

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 February 2009 (05.02.2009)

PCT

(10) International Publication Number
WO 2009/018538 A2

(51) International Patent Classification:
G05B 19/00 (2006.01)

(21) International Application Number:
PCT/US2008/071983

(22) International Filing Date: 1 August 2008 (01.08.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/833,187 2 August 2007 (02.08.2007) US

(71) Applicant (for all designated States except US): **BRAIN-TECH CANADA, INC.** [CA/CA]; 102 - 930 West 1st Street, North Vancouver, British Columbia V7P 3N4 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BOCA, Remus, F.W.** [CA/CA]; 2009-2024 Fullerton Avenue, North Vancouver, British Columbia V7P 3G4 (CA). **BEIS, Jeffrey, Scott** [CA/CA]; 5425 Indian River Drive, North Vancouver, British Columbia V7G 2T7 (CA). **PESCARU,**

Simona, Liliana [CA/CA]; 507-1263 Barclay Street, Vancouver, British Columbia V6E 1H5 (CA).

(74) Agents: **ABRAMONTE, Frank** et al.; Seed Intellectual Property Law Group PLLC, Suite 5400, 701 Fifth Avenue, Seattle, Washington 98104-7064 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

(54) Title: SYSTEM AND METHOD OF THREE-DIMENSIONAL POSE ESTIMATION

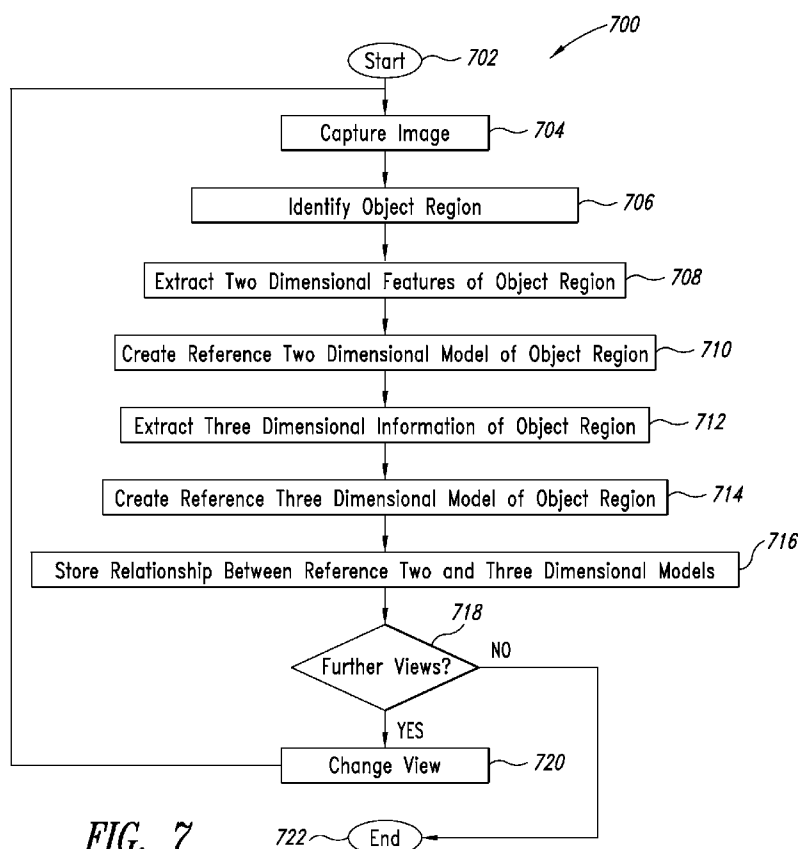


FIG. 7

(57) Abstract: A system and method for identifying objects using a machine-vision based system are disclosed. Briefly described, one embodiment is a method that captures a first image of at least one object with an image capture device, processes the first captured image to find an object region based on a reference two-dimensional model and determines a three-dimensional pose estimation based on a reference three-dimensional model that corresponds to the reference two-dimensional model and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known. Thus, two-dimensional information or data is used to segment an image and three-dimensional information or data used to perform three-dimensional pose estimation on a segment of the image.



European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

SYSTEM AND METHOD OF THREE-DIMENSIONAL POSE ESTIMATION

BACKGROUND OF THE INVENTION

Field of the Invention

This disclosure generally relates to systems and methods of
5 three-dimensional pose estimation employing machine vision, for example
useful in robotic systems.

Description of the Related Art

The ability to determine a three-dimensional pose (*i.e.*, three-
dimensional position and orientation) of an object can be useful in a number of
10 settings. For example, three-dimensional pose estimation may be useful in
various robotic systems that employ machine-vision.

One type of machine-vision problem is known as bin picking. Bin
picking typically takes the form of identifying an object collocated in a group of
identical or similar objects, for example objects such as parts collocated in a bin
15 or other container. Identification may include three-dimensional pose
estimation of the object to allow engagement of the object by a robot member
and removal of the object from the group of objects.

