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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 November 2022 (24.11.2022)

(10) International Publication Number
WO 2022/242001 A1

(51) International Patent Classification:
G06T 7/00 (2017.01)

(21) International Application Number:
PCT/CN2021/120370

(22) International Filing Date:
24 September 2021 (24.09.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
202110552582.4 20 May 2021 (20.05.2021) CN

(71) Applicant: **ZHEJIANG DAHUA TECHNOLOGY CO., LTD.** [CN/CN]; No. 1187 Binan Road, Binjiang District, Hangzhou, Zhejiang 310053 (CN).

(72) Inventors: **ZHANG, Xingming**; No. 1187 Binan Road, Binjiang District, Hangzhou, Zhejiang 310053 (CN). **ZHENG, Lei**; No. 1187 Binan Road, Binjiang District, Hangzhou, Zhejiang 310053 (CN).

(74) Agent: **METIS IP (CHENGDU) LLC**; Room 401, 4th Floor, Unit 1 of Building 1, No. 269 North Section of Hupan Road, Tianfu New District, Chengdu City, China Free Trade Zone (Sichuan), Chengdu, Sichuan 610213 (CN).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

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

(54) **Title:** SYSTEMS AND METHODS FOR SUBJECT DETECTION

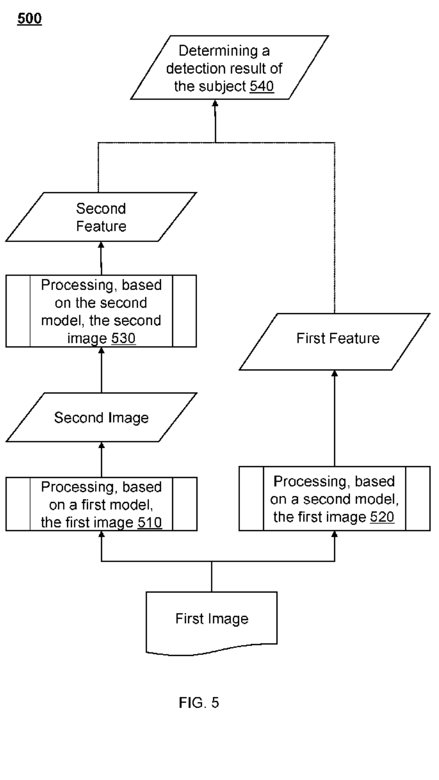

FIG. 5

(57) **Abstract:** Systems and methods for subject detection are provided. The system may obtain a second image by processing, based on a first model, a first image of a subject to be detected. The first model may be configured to process a first representation of the subject in the first image to obtain a second representation of the subject in the second image. The system may also obtain a first feature by processing, based on a second model, the first image. The second model may be configured to extract features from an image. The system may obtain a second feature by processing, based on the second model, the second image. Further, the system may determine, based on the first feature and the second feature, a detection result of the subject. The detection result may indicate whether the subject includes an anomaly.

WO 2022/242001 A1

SYSTEMS AND METHODS FOR SUBJECT DETECTION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority of Chinese Patent Application No. 202110552582.4 filed on May 20, 2021, the contents of each of which are hereby incorporated by reference.

TECHNICAL FIELD

[0002] The present disclosure generally relates to image processing, and in particular, to systems and methods for subject detection.

BACKGROUND

[0003] In recent years, as a development speed of the logistics industry has been accelerated, logistics-related companies have paid more attention to monitoring a state of a subject, so as to detect and/or track a damage of the subject in real time. The damage of the subject may occur at a beginning of a transportation or during the transportation. The damage of the subject may include deformation, wrinkles, pollution, cracks, etc., of the subject. Therefore, in a complex scenario including a subject, how to detect an occurrence of the damage of the subject in time and effectively has become a problem to be solved.

[0004] Methods and systems for subject detection are provided, so that the damage of the subject may be detected in time and accurately.

SUMMARY

[0005] An aspect of the present disclosure relates to a method for subject detection. The method may include obtaining a second image by processing, based on a first model, a first image of a subject to be detected. The first model may be configured to process a first representation of the subject in the first image to obtain a second representation of the subject in the second image. The method may also include obtaining a first feature by processing, based on a second model, the first image.

The second model may be configured to extract features from an image. The method may include obtaining a second feature by processing, based on the second model, the second image. The method may further include determining, based on the first feature and the second feature, a detection result of the subject. The detection result may indicate whether the subject includes an anomaly.

[0006] In some embodiments, the first model may be obtained according to a first training process. The first training process may include obtaining a plurality of first training samples. Each of the plurality of first training samples may include a first sample image of a first sample subject with an anomaly and a reference image corresponding to the first sample image. The reference image may include the sample subject without the anomaly. The first training process may include generating the first model by training a first preliminary model using the plurality of first training samples.

[0007] In some embodiments, the second model may be obtained according to a second training process. The second training process may include obtaining a plurality of second training samples. Each of the plurality of second training samples may include a second sample image of a second sample subject without an anomaly and a label of the second sample image indicating a type of the second sample image. The type of the second sample image may include one of an actual image acquired by an imaging device and a simulated image generated by the first model. The second training process may include generating the second model by training a second preliminary model using the plurality of second training samples.

[0008] In some embodiments, the first model and the second model may be two portions of a joint model. The joint model may be obtained according to a third training process. The third training process may include obtaining a plurality of third training samples. Each of the plurality of third training samples may include a third sample image of a third sample subject and a sample feature of a reference image of the third sample subject without anomaly corresponding to the third sample subject. The third training process may include generating the joint model by training a preliminary joint model including a preliminary first model and a preliminary second model using the plurality of third training samples. In the third training process, a simulated image of the third sample subject without anomaly output by the preliminary first model may be input in the preliminary second model, and the preliminary second model may output a feature of the simulated image of the sample subject without anomaly.

[0009] In some embodiments, the second model may output at least one of the first feature or the second feature based on an average value of a plurality of elements in a feature vector.

[0010] In some embodiments, a dimension of the feature vector outputted by the second model may be greater than 1.

[0011] In some embodiments, the first image may include a depth image.

[0012] In some embodiments, the method may further include obtaining the first image by acquiring, via an imaging device, images of the subject from a plurality of angles.

[0013] In some embodiments, the method may further include moving, via a driving assembly, the subject to acquire, via the imaging device, the images of the subject to be detected from the plurality of angles.

[0014] In some embodiments, the obtaining the first image by acquiring, via an imaging device, images of the subject to be detected from a plurality of angles may include obtaining the first image by reconstructing, based on a reconstruction algorithm, the images of the subject acquired from the plurality of angles, wherein the first image includes a three-dimensional image.

[0015] In some embodiments, the determining, based on the first feature and the second feature, a detection result of the subject may include obtaining, based on the first feature and the second feature, a similarity degree between the first feature and the second feature; and determining, based on the similarity degree, the detection result of the subject.

[0016] Another aspect of the present disclosure relates to a system for subject detection. The system may include at least one storage device and at least one processor. The at least one storage device may include a set of instructions. The at least one processor may be in communication with the at least one storage device. When executing the set of instructions, the at least one processor may be directed to perform operations. The operations may include obtaining a second image by processing, based on a first model, a first image of a subject to be detected. The first model may be configured to process a first representation of the subject in the first image to obtain a second representation of the subject in the second image. The method may also include obtaining a first feature by processing, based on a second model, the first image. The second model may be configured to extract features from an image. The method may include obtaining a second feature by processing, based on the second model, the second image. The method may further include determining, based on the first feature and the second feature, a detection result of the subject. The detection result may indicate whether the subject includes an anomaly.

[0017] In some embodiments, the first model may be obtained according to a first training process. The first training process may include obtaining a plurality of first training samples. Each of the plurality of first training samples may include a first sample image of a first sample subject with an anomaly and a reference image corresponding to the first sample image. The reference image may include the sample subject without the anomaly. The first training process may include generating the first model by training a first preliminary model using the plurality of first training samples.

[0018] In some embodiments, the second model may be obtained according to a second training process. The second training process may include obtaining a plurality of second training samples. Each of the plurality of second training samples may include a second sample image of a second sample subject without an anomaly and a label of the second sample image indicating a type of the second sample image. The type of the second sample image may include one of an actual image acquired by an imaging device and a simulated image generated by the first model. The second training process may include generating the second model by training a second preliminary model using the plurality of second training samples.

[0019] In some embodiments, the first model and the second model may be two portions of a joint model. The joint model may be obtained according to a third training process. The third training process may include obtaining a plurality of third training samples. Each of the plurality of third training samples may include a third sample image of a third sample subject and a sample feature of a reference image of the third sample subject without anomaly corresponding to the third sample subject. The third training process may include generating the joint model by training a preliminary joint model including a preliminary first model and a preliminary second model using the plurality of third training samples. In the third training process, a simulated image of the third sample subject without anomaly output by the preliminary first model may be input in the preliminary second model, and the preliminary second model may output a feature of the simulated image of the sample subject without anomaly.

[0020] In some embodiments, the second model may output at least one of the first feature or the second feature based on an average value of a plurality of elements in a feature vector.

[0021] In some embodiments, a dimension of the feature vector outputted by the second model may be greater than 1.

[0022] In some embodiments, the first image may include a depth image.

[0023] In some embodiments, the operations may further include obtaining the first image by acquiring, via an imaging device, images of the subject from a plurality of angles.

[0024] In some embodiments, the operations may further include moving, via a driving assembly, the subject to acquire, via the imaging device, the images of the subject to be detected from the plurality of angles.

[0025] In some embodiments, the obtaining the first image by acquiring, via an imaging device, images of the subject to be detected from a plurality of angles may include obtaining the first image by reconstructing, based on a reconstruction algorithm, the images of the subject acquired from the plurality of angles, wherein the first image includes a three-dimensional image.

[0026] In some embodiments, the determining, based on the first feature and the second feature, a detection result of the subject may include obtaining, based on the first feature and the second feature, a similarity degree between the first feature and the second feature; and determining, based on the similarity degree, the detection result of the subject.

[0027] Still another aspect of the present disclosure relates to a non-transitory computer readable medium. The non-transitory computer readable medium may include executable instructions that, when executed by at least one processor, direct the at least one processor to perform a method for subject detection. The method may include obtaining a second image by processing, based on a first model, a first image of a subject to be detected. The first model may be configured to process a first representation of the subject in the first image to obtain a second representation of the subject in the second image. The method may also include obtaining a first feature by processing, based on a second model, the first image. The second model may be configured to extract features from an image. The method may include obtaining a second feature by processing, based on the second model, the second image. The method may further include determining, based on the first feature and the second feature, a detection result of the subject. The detection result may indicate whether the subject includes an anomaly.

[0028] Additional features will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon examination of the following and the accompanying drawings or may be learned by production or operation of the examples. The features of the present disclosure may be realized and attained by practice or use of various aspects of the methodologies, instrumentalities, and combinations set forth in the detailed examples discussed below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] The present disclosure is further described in terms of exemplary embodiments. These exemplary embodiments are described in detail with reference to the drawings. These embodiments are non-limiting exemplary embodiments, in which like reference numerals represent similar structures throughout the several views of the drawings, and wherein:

[0030] FIG. 1 is a schematic diagram illustrating an exemplary detection system according to some embodiments of the present disclosure;

[0031] FIG. 2A is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary computing device 200 according to some embodiments of the present disclosure;

[0032] FIG. 2B is a schematic diagram illustrating an exemplary structure of a computer-readable storage medium according to some embodiments of the present disclosure;

[0033] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary mobile device according to some embodiments of the present disclosure;

[0034] FIG. 4 is a block diagram illustrating an exemplary processing device according to some embodiments of the present disclosure;

[0035] FIG. 5 is a flowchart illustrating an exemplary process for subject detection according to some embodiments of the present disclosure;

[0036] FIG. 6 is a flowchart illustrating an exemplary process for generating one or more models according to some embodiments of the present disclosure; and

[0037] FIG. 7 is a flowchart illustrating an exemplary process for subject detection according to some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0038] In the following detailed description, numerous specific details are set forth by way of examples in order to provide a thorough understanding of the relevant disclosure. However, it should be apparent to those skilled in the art that the present disclosure may be practiced without such details. In other instances, well-known methods, procedures, systems, components, and/or circuitry have been described at a relatively high-level, without detail, in order to avoid unnecessarily obscuring aspects of the present disclosure. Various modifications to the disclosed embodiments will be readily apparent to those skilled in the art, and the general principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the present disclosure. Thus, the present disclosure is not limited to the embodiments shown, but to be accorded the widest scope consistent with the claims.

[0039] It will be understood that the terms “system,” “engine,” “unit,” “module,” and/or “block” used herein are one method to distinguish different components, elements, parts, sections, or assemblies of different levels in ascending order. However, the terms may be displaced by other expressions if they may achieve the same purpose.

[0040] Generally, the words “module,” “unit,” or “block” used herein, refer to logic embodied in hardware or firmware, or to a collection of software instructions. A module, a unit, or a block described herein may be implemented as software and/or hardware and may be stored in any type of non-transitory computer-readable medium or other storage devices. In some embodiments, a software module/unit/block may be compiled and linked into an executable program. It will be appreciated that software modules can be callable from other modules/units/blocks or from themselves, and/or may be invoked in response to detected events or interrupts. Software modules/units/blocks configured for execution on computing devices (e.g., processor 210 illustrated in FIG. 2A) may be provided on a computer-readable medium, such as a compact disc, a digital video disc, a flash drive, a magnetic disc, or any other tangible medium, or as a digital download (and can be originally stored in a compressed or installable format that needs installation, decompression, or decryption prior to execution). Such software code may be stored, partially or fully, on a storage device of the executing computing device, for execution by the computing device. Software instructions may be embedded in firmware, such as an EPROM. It will be further appreciated that hardware modules (or units or blocks) may be included in connected logic components, such as gates and flip-flops, and/or can be included in programmable units, such as programmable gate arrays or processors. The modules (or units or blocks) or computing device functionality described herein may be implemented as software modules (or units or blocks), but may be represented in hardware or firmware. In general, the modules (or units or blocks) described herein refer to logical modules (or units or blocks) that may be combined with other modules (or units or blocks) or divided into sub-modules (or sub-units or sub-blocks) despite their physical organization or storage.

[0041] It will be understood that when a unit, an engine, a module, or a block is referred to as being “on,” “connected to,” or “coupled to” another unit, engine, module, or block, it may be directly on, connected or coupled to, or communicate with the other unit, engine, module, or block, or an intervening unit, engine, module, or block may be present, unless the context clearly indicates otherwise. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0042] The terminology used herein is for the purposes of describing particular examples and embodiments only and is not intended to be limiting. As used herein, the singular forms “a,” “an,” and “the” may be intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “include” and/or “comprise,” when used in this disclosure, specify the presence of integers, devices, behaviors, stated features, steps, elements, operations, and/or components, but do not exclude the presence or addition of one or more other integers, devices, behaviors, features, steps, elements, operations, components, and/or groups thereof.

