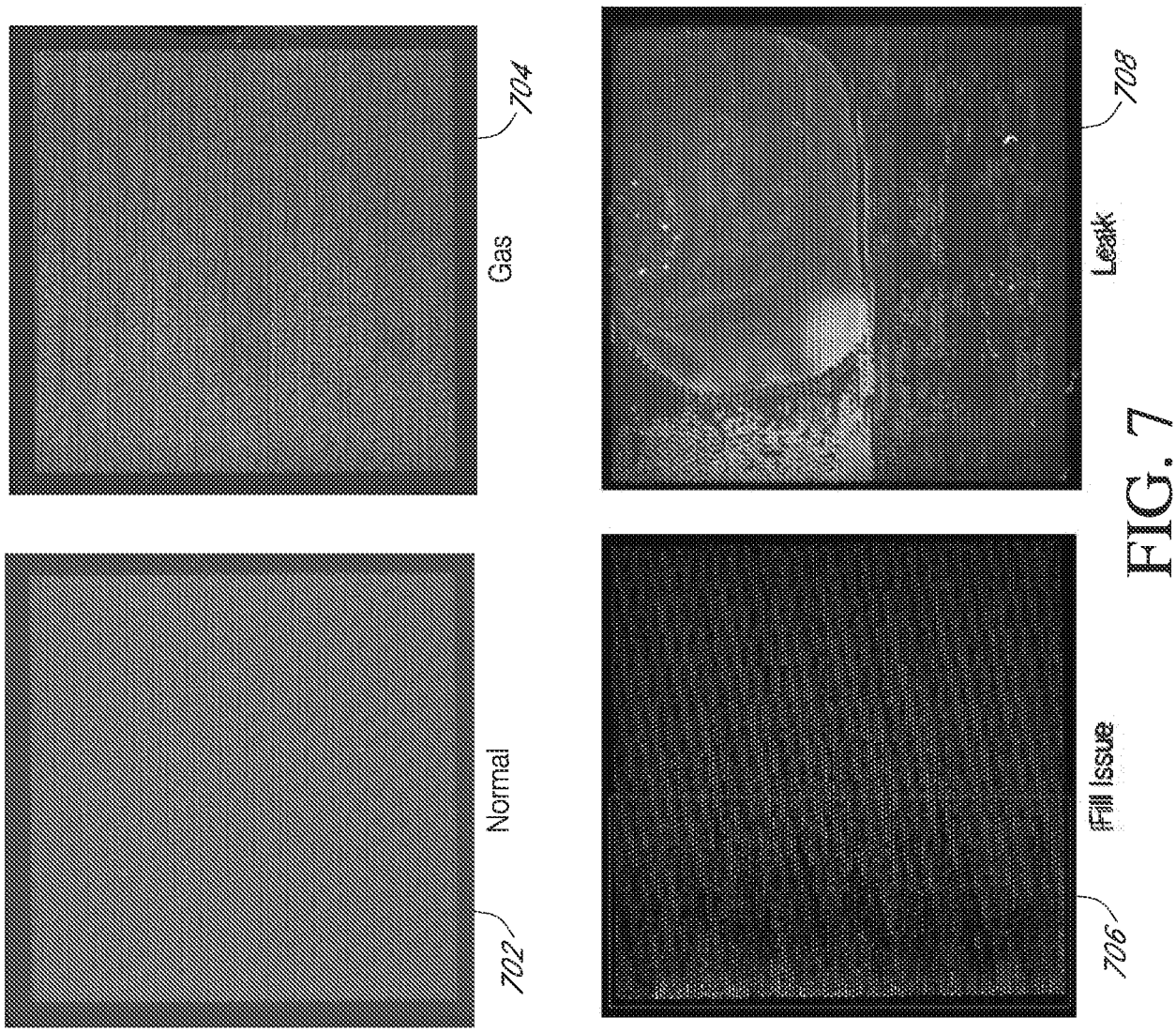

FIG. 7

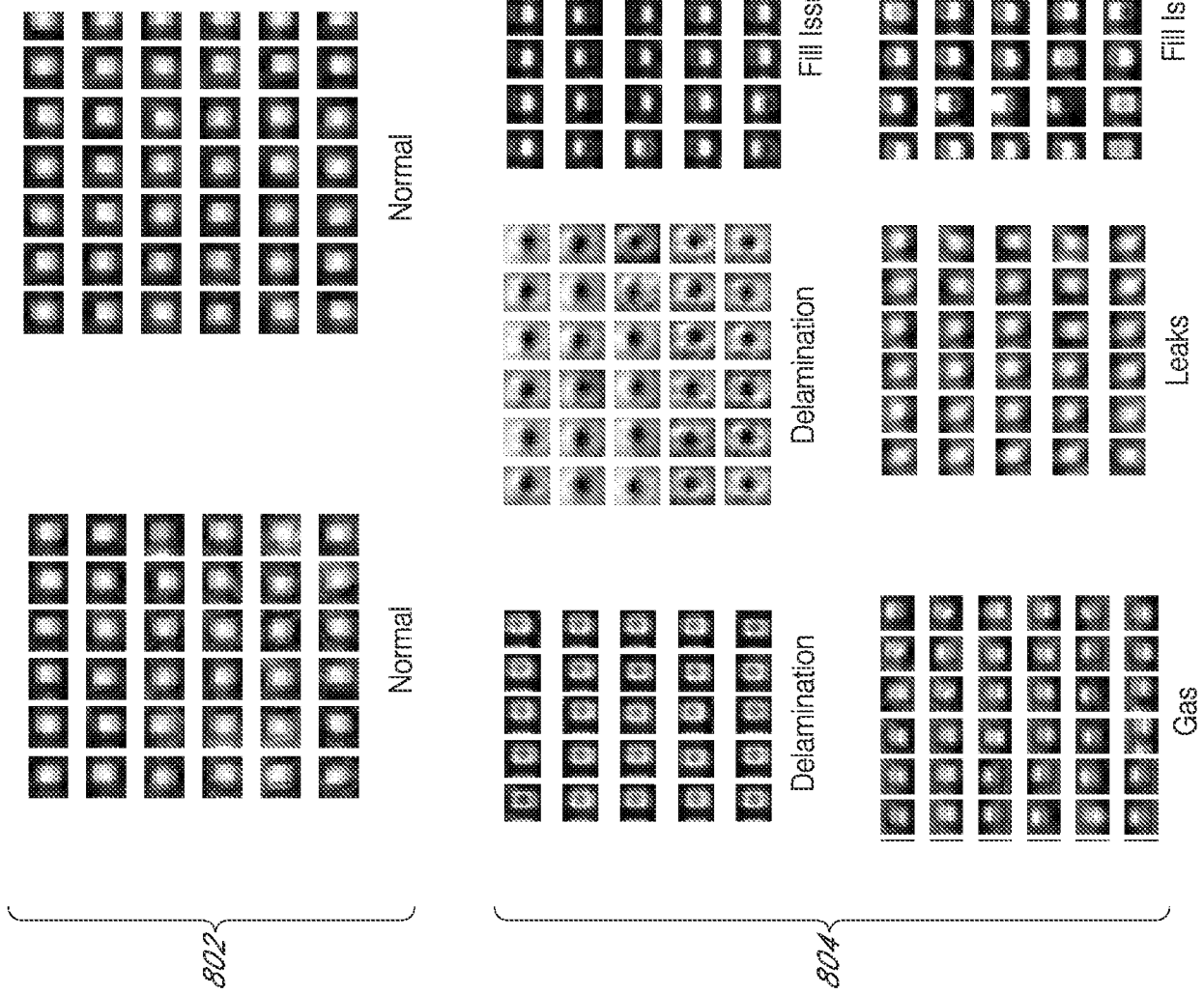
802
804
Normal
Normal
Delamination
Fill Issues
Delamination
Leaks
Gas
Fill Issues