There are many object recognition methods available for locating
complex industrial parts having a large number of machine-vision detectable
20 features. A complex part with a large number of features provides redundancy,
and typically can be reliably recognized even when some fraction of the
features are not properly detected. However, many parts are simple parts and
do not have a sufficient level of redundancy in machine-vision detectable
features and/or which have rough edges or other geometric features which are
25 not clear. In addition, the features typically used for recognition, such as edges
detected in captured images, are notoriously difficult to extract consistently from
image to image when a large number of parts are jumbled together in a bin.
The parts therefore cannot be readily located, especially given the potentially

harsh nature of the environment, *e.g.*, uncertain lighting conditions, varying amounts of occlusions, etc.

The problem of recognizing a simple part among many parts lying jumbled in a bin, such that a robotic system is able to grasp and manipulate the part in an industrial or other process, is quite different from the problem of recognizing a complex part having many detectable features. Machine-vision based systems recognizing and locating three-dimensional objects, using either (a) two-dimensional data from a single image or (b) three-dimensional data from stereo images or range scanners, are known. Single image methods can be subdivided into model-based and appearance-based approaches.

The model-based approaches suffer from difficulties in feature extraction under harsh lighting conditions, including significant shadowing and specularities. Furthermore, simple parts do not contain a large number of machine-vision detectable features, which degrades the accuracy of a model-based fit to noisy image data.

The appearance-based approaches have no knowledge of the underlying three-dimensional structure of the object, merely knowledge of two-dimensional images of the object. These approaches have problems in segmenting out the object for recognition, have trouble with occlusions, and may not provide a three-dimensional pose estimation that is accurate enough for grasping purposes.

Approaches that use three-dimensional data for recognition have somewhat different issues. Lighting effects cause problems for stereo reconstruction, and specularities can create spurious data both for stereo and laser range finders. Once the three-dimensional data is generated, there are the issues of segmentation and representation. On the representation side, more complex models are often used than in the two-dimensional case (*e.g.*, superquadrics). These models contain a larger number of free parameters, which can be difficult to fit to noisy data.

Assuming that a part can be located, it must be picked up by the robotic system. The current standard for motion trajectories leading up to the grasping of an identified part is known as image based visual servoing (IBVS).

A key problem for IBVS is that image based servo systems control image error, but do not explicitly consider the physical camera trajectory. Image error results when image trajectories cross near the center of the visual field (*i.e.*, requiring a large scale rotation of the camera). The conditioning of the image Jacobian results in a phenomenon known as camera retreat. Namely, the robotic system is also required to move the camera back and forth along the optical axis direction over a large distance, possibly exceeding the robotic system range of motion. Hybrid approaches decompose the robotic system motion into translational and rotational components either through identifying homeographic relationships between sets of images, which is computationally expensive, or through a simplified approach which separates out the optical axis motion. The more simplified hybrid approaches introduce a second key problem for visual servoing, which is the need to keep features within the image plane as the robotic system moves.

Conventional bin picking systems are relatively deficient in at least one of the following: robustness, accuracy, and speed. Robustness is required since there may be no cost savings to the manufacturer if the error rate of correctly picking an object from a bin is not close to zero (as the picking station will still need to be manned). Location accuracy is necessary so that the grasping operation will not fail. And finally, solutions which take too long between picks would slow down entire production lines, and would not be cost effective.

BRIEF SUMMARY OF THE INVENTION

In one aspect, an embodiment of a method of object pose estimation using machine-vision may be summarized as including identifying an object region of an image on which pose estimation is being performed based on a correspondence between at least a portion of a representation of an object in the object region of the image and at least a corresponding one of a plurality of reference two-dimensional models of the object, the object region being a portion of the image that contains the representation of at least a portion of the object; and determining a three-dimensional pose of the object based on at

least one of a plurality of reference three-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is
5 not necessarily previously known.

In another aspect, an embodiment of a computer-readable medium that stores instructions for causing a computer to perform object pose estimation using machine-vision may be summarized as including identifying an object region of an image based on a correspondence between at least a
10 portion of a representation of an object in the object region of the image and at least a corresponding one of a plurality of reference two-dimensional models of the object, the object region being a portion of the image that contains the representation of at least a portion of the object; and determining a three-dimensional pose of the object based on at least one of a plurality of reference
15 three-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known.

In further aspect, an embodiment of a system to perform three-dimensional pose estimation may be summarized as including at least one sensor; at least one processor; and at least one memory storing processor executable instructions that cause the at least one processor to segment an image captured by the at least one sensor into a number of object regions
25 based at least in part on a correspondence between at least a portion of a representation of an object in the object region of the image and at least a corresponding one of a plurality of reference two-dimensional models of the object and to cause the at least one processor to determine a three-dimensional pose of the object based on at least one of a plurality of reference
30 three-dimensional models of the object that is related to the corresponding one of the plurality of reference two-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point

relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING(S)

5 In the drawings, identical reference numbers identify similar elements or acts. The sizes and relative positions of elements in the drawings are not necessarily drawn to scale. For example, the shapes of various elements and angles are not drawn to scale, and some of these elements are arbitrarily enlarged and positioned to improve drawing legibility. Further, the
10 particular shapes of the elements as drawn, are not intended to convey any information regarding the actual shape of the particular elements, and have been solely selected for ease of recognition in the drawings.