[0043] In addition, it should be understood that in the description of the present disclosure, the terms "first," "second," or the like, are only used for the purpose of differentiation, and cannot be interpreted as indicating or implying relative importance, nor can be understood as indicating or implying the order.

[0044] The flowcharts used in the present disclosure illustrate operations that systems implement according to some embodiments of the present disclosure. It is to be expressly understood, the operations of the flowcharts may be implemented not in order. Conversely, the operations may be implemented in an inverted order, or simultaneously. Moreover, one or more other operations may be added to the flowcharts. One or more operations may be removed from the flowcharts.

[0045] The present disclosure relates to systems and methods for subject detection. The systems may include obtaining a second image by processing, based on a first model, a first image of a subject to be detected. The first model may be configured to process a first representation of the subject in the first image to obtain a second representation of the subject in the second image. The systems may include obtaining a first feature by processing, based on a second model, the first image. The second model may be configured to extract features from an image. The systems may also include obtaining a second feature by processing, based on the second model, the second image. The systems may further include determining, based on the first feature and the second feature, a detection result of the subject. The detection result may indicate whether the subject includes an anomaly. The systems for subject detection may be simple, which may quickly and accurately determine whether the subject includes an anomaly. The anomaly may include a damage of the subject, position change of the subject, etc. In addition, the systems may include determining a degree of damage to the subject to be detected based on a similarity degree between the first feature and the second feature, which may facilitate a user to perform different treatments on different degrees of damage.

[0046] In addition, the first model and the second model may be two portions of a joint model. The joint model may be obtained according to a jointly training process. By jointly training the joint model, the trained first model and the trained second model may include better performance. That is, the trained joint model may generate a simulated image that is realistic enough, and extract image features more quickly and accurately, thereby improving the accuracy of the detection result of the subject.

[0047] FIG. 1 is a schematic diagram illustrating an exemplary detection system 100 according to some embodiments of the present disclosure. As shown, the detection system 100 may include a server 110, a network 120, an imaging device 130, a user device 140, and a storage device 150.

[0048] The detection system 100 may be applied to a variety of application scenarios for subject detection. For example, the detection system 100 may determine a detection result of a courier package based on image data of one or more surfaces of the courier package. As another example, the detection system 100 may determine a detection result of a product package based on image data of one or more surfaces of the product package. In some embodiments, the detection system 100 may determine whether a subject (e.g., a subject 160) includes an anomaly according to some embodiments of the present disclosure.

[0049] The server 110 may be configured to manage resources and processing data and/or information from at least one component or external data source of the detection system 100. In some embodiments, the server 110 may be a single server or a server group. The server group may be centralized or distributed (e.g., the server 110 may be a distributed system). In some embodiments, the server 110 may be local or remote. For example, the server 110 may access information and/or data stored in the imaging device 130, the user device 140, and/or the storage device 150 via the network 120. As another example, the server 110 may be directly connected to the imaging device 130, the user device 140, and/or the storage device 150 to access stored information and/or data. In some embodiments, the server 110 may be implemented on a cloud platform. Merely by way of example, the cloud platform may include a private cloud, a public cloud, a hybrid cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, or the like, or any combination thereof. In some embodiments, the server 110 may be implemented on a computing device 200 including one or more components illustrated in FIG. 2A of the present disclosure.

[0050] In some embodiments, the server 110 may include a processing device 112. The processing device 112 may process information and/or data relating to monitoring to perform one or more functions described in the present disclosure. For example, the processing device 112 may obtain a second image by processing, based on a first model, a first image of a subject 160 to be detected. The first model may be configured to process a first representation of the subject 160 in the first image to obtain a second representation of the subject 160 in the second image. As another example, the processing device 112 may obtain a first feature by processing, based on a second model, the first image. The second model may be configured to extract features from an image. As still another example, the processing device 112 may obtain a second feature by processing, based on the second model, the second image. As still another example, the processing device 112 may determine, based on the first feature and the second feature, a detection result of the subject 160. The detection result may indicate whether the subject 160 includes an anomaly. In some embodiments, the processing device 112 may include one or more processing devices (e.g., single-core processing device(s) or multi-core processor(s)).

[0051] In some embodiment, the server 110 may be unnecessary and all or part of the functions of the server 110 may be implemented by other components (e.g., the imaging device 130, the user device 140) of the detection system 100. For example, the processing device 112 may be integrated into the imaging device 130 and the functions (e.g., determining the detection result of the subject 160) of the processing device 112 may be implemented by the imaging device 130.

[0052] The network 120 may facilitate the exchange of information and/or data for the detection system 100. In some embodiments, one or more components (e.g., the server 110, the imaging device 130, the user device 140, the storage device 150) of the detection system 100 may transmit information and/or data to other component(s) of the detection system 100 via the network 120. For example, the server 110 may obtain images (e.g., the first image and/or the second image) from the imaging device 130 via the network 120. As another example, the server 110 may transmit the detection result of the subject 160 to the user device 140 via the network 120. In some embodiments, the network 120 may be any type of wired or wireless network, or a combination thereof.

[0053] In some embodiments, the network 120 may be configured to connect to each component of the detection system 100 and/or connect the detection system 100 and an external resource portion. The network 120 may be configured to implement communication between components of the detection system 100 and/or between each component of the detection system 100 and an external resource portion. In some embodiments, the network 120 may include a wired network, a wireless network, or a combination thereof. For example, the network 120 may include a cable network, a fiber network, a telecommunication network, Internet, a local area network (LAN), a wide area network (WAN), a wireless local area network (WLAN), a metropolitan area network (MAN), a public switched telephone network (PSTN), a Bluetooth network, Zigbee, near field communication (NFC), an intra-device bus, an intra-device line, a cable connection, or the like, or any combination thereof. In some embodiments, the network 120 may include a point-to-point topology structure, a shared topology structure, a centralized topology structure, or the like, or any combination thereof. In some embodiments, the network 120 may include one or more network access points. For example, the network 120 may include a wired or wireless network access point, such as base station and/or network exchange points 120-1, 120-2, etc. One or more components of the detection system 100 may be connected to the network 120 to exchange data and/or information through these network access points.

[0054] The imaging device 130 may be configured to acquire image data. In some embodiments, the imaging device 130 may be configured to acquire image data of a subject to be detected. For example, the imaging device 130 may be configured to acquire a first image of the subject 160 to be detected. As another example, the imaging device 130 may be configured to acquire one or more images of one or more surfaces of the subject 160. In some embodiments, the imaging device 130 may include one or more image acquisition devices and one or more driving assemblies (not shown in FIG. 1). In some embodiments, the image data of the subject 160 may be acquired via the one or more image acquisition devices. The image acquisition device may include a camera, a video recorder, an image sensor, etc. The camera may include a gun camera, a dome camera, an integrated camera, a monocular camera, a binocular camera, a multi-view camera, a visible light camera, a thermal imaging camera, or the like, or any combination thereof. The video recorder may include a PC digital video recorder (DVR), an embedded DVR, a visible light DVR, a thermal imaging DVR, or the like, or any combination thereof. The image sensor may include a charge coupled device (CCD) image sensor, a complementary metal oxide semiconductor (CMOS) image sensor, or the like, or any combination thereof. In some embodiments, the one or more driving assemblies may be configured to cause at least one portion of the imaging device 130 to move. For example, the subject 160 to be detected may be moved by the one or more driving assemblies to acquire the first image of the subject 160 from a plurality of angles. In some embodiments, the driving assembly may include at least one driving member and at least one transmission member. Exemplary driving members may include a pneumatic driving member, a hydraulic driving member, an electric driving member, a mechanical driving member, or the like, or any combination thereof. Exemplary transmission members may include a turntable, a belt, a chain, a shaft, a bearing, or the like, or any combination thereof. In some embodiments, the imaging device 130 may transmit the acquired image data (e.g., the first image) to one or more components (e.g., the server 110, the user device 140, the storage device 150) of the detection system 100 via the network 120.

[0055] The subject 160 to be detected may refer to a subject for detecting whether the subject includes an anomaly. In some embodiments, the anomaly may include deformation, contamination, rupture, wrinkles, depressions, or the like, or any combination thereof. In some embodiments, the subject 160 to be detected may include a courier package, a product package, or the like, or any combination thereof.

[0056] The user device 140 may be configured to receive information and/or data from the server 110, the imaging device 130, and/or the storage device 150, via the network 120. For example, the user device 140 may receive the images (e.g., the first image, the second image, etc.) from the imaging device 130. As another example, the user device 140 may receive the first feature, the second feature, and/or the detection result from the server 110. In some embodiments, the user device 140 may process information and/or data received from the server 110, the imaging device 130, and/or the storage device 150, via the network 120. In some embodiments, the user device 140 may provide a user interface via which a user may view information and/or input data and/or instructions to the detection system 100. For example, the user may view the images (e.g., the first image, the second image, and/or the detection result) via the user interface. As another example, the user may input an instruction associated with the subject detection via the user interface. In some embodiments, the user device 140 may include a mobile phone 140-1, a computer 140-2, a wearable device 140-3, or the like, or any combination thereof. In some embodiments, the user device 140 may include a display that can display information in a human-readable form, such as text, image, audio, video, graph, animation, or the like, or any combination thereof. The display of the user device 140 may include a cathode ray tube (CRT) display, a liquid crystal display

(LCD), a light-emitting diode (LED) display, a plasma display panel (PDP), a three-dimensional (3D) display, or the like, or a combination thereof.

[0057] The storage device 150 may be configured to store data and/or instructions. The data and/or instructions may be obtained from, for example, the server 110, the imaging device 130, and/or any other component of the detection system 100. In some embodiments, the storage device 150 may store data and/or instructions that the server 110 may execute or use to perform exemplary methods described in the present disclosure. In some embodiments, the storage device 150 may include a mass storage, a removable storage, a volatile read-and-write memory, a read-only memory (ROM), or the like, or any combination thereof. In some embodiments, the storage device 150 may be implemented on a cloud platform. Merely by way of example, the cloud platform may include a private cloud, a public cloud, a hybrid cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, or the like, or any combination thereof.

[0058] In some embodiments, the storage device 150 may be connected to the network 120 to communicate with one or more components (e.g., the server 110, the imaging device 130, the user device 140) of the detection system 100. One or more components of the detection system 100 may access the data or instructions stored in the storage device 150 via the network 120. In some embodiments, the storage device 150 may be directly connected to or communicate with one or more components (e.g., the server 110, the imaging device 130, the user device 140) of the detection system 100. In some embodiments, the storage device 150 may be part of other components of the detection system 100, such as the server 110, the imaging device 130, or the user device 140.

[0059] It should be noted that the above description is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure.

[0060] FIG. 2A is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary computing device 200 according to some embodiments of the present disclosure. In some embodiments, the server 110 may be implemented on the computing device 200. For example, the processing device 112 may be implemented on the computing device 200 and configured to perform methods as disclosed in this disclosure. It should be noted that the description of the computing device 200 in FIG. 2A is intended to be illustrative, and not to limit the scope of the present disclosure. For example, the detection device 200 may be any device with an image processing function, such as a mobile phone, a desktop computer, a tablet computer, etc., which is not limited herein.

[0061] As illustrated in FIG. 2A, the computing device 200 may include at least one processor 210, at least one storage device 220, a communication circuit 230, or the like, or any combination thereof.

[0062] The communication circuit 230 may be configured to connect other components in the computing device 200 (e.g., the processors 210, the storage device 220, etc.). The communication circuit 230 may represent one or more bus structures. Exemplary bus structures may include a memory bus, a memory controller, a peripheral bus, a graphical acceleration port, a processor, or a local bus that uses any of several bus structures. For example, these bus structures may include an industry standards architecture (ISA) bus, a microchannel architecture (MAC) bus, an enhanced ISA bus, a video electronics standards association (VESA) local bus, a peripheral component interconnection (PCI) bus, or the like, or any combination thereof.

[0063] The at least one processor 210 may execute computer instructions (e.g., program codes) and perform functions of the processing device 112 in accordance with techniques described herein. The computer instructions may include, for example, routines, programs, objects, components, data structures, procedures, modules, etc., which perform particular functions described herein. For example, the at least one processor 210 may process data obtained from the server 110, the imaging device 130, the user device 140, the storage device 150, and/or any other component of the detection system 100. In some embodiments, the at least one processor 210 may include one or more hardware processors, such as a microcontroller, a microprocessor, a reduced instruction set computer (RISC), an application specific integrated circuits (ASICs), an application-specific instruction-set processor (ASIP), a central processing unit (CPU), a graphics processing unit (GPU), a physics processing unit (PPU), a microcontroller unit, a digital signal processor (DSP), a field programmable gate array (FPGA), an advanced RISC machine (ARM), a programmable logic device (PLD), any circuit or processor capable of executing one or more functions, or the like, or combinations thereof.

[0064] Merely for illustration, only one processor is described in the computing device 200. However, it should be noted that the computing device 200 in the present disclosure may also include multiple processors, thus operations and/or method steps that are performed by one processor as described in the present disclosure may also be jointly or separately performed by the multiple processors. For example, if in the present disclosure the processor of the computing device 200 executes both operation A and operation B, it should be understood that operation A and operation B may also be performed by two or more different processors jointly or separately in the computing device 200 (e.g., a first processor executes operation A and a second processor executes operation B, or the first and second processors jointly execute operations A and B).

[0065] The at least one storage device 220 may store data/information obtained from the server 110, the imaging device 130, the user device 140, the storage device 150, and/or any other component of the detection system 100. The storage device 220 may include a computer readable medium in the form of a volatile memory, such as a random access memory (RAM), a cache memory, and/or a read-only memory (ROM). In some embodiments, the at least one storage device 220 may include a program/utility including at least one set of program modules. Such a program module may include an operating system, one or more applications, other program modules, program data, etc. Each or some combination of these embodiments may include an implementation of a network environment. The program module may perform functions and/or methods described in the embodiments of the present disclosure.