FIG. 8

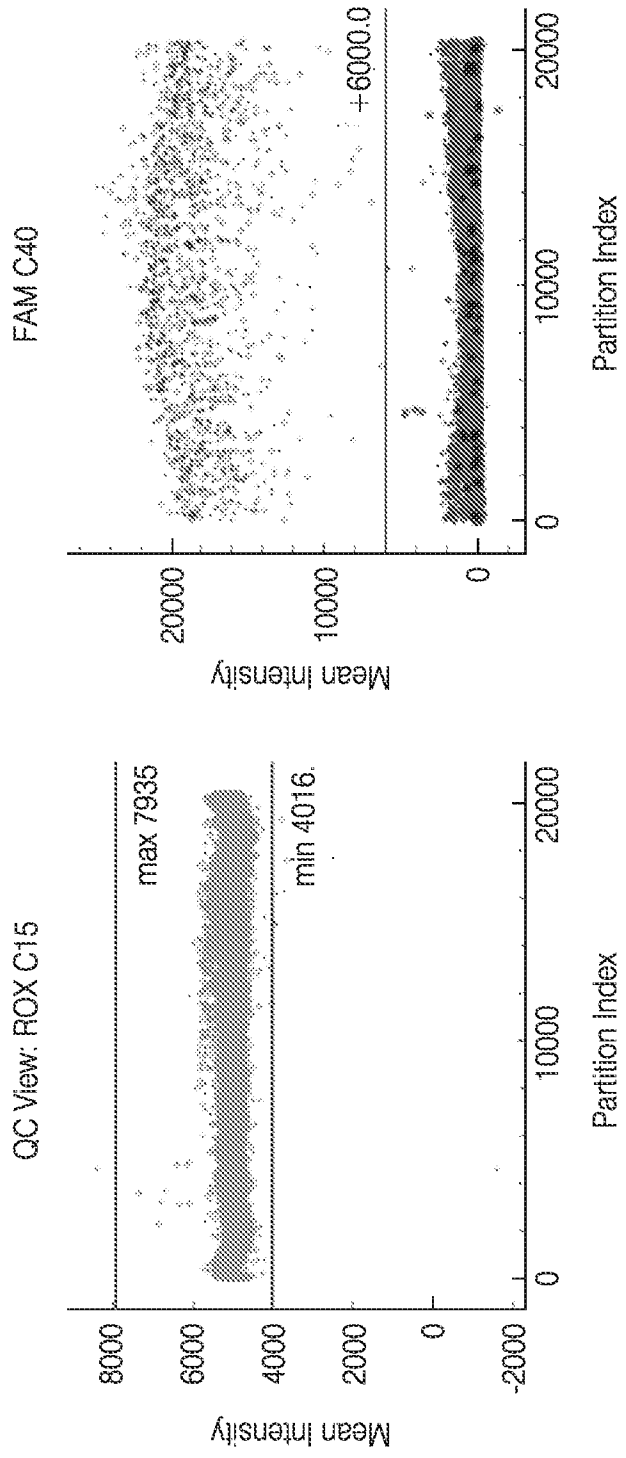
QC View: ROX C15
max 7935
min 4016.
Mean Intensity
8000
6000
4000
2000
0
-2000
0
10000
20000
Partition Index
FAM C40
+6000.0
Mean Intensity
20000
10000
0
0
10000
20000
Partition Index
Normal Data

FIG. 9

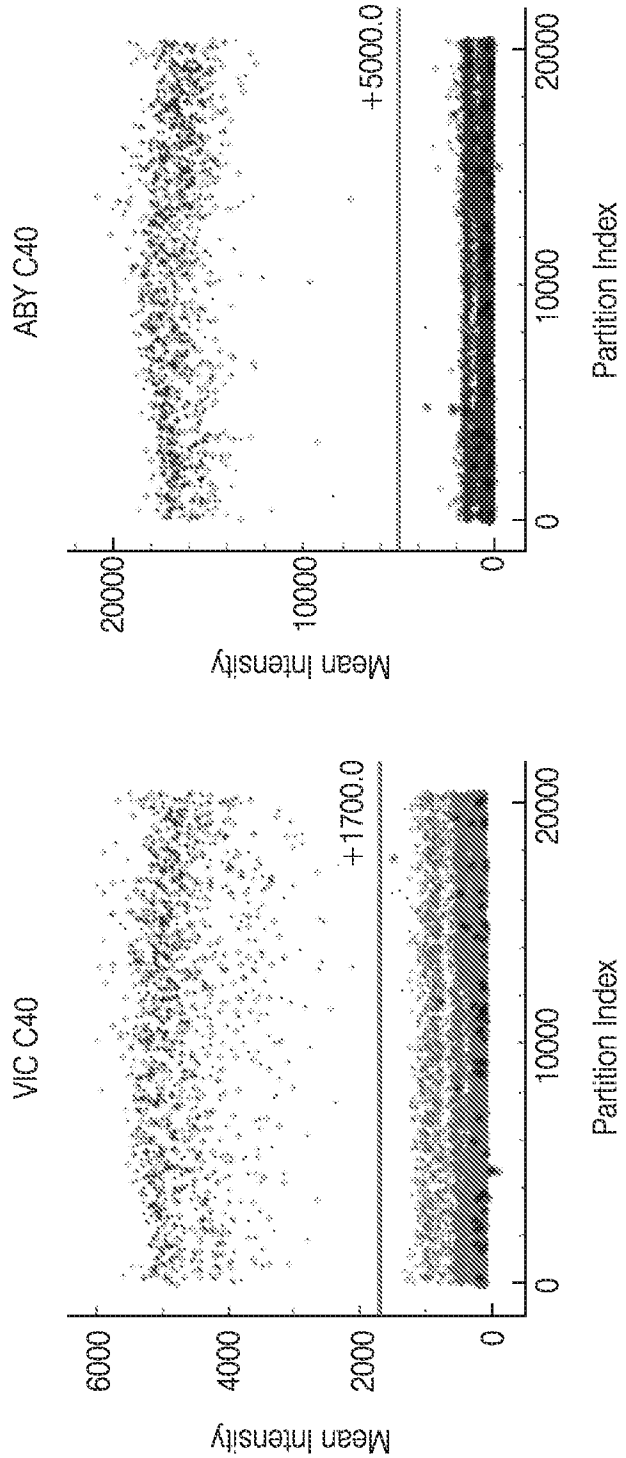
ABY C40
+5000.0
Mean Intensity
20000
10000
0
0
10000
20000
Partition Index
VIC C40
+1700.0
Mean Intensity
6000
4000
2000
0
0
10000
20000
Partition Index
Normal Data

FIG. 9 (Cont.)