Figure 1 is an isometric view of a machine-vision based system including a control system, a sensor system and robotic system operating on a
15 bin containing parts, according to one illustrated embodiment, the sensor system including a camera mounted for movement with respect the parts.

Figure 2 is an isometric view of a sensor system according to another illustrated embodiment, the sensor system including a pair of cameras in a stereo configuration positioned to capture stereo images of the parts in the
20 bin.

Figure 3 is an isometric view of a sensor system according to still another illustrated embodiment, the sensor system including at least one camera positioned to capture images of the parts in the bin and a range finding system positioned to determine range information indicative of a distance to
25 parts in the bin.

Figure 4 is an isometric view of a sensor system according to yet another illustrated embodiment, the sensor system including a camera and a structure light system positioned to capture images of the parts in the bin.

Figure 5 is a block diagram illustrating a control system, according
30 to one illustrated embodiment.

Figure 6 is a flow diagram of a method of operating a machine-vision system to perform three-dimensional pose estimation according to one illustrated embodiment, the method including calibrating, training in a training mode or time before a runtime or runtime mode, and three-dimensional pose
5 estimating during the runtime or runtime mode.

Figure 7 is a flow diagram of a method of training a machine-vision system according to one illustrated embodiment, the method including extracting two-dimensional feature information, creating reference two-dimensional models, extracting three-dimensional information and creating
10 reference three-dimensional model.

Figure 8 is a flow diagram of a method of extracting two- and three-dimensional information according to one illustrated embodiment in which the two- and three-dimensional feature information is extracted by accessing an existing computer or digital model of the object.

15 Figure 9 is a flow diagram of a method of extracting two- and three-dimensional information according to another illustrated embodiment in which the two- and three-dimensional feature information is extracted from data sensed from a representative of training object.

Figure 10 is a flow diagram of a method of performing runtime
20 three-dimensional pose estimation according to one illustrated embodiment which includes capturing an image, identifying a object region, identifying a reference three-dimensional model that corresponds to the identified object region, and determining a three-dimensional pose estimation for the object.

Figure 11 is a flow diagram of a method of identifying object
25 regions in an image, according to one illustrated embodiment.

Figure 12 is a flow diagram of a method illustrating types of data on which may be used to identify the object region according to one illustrated embodiment.

Figure 13 is a flow diagram of a method illustrating a variety of
30 approaches for identifying the object region according to one illustrated embodiment.

Figure 14 is a flow diagram of a method of determining a three-dimensional pose estimation for the object according to one illustrated embodiment which includes performing a registration.

Figure 15 is a flow diagram of illustrating a method of performing
5 a registration according to one illustrated embodiment.

DETAILED DESCRIPTION OF THE INVENTION

In the following description, certain specific details are set forth in order to provide a thorough understanding of various embodiments. However, one skilled in the art will understand that the invention may be practiced without
10 these details. In other instances, well-known structures associated with robotic systems, cameras and other image capture devices, range finders, lighting, as well as control systems including computers and networks, have not been shown or described in detail to avoid unnecessarily obscuring descriptions of the embodiments.

15 Unless the context requires otherwise, throughout the specification and claims which follow, the word “comprise” and variations thereof, such as, “comprises” and “comprising” are to be construed in an open, inclusive sense, that is as “including, but not limited to.”

Reference throughout this specification to “one embodiment” or
20 “an embodiment” means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment. Thus, the appearances of the phrases “in one embodiment” or “in an embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment. Further more, the particular
25 features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

As used in this specification and the appended claims, the singular forms “a,” “an,” and “the” include plural referents unless the content clearly dictates otherwise. It should also be noted that the term “or” is generally
30 employed in its sense including “and/or” unless the content clearly dictates otherwise.

The headings and Abstract of the Disclosure provided herein are for convenience only and do not interpret the scope or meaning of the embodiments.

Figure 1 shows a machine-vision based system 100, according to one illustrated embodiment.

The machine-vision based system 100 may include a sensor system 102, a robotic system 104, a control system 106 and a network 108 communicatively coupling the sensor system 102, robotic system 104 and control system 106. The machine-vision based system 100 may be employed to recognize a pose of and manipulate one or more work pieces, for example one or more objects such as parts 110. The parts 110 may be collocated, for example in a container such as a bin 112.