[0066] The computing device 200 may communicate with one or more external devices (e.g., a keyboard, a pointing device, a display, etc.). The computing device 200 may communicate with one or more devices that enable a user to interact with the computing device 200, and/or with any device (e.g., a network card, a modem, etc.) that enables the computing device 200 to communicate with one or more other computing devices. The communication may be performed through an input/output (I/O) interface. In addition, the computing device 200 may also communicate with one or more networks (e.g., a local area network (LAN), a wide area network (WAN), and/or a public network, such as the Internet) through a network adapter. It should be noted that, although not shown in FIG. 2A, other hardware and/or software modules may be used in accordance with the computing device 200. The hardware and/or software modules may include, but not be limited to, a microcode, a device driver, a redundant processing unit, a drive array of external disks, a redundant array of independent disks (RAID) system, a tape drive, a data backup storage device, or the like, or any combination thereof. It may be considered that those skilled in the art may also be familiar with such structures, programs, or general operations of this type of computing device.

[0067] FIG. 2B is a schematic diagram illustrating an exemplary structure of a computer-readable storage medium according to some embodiments of the present disclosure. The computer-readable storage medium 250 may store a computer program 251. The computer program 251 may be executed by a processor to implement the operations in any of the methods disclosed in the present disclosure.

[0068] The computer-readable storage medium 250 may include a U disk, a mobile hard disk, a read-only memory (ROM), a random access memory (RAM), a magnetic disk, an optical disk, or the like, or any combination thereof, which may store the computer program 251. The computer-readable storage medium 250 may also include a server storing the computer program 251. In some embodiments, the computer-readable storage medium 250 may send the stored computer program 251 to other devices to execute. Alternatively, the computer-readable storage medium 250 may execute the stored computer program 251.

[0069] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary mobile device 300 according to some embodiments of the present disclosure. In some embodiments, the user device 140 may be implemented on the mobile device 300 shown in FIG. 3.

[0070] As illustrated in FIG. 3, the mobile device 300 may include a communication platform 310, a display 320, a graphic processing unit (GPU) 330, a central processing unit (CPU) 340, an I/O 350, a memory 360, and a storage 390. In some embodiments, any other suitable component, including but not limited to a system bus or a controller (not shown), may also be included in the mobile device 300.

[0071] In some embodiments, an operating system 370 (e.g., iOS™, Android™, Windows Phone™) and one or more applications (Apps) 380 may be loaded into the memory 360 from the storage 390 in order to be executed by the CPU 340. The applications 380 may include a browser or any other suitable mobile apps for receiving and rendering information relating to monitoring or other information from the processing device 112. User interactions may be achieved via the I/O 350 and provided to the processing device 112 and/or other components of the detection system 100 via the network 120.

[0072] FIG. 4 is a block diagram illustrating an exemplary processing device according to some embodiments of the present disclosure. The processing device 112 may include an obtaining module 410 and a determination module 420.

[0073] The obtaining module 410 may be configured to obtain information and/or data associated with the detection system 100. In some embodiments, the obtaining module 410 may include a processing module 411, a first feature obtaining module 412, and a second feature obtaining module 413. The processing module 411 may be configured to obtain a second image by processing, based on a first model, a first image of a subject to be detected. More descriptions regarding the obtaining of the second image may be found elsewhere in the present disclosure, for example, operation 510 in FIG. 5 and relevant descriptions thereof. The first feature obtaining module 412 may be configured to obtain a first feature by processing, based on a second model, the first image. More descriptions regarding the obtaining of the first feature may be found elsewhere in the present disclosure, for example, operation 520 in FIG. 5 and relevant descriptions thereof. The second feature obtaining module 413 may be configured to obtain a second feature by processing, based on the second model, the second image. More descriptions regarding the obtaining of the second feature may be found elsewhere in the present disclosure, for example, operation 530 in FIG. 5 and relevant descriptions thereof.

[0074] The determination module 420 may be configured to determine, based on the first feature and the second feature, a detection result of the subject. The detection result may indicate whether the subject includes an anomaly. The anomaly may include deformation, contamination, rupture, wrinkles, depressions, or the like, or any combination thereof. More descriptions regarding determining the detection result of the subject may be found elsewhere in the present disclosure, for example, operation 540 in FIG. 5 and relevant descriptions thereof.

[0075] The modules in the processing device 112 may be connected to or communicate with each other via a wired connection or a wireless connection. The wired connection may include a metal cable, an optical cable, a hybrid cable, or the like, or any combination thereof. The wireless connection may include a Local Area Network (LAN), a Wide Area Network (WAN), a Bluetooth, a ZigBee, a Near Field Communication (NFC), or the like, or any combination thereof. In some embodiments, two or more of the modules may be combined as a single module, and any one of the modules may be divided into two or more units. For example, the first feature obtaining module 412 and the second feature obtaining module 413 may be combined as a feature obtaining module. The feature obtaining module may be configured to obtain the first feature and the second feature. In some embodiments, the processing device 112 may include one or more additional modules. For example, the processing device 112 may also include a transmission module configured to transmit signals (e.g., electrical signals, electromagnetic signals) to one or more components (e.g., the imaging device 130, the user device 140) of the detection system 100. As another example, the processing device 112 may include a storage module (not shown) used to store information and/or data (e.g., the first image, the second image, the detection result, etc.).

[0076] FIG. 5 is a flowchart illustrating an exemplary process for subject detection according to some embodiments of the present disclosure. In some embodiments, process 500 may be executed by the detection system 100. For example, the process 500 may be implemented as a set of instructions (e.g., an application) stored in a storage device (e.g., the storage device 150, the storage device 220, and/or the storage 390). In some embodiments, the processing device 112 (e.g., the processor 210 of the computing device 200, the CPU 340 of the mobile device 300, and/or one or more modules illustrated in FIG. 4) may execute the set of instructions and may accordingly be directed to perform the process 500. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 500 may be accomplished with one or more additional operations not described and/or without one or more of the operations discussed. Additionally, the order of the operations of process 500 illustrated in FIG. 5 and described below is not intended to be limiting.

[0077] In 510, the processing device 112 (e.g., the obtaining module 410, the processing module 411) may obtain a second image by processing, based on a first model, a first image of a subject to be detected.

[0078] The subject to be detected may refer to a subject (e.g., the subject 160 in FIG. 1) that needs to be detected whether the subject includes an anomaly. In some embodiments, the subject to be detected may include a package. In some embodiments, the subject to be detected may also include other subjects, such as a packaging bottle, a packaging bag, a box, or the like, or any combination thereof.

[0079] In some embodiments, the first image may refer to an image obtained by shooting the subject to be detected. In some embodiments, the first image may also be obtained by scanning, reconstruction, or the like.

[0080] In some embodiments, a count of the first image may be one. In some embodiments, the processing device 112 may obtain more than one first image. For example, the count of the more than one first image may be two, three, six, or more based on a shape and/or volume of the subject to be detected. In some embodiments, the first image may include a two-dimensional image or a three-dimensional image. In some embodiments, the first image may include one surface of the subject to be detected. In some embodiments, the first image may include a plurality of surfaces of the subject to be detected, such as a portion of surfaces or all surfaces.

[0081] In some embodiments, the first image may include a grayscale image, a color image, a depth image, a binary image, or the like, or any combination thereof. For example, the first image may include a color image and a depth image at a same location by shooting the subject to be detected.

[0082] In some embodiments, the first image may be obtained by an imaging device (e.g., the imaging device 140 in FIG. 1). The imaging device may include a camera, a video recorder, an image sensor, etc. More descriptions regarding the imaging device may be found elsewhere in the present disclosure (e.g., FIG. 1 and the descriptions thereof). For example, the first image may be obtained by shooting the subject to be detected via a camera. As another example, a video of the object to be detected may be shoot via a video recorder. The first image may be obtained by intercepting a video frame from the video. In some embodiments, the three-dimensional image may be constructed based on the two-dimensional image captured by the imaging device. The three-dimensional image may be used as the first image. More descriptions regarding the constructing the three-dimensional image may be found elsewhere in the present disclosure (e.g., FIG. 5 and the descriptions thereof).

[0083] In some embodiments, the first image may be acquired by the imaging device, and the processing device 112 may obtain the first image from the imaging device. In some embodiments, the first image may be acquired by the imaging device and stored in a storage device (e.g., the storage device 150, the storage device 220, and/or the storage 390). The processing device 112 may obtain the first image from the storage device via a network (e.g., the network 120).

[0084] The first model may include a trained machine learning model that can process the first image of the subject to be detected to obtain the second image. That is, the first model may be configured to process a first representation of the subject in the first image to obtain a second representation of the subject in the second image. The first representation of the subject may indicate a current appearance of the subject acquired by the imaging device. For example, the current appearance of the subject may include an anomaly (e.g., a damage). The second representation of the subject may indicate a processed appearance (e.g., a restored appearance) of the subject without an anomaly. In some embodiments, the first image of the subject to be detected may be input into the first model, and the first model may output the second image. More descriptions regarding the first model may be found elsewhere in the present disclosure (e.g., FIG. 6 and the descriptions thereof).

[0085] The second image may be obtained after the first image of the subject to be detected is processed through the first model. The second image may correspond to the first image. If the processing device 112 obtained more than one first image, the processing device 112 may determine more than one second image each of which may correspond to one of the more than one first image. For example, a count of the more than one second image may be the same as the count of the more than one first image. As another example, the second representation of the subject in the second image may be located at a same position as the first representation of the subject in the first image.

[0086] In some embodiments, if the subject to be detected includes an anomaly, the second image may be different from the first image. That is, the processing device 112 may process the first representation of the subject in the first image to obtain the second representation of the subject in the second image via the first model. In some embodiments, if the subject to be detected includes no anomaly, the second image may be the same as or similar to the first image. That is, the first representation of the subject in the first image may be the same as or similar to the second representation of the subject in the second image.

[0087] In 520, the processing device 112 (e.g., the obtaining module 410, the first feature obtaining module 412) may obtain a first feature by processing, based on a second model, the first image.

[0088] The second model may be configured to process an image to extract one or more features from the image. In some embodiments, the first image may be input into the second model, and the second model may output the one or more first features. More descriptions regarding the second model may be found elsewhere in the present disclosure (e.g., FIG. 6 and the descriptions thereof).

[0089] The first feature may include an image feature of the subject to be detected in the first image. In some embodiments, the image feature of the subject to be detected in the first image may include a texture feature, a shape feature, a crack feature, a spatial relationship feature, or the like, or any combination thereof. Merely by way of example, the image feature of the subject to be detected in the first image may be a crack feature of the subject to be detected in the first image. In some embodiments, if the subject to be detected in the first image includes a crack, the processing device 112 may designate value “1” as a feature value of the first feature that is detected and outputted by the second model. When the subject to be detected in the first image includes no crack, the processing device 112 may designate value “0” as a feature value of the first feature that is detected and outputted by the second model.

[0090] In some embodiments, a dimension of the first feature may be greater than 1. For example, the dimension of the first feature may be 6. That is, each dimension may represent an image feature of one of six surfaces of a rectangular package.

[0091] In some embodiments, the first image may be divided into a plurality of regions. Each region may correspond to one dimension. The processing device 112 may designate value “0” or “1” as a feature value of the image feature of each region. For example, a packaging bag may include two faces. Each of the two faces may be divided into four regions. That is, the packaging bag may include eight regions. A dimension of the first feature of the packaging bag may be 8 each of which corresponds to one of the eight regions of the packaging bag, respectively. If the first feature of the packaging bag extracted from the first image are (0, 0, 0, 0, 1, 1, 0, 1), first four elements (0, 0, 0, 0) may indicate that the four regions of a front surface of the package bag include no anomaly. Last four elements (1, 1, 0, 1) may indicate that three regions in the four regions of a rear surface of the package bag include anomalies.

[0092] By setting the dimension of the first feature to be greater than one, the description of the image features of the subject to be detected in the first image by the second model may be increased, which may improve the accuracy of the description of the image features of the subject to be detected in the first image.

[0093] In some embodiments, the processing device 112 may obtain the first feature based on an average value of a plurality of elements in a first feature vector via the second model. For example, when the first feature vector output by the second model is (0, 0, 0, 1, 1, 1), the average value of the plurality of elements may be calculated to be 0.5. The processing device 112 may designate value “0.5” as the first feature.

[0094] It should be noted that since the first feature is the average value of the plurality of elements in the first feature vector, an overall case of the first feature vector may be reflected by the first feature.

[0095] In some embodiments, a value with a most repeated time of the plurality of elements in the first feature vector may be determined as the first feature. If there are a plurality of values with the most repeated time, the plurality of values may be calculated to obtain an average value. For example, if the first feature vector is represented as (0, 0, 1, 1, 1, 1), a value with a most repeated time of the plurality of elements in the first feature vector may be 1 (4 times in total). Therefore, the processing device 112 may designate value "1" as the first feature.

[0096] In 530, the processing device 112 (e.g., the obtaining module 410, the second feature obtaining module 413) may obtain a second feature by processing, based on the second model, the second image.

[0097] The second feature may include an image feature of the subject to be detected in the second image. In some embodiments, the second feature may correspond to the first feature. For example, if the dimension of the first feature is greater than 1, a dimension of the second feature may be greater than 1. It should be noted that in some embodiments, the dimension of the first feature and the dimension of the second feature may be the same.

[0098] In some embodiments, the processing device 112 may obtain the second feature based on an average value of a plurality of elements in a second feature vector via the second model. For example, when the second feature vector output by the second model is (0, 0, 0, 1, 1, 1), the average value of the plurality of elements may be calculated to be 0.5. The processing device 112 may designate value "0.5" as the second feature.

[0099] It should be noted that since the second feature is the average value of the plurality of elements in the second feature vector, an overall case of the second feature vector may be reflected by the second feature.

[0100] In some embodiments, a value with a most repeated time of the plurality of elements in the second feature vector may be determined as the second feature. If there are a plurality of values with the most repeated time, the plurality of values may be calculated to obtain an average value. For example, if the second feature vector is represented as (0, 0, 1, 1, 1, 1), a value with a most repeated time of the plurality of elements in the second feature vector may be 1 (4 times in total). Therefore, the processing device 112 may designate value “1” as the second feature. That is, the second model may obtain the first feature based on the average value of the plurality of elements in the first feature vector or the second feature based on the average value of the plurality of elements in the second feature vector.

[0101] In some embodiments, the first feature and the second feature may belong to a same type of image features. For example, the first feature and the second feature may include texture features.

[0102] In 540, the processing device 112 (e.g., the determination 420) may determine, based on the first feature and the second feature, a detection result of the subject.

[0103] The detection result may indicate whether the subject includes an anomaly. The anomaly (e.g., a damage) may include deformation, contamination, rupture, wrinkles, depressions, or the like, or any combination thereof.

[0104] In some embodiments, the processing device 112 may obtain, based on the first feature and the second feature, a similarity degree between the first feature and the second feature.

[0105] In some embodiments, the similarity degree may be determined based on a vector distance between the first feature and the second feature. In some embodiments, the vector distance between the first feature and the second feature may be negatively correlated with the similarity degree. That is, the greater the vector distance between the first feature and the second feature is, the less the similarity degree may be.