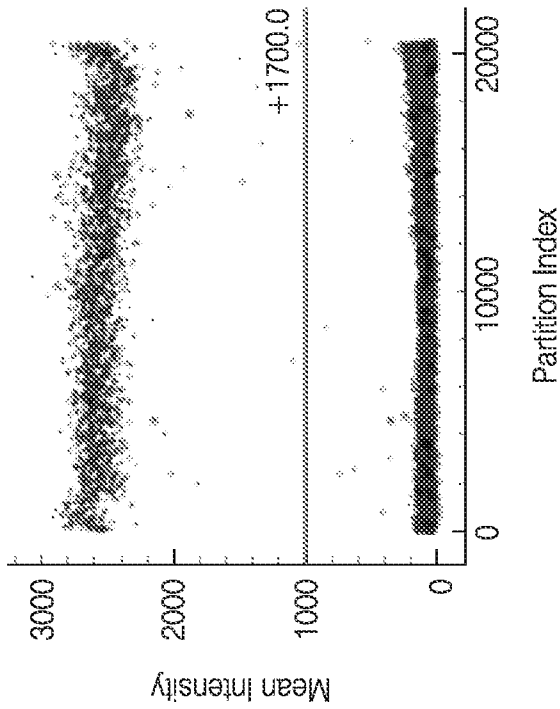
JUN C40
+1700.0
Mean Intensity
3000
2000
1000
0
0
10000
20000
Partition Index
Normal Data

FIG. 9 (Cont.)

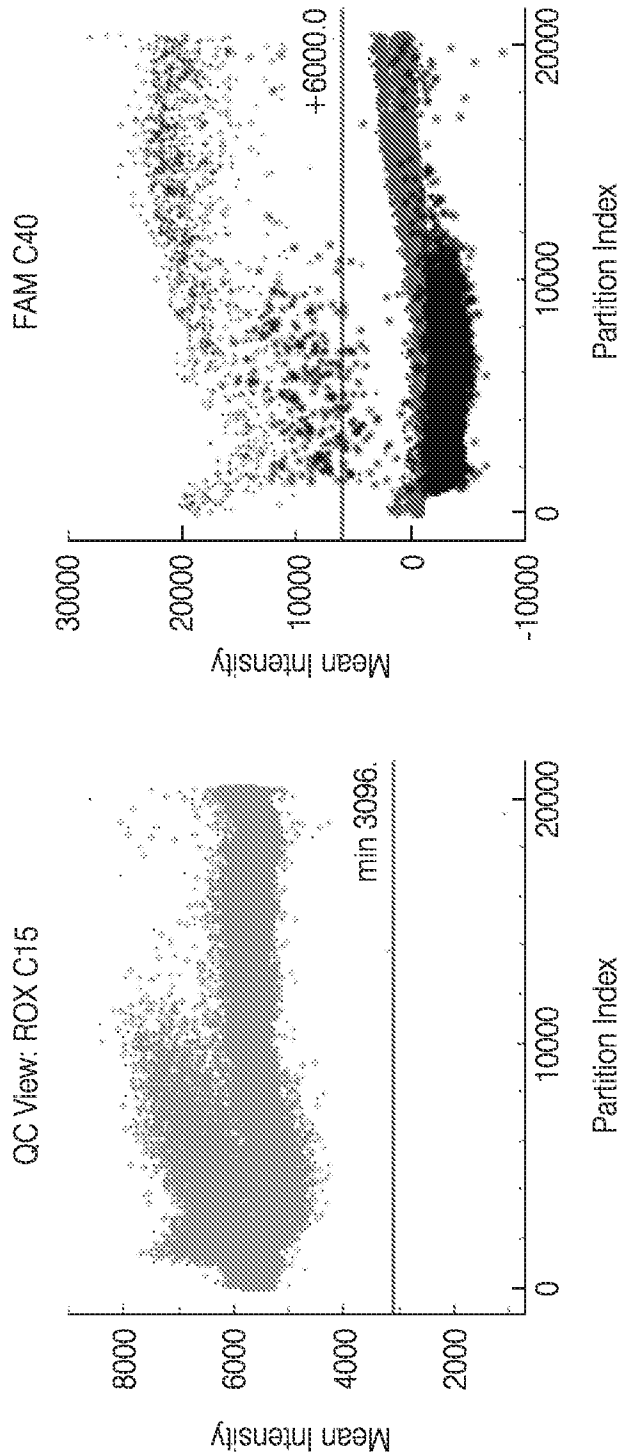
QC View: ROX C15
min 3096.
Mean Intensity
8000
6000
4000
2000
0
10000
20000
Partition Index
FAM C40
+6000.0
Mean Intensity
30000
20000
10000
0
-10000
0
10000
20000
Partition Index
QC-afflicted Data

FIG. 9 (Cont.)

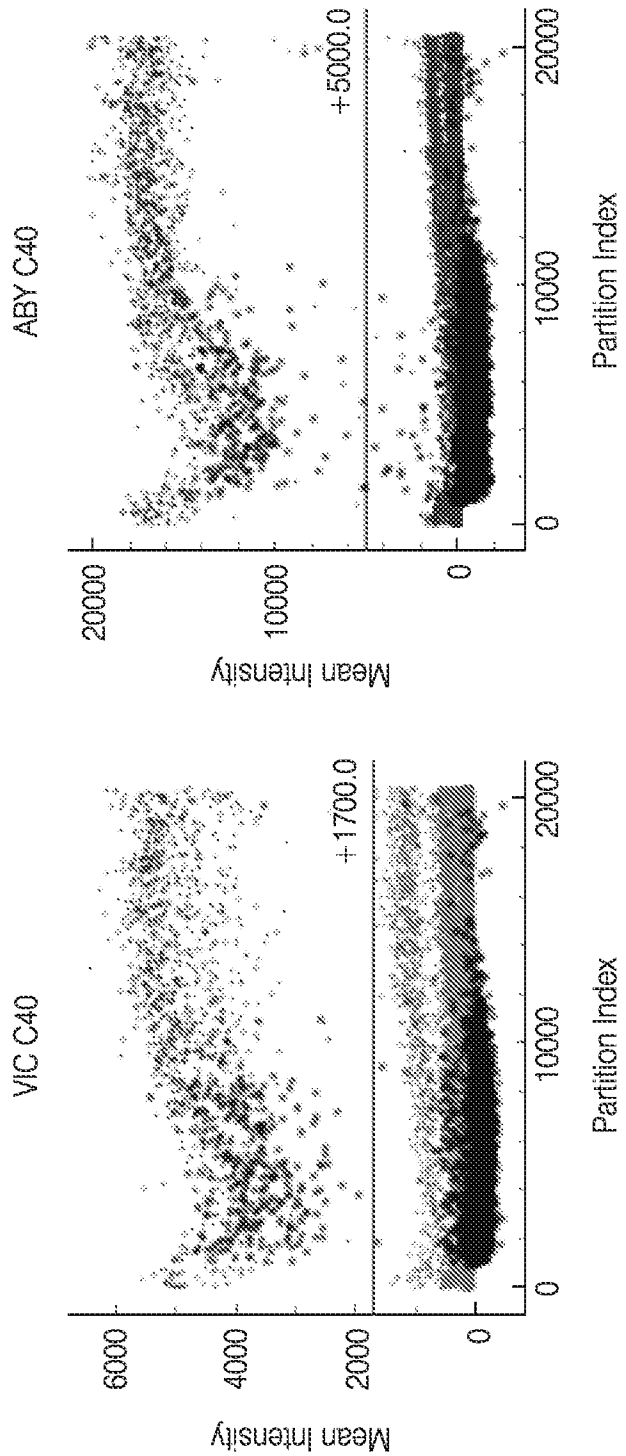
ABY C40
+5000.0
Mean Intensity
20000
10000
0
0
10000
20000
Partition Index
VIC C40
+1700.0
Mean Intensity
6000
4000
2000
0
0
10000
20000
Partition Index
QC-afflicted Data

FIG. 9 (Cont.)