While illustrated as a machine-vision based system 100, aspects of the present disclosure may be employed in other systems, for example non-machine-vision based systems. Such non-machine-vision based systems may, for example, take the form of inspection systems. Also, while illustrated as operating in a bin picking environment, aspects of the present disclosure may be employed in other environments, for example non-bin picking environments in which the objects are not collocated or jumbled.

As illustrated in Figure 1, the sensor system 102 includes an image capture device 114. The image capture device 114 may take a variety of forms, for example an analog camera or digital camera. The image capture device 114 may, for example, take the form an array of charge coupled devices (CCDs) or complementary metal oxide semiconductor (CMOS) sensors, a Vidicon and other image capture devices.

In some embodiments, the image capture device 114 may be mounted for movement relative to the parts 110. For example, the image capture device 114 may be mounted to a sensor robotic system 116, which may include a base 116a, one or more arms 116b-116e, and one or more servomotors and other suitable actuators (not shown) which are operable to move the various arms 116b-116e and/or base 116a. It is noted that the sensor robotic system 116 may include a greater or less number of arms and/or

different types of members such that any desirable range of rotational and/or translational movement of the image capture device 114 may be provided.

Accordingly, the image capture device 114 may be positioned and/or oriented in any desirable pose to capture images of the pile of objects 112. Such permits

5 the capture of images of two or more views of a given part 110, allowing the generation or derivation of three-dimensional data or information regarding the part 110.

In typical embodiments, the position and/or orientation or pose of the various components of the sensor robotic system 116 may be known or
10 ascertainable to the control system 106. For example, the sensor robotic system 116 may include one or more sensors (*e.g.*, encoders, Reed switches, position sensors, contact switches, accelerometers, etc.) or other devices positioned and configured to sense, measure or otherwise determine information indicative of a current position, speed, acceleration, and/or
15 orientation or pose of the image capture device 114 in a defined coordinate frame (*e.g.*, sensor robotic system coordinate frame, real world coordinate frame, etc.). The control system 106 may receive information from the various sensors or devices, and/or from actuators indicating position and/or orientation of the arms 116b-116e. Alternatively, or additionally, the control system 106
20 may maintain the position and/or orientation or pose information based on movements of the arms 116b-116e made from an initial position and/or orientation or pose of the sensor robotic system 116. The control system 106 may computationally determine a position and/or orientation or pose of the image capture device 114 with respect to a reference coordinate system 122.
25 Any suitable position and/or orientation or pose determination methods, systems or devices may be used by the various embodiments. Further, the reference coordinate system 122 is illustrated for convenience as a Cartesian coordinate system using an x-axis, a y-axis, and a z-axis. Alternative embodiments may employ other reference systems, for example a polar
30 coordinate system.

The robotic system 104 may include a base 104a, an end effector 104b, and a plurality of intermediate members 104c-104e. End effector 104b is

illustrated for convenience as a grasping device operable to grasp a selected one of the objects 110 from the pile of objects 110. Any device that can engage a part 110 may be suitable as an end effector device(s).

In typical embodiments, the position and/or orientation or pose of the various components of the robotic system 104 may be known or ascertainable to the control system 106. For example, the robotic system 104 may include one or more sensors (e.g., encoders, Reed switches, position sensors, contact switches, accelerometers, etc.) or other devices positioned and configured to sense, measure or otherwise determine information indicative of a current position and/or orientation or pose of the end effector 104b in a defined coordinate frame (e.g., robotic system coordinate frame, real world coordinate frame, etc.). The control system 106 may receive information from the various sensors or devices, and/or from actuators indicating position and/or orientation of the arms 104c-104e. Alternatively, or additionally, the control system 106 may maintain the position and/or orientation or pose information based on movements of the arms 104c-104e made from an initial position and/or orientation or pose of the robotic system 104. The control system 106 may computationally determine a position and/or orientation or pose of the end effector 104b with respect to a reference coordinate system 122. Any suitable position and/or orientation or pose determination methods, systems or devices may be used by the various embodiments.

Figure 2 shows a sensor system 202 positioned to capture images of parts 210 which may, for example, be collocated in a bin 212, according to another embodiment.

In particular, the sensor system 202 includes a pair of cameras 214 to produce stereo images. The pair of cameras 214 may be packaged as a stereo sensor (commercially available) or may be separate cameras positioned to provide stereo images. Such permits the capture of stereo images of a given part 210 from two different views, allowing the generation or derivation of three-dimensional data or information regarding the part 210.

Figure 3 shows a sensor system 302 positioned to capture three-dimensional information or data regarding parts 310 which may, for example, be collocated in a bin 312, according to another embodiment.

In particular, the sensor system 302 includes at least one image capture device 314a, 314b and at least one range finding device 316, which may, for example, include a transmitter 316a and receiver 316b. The range finding device 316 may, for example, take the form of a laser range finding device, infrared range finding device or ultrasonic range finding device. Other range finding devices may be employed. Such permits the capture of images of a given part 310 along with distance data, allowing the generation or derivation of three-dimensional data or information regarding the part 310.