[0106] In some embodiments, the processing device 112 may determine whether the first feature and the second feature are similar by setting a distance threshold. In some embodiments, when the first feature and the second feature are similar, the similarity degree of between the first feature and the second feature may be determined as 1. When the first feature and the second feature are not similar, the similarity degree between the first feature and the second feature may be determined as 0. Merely by way of example, when the vector distance between the first feature and the second feature is less than or equal to the distance threshold, the processing device 112 may determine that the first feature and the second feature are similar. Accordingly, the processing device 112 may determine the similarity degree between the first feature and the second feature as 1. When the vector distance between the first feature and the second feature is greater than the distance threshold, the processing device 112 may determine that the first feature and the second feature are not similar. Accordingly, the processing device 112 may determine the similarity degree between the first feature and the second feature as 0. For example, when the vector distance between the first feature and the second feature is 0.8 and the distance threshold is 0.5, the vector distance between the first feature and the second feature may be greater than the distance threshold. The processing device 112 may determine that the first feature and the second feature are not similar. Accordingly, the processing device 112 may determine the similarity degree between the first feature and the second feature as 0.

[0107] In some embodiments, the processing device 112 may represent the similarity degree between the first feature and the second feature using numbers from 0 to 1. For example, elements in the image feature may be set to be 0 or 1. The first feature and the second feature may be obtained. Accordingly, the similarity degree between the first feature and the second feature may be determined based on the vector distance between the first feature and the second feature.

[0108] In some embodiments, the vector distance between the first feature and the second feature may be determined according to Equation (1):

$$d(A,B) = \sqrt{\sum(a_i - b_i)^2}, \qquad (1)$$

where, d may indicate the vector distance between the first feature A and the second feature B; a_i may indicate an element value of the first feature A in an i-th dimension; and b_i may indicate an element value of the second feature B in an i-th dimension.

[0109] In some embodiments, the similarity degree between the first feature and the second feature may be calculated according to Equation (2):

$$S = \frac{d - d_{min}}{d_{max} - d_{min}}, \qquad (2)$$

where, S may indicate the similarity degree between the first feature and the second feature; d_{min} may indicate a minimum value of the vector distance between the first feature and the second feature; and d_{max} may indicate a maximum value of the vector distance between the first feature and the second feature.

[0110] Merely by way of example, the first feature may be denoted as a first feature vector of (0, 0, 1, 1), and the second feature may be denoted as a second feature vector of (0, 0, 1, 0). The vector distance between the first feature and the second feature may be determined to be 1 based on Equation (1). Since all elements in the image features (i.e., the first feature and the second feature) are 0 or 1, the maximum value of the vector distance between the first feature and the second feature may be a vector distance between (0, 0, 0, 0) and (1, 1, 1, 1). The minimum value of the vector distance between the first feature and the second feature may be a vector distance between two identical feature vectors (e.g., between (0, 0, 0, 0) and (0, 0, 0, 0) or between and (1, 1, 1, 1) and (1, 1, 1, 1). The maximum value of the vector distance between the first feature and the second feature may be determined to be 2 based on Equation (1). The minimum value of the vector distance between the first feature and the second feature may be determined to be 0. The similarity degree between the first feature and the second feature may be calculated to be 0.5 based on the vector distance, the maximum value, and the minimum value between the first feature and the second feature, and Equation (2).

[0111] In some embodiments, the similarity degree between the first feature and the second feature may include a corresponding relationship with the detection result of the subject to be detected. Based on the above description, more similar between the first image (i.e., the subject to be detected) and the second image (i.e., the first image processed by the first model), the lower probability that the subject to be detected includes an anomaly.

[0112] In some embodiments, the processing device 112 may determine, based on the similarity degree, the detection result of the subject. In some embodiments, the processing device 112 may determine whether the subject to be detected includes an anomaly by setting a similarity threshold. When the similarity degree between the first feature and the second feature is less than the similarity threshold, the processing device 112 may determine that the subject to be detected includes an anomaly. When the similarity degree between the first feature and the second feature is greater than or equal to the similarity threshold, the processing device 112 may determine that the subject to be detected includes no anomaly. For example, if the similarity degree of between the first feature and the second feature is 0.2, and the similarity threshold is 0.5, the processing device 112 may determine that the subject to be detected includes an anomaly.

[0113] In some embodiments, the detection result may further include a degree of damage of the subject to be detected. For example, the detection result may reflect that there is a slight damage to the subject to be detected. In some embodiments, the similarity degree between the first feature and the second feature may have a corresponding relationship with the degree of damage of the subject to be detected.

[0114] Merely by way of example, the processing device 112 may determine the degree of damage to the subject to be detected by setting a plurality of similarity thresholds between the first feature and the second feature. When the similarity degree between the first feature and the second feature is greater than or equal to a first similarity threshold, the processing device 112 may determine that there is no damage to the subject to be detected. When the similarity degree between the first feature and the second feature is less than the first similarity threshold and greater than or equal to a second similarity threshold, the processing device 112 may determine that there is a slight damage to the subject to be detected. When the similarity degree between the first feature and the second feature is smaller than the second similarity threshold, the processing device 112 may determine that the subject to be detected is severely damaged. The first threshold may be greater than the second threshold. For example, if the first threshold is 0.80, the second threshold is 0.65, and when the similarity degree between the first feature and the second feature is 0.60, the processing device 112 may determine that there is a slight damage to the subject to be detected.

[0115] The first image may be obtained by shooting by the subject to be detected. Since an original state of the subject to be tested is not clear, the detection result of the subject to be detected directly determined based on the first image may be inaccurate. For example, if a subject is changed by water soaking, since the original color is unclear, the detection result of the subject may be inaccurate.

[0116] According to some embodiments of the present disclosure, the second image may be obtained based on the first image of the subject to be detected. The processing device 112 may obtain the first feature and the second feature by processing, based on the second model, the second image, and the first image, respectively. Finally, the processing device 112 may determine, based on the first feature and the second feature, the detection result of the subject. The detection result may indicate whether the subject includes an anomaly. The method for subject detection may be simple, which may quickly and accurately determine whether the subject includes an anomaly. In addition, the processing device 112 may determine the degree of damage to the subject to be detected based on the similarity degree between the first feature and the second feature, which may facilitate a user to perform different treatments on different degrees of damage.

[0117] It should be noted that the above description is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure.

For persons having ordinary skills in the art, multiple variations or modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. For example, operation 520 and operation 530 may be performed simultaneously. Alternatively, operation 530 may be performed before operation 520.

[0118] FIG. 6 is a flowchart illustrating an exemplary process for generating one or more models according to some embodiments of the present disclosure. In some embodiments, process 600 may be executed by the detection system 100. For example, the process 600 may be implemented as a set of instructions (e.g., an application) stored in a storage device (e.g., the storage device 150, the storage device 220, and/or the storage 390). In some embodiments, the processing device 112 (e.g., the processor 210 of the computing device 200, the CPU 340 of the mobile device 300, and/or one or more modules illustrated in FIG. 4) may execute the set of instructions and may accordingly be directed to perform the process 600. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 600 may be accomplished with one or more additional operations not described and/or without one or more of the operations discussed. Additionally, the order of the operations of process 600 illustrated in FIG. 6 and described below is not intended to be limiting. In some embodiments, the first model described in connection with operation 510 in FIG. 5 and the second model described in connection with operations 520 and 530 in FIG. 5 may be obtained according to the process 600. In some embodiments, the process 600 may be performed by another device or system other than the imaging system 100, e.g., a device or system of a vendor of a manufacturer. For illustration purposes, the implementation of the process 600 by the processing device 112 is described as an example.

[0119] As shown in FIG. 6, in some embodiments, an image 610 of a subject to be detected (e.g., the first image or the second image in FIG. 5) may be input into a joint model 620, the joint model 620 may output a feature 630 (e.g., the first feature and the second feature in FIG. 5) that indicates whether the subject includes an anomaly. In some embodiments, the joint model 620 may include a first model 621 and a second model 622.

[0120] In some embodiments, the first model 621 may be obtained according to a first training process. The processing device 112 (e.g., the obtaining module 410, the processing module 411) may obtain a plurality of first training samples 640. Each of the plurality of first training samples 640 may include a first sample image 641 of a first sample subject with an anomaly and a reference image 642 corresponding to the first sample image. The reference image 642 may include the first sample subject without the anomaly.

[0121] As used herein, a first sample subject refers to an object that is used for training the first model 621. The first sample subject may be of the same type or a different type of object as the subject as described in connection with FIG. 5. For example, if the first model is used to process a first image of a subject (or a portion thereof), the first sample subject may be another subject. The first sample image 641 of the first sample subject refers to a first image of the sample subject. More descriptions regarding the first sample image may be found elsewhere in the present disclosure (e.g., FIG. 5 and the descriptions thereof). The reference image 642 refers to a ground truth first image that is obtained when the first sample subject includes no anomaly. For example, the reference image 642 may be acquired by an imaging device when the first sample subject is packaged.

[0122] In some embodiments, the first training sample 640 may be acquired by shooting the first sample subject before and after a damage. For example, the reference image 642 may be acquired by an imaging device when a first sample subject is packaged. Then, the first sample subject may be damaged. The first sample image 641 of the first sample subject with damage may be acquired by the imaging device.

[0123] In some embodiments, the first training sample 640 may be previously generated and stored in a storage device (e.g., the storage device 130, the storage device 220, the storage 390, or an external database). The processing device 112 may retrieve the first training samples 640 directly from the storage device. In some embodiments, at least a portion of the first training sample 640 may be generated by the processing device 112. Merely by way of example, a video may be acquired by shooting a first sample subject. The processing device 112 may acquire the video of the first sample subject from a storage device where the video is stored. The processing device 112 may determine a first sample image 641 of the first sample subject. The first sample image 641 of the first sample subject may be determined by performing a framing operation on the video.

[0124] In some embodiments, the first training samples 640 (or a portion thereof) may need to be preprocessed before being used in training the first model 621. For example, for a first training sample, the processing device 112 may perform image resizing, image resampling, and image normalization on the first sample image 641.

[0125] In some embodiments, the processing device 112 may generate the first model 621 by training a first preliminary model using the plurality of first training samples 640.

[0126] In some embodiments, the first preliminary model may be an initial model (e.g., a machine learning model) before being trained. Exemplary machine learning models may include a convolutional neural network (CNN) model, a recurrent neural network (RNN) model, a long short term memory (LSTM) network model, a fully convolutional neural network (FCN) model, a generative adversarial network (GAN) model, a radial basis function (RBF) machine learning model, a DeepMask model, a SegNet model, a dilated convolution model, a conditional random fields as recurrent neural networks (CRFasRNN) model, a pyramid scene parsing network (pspnet) model, or the like, or any combination thereof.

[0127] In some embodiments, the first preliminary model may include a multi-layer structure. For example, the first preliminary model may include an input layer, an output layer, and one or more hidden layers between the input layer and the output layer. In some embodiments, the hidden layers may include one or more convolution layers, one or more rectified-linear unit layers (ReLU layers), one or more pooling layers, one or more fully connected layers, or the like, or any combination thereof. As used herein, a layer of a model refers to an algorithm or a function for processing input data of the layer. Different layers may perform different kinds of processing on their respective input. A successive layer may use output data from a previous layer of the successive layer as input data. In some embodiments, the convolutional layer may include a plurality of kernels, which may be used to extract a feature. In some embodiments, each kernel of the plurality of kernels may filter a portion (i.e., a region). The pooling layer may take an output of the convolutional layer as an input. The pooling layer may include a plurality of pooling nodes, which may be used to sample the output of the convolutional layer, so as to reduce the computational load of data processing and accelerate the speed of data processing speed. In some embodiments, the size of the matrix representing the inputted data may be reduced in the pooling layer. The fully connected layer may include a plurality of neurons. The neurons may be connected to the pooling nodes in the pooling layer. In the fully connected layer, a plurality of vectors corresponding to the plurality of pooling nodes may be determined based on a training sample, and a plurality of weighting coefficients may be assigned to the plurality of vectors. The output layer may determine an output based on the vectors and the weighting coefficients obtained from the fully connected layer.

[0128] In some embodiments, each of the layers may include one or more nodes. In some embodiments, each node may be connected to one or more nodes in a previous layer. The number (or count) of nodes in each layer may be the same or different. In some embodiments, each node may correspond to an activation function. As used herein, an activation function of a node may define an output of the node given input or a set of inputs. In some embodiments, each connection between two of the plurality of nodes in the first preliminary model may transmit a signal from one node to another node. In some embodiments, each connection may correspond to a weight coefficient. A weight coefficient corresponding to a connection may be used to increase or decrease the strength or impact of the signal at the connection.

[0129] The first preliminary model may include one or more model parameters, such as architecture parameters, learning parameters, etc. For example, the first preliminary model may be a CNN model and exemplary model parameters of the preliminary model may include the number (or count) of layers, the number (or count) of kernels, a kernel size, a stride, a padding of each convolutional layer, a loss function, or the like, or any combination thereof. Before training, the model parameter(s) of the first preliminary model may have their respective initial values. For example, the processing device 112 may initialize parameter value(s) of the model parameter(s) of the first preliminary model.

[0130] In some embodiments, the first preliminary model may be trained according to a machine learning algorithm. For example, the processing device 112 may generate the first model 621 according to a supervised machine learning algorithm by performing one or more iterations to iteratively update the model parameter(s) of the first preliminary model. As another example, the first sample image 641 of the first sample subject may be input into the first preliminary model. The reference image 642 corresponding to the first sample image 641 may serve as a supervisor to train the first model 621. Parameters (i.e., parameter values) of the first model 621 may be updated by a learning algorithm (e.g., a random gradient drop (SGD) algorithm) so as to minimize a first loss function.

[0131] In some embodiments, a perceived loss function may be determined as the first loss function during the first training process. In some embodiments, the first loss function may include a square loss function, a logic regression loss function, or the like, or any combination thereof.

[0132] In some embodiments, the second model 622 may be obtained according to a second training process. The processing device 112 (e.g., the obtaining module 410) may obtain a plurality of second training samples 650. Each of the plurality of second training samples 650 may include a second sample image of a second sample subject without an anomaly and a label of the second sample image indicating a type 653 of the second sample image.

[0133] As used herein, a second sample subject refers to an object that is used for training the second model 622. The second sample subject may be of the same type or a different type of object as the subject as described in connection with FIG. 5 or the first sample subject. For example, if the second model is used to obtain a feature from an image (e.g., the first image, the second image) of a subject (or a portion thereof), the second sample subject may be another subject. The second sample image of the second sample subject without the anomaly may include an actual image 651 acquired by an imaging device or a simulated image 652 generated by the first model. The type 653 of the second sample image may include one of an actual image acquired by an imaging device and a simulated image generated by the first model.