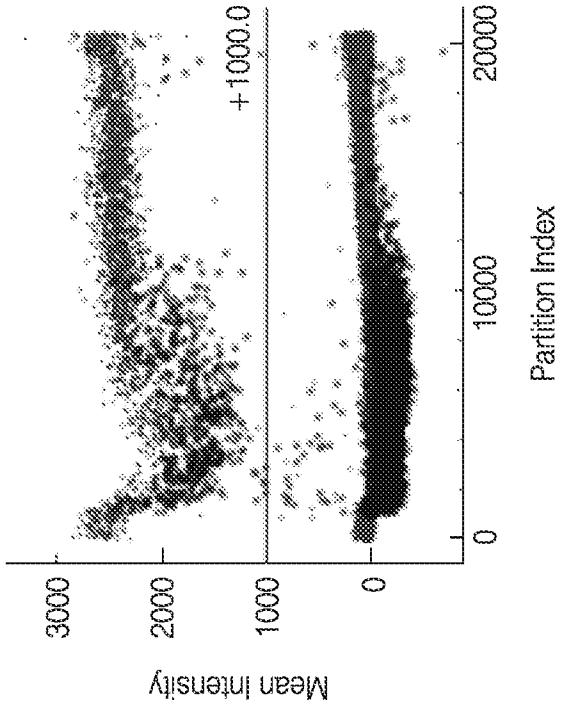
JUN C40
+1000.0
Mean Intensity
3000
2000
1000
0
0
10000
20000
Partition Index
QC-afflicted Data

FIG. 9 (Cont.)

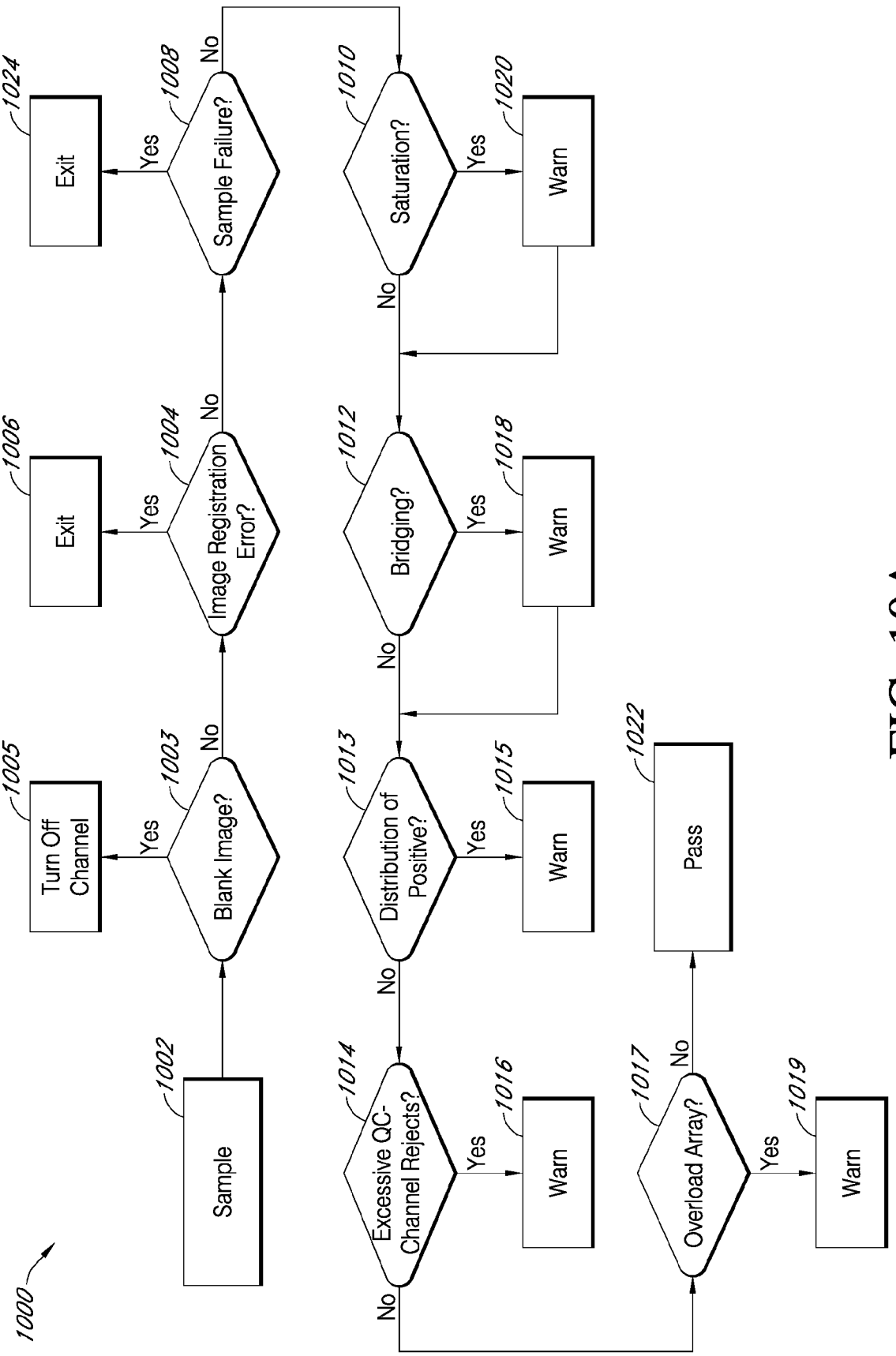
1000
Sample
1002
Blank Image?
1003
Yes
No
Turn Off Channel
1005
Image Registration Error?
1004
Yes
No
Exit
1006
Sample Failure?
1008
Yes
No
Exit
1024
Distribution of Positive?
1013
Yes
No
Warn
1015
Bridging?
1012
Yes
No
Warn
1018
Saturation?
1010
Yes
No
Warn
1020
Excessive QC-Channel Rejects?
1014
Yes
No
Warn
1016
Overload Array?
1017
Yes
No
Warn
1019
Pass
1022