Figure 4 shows a sensor system 402 positioned to capture three-dimensional information or data regarding parts 410 which may, for example, be collocated in a bin 412, according to another embodiment.

In particular, the sensor system 402 includes at least one image capture device 414 and structured lighting 418. The structure lighting 418 may, for example, include one or more light sources 418a, 418b. Such permits the capture of images of a given part 410 from two or more different lighting perspectives, allowing the generation or derivation of three-dimensional data or information regarding the part 410.

As will be described in more detail below with reference to Figure 5, the control system 106 may take a variety of forms including one or more controllers, processors, microprocessors, digital signal processors (DSPs), application specific integrated circuits (ASICs), field programmable gate arrays (FPGAs) and associated devices and buses.

Discussion of a Suitable Computing Environment

Figure 5 and the following discussion provide a brief, general description of a suitable control system 504 in which the various illustrated embodiments can be implemented. The control system 504 may, for example, implement the control system 106 (Figure 1). Although not required, some portion of the embodiments will be described in the general context of

computer-executable instructions or logic, such as program application modules, objects, or macros being executed by a computer. Those skilled in the relevant art will appreciate that the illustrated embodiments as well as other embodiments can be practiced with other computer system configurations, including handheld devices, multiprocessor systems, microprocessor-based or programmable consumer electronics, personal computers ("PCs"), network PCs, minicomputers, mainframe computers, and the like. The embodiments can be practiced in distributed computing environments where tasks or modules are performed by remote processing devices, which are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

The control system 504 may take the form of a conventional PC, which includes a processing unit 506, a system memory 508 and a system bus 510 that couples various system components including the system memory 508 to the processing unit 506. The control system 504 will at times be referred to in the singular herein, but this is not intended to limit the embodiments to a single system, since in certain embodiments, there will be more than one system or other networked computing device involved. Non-limiting examples of commercially available systems include, but are not limited to, an 80x86 or Pentium series microprocessor from Intel Corporation, U.S.A., a PowerPC microprocessor from IBM, a Sparc microprocessor from Sun Microsystems, Inc., a PA-RISC series microprocessor from Hewlett-Packard Company, or a 68xxx series microprocessor from Motorola Corporation.

The processing unit 506 may be any logic processing unit, such as one or more central processing units (CPUs), microprocessors, digital signal processors (DSPs), application-specific integrated circuits (ASICs), field programmable gate arrays (FPGAs), etc. Unless described otherwise, the construction and operation of the various blocks shown in Figure 5 are of conventional design. As a result, such blocks need not be described in further detail herein, as they will be understood by those skilled in the relevant art.

The system bus 510 can employ any known bus structures or architectures, including a memory bus with memory controller, a peripheral bus,

and a local bus. The system memory 508 includes read-only memory ("ROM") 512 and random access memory ("RAM") 514. A basic input/output system ("BIOS") 516, which can form part of the ROM 512, contains basic routines that help transfer information between elements within the user system 104a, such as during start-up. Some embodiments may employ separate buses for data, instructions and power.

The control system 504 also includes a hard disk drive 518 for reading from and writing to a hard disk 520, and an optical disk drive 522 and a magnetic disk drive 524 for reading from and writing to removable optical disks 526 and magnetic disks 528, respectively. The optical disk 526 can be a CD or a DVD, while the magnetic disk 528 can be a magnetic floppy disk or diskette. The hard disk drive 518, optical disk drive 522 and magnetic disk drive 524 communicate with the processing unit 506 via the system bus 510. The hard disk drive 518, optical disk drive 522 and magnetic disk drive 524 may include interfaces or controllers (not shown) coupled between such drives and the system bus 510, as is known by those skilled in the relevant art. The drives 518, 522, 524, and their associated computer-readable media 520, 526, 528, provide nonvolatile storage of computer-readable instructions, data structures, program modules and other data for the user system 504. Although the depicted user system 504 employs hard disk 520, optical disk 526 and magnetic disk 528, those skilled in the relevant art will appreciate that other types of computer-readable media that can store data accessible by a computer may be employed, such as magnetic cassettes, flash memory cards, Bernoulli cartridges, RAMs, ROMs, smart cards, etc.

Program modules can be stored in the system memory 508, such as an operating system 530, one or more application programs 532, other programs or modules 534, drivers 536 and program data 538.

The application programs 532 may, for example, include pose estimation logic 532a, sensor device logic 532b, and robotic system control logic 532c. The logic 532a-532c may, for example, be stored as one or more executable instructions. As discussed in more detail below, the pose estimation logic 532a may include logic or instructions to perform initialization, training and

runtime three-dimensional pose estimation, and may include matching or registration logic. The sensor device logic 532b may include logic to operate image capture devices, range finding devices, and light sources, such as structured light sources. As discussed in more detail below, the sensor device logic 532b may also include logic to convert information captured by the image capture devices and range finding devices into two-dimensional and/or three-dimensional information or data, for example two dimension and/or three-dimensional models of objects. In particular, the sensor device logic 532b may include image processing or machine-vision logic to extract features from image data captured by one or more image capture devices 114, 214, 314a, 314b, 414 into two or three-dimensional information, data or models. The sensor device logic 532b may also include logic to convert range information captured by the range finding device 316 into three-dimensional information or models of objects. The robotic system logic may include logic to convert three-dimensional pose estimations into drive signals to control the robotic system 104 or to provide appropriate information (e.g., transformations) to suitable drivers of the robotic system 104.