[0134] In some embodiments, the second training sample 650 may be previously generated and stored in a storage device (e.g., the storage device 130, the storage device 220, the storage 390, or an external database). The processing device 112 may retrieve the second training sample 650 directly from the storage device. In some embodiments, at least a portion of a second training sample 650 may be generated by the processing device 112. For example, the actual image 651 may be acquired by an imaging device. The processing device 112 may acquire the actual image 651 from a storage device where the actual image 651 is stored. Additionally or alternatively, the processing device 112 may determine a type 653 of the actual image 651. That is, a label of the actual image 651 may be determined to be 0. As another example, the simulated image 652 may be generated by the first model. The processing device 112 may acquire the simulated image 652 from a storage device where the simulated image 652 is stored. Additionally or alternatively, the processing device 112 may determine the type 653 of the simulated image 652. That is, a label of the simulated image 652 may be determined to be 1.

[0135] In some embodiments, the second training samples 650 (or a portion thereof) may need to be preprocessed before being used in training the second model 621. For example, for a second training sample, the processing device 112 may perform image resizing, image resampling, and image normalization on the second sample image of the second sample subject without the anomaly.

[0136] In some embodiments, the processing device 112 may generate the second model 622 by training a second preliminary model using the plurality of second training samples 650.

[0137] In some embodiments, the second preliminary model may be an initial model (e.g., a machine learning model) before being trained. Exemplary machine learning models may include a convolutional neural network (CNN) model, a recurrent neural network (RNN) model, a long short term memory (LSTM) network model, a fully convolutional neural network (FCN) model, a generative adversarial network (GAN) model, a radial basis function (RBF) machine learning model, a DeepMask model, a SegNet model, a dilated convolution model, a conditional random fields as recurrent neural networks (CRFasRNN) model, a pyramid scene parsing network (pspnet) model, or the like, or any combination thereof.

[0138] In some embodiments, the second preliminary model may include a multi-layer structure. For example, the second preliminary model may include an input layer, an output layer, and one or more hidden layers between the input layer and the output layer. In some embodiments, the hidden layers may include one or more convolution layers, one or more rectified-linear unit layers (ReLU layers), one or more pooling layers, one or more fully connected layers, or the like, or any combination thereof. As used herein, a layer of a model refers to an algorithm or a function for processing input data of the layer. Different layers may perform different kinds of processing on their respective input. A successive layer may use output data from a previous layer of the successive layer as input data. In some embodiments, the convolutional layer may include a plurality of kernels, which may be used to extract a feature. In some embodiments, each kernel of the plurality of kernels may filter a portion (i.e., a region). The pooling layer may take an output of the convolutional layer as an input. The pooling layer may include a plurality of pooling nodes, which may be used to sample the output of the convolutional layer, so as to reduce the computational load of data processing and accelerate the speed of data processing speed. In some embodiments, the size of the matrix representing the inputted data may be reduced in the pooling layer. The fully connected layer may include a plurality of neurons. The neurons may be connected to the pooling nodes in the pooling layer. In the fully connected layer, a plurality of vectors corresponding to the plurality of pooling nodes may be determined based on a training sample, and a plurality of weighting coefficients may be assigned to the plurality of vectors. The output layer may determine an output based on the vectors and the weighting coefficients obtained from the fully connected layer.

[0139] In some embodiments, each of the layers may include one or more nodes. In some embodiments, each node may be connected to one or more nodes in a previous layer. The number (or count) of nodes in each layer may be the same or different. In some embodiments, each node may correspond to an activation function. As used herein, an activation function of a node may define an output of the node given input or a set of inputs. In some embodiments, each connection between two of the plurality of nodes in the second preliminary model may transmit a signal from one node to another node. In some embodiments, each connection may correspond to a weight coefficient. A weight coefficient corresponding to a connection may be used to increase or decrease the strength or impact of the signal at the connection.

[0140] The second preliminary model may include one or more model parameters, such as architecture parameters, learning parameters, etc. For example, the second preliminary model may be a CNN model and exemplary model parameters of the preliminary model may include the number (or count) of layers, the number (or count) of kernels, a kernel size, a stride, a padding of each convolutional layer, a loss function, or the like, or any combination thereof. Before training, the model parameter(s) of the second preliminary model may have their respective initial values. For example, the processing device 112 may initialize parameter value(s) of the model parameter(s) of the second preliminary model.

[0141] In some embodiments, the second preliminary model may be trained according to a machine learning algorithm. For example, the processing device 112 may generate the second model 622 according to a supervised machine learning algorithm by performing one or more iterations to iteratively update the model parameter(s) of the second preliminary model. As another example, the second sample image of the second sample subject may be input into the second preliminary model. The label of the second sample image indicating the type 653 of the second sample image may serve as a supervisor to train the second model 622. Parameters (i.e., parameter values) of the second model 622 may be updated by a learning algorithm (e.g., a random gradient drop (SGD) algorithm) so as to minimize a second loss function.

[0142] In some embodiments, a determination loss function may be determined as the second loss function during the second training process. In some embodiments, the second loss function may be the same as the first loss function, such as a square loss functions, a logic regression loss function, or the like, or any combination thereof.

[0143] During the training of the second model, a feature vector whose elements are 0 may be designated as a feature of the simulated image. A feature vector whose elements are 1 may be designated as a feature of the actual image. The second model may output the feature of the simulated image or the actual image. In some embodiments, a dimension of the feature may be greater than 1.

[0144] In addition, considering an error factor, the elements in the feature vector output by the second model may not be the same. For example, a portion of elements in the feature vector output by the second model may be 0, and a portion of elements in the feature vector output by the second model may be 1. At this time, it may be inconvenient to compare the feature vector of the second model output by the second model with the label for supervisory during the second training process. Therefore, when the second model is trained, an average value of the elements in the image feature output by the second model may be determined. If the average value is greater than or equal to a threshold, all elements in the image feature output by the second model may be designated as 1. If the average value is less than the threshold, all elements in the image feature output by the second model may be designated as 0.

[0145] For example, the threshold may be determined to be 0.5. If a feature vector output by the second model is (0, 1, 0, 0, 0, 0, 1) during the second training process, the average value may be determined to be two sevenths. Since the threshold is 0.5, the average value may be less than the threshold. All elements in the image feature output by the second model may be designated as 0. That is, the second model may recognize the image as a simulated image.

[0146] It should be noted that designating all elements in the image feature of the second model to be 0 or 1 may be performed during the training of the second model based on the comparison between the average value and the threshold. During an actual application, the second model may directly use the output feature vector as the first feature or the second feature when the second model performs feature extraction on the first image or the second image.

[0147] In some embodiments, the second model may be implemented based on a 3D convolution (e.g., a POINTNET++ algorithm or other image feature extraction algorithm).

[0148] In some embodiments, the trained first model may output a simulated image based on a first image of the subject to be detected. However, the simulated image may be obviously different from the actual image. Therefore, the simulated image may be input into the second model, and the parameters of the first model may be continuously trained based on results of the second model. At the same time, as the similarity degree between the simulated image output by the first model and the actual image are increasing, the second model may be constantly trained to improve an identification of the second model. In some embodiments, the first model and the second model may be jointly trained.

[0149] In some embodiments, the joint model 620 may include the first model 621 and the second model 622. The processing device 112 may obtain a plurality of third training samples. Each of the plurality of third training samples may include a third sample image 661 of a third sample subject and a sample feature 662 of a reference image of the third sample subject without anomaly corresponding to the third sample subject.

[0150] As used herein, a third sample subject refers to an object that is used for training the third model 620. The third sample subject may be of the same type or a different type of object as the subject as described in connection with FIG. 5, the first sample subject, or the second sample subject. For example, if the joint model 620 is used to obtain a feature from an image (e.g., the first image, the second image) of a subject (or a portion thereof), the third sample subject may be another subject. The third sample image 661 of the third sample subject may be acquired by an imaging device. The reference image of the third sample subject without anomaly refers to a truth third image that is obtained when the third sample subject includes no anomaly. For example, the reference image of the third sample subject without anomaly may be acquired by an imaging device when the third sample subject is packaged. The sample feature 662 may be obtained by extract an image feature from the reference image of the third sample subject without anomaly. For example, the processing device 112 may perform an image feature extraction algorithm on the reference image of the third sample subject without anomaly.

[0151] In some embodiments, the third training sample 660 may be previously generated and stored in a storage device (e.g., the storage device 130, the storage device 220, the storage 390, or an external database). The processing device 112 may retrieve the third training sample 660 directly from the storage device. In some embodiments, at least a portion of a third training sample 660 may be generated by the processing device 112. Merely by way of example, the reference image of the third sample subject without anomaly may be acquired by an imaging device when the third sample subject is packaged. Then, the third sample subject may be damaged. The third sample image 661 of the third sample subject may be acquired by the imaging device.

[0152] In some embodiments, the third training samples 660 (or a portion thereof) may need to be preprocessed before being used in training the joint model 620. For example, for a third training sample 661, the processing device 112 may perform image resizing, image resampling, and image normalization on the third sample image 661.

[0153] In some embodiments, the processing device 112 may generate the joint model 620 by training a preliminary joint model including a preliminary first model and a preliminary second model using the plurality of third training samples. A simulated image of the third sample subject without anomaly output by the preliminary first model may be input in the preliminary second model, and the preliminary second model may output a feature of the simulated image of the sample subject without anomaly.

[0154] In some embodiments, parameters (e.g., parameter values) of the preliminary joint model may be updated to decrease a difference between the feature of the simulated image of the sample subject without anomaly and the feature 662 of the reference image of the sample subject without anomaly.

[0155] In some embodiments, the joint model 620 may be trained after a certain degree of training for the first model and/or the second model. Therefore, the joint model 620 may be trained efficiently. In some embodiments, the joint model 620 may be jointly trained directly based on the third training sample 660.

[0156] In some embodiments, the joint model 620 may be implemented based on a generative adversarial network (GAN). The first model may be a generator in the GAN, and the second model may be a discriminator in the GAN.

[0157] During training the GAN, the generator may be configured to generate a simulated image which has a high similarity degree with the reference image that is inputted into the discriminator, so that the discriminator may not distinguish the reference image or the simulated image. A target of the discriminator may be to distinguish the reference image or the simulated image generated by the generator. Therefore, the generator and the discriminator may constitute a dynamic game process.

[0158] In some embodiments, a plurality of stages may be performed when the joint model 620 is trained. For example, in a first stage, the processing device 112 may keep parameters of the second model unchanged and train the first model. During the first stage, a second model with some discriminant ability may be used, so that the first model may constantly generate simulated images based on the subject to be detected. Since a generation ability of the first model is weak, the simulated images generated by the first model may be easily distinguished by the second model. However, as the first model is continuously trained, a similarity degree between the simulated image output by the first model and the reference image may be improved until the second model may not determine whether an input image is a simulated image or a reference image. In a second stage, the processing device 112 may keep the parameters of the first model unchanged and train the second model. After continuous training at the second stage, the discriminant ability of the second model may be improved. The second model may accurately determine all simulated images. At this time, the simulated image output by the first model may be distinguished. Accordingly, the first and second stages may be cycled. Finally, the capabilities of the first model and the second model may become stronger and stronger through the continuous cycles. The GAN may achieve convergence to obtain a trained joint model 620. When the GAN finally converges, the second model may determine the simulated image as the reference image, i.e., the second model cannot or may be difficult to distinguish the simulated image and the reference image corresponding to the simulated image.

[0159] When jointly training the first model and the second model, a joint loss function may be determined based on the second loss function and the first loss function. The joint model 620 may be trained to minimize the joint loss function. In some embodiments, when the second loss function and the first loss function are processed, the second loss function and the first loss function may be weighted. For example, the second loss function and the first loss function may be weighted equally. In some embodiments, the weight may be set in advance to react to the extents of the first model and the second model. For example, when more attention is paid to whether the simulated image output by the first model is close to the reference image, the second loss function may be given a larger weight in advance during the training. Merely by way of example, a weight parameter of the first loss function may be set to be 0.5, and a weight parameter of the second loss function may be set to be 1.

[0160] An image of the subject to be detected may be input into the jointly trained first model to generate a simulated image that is realistic enough. The simulated image output by the first model may be input into the jointly trained second model. A feature of the simulated image may be obtained.

[0161] By jointly training the joint model, the trained first model and the trained second model may include better performance. That is, the trained joint model may generate a simulated image that is realistic enough, and extract image features more quickly and accurately, thereby determining the detection result of the subject.

[0162] It should be noted that the above description regarding process 600 is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. In some embodiments, one or more operations may be added or omitted. For example, the model (e.g., the first model, the second model, the joint model) may be stored in a storage device (e.g., the storage device 130) disclosed elsewhere in the present disclosure for further use. As another example, after the model (e.g., the first model, the second model, the joint model) is generated, the processing device 112 may further test the model (e.g., the first model, the second model, the joint model) using a set of testing images. Additionally or alternatively, the processing device 112 may update the model (e.g., the first model, the second model, the joint model) periodically or irregularly based on one or more training images that become available (e.g., new first training samples, new second training samples, new third training samples).

[0163] FIG. 7 is a flowchart illustrating an exemplary process for subject detection according to some embodiments of the present disclosure. In some embodiments, process 700 may be executed by the detection system 100. For example, the process 700 may be implemented as a set of instructions (e.g., an application) stored in a storage device (e.g., the storage device 150, the storage device 220, and/or the storage 390). In some embodiments, the processing device 112 (e.g., the processor 210 of the computing device 200, the CPU 340 of the mobile device 300, and/or one or more modules illustrated in FIG. 4) may execute the set of instructions and may accordingly be directed to perform the process 700. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, the process 700 may be accomplished with one or more additional operations not described and/or without one or more of the operations discussed. Additionally, the order of the operations of process 700 illustrated in FIG. 7 and described below is not intended to be limiting.

[0164] In 710, the processing device 112 (e.g., the obtaining module 410) may acquire, via an imaging device, images of the subject from a plurality of angles. The imaging device may include a camera, a video recorder, an image sensor, etc. More descriptions regarding the imaging device may be found elsewhere in the present disclosure (e.g., FIG. 1 and the descriptions thereof).

[0165] The image of the subject to be detected may include surface information (e.g., a size, a texture, an anomaly, etc.) of one or more surfaces of the subject to be detected. In some embodiments, the image of the subject to be detected may include a two-dimensional (2D) image.