FIG. 10A

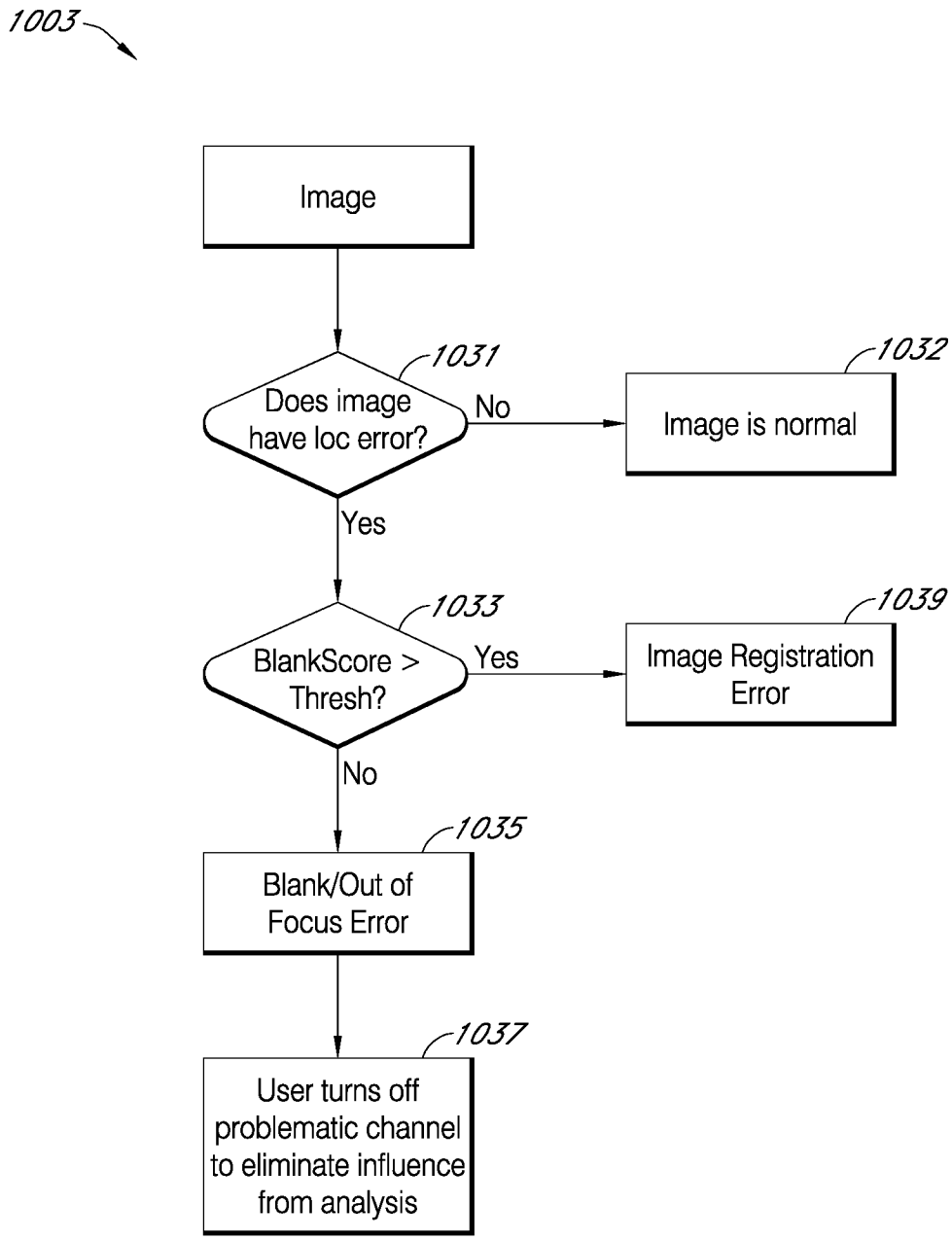
1003
Image
1031
Does image have loc error?
No
1032
Image is normal
Yes
1033
BlankScore > Thresh?
Yes
1039
Image Registration Error
No
1035
Blank/Out of Focus Error
1037
User turns off problematic channel to eliminate influence from analysis

FIG. 10B

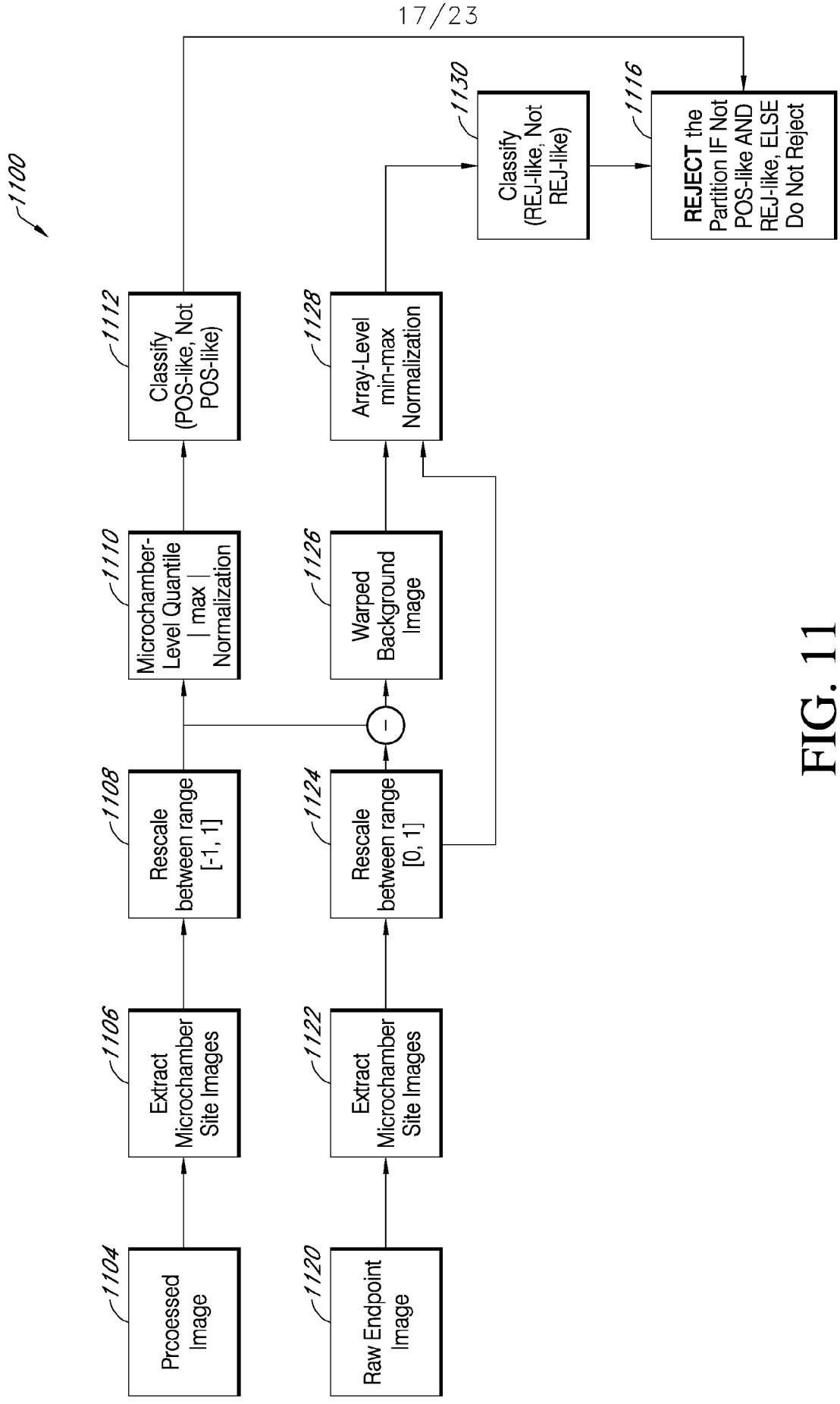
1100
1104
Prcoessed Image
1106
Extract Microchamber Site Images
1108
Rescale between range [-1, 1]
1110
Microchamber-Level Quantile | max | Normalization
1112
Classify (POS-like, Not POS-like)
1120
Raw Endpoint Image
1122
Extract Microchamber Site Images
1124
Rescale between range [0, 1]
1126
Warped Background Image
1128
Array-Level min-max Normalization
1130
Classify (REJ-like, Not REJ-like)
1116
REJECT the Partition IF Not POS-like AND REJ-like, ELSE Do Not Reject