The system memory 508 may also include communications programs 540, for example a server and/or a Web client or browser for permitting the user system 504 to access and exchange data with sources such as Web sites on the Internet, corporate intranets, or other networks as described below. The communications programs 540 in the depicted embodiment is markup language based, such as Hypertext Markup Language (HTML), Extensible Markup Language (XML) or Wireless Markup Language (WML), and operates with markup languages that use syntactically delimited characters added to the data of a document to represent the structure of the document. A number of servers and/or Web clients or browsers are commercially available such as those from Mozilla Corporation of California and Microsoft of Washington.

While shown in Figure 5 as being stored in the system memory 508, the operating system 530, application programs 532, other programs/modules 534, drivers 536, program data 538 and browser 540 can be

stored on the hard disk 520 of the hard disk drive 518, the optical disk 526 of the optical disk drive 522 and/or the magnetic disk 528 of the magnetic disk drive 524. A user can enter commands and information into the control system 504 through input devices such as a touch screen or keyboard 542 and/or a pointing
5 device such as a mouse 544. Other input devices can include a microphone, joystick, game pad, tablet, scanner, biometric scanning device, etc. These and other input devices are connected to the processing unit 506 through an interface 546 such as a universal serial bus ("USB") interface that couples to the system bus 510, although other interfaces such as a parallel port, a game port or a
10 wireless interface or a serial port may be used. A monitor 548 or other display device is coupled to the system bus 510 via a video interface 550, such as a video adapter. Although not shown, the control system 504 can include other output devices, such as speakers, printers, etc.

The control system 504 operates in a networked environment
15 using one or more of the logical connections 502a-502c to communicate with one or more remote computers, servers and/or devices via one or more communications channels, for example a network 514. These logical connections 502a-502c may facilitate any known method of permitting computers to communicate, such as through one or more LANs and/or WANs,
20 such as the Internet. Such networking environments are well known in wired and wireless enterprise-wide computer networks, intranets, extranets, and the Internet. Other embodiments include other types of communication networks including telecommunications networks, cellular networks, paging networks, and other mobile networks.

25 When used in a WAN networking environment, the control system 504 may include a modem 554 for establishing communications over the WAN 514. The modem 554 is shown in Figure 5 as communicatively linked between the interface 546 and the WAN 514. Additionally or alternatively, another device, such as a network interface 552a-552c, that is communicatively linked
30 to the system bus 510, may be used for establishing communications over the WAN 514. In particular, a sensor interface 522a may provide communications with a sensor system (e.g., sensor system 102 of Figure 1; sensor system 202

of Figure 2, sensor system 302 of Figure 3; sensor system 402 of Figure 4). A robot interface 552b may provide communications with a robotic system (e.g., robotic system 104 of Figure 1). A lighting interface 552c may provide communications with specific lights or a lighting system (e.g., lighting system 418 of Figure 4).

In a networked environment, program modules, application programs, or data, or portions thereof, can be stored in a server computing system (not shown). Those skilled in the relevant art will recognize that the network connections shown in Figure 5 are only some examples of ways of establishing communications between computers, and other connections may be used, including wirelessly.

For convenience, the processing unit 506, system memory 508, and interfaces 546, 552a-552c8 are illustrated as communicatively coupled to each other via the system bus 510, thereby providing connectivity between the above-described components. In alternative embodiments of the control system 504, the above-described components may be communicatively coupled in a different manner than illustrated in Figure 5. For example, one or more of the above-described components may be directly coupled to other components, or may be coupled to each other, via intermediary components (not shown). In some embodiments, system bus 510 is omitted and the components are coupled directly to each other using suitable connections.

Discussion of Exemplary Operation

Operation of an exemplary embodiment of the machine-vision based system 100 will now be described in greater detail. While reference is made throughout the following discussion to the embodiment of Figure 1, the method may be employed with the other described embodiments, as well as even other embodiments, with or without modification.

Figure 6 shows a method 600 of operating a machine-vision based system 100, according to one illustrated embodiment.

The method 600 starts at 602. The method 600 may start, for example, when power is supplied to the machine-vision based system 100 or in

response to activation by a user or by an external system, for example the robotic system 104.