[0166] In some embodiments, the image of the subject to be detected may include a red-green-blue-depth (RGB-D) image. The RGB-D image may include an RGB image and a depth image. In some embodiments, the RGB-D image may be directly acquired by an imaging device. In some embodiments, an RGB image and a depth image of the subject to be detected at a same position may be obtained by a set of imaging devices (e.g., two imaging devices), respectively. The processing device 112 may generate the RGB-D image based on the RGB image and the depth image of the subject to be detected.

[0167] In some embodiments, an image of the subject to be detected corresponding to a surface may be obtained by one or more (sets of) imaging devices. For example, when the subject to be detected is a package, images of six surfaces of the package may be obtained by one or more (sets of) imaging devices. In some embodiments, there may be an included angle between the imaging device and a corresponding surface of the subject to be inspected. For example, the included angle between an optical axis of the imaging device and the corresponding surface of the subject to be detected may include 30 degrees, 60 degrees, 90 degrees, or the like, or any combination thereof.

[0168] In some embodiments, the processing device 112 may cause to move, via a driving assembly, the subject to acquire the images of the subject to be detected from the plurality of angles via the imaging device. The driving assembly may be configured to cause at least one portion of the imaging device to move. More descriptions regarding the driving assembly may be found elsewhere in the present disclosure (e.g., FIG. 1 and the descriptions thereof). Merely by way of example, the subject to be detected may be a package. In some embodiments, three images corresponding to three adjacent surfaces may be obtained at a time by setting three imaging devices. The processing device 112 may cause to flip, via the driving assembly, the package to acquire three other images of the package via the imaging device. Therefore, six images corresponding to six surfaces of the package may be obtained. In some embodiments, at least one of the three imaging devices may be perpendicular to a surface of the package (i.e., forming an angle of 90 degrees), which may improve the image integrity. For example, an optical axis of one imaging device may be perpendicular to a surface of the package. Optical axes of the other imaging devices and the surface of the package may form other angles (e.g., 30 degrees, 60 degrees, etc.). In some embodiments, an image corresponding to a single surface of the package at a time may be acquired via an imaging device. The package may be flipped five times via the driving assembly to obtain five images corresponding to the other five surfaces. In some embodiments, images of the subject to be detected may also be obtained by other manners. For example, an imaging device may be disposed at each of the six surfaces of the package to obtain an image of the corresponding surface, which may not be limited herein.

[0169] According to some embodiments of the present disclosure, by acquiring images of the subject to be detected by the imaging device from a plurality of angles and moving, via the driving assembly, the subject to acquire, via the imaging device, the images of the subject to be detected from the plurality of angles, the surface information included in the images of the subject to be detected may be obtained to accurately determine whether the subject includes an anomaly (e.g., rupture, wrinkles, depressions). At the same time, the image corresponding to all surfaces of the subject to be detected may be obtained by moving, via the driving assembly, the subject to be detected, which may reduce a count of imaging devices, reduce costs, and improve the efficiency of the image acquisition.

[0170] In some embodiments, the image of the subject to be detected may be directly designated as the first image. Alternatively, the image of the subject to be detected may be further processed. For example, the first image may be generated by processing an image of the subject to be detected and a depth image. More descriptions regarding the first image may be found elsewhere in the present disclosure (e.g., FIG. 5 and the descriptions thereof).

[0171] In 720, the processing device 112 (e.g., the obtaining module 410) may obtain the first image by performing a reconstruction process on the images of the subject to be detected from the plurality of angles based on a reconstruction algorithm.

[0172] The reconstruction process refers to a process that performs a three-dimensional (3D) reconstruction on the image of the subject to be detected based on the reconstruction algorithm. In some embodiments, the reconstruction algorithm may include a pixel-aligned implicit function for a high-resolution 3D (PiFuHD) algorithm, a differentiable renderer (DIB-R) algorithm, or the like. In some embodiments, the 3D image may be obtained after reconstructing the image of the subject to be detected.

[0173] The 3D image refers to an image having a stereo effect. In some embodiments, the 3D image may include a depth image. A pixel value in the depth image may indicate a distance from the imaging device to a position of a surface of the subject to be detected. In some embodiments, the depth image may reflect depth information of the surface of the subject to be detected so as to reflect an anomaly such as a recess, a deformation, etc. That is, the depth image may reflect an anomaly of the surface of the subject to be detected. In some embodiments, the 3D image of the subject to be detected may be acquired in other manners. For example, the 3D image may be obtained by directly scanning, via a 3D laser scanner, the subject to be detected, which may not be limited herein.

[0174] The first image obtained by reconstructing the image of the subject to be detected based on the reconstruction algorithm may include a higher resolution, which may accurately reflect the surface information of the subject to be detected.

[0175] Since the first image also includes the depth image, the first image may intuitively and clearly reflect the anomaly of the surface of the subject to be detected.

[0176] In 730, the processing device 112 (e.g., the obtaining module 410, the processing module 411) may obtain a second image by processing, based on a first model, the first image of the subject to be detected.

[0177] The first model refers to a machine learning model that processes a first representation of the subject in the first image to obtain a second representation of the subject in the second image. In some embodiments, the first model may include a generator of a generative adversarial network (GAN). More descriptions regarding the first model and a first training process may be found elsewhere in the present disclosure (e.g., FIG. 5, FIG. 6, and the descriptions thereof).

[0178] In 740, the processing device 112 (e.g., the obtaining module 410, the first feature obtaining module 412) may obtain a first feature by processing, based on a second model, the first image.

[0179] The second model may be configured to extract features from an image. In some embodiments, the second model may include a discriminator in the GAN. More descriptions regarding the second model and a second training process may be found elsewhere in the present disclosure (e.g., FIG. 5, FIG. 6, and the descriptions thereof).

[0180] In 750, the processing device 112 (e.g., the obtaining module 410, the second feature obtaining module 413) may obtain a second feature by processing, based on the second model, the second image.

[0181] More descriptions regarding the first feature and the second feature may be found elsewhere in the present disclosure (e.g., FIG. 5 and the descriptions thereof).

[0182] In 760, the processing device 112 (e.g., the determination module 420) may determine, based on the first feature and the second feature, a detection result of the subject.

[0183] More descriptions regarding the detection result may be found elsewhere in the present disclosure (e.g., FIG. 5 and the descriptions thereof).

[0184] In 770, the processing device 112 may output a prompt based on the detection result.

[0185] In some embodiments, the processing device 112 may output the prompt based on the detection result determined in operation 760. For example, when the detection result is 1, the output prompt may be that the subject includes an anomaly. When the detection result is 0, the output prompt may be that the subject includes no anomaly.

[0186] It should be noted that the above description is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations or modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. For example, operation 740 and operation 750 may be performed simultaneously. Alternatively, operation 750 may be performed before operation 740.

[0187] Having thus described the basic concepts, it may be rather apparent to those skilled in the art after reading this detailed disclosure that the foregoing detailed disclosure is intended to be presented by way of example only and is not limiting. Various alterations, improvements, and modifications may occur and are intended to those skilled in the art, though not expressly stated herein. These alterations, improvements, and modifications are intended to be suggested by this disclosure, and are within the spirit and scope of the exemplary embodiments of this disclosure.

[0188] Moreover, certain terminology has been used to describe embodiments of the present disclosure. For example, the terms “one embodiment,” “an embodiment,” and/or “some embodiments” mean that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the present disclosure. Therefore, it is emphasized and should be appreciated that two or more references to “an embodiment” or “one embodiment” or “an alternative embodiment” in various portions of this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined as suitable in one or more embodiments of the present disclosure.

[0189] Further, it will be appreciated by one skilled in the art, aspects of the present disclosure may be illustrated and described herein in any of a number of patentable classes or context including any new and useful process, machine, manufacture, or colocation of matter, or any new and useful improvement thereof. Accordingly, aspects of the present disclosure may be implemented entirely hardware, entirely software (including firmware, resident software, micro-code, etc.) or combining software and hardware implementation that may all generally be referred to herein as a “unit,” “module,” or “system.” Furthermore, aspects of the present disclosure may take the form of a computer program product embodied in one or more computer readable media having computer-readable program code embodied thereon.

[0190] A computer readable signal medium may include a propagated data signal with computer readable program code embodied therein, for example, in baseband or as part of a carrier wave. Such a propagated signal may take any of a variety of forms, including electromagnetic, optical, or the like, or any suitable combination thereof. A computer readable signal medium may be any computer readable medium that is not a computer readable storage medium and that may communicate, propagate, or transport a program for use by or in connection with an instruction execution system, apparatus, or device. Program code embodied on a computer readable signal medium may be transmitted using any appropriate medium, including wireless, wireline, optical fiber cable, RF, or the like, or any suitable combination of the foregoing.

[0191] Computer program code for carrying out operations for aspects of the present disclosure may be written in a combination of one or more programming languages, including an object oriented programming language such as Java, Scala, Smalltalk, Eiffel, JADE, Emerald, C++, C#, VB. NET, Python or the like, conventional procedural programming languages, such as the "C" programming language, Visual Basic, Fortran 2003, Perl, COBOL 2002, PHP, ABAP, dynamic programming languages such as Python, Ruby, and Groovy, or other programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer, and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider) or in a cloud computing environment or offered as a service such as a Software as a Service (SaaS).

[0192] Furthermore, the recited order of processing elements or sequences, or the use of numbers, letters, or other designations thereof, are not intended to limit the claimed processes and methods to any order except as may be specified in the claims. Although the above disclosure discusses through various examples what is currently considered to be a variety of useful embodiments of the disclosure, it is to be understood that such detail is solely for that purpose, and that the appended claims are not limited to the disclosed embodiments, but, on the contrary, are intended to cover modifications and equivalent arrangements that are within the spirit and scope of the disclosed embodiments. For example, although the implementation of various components described above may be embodied in a hardware device, it may also be implemented as a software only solution, e.g., an installation on an existing server or mobile device.

[0193] Similarly, it should be appreciated that in the foregoing description of embodiments of the present disclosure, various features are sometimes grouped together in a single embodiment, figure, or description thereof for the purpose of streamlining the disclosure aiding in the understanding of one or more of the various embodiments. This method of disclosure, however, is not to be interpreted as reflecting an intention that the claimed subject matter requires more features than are expressly recited in each claim. Rather, claimed subject matter may lie in less than all features of a single foregoing disclosed embodiment.

WHAT IS CLAIMED IS:

1. A method for subject detection, comprising:

obtaining a second image by processing, based on a first model, a first image of a subject to be detected, the first model being configured to process a first representation of the subject in the first image to obtain a second representation of the subject in the second image;

obtaining a first feature by processing, based on a second model, the first image, the second model being configured to extract features from an image;

obtaining a second feature by processing, based on the second model, the second image; and

determining, based on the first feature and the second feature, a detection result of the subject, the detection result indicating whether the subject includes an anomaly.

2. The method of claim 1, wherein the first model is obtained according to a first training process, the first training process including:

obtaining a plurality of first training samples, each of the plurality of first training samples including a first sample image of a first sample subject with an anomaly and a reference image corresponding to the first sample image, the reference image including the sample subject without the anomaly; and

generating the first model by training a first preliminary model using the plurality of first training samples.

3. The method of claim 1, wherein the second model is obtained according to a second training process, the second training process including:

obtaining a plurality of second training samples, each of the plurality of second training samples including a second sample image of a second sample subject without an anomaly and a label of the second sample image indicating a type of the second sample image, the type of the second sample image including one of an actual image acquired by an imaging device and a simulated image generated by the first model; and

generating the second model by training a second preliminary model using the plurality of second training samples.

4. The method of claim 1, wherein the first model and the second model are two portions of a joint model, the joint model is obtained according to a third training process, the third training process including:

obtaining a plurality of third training samples, each of the plurality of third training samples including a third sample image of a third sample subject and a sample feature of a reference image of the third sample subject without anomaly corresponding to the third sample subject; and

generating the joint model by training a preliminary joint model including a preliminary first model and a preliminary second model using the plurality of third training samples, wherein in the third training process,

a simulated image of the third sample subject without anomaly output by the preliminary first model is input in the preliminary second model, and the preliminary second model outputs a feature of the simulated image of the sample subject without anomaly.

5. The method of claim 1, wherein the second model outputs at least one of the first feature or the second feature based on an average value of a plurality of elements in a feature vector.

6. The method of claim 5, wherein a dimension of the feature vector outputted by the second model is greater than 1.

7. The method of claim 1, wherein the first image includes a depth image.

8. The method of claim 1, further including:

obtaining the first image by acquiring, via an imaging device, images of the subject from a plurality of angles.

9. The method of claim 8, further including:

moving, via a driving assembly, the subject to acquire, via the imaging device, the images of the subject to be detected from the plurality of angles.

10. The method of claim 8, wherein the obtaining the first image by acquiring, via an imaging device, images of the subject to be detected from a plurality of angles includes:

obtaining the first image by reconstructing, based on a reconstruction algorithm, the images of the subject acquired from the plurality of angles, wherein the first image includes a three-dimensional image.

11. The method of claim 1, wherein the determining, based on the first feature and the second feature, a detection result of the subject includes:

obtaining, based on the first feature and the second feature, a similarity degree between the first feature and the second feature; and

determining, based on the similarity degree, the detection result of the subject.

12. A system for subject detection, comprising:

at least one storage device including a set of instructions; and

at least one processor in communication with the at least one storage device, wherein when executing the set of instructions, the at least one processor is directed to perform operations including:

obtaining a second image by processing, based on a first model, a first image of a subject to be detected, the first model being configured to process a first representation of the subject in the first image to obtain a second representation of the subject in the second image;

obtaining a first feature by processing, based on a second model, the first image, the second model being configured to extract features from an image;

obtaining a second feature by processing, based on the second model, the second image; and

determining, based on the first feature and the second feature, a detection result of the subject, the detection result indicating whether the subject includes an anomaly.

13. The system of claim 12, wherein the first model is obtained according to a first training process, the first training process including:

obtaining a plurality of first training samples, each of the plurality of first training samples including a first sample image of a first sample subject with an anomaly and a reference image corresponding to the first sample image, the reference image including the sample subject without the anomaly; and

generating the first model by training a first preliminary model using the plurality of first training samples.

14. The system of claim 12, wherein the second model is obtained according to a second training process, the second training process including:

obtaining a plurality of second training samples, each of the plurality of second training samples including a second sample image of a second sample subject without an anomaly and a label of the second sample image indicating a type of the second sample image, the type of the second sample image including one of an actual image acquired by an imaging device and a simulated image generated by the first model; and

generating the second model by training a second preliminary model using the plurality of second training samples.