FIG. 11

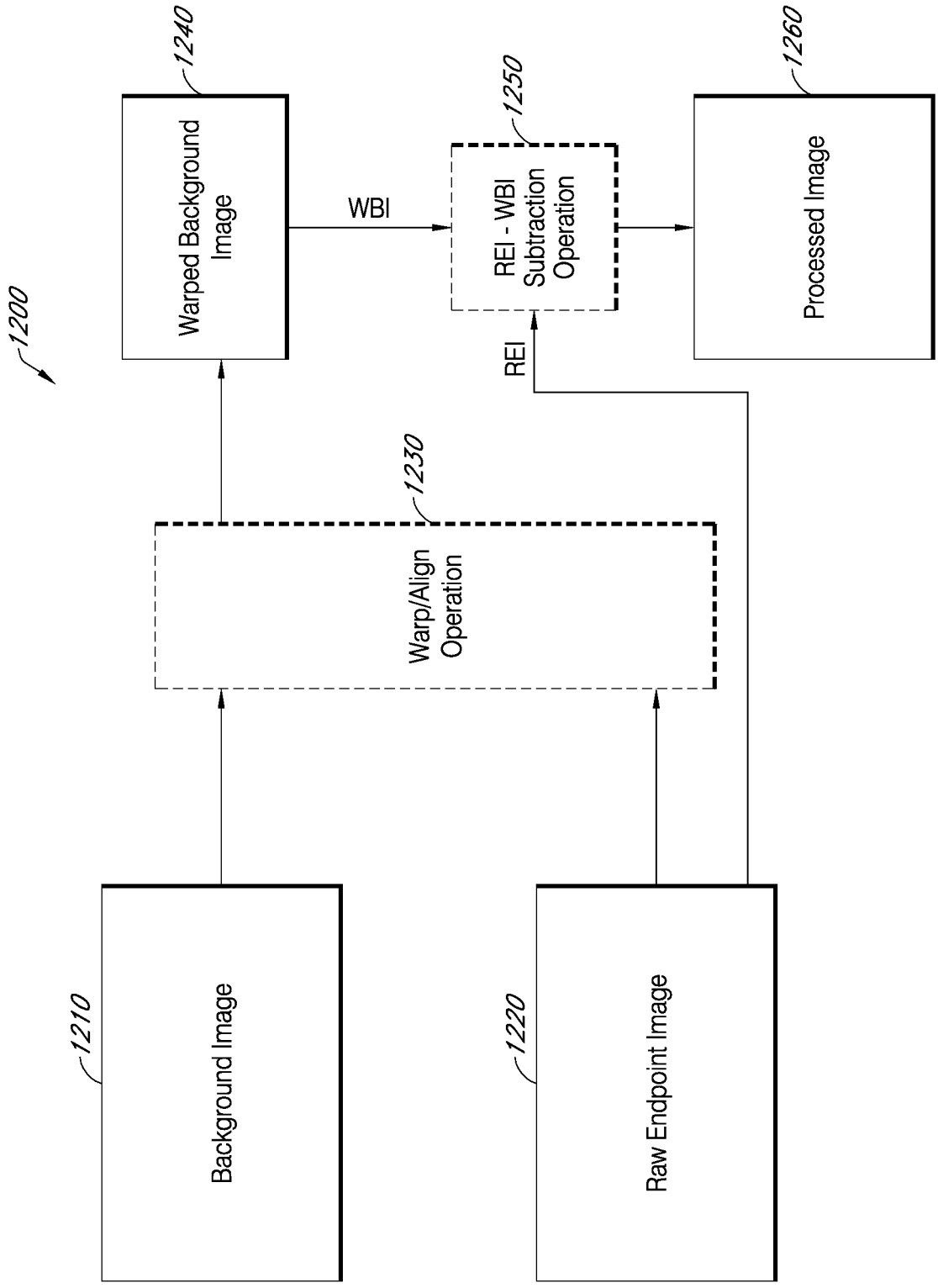
1200
1210
Background Image
1220
Raw Endpoint Image
1230
Warp/Align Operation
1240
Warped Background Image
WBI
1250
REI - WBI Subtraction Operation
REI
1260
Processed Image

FIG. 12

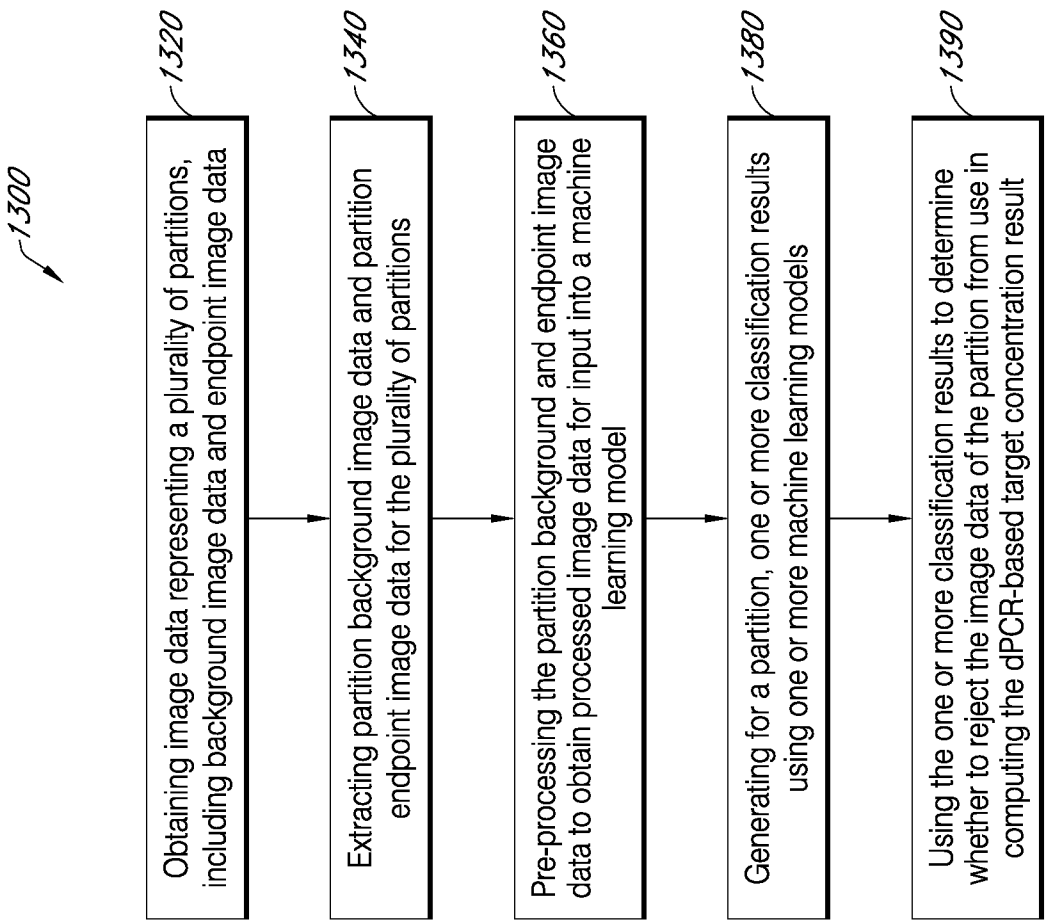
1300
1320
Obtaining image data representing a plurality of partitions, including background image data and endpoint image data
1340
Extracting partition background image data and partition endpoint image data for the plurality of partitions
1360
Pre-processing the partition background and endpoint image data to obtain processed image data for input into a machine learning model
1380
Generating for a partition, one or more classification results using one or more machine learning models
1390
Using the one or more classification results to determine whether to reject the image data of the partition from use in computing the dPCR-based target concentration result