At 604, the machine-vision based system 100 and in particular the sensor system 102 are calibrated in a setup mode or time. The setup mode or
5 time typically occurs before a training mode or time, and before a runtime or runtime mode. The calibration 604 may include intrinsic and/or extrinsic calibration of image capture devices 114 as well as calibration of range finding devices 316 and/or lighting 418. The calibration 604 may include any one or more of a variety of acts or operations.

10 For example, intrinsic calibration may be performed for all the image capture devices 114, and may involve the determination of the internal parameters such as focal length, image sensor center and distortion factors. An explanation of the preferred calibration algorithms and descriptions of the variables to be calculated can be found in commonly assigned U.S. Patent
15 6,816,755 issued on November 9, 2004, and pending applications Serial No. 10/634,874 and 11/183,228. The method 600 may employ any of the many other known techniques for performing the intrinsic calibration. In some embodiments, the intrinsic calibration of the image capture devices 114 may be performed before installation in the field. In such situations, the calibration data
20 is stored and provided for each image capture devices 114. It is also possible to use typical internal parameters for a specific image sensor, for example parameters associate with particular camera model-lens combinations. Where a pair of cameras 314 are in a stereo configuration, camera-to-camera calibration may be performed.

25 For example, extrinsic calibration may be preformed by determining the pose of one or more of the image capture devices 114. For example, one of the image capture devices 114 may be calibrated relative to a robotic coordinate system, while the other image capture devices 114 are not calibrated. Through extrinsic calibration the relationship (*i.e.*, three-dimensional
30 transformation) between an image sensor coordinate reference frame and an external coordinate system (*e.g.*, robotic system coordinate reference system) is determined, for example by computation. In at least one embodiment,

extrinsic calibration is performed for at least one image capture devices 114 to a preferred reference coordinate frame, typically that of the robotic system 104. An explanation of the preferred extrinsic calibration algorithms and descriptions of the variables to be calculated can be found in commonly assigned U.S.

- 5 Patent 6,816,755 issued on November 9, 2004 and pending applications Serial No. 10/634,874 and 11/183,228. The method may employ any of the many other known techniques for performing the extrinsic calibration.

Some embodiments may omit extrinsic calibration of the image capture devices 114, for example where the method 600 is employed only to
10 create a comprehensive object model without driving the robotic system 104.

At 606, the machine-vision based system is trained in a training mode or time. In particular, the machine-vision based system 100 is trained to recognize work pieces or objects, for example parts 110. Training is discussed in more detail below with reference to Figures 7-9.

- 15 At 608, the machine-vision based system 100 performs three-dimensional pose estimation in at runtime or in a runtime mode. In particular, the machine-vision based system 100 employs reference two-dimensional information or models to identify object regions in an image, and employs reference three-dimensional information or models to determine a three-
20 dimensional pose of an object represented in the object region. The three-dimensional pose of the object may be determined based on at least one of a plurality of reference three-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and
25 the runtime three-dimensional representation of the object region is not necessarily previously known. Three-dimensional pose estimation is discussed in more detail below with reference to Figures 10-15.

- Optionally, at 610 the machine-vision based system 100 drives the robotic system 102. For example, the machine-vision based system may
30 provide control signals to the robotic system or to an intermediary robotic system controller to cause the robotic system to move from one pose to another pose. The signals may, for example, encode a transformation.

The method 600 terminates at 612. The method 600 may terminate, for example, in response to a disabling of the machine-vision based system 100 by a user, the interruption of power, or an absence of parts 110 in an image of the bin 112.

5 Figure 7 shows a method of training 700 the machine-vision based system 100, according to one illustrated embodiment. Training refers to the process whereby a training, sample, or reference object (e.g., part 110) and its attributes are introduced to the machine-vision system 100. During the training process, various views of the training object are captured or acquired
10 and various landmark features are selected whose geometrical properties are determined and stored. In some embodiments the views may be stored along with sparse model information, while in other embodiments feature information extracted from the views may be stored along with the sparse model information.

15 The method 700 starts at 702, for example in response to an appropriate input by a user. The method 700 may be performed manually, automatically or a combination of manually and automatically.

 Optionally at 704, the image sensor 114 captures an image of a first view of a work piece or training object such as a part 110. As explained
20 below, some embodiments may employ existing information, for example existing digital models of the object or existing images of the object for training.

 At 706, an object region is identified, the object region including a representation of at least a portion of the training object. The object region may be identified manually or automatically or a combination of manually and
25 automatically, for example by application of one or more rules to a computer or digital model of training the object.

 At 708, the control system 504 extracts reference two-dimensional information or data, for example in the form of features. The reference two-dimensional information or data may be features that are discernable in the
30 captured image, and which are good subjects for machine-vision algorithms. Alternatively, the reference two-dimensional information or data may be features that are discernable in a two-dimensional projection of a computer

model of the training object. The features may, for example, include points, lines, edges, contours, circles, corners, centers, radia, image patches, etc. In some embodiments, the feature on the object (e.g., part 110) is an artificial feature. The artificial feature may be painted on the object or may be a decal or
5 the like affixed to the object.