15. The system of claim 12, wherein the first model and the second model are two portions of a joint model, the joint model is obtained according to a third training process, the third training process including:

obtaining a plurality of third training samples, each of the plurality of third training samples including a third sample image of a third sample subject and a sample feature of a reference image of the third sample subject without anomaly corresponding to the third sample subject; and

generating the joint model by training a preliminary joint model including a preliminary first model and a preliminary second model using the plurality of third training samples, wherein in the third training process,

a simulated image of the third sample subject without anomaly output by the preliminary first model is input in the preliminary second model, and the preliminary second model outputs a feature of the simulated image of the sample subject without anomaly.

16. The system of claim 12, wherein the second model outputs at least one of the first feature or the second feature based on an average value of a plurality of elements in a feature vector.

17. The system of claim 16, wherein a dimension of the feature vector outputted by the second model is greater than 1.

18. The system of claim 12, wherein the first image includes a depth image.

19. The system of claim 12, the operations further include:

obtaining the first image by acquiring, via an imaging device, images of the subject from a plurality of angles.

20. The system of claim 19, the operations further include:

moving, via a driving assembly, the subject to acquire, via the imaging device, the images of the subject to be detected from the plurality of angles.

21. The system of claim 19, wherein the obtaining the first image by acquiring, via an imaging device, images of the subject to be detected from a plurality of angles includes:

obtaining the first image by reconstructing, based on a reconstruction algorithm, the images of the subject acquired from the plurality of angles, wherein the first image includes a three-dimensional image.

22. The system of claim 12, wherein the determining, based on the first feature and the second feature, a detection result of the subject includes:

obtaining, based on the first feature and the second feature, a similarity degree between the first feature and the second feature; and

determining, based on the similarity degree, the detection result of the subject.

23. A non-transitory computer readable medium, comprising executable instructions that, when executed by at least one processor, direct the at least one processor to perform a method for subject detection, the method comprising:

obtaining a second image by processing, based on a first model, a first image of a subject to be detected, the first model being configured to process a first representation of the subject in the first image to obtain a second representation of the subject in the second image;

obtaining a first feature by processing, based on a second model, the first image, the second model being configured to extract features from an image;

obtaining a second feature by processing, based on the second model, the second image; and

determining, based on the first feature and the second feature, a detection result of the subject, the detection result indicating whether the subject includes an anomaly.

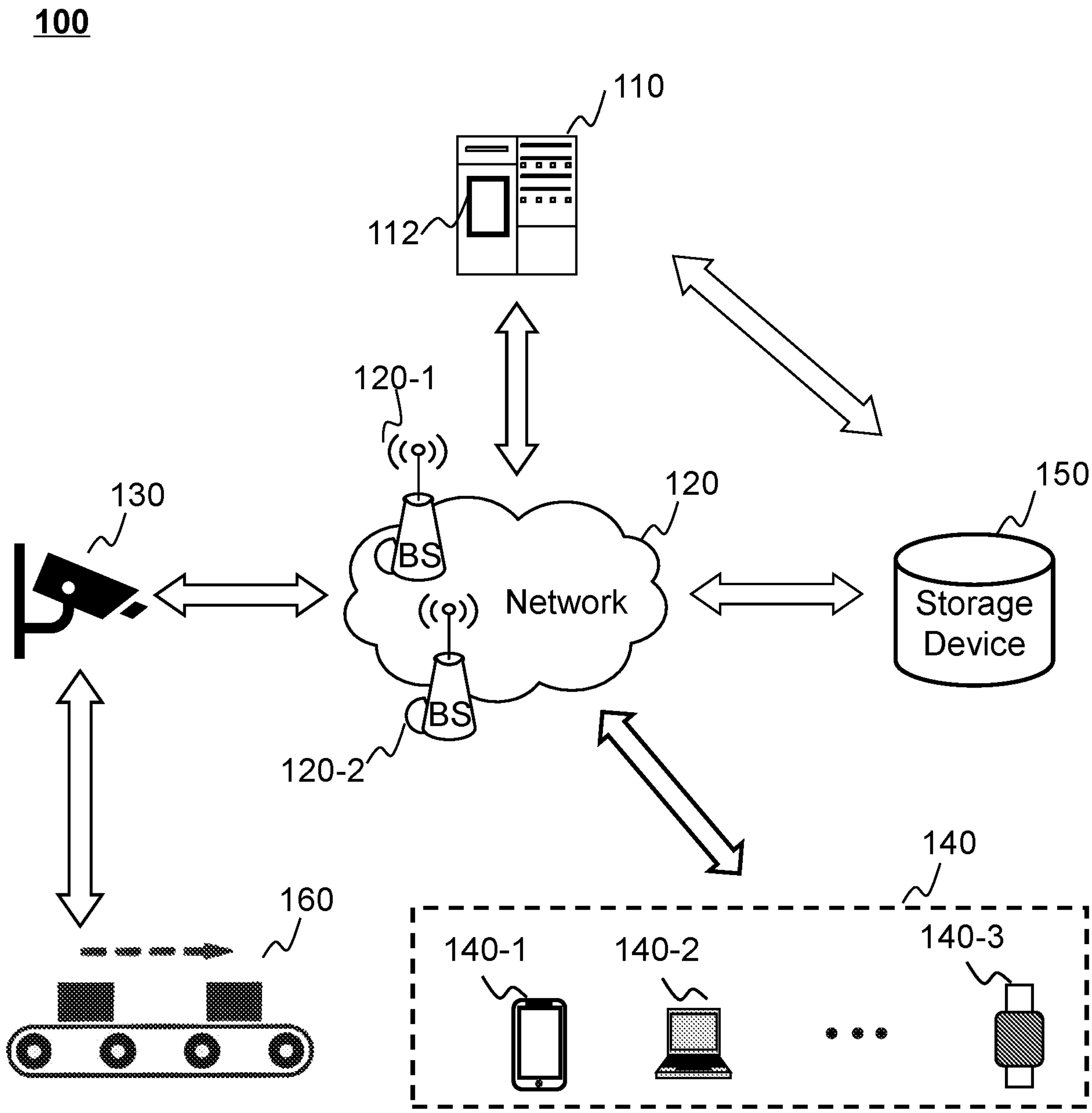
100
110
112
120-1
130
120
150
BS
Network
Storage Device
BS
120-2
140
160
140-1
140-2
140-3

FIG. 1

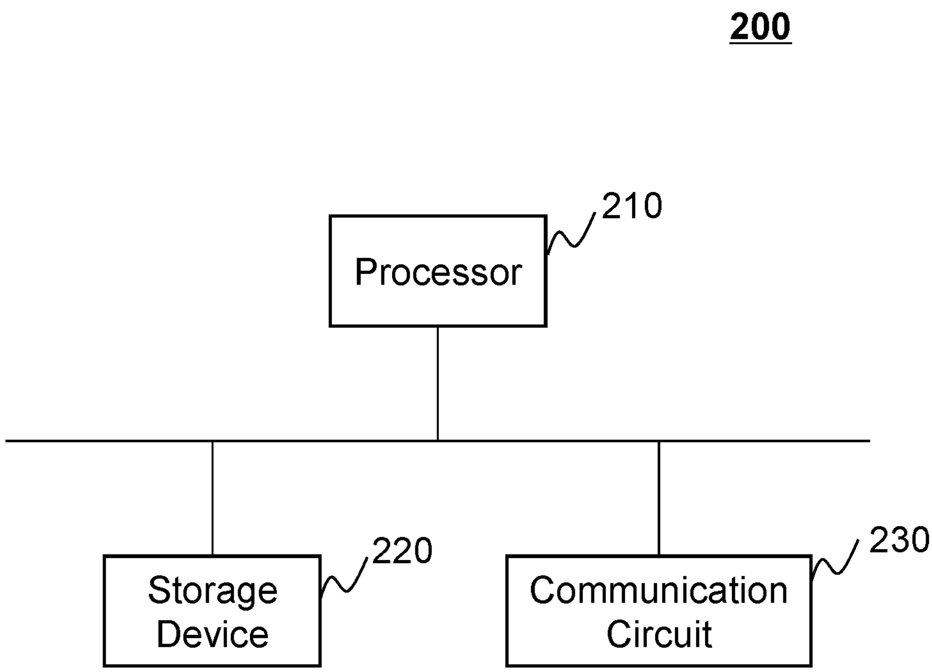
200
210
Processor
220
Storage Device
230
Communication Circuit

FIG. 2A

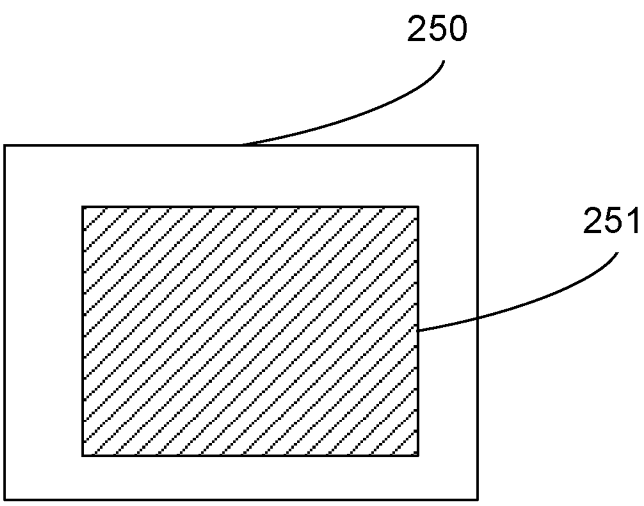
250
251

FIG. 2B

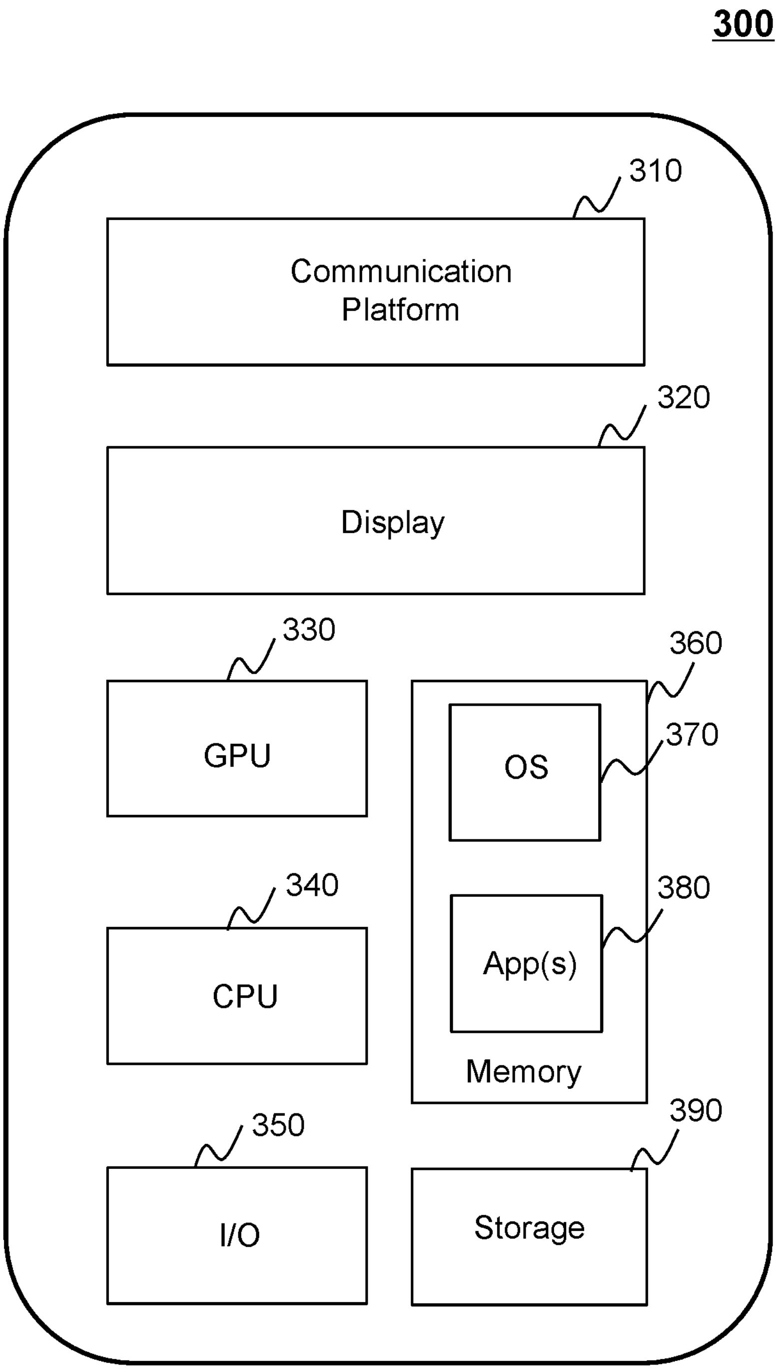
300
310
Communication Platform
320
Display
330
GPU
340
CPU
350
I/O
360
OS
370
App(s)
380
Memory
390
Storage

FIG. 3

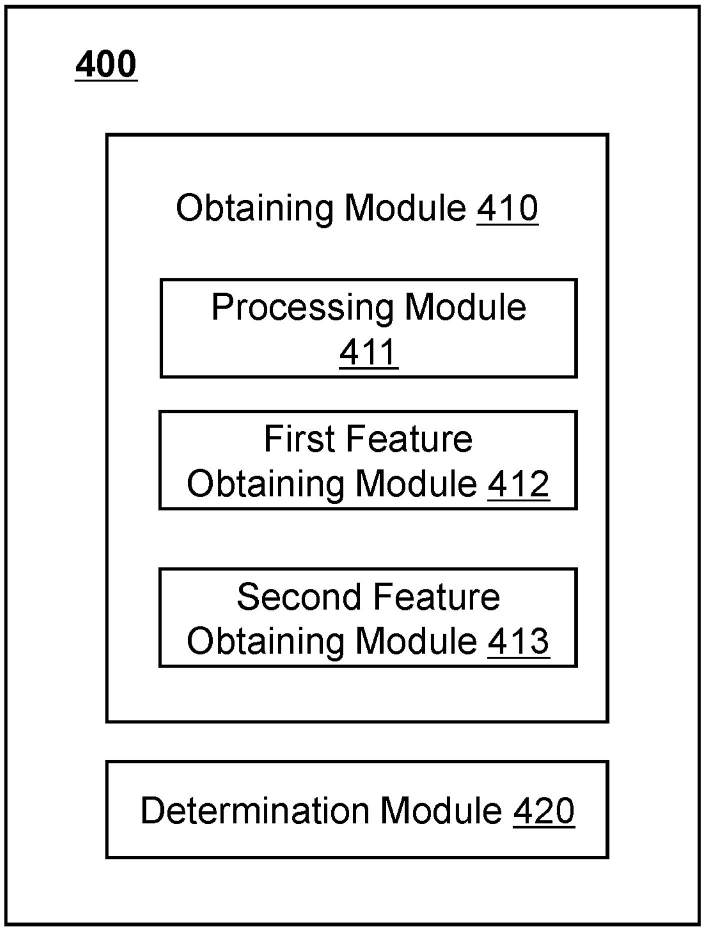
400
Obtaining Module 410
Processing Module 411
First Feature Obtaining Module 412
Second Feature Obtaining Module 413
Determination Module 420