FIG. 13

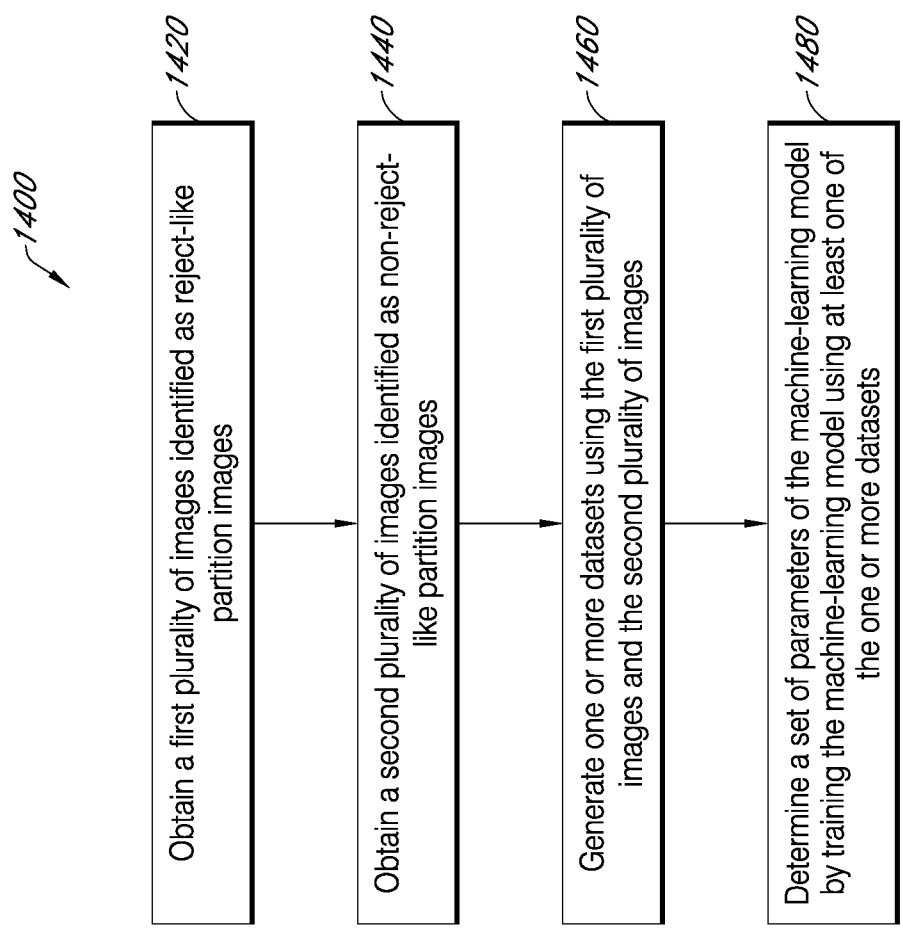
1400
1420
Obtain a first plurality of images identified as reject-like partition images
1440
Obtain a second plurality of images identified as non-reject-like partition images
1460
Generate one or more datasets using the first plurality of images and the second plurality of images
1480
Determine a set of parameters of the machine-learning model by training the machine-learning model using at least one of the one or more datasets

FIG. 14

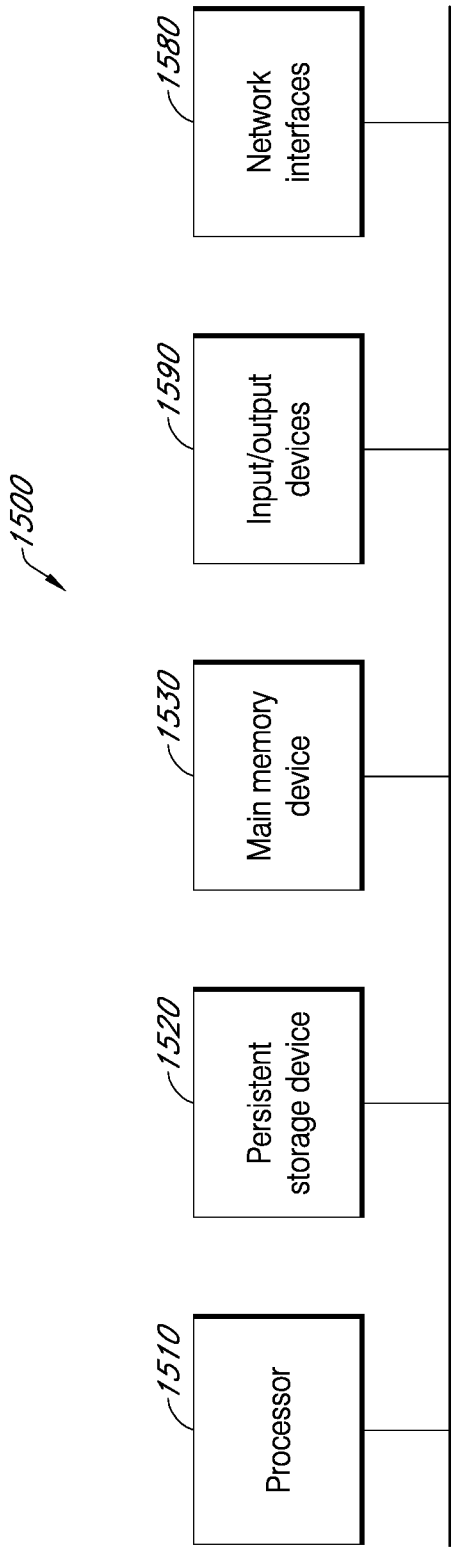
1500
1510
Processor
1520
Persistent storage device
1530
Main memory device
1590
Input/output devices
1580
Network interfaces

FIG. 15

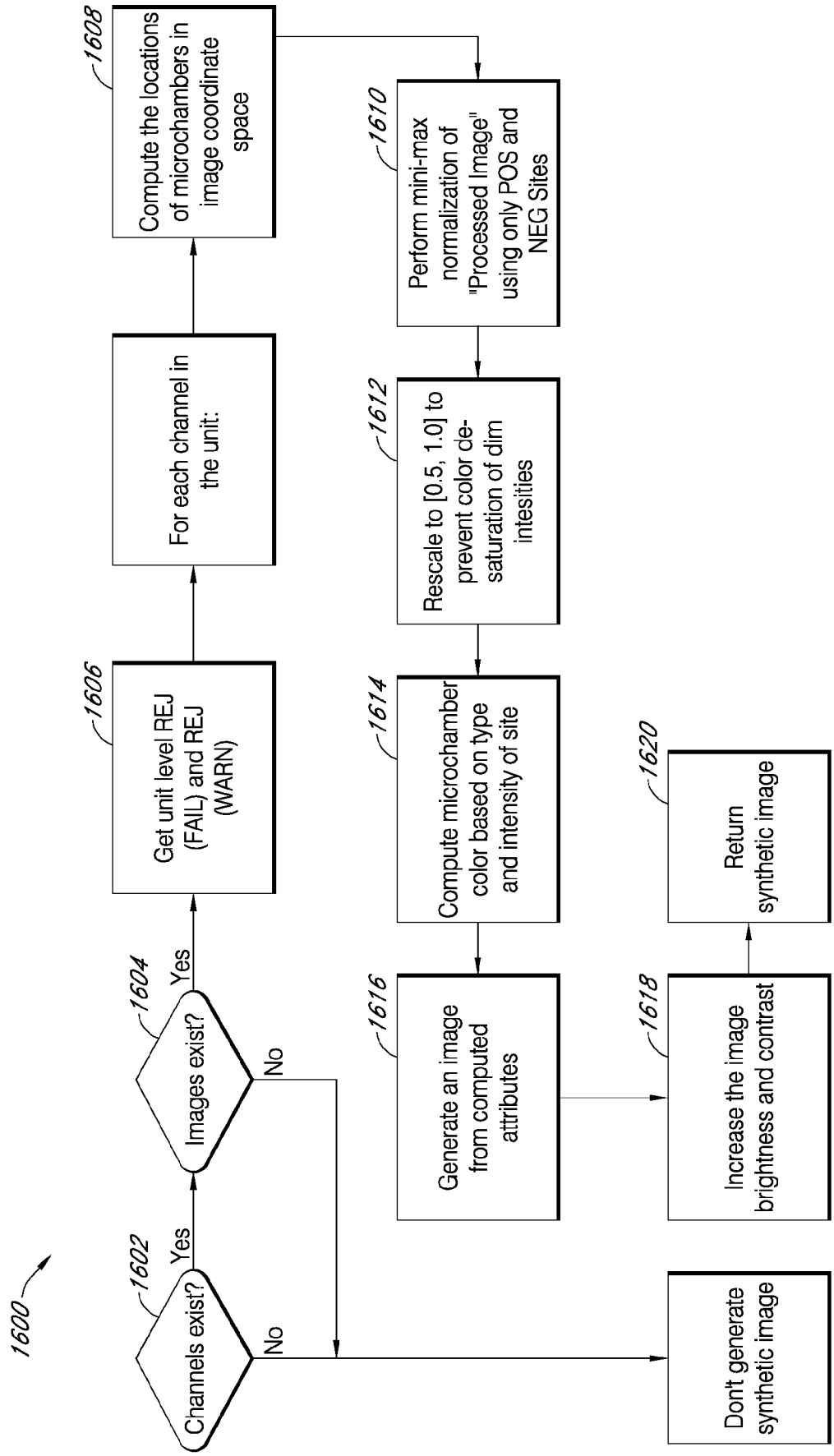
1600
1602
Channels exist?
Yes
No
1604
Images exist?
Yes
No
1606
Get unit level REJ (FAIL) and REJ (WARN)
For each channel in the unit:
1608
Compute the locations of microchambers in image coordinate space
1610
Perform mini-max normalization of "Processed Image" using only POS and NEG Sites
1612
Rescale to [0.5, 1.0] to prevent color de-saturation of dim intesities
1614
Compute microchamber color based on type and intensity of site
1616
Generate an image from computed attributes
1618
Increase the image brightness and contrast
1620
Return synthetic image
Don't generate synthetic image

FIG. 16A

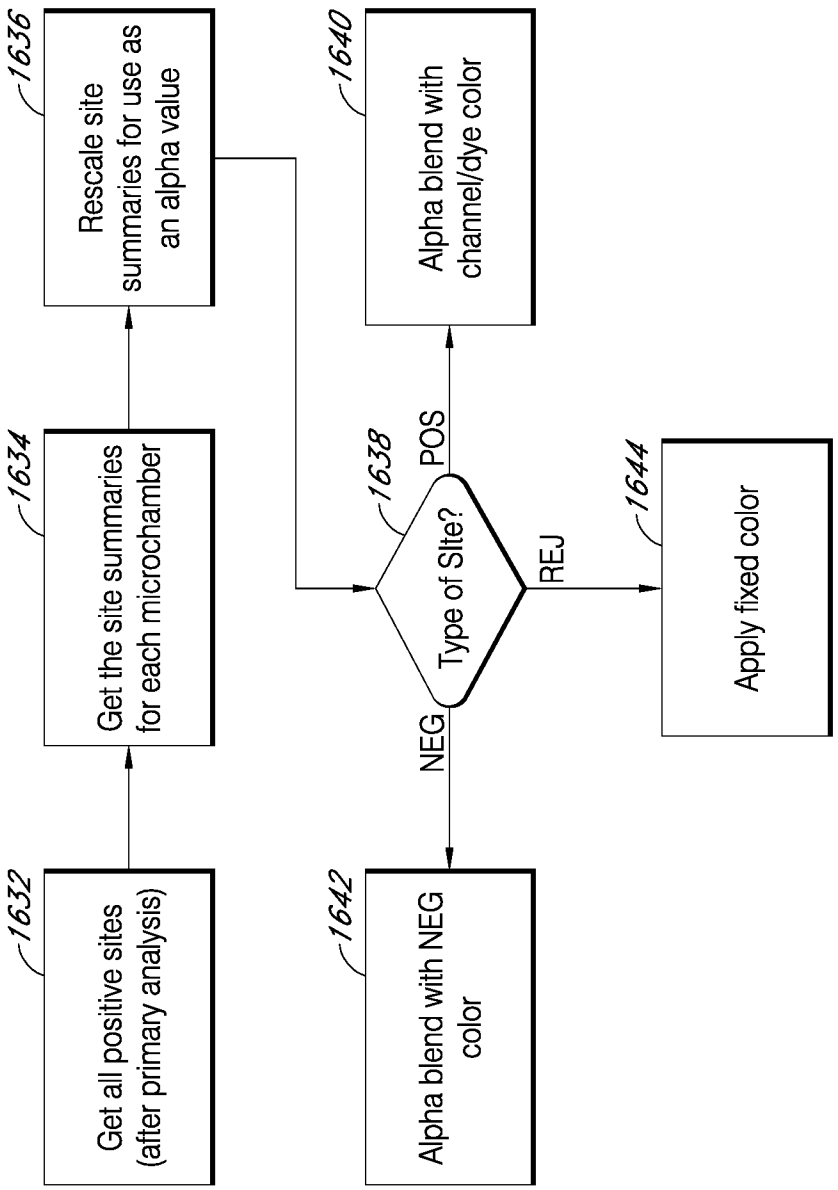
1632
Get all positive sites (after primary analysis)
1634
Get the site summaries for each microchamber
1636
Rescale site summaries for use as an alpha value
1638
Type of Site?
POS
NEG
REJ
1640
Alpha blend with channel/dye color
1642
Alpha blend with NEG color
1644
Apply fixed color

FIG. 16B

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/024967

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06V20/69
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)
G06V

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2023/043851 A1 (COMBINATI INC [US]) 23 March 2023 (2023-03-23) 71, 81, 102-113, 129-133; figures 3, 13, 14, 15, 16	1-51

☐ 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
25 June 2024	08/07/2024

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Grigorescu, Simona

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2024/024967

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2023043851 A1	23-03-2023	NONE	