FIG. 4

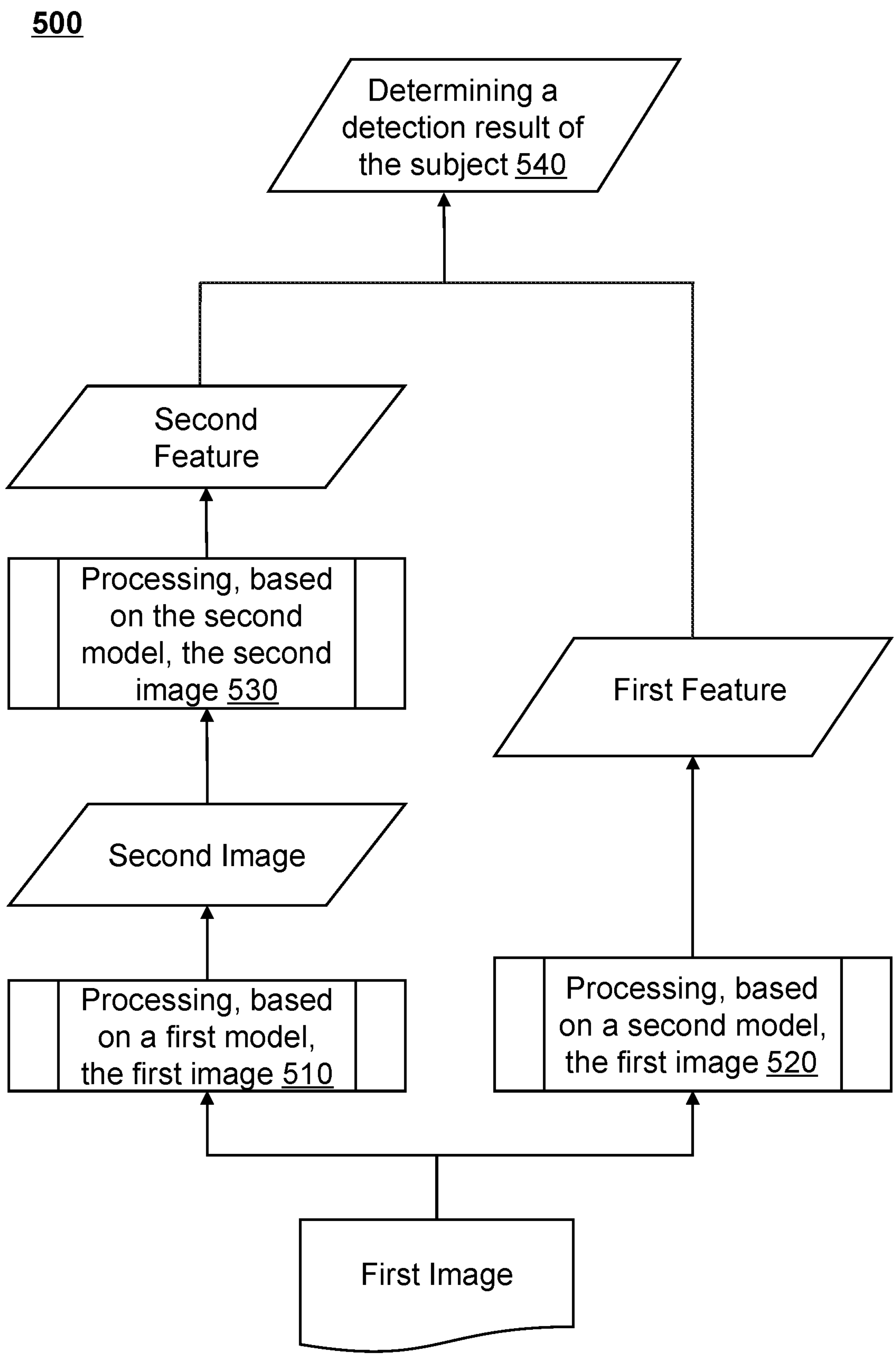
500
Determining a detection result of the subject 540
Second Feature
Processing, based on the second model, the second image 530
Second Image
Processing, based on a first model, the first image 510
First Feature
Processing, based on a second model, the first image 520
First Image

FIG. 5

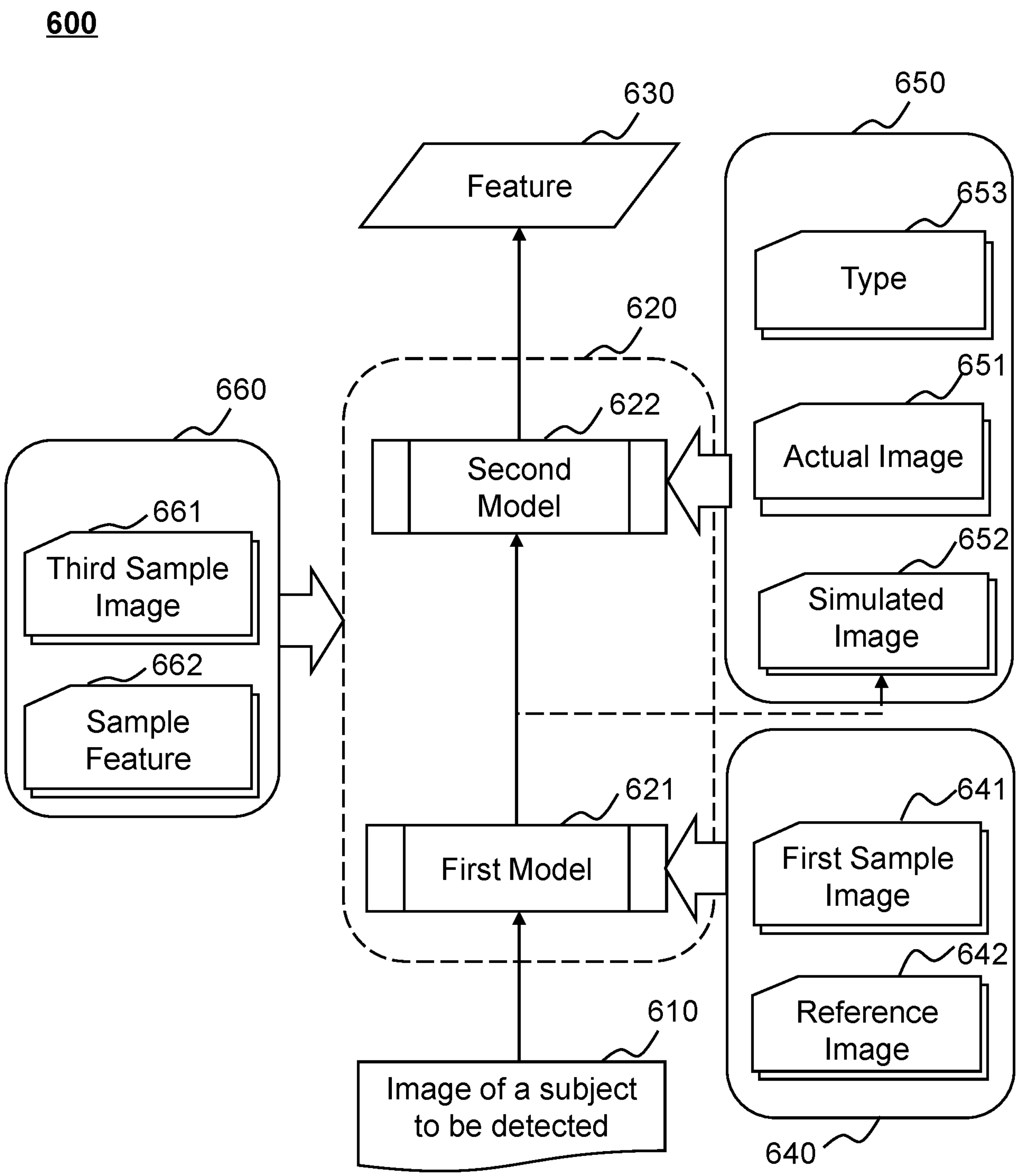
600
630
Feature
650
653
Type
620
651
Actual Image
660
622
Second Model
661
Third Sample Image
652
Simulated Image
662
Sample Feature
621
641
First Model
First Sample Image
642
Reference Image
610
Image of a subject to be detected
640

FIG. 6

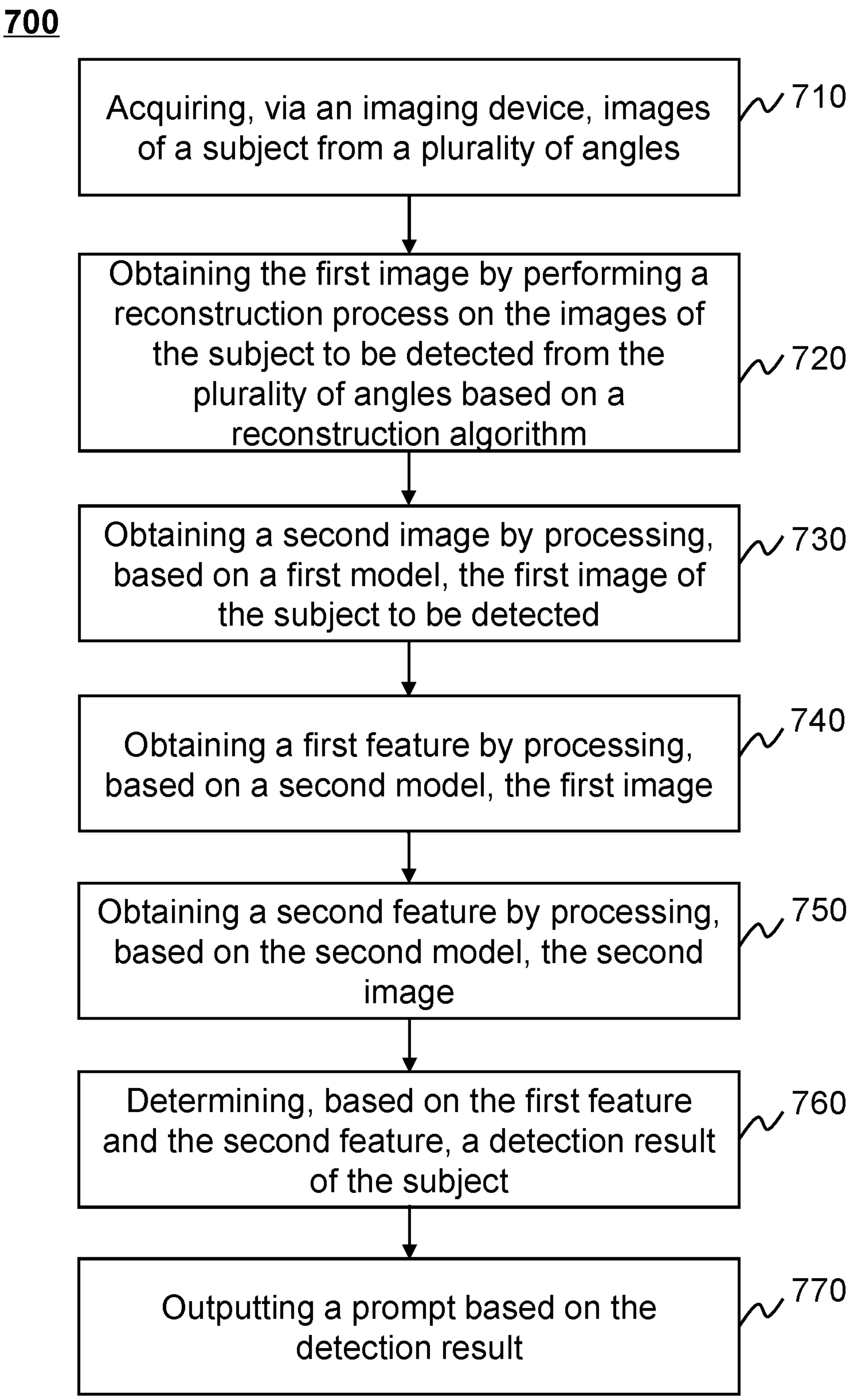
700
Acquiring, via an imaging device, images of a subject from a plurality of angles
710
Obtaining the first image by performing a reconstruction process on the images of the subject to be detected from the plurality of angles based on a reconstruction algorithm
720
Obtaining a second image by processing, based on a first model, the first image of the subject to be detected
730
Obtaining a first feature by processing, based on a second model, the first image
740
Obtaining a second feature by processing, based on the second model, the second image
750
Determining, based on the first feature and the second feature, a detection result of the subject
760
Outputting a prompt based on the detection result
770

FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2021/120370

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/00(2017.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T

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)

CNKI, CNPAT, WPI, EPODOC, IEEE: subject, detection, image, model, feature, abnormal, train, depth, similarity

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 112767331 A (BEIHANG UNIVERSITY) 07 May 2021 (2021-05-07) description, paragraphs [0022]-[0064]	1-3, 11-14, 22-23
A	CN 111767802 A (JD DIGITAL TECHNOLOGY HOLDING CO., LTD.) 13 October 2020 (2020-10-13) the whole document	1-23
A	CN 109740617 A (GUOXIN YOUE DATA CO., LTD.) 10 May 2019 (2019-05-10) the whole document	1-23
A	CN 111382785 A (WUHAN JINGLI ELECTRONIC TECHNOLOGY CO., LTD. et al.) 07 July 2020 (2020-07-07) the whole document	1-23
A	US 2021109036 A1 (CANON KABUSHIKI KAISHA) 15 April 2021 (2021-04-15) the whole document	1-23

☐ 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
29 January 2022	**16 February 2022**
Name and mailing address of the ISA/CN **National Intellectual Property Administration, PRC** **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088, China**	Authorized officer **MU,Ying**
Facsimile No. **(86-10)62019451**	Telephone No. **86-(10)-53961531**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2021/120370

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112767331	A	07 May 2021	None			
CN	111767802	A	13 October 2020	None			
CN	109740617	A	10 May 2019	None			
CN	111382785	A	07 July 2020	None			
US	2021109036	A1	15 April 2021	JP	2021063739	A	22 April 2021
				KR	20210044698	A	23 April 2021
				TW	202129260	A	01 August 2021