The extraction 708 may include the manual identification of suitable features by a user and/or automatic identification of suitable features, for example defined features in a computer model, such as a digital model. As illustrated in Figure 8, extraction 708 of the features may include performing a
10 method 800, which accesses an existing computer digital model (e.g., computer aided design or CAD model) of the part 110 at 802. As illustrated in Figure 9, extraction 708 of the features may include performing a method 800, which employs sensed data or information, for example at 902 accessing the image captured at 704 (Figure 7).

At 710, the machine-vision based system 100 creates a reference two-dimensional model using the extracted reference two-dimensional information. The reference two-dimensional model may include information or data representative of some or all of the extracted features. For example, the reference two-dimensional model may include points defining a line, edge or
20 contour, or a point defining a center of an opening and a radius defining a perimeter of the opening. Various approaches to defining and/or storing the reference two-dimensional information representing a feature may be employed.

At 712, the control system 106 extracts reference three-dimensional information or data. For example, the control system 106 may extract reference three-dimensional information or data in the form of a reference three-dimensional point cloud (also referred to as a dense point cloud or stereo dense point cloud) for all or some of the image points in the object
25 region.

The reference three-dimensional information may be extracted in a variety of ways, at least partially based on the particular components of the machine-vision based system 100. For example, the machine-vision based
30

system 100 may employ two or more images of slightly different views of an object such as a part 110. For instance, in the embodiment of Figure 1, the machine-vision based system 100 may cause the image capture device 114 to capture an image at a first view, then cause the sensor robotic system 116 to
5 move the image capture device 114 to capture an image of the part 110 at second view, to produce a stereo pair of images. The machine-vision based system 100 may perform stereo processing on the stereo pair of images to derive and/or extract the reference three-dimensional data or information. Also for example, in the embodiment of Figure 2, the machine-vision based system
10 100 may employ the pair of cameras 214 to produce the stereo pair of images, and process the same accordingly to produce a stereo dense point cloud. Suitable packaged stereo pairs of camera 214 with suitable processing software are commercially available from a variety of sources. As a further example, in the embodiment of Figure 3 the machine-vision based system 100
15 may employ an image captured by the image capture device 314a, 314b along with range finding data or information acquired by the range finding device 316 to derive or extract the reference three-dimensional information or data. Such range finding information or data may, for example, be determined using laser triangulation or laser time of flight techniques. Such range finding information
20 or data may, for example, be determined using ultrasonic or infrared range finding techniques. As yet a further example, the machine-vision based system 100 embodiment of Figure 4 may employ two or more images captured with different lighting to derive or extract the reference three-dimensional information or data. As an even further example, the machine-vision based system 100
25 may employ a computer or digital model of the training object to extract the reference three-dimensional information or data.

At 714, the machine-vision based system 100 creates a reference three-dimensional model using the extracted reference three-dimensional information or data. The reference three-dimensional model may include three-
30 dimensional information or data representative of some or all of the extracted reference information or data. For example, the reference three-dimensional model may include a point cloud, dense point cloud, or stereo dense point

cloud. Various approaches to defining and/or storing the reference three-dimensional information or data may be employed.

At 716, the machine-vision based system 100 may store relationships between the reference two- and three-dimensional models. Such
5 may be an explicit act, or may be inherent in the way the reference two- and three-dimensional models are themselves stored. For example, the machine-vision based system 100 may store information that is indicative or reflects the relative pose between the image capture device and the object or part 110, and between the reference two- and three-dimensional models of the particular
10 view.

At 718, the machine-vision based system 100 determines whether additional views of the training object (e.g., training part) are to be trained. Typically, several views of each stable pose of an object (e.g., part 110) are desired. If so, control passes to 720, if not control passes to 722.

At 720, the machine-vision based system 100 changes the view of
15 the training object, for example changing the pose of the image capture device 114 with respect to the training object (e.g., training part). The machine-vision based system 100 may change the pose using one or more of a variety of approaches. For example, in the machine-vision based system 100 of the
20 embodiment of Figure 1, the control system 106 may cause the sensor robot system 116 to move the image capture device 114 with respect to the training part 110. Also for example, in the machine-vision based system 100 of the embodiment of Figure 3, the machine-vision based system 100 may employ multiple image capture devices 314a, 314b which are positioned at various
25 locations to provide different views of the training object (e.g., training part). Alternatively or additionally, the machine-vision based system 100 may cause the training part to be moved, for example where the training part is on a conveyor or table or is held by the end effector 104b of a robotic system 104.

At 722, the method 700 terminates. An appropriate indication
30 may be provided to a user, for example prompting the user to enter runtime or the runtime mode. Control may pass back to a calling routine or program, or may automatically or manually enter a runtime routine or program.

Figure 10 shows a method 1000 of performing three-dimensional pose estimation, according to one illustrated embodiment.

The method 1000 starts at 1002 during runtime or in the runtime mode. For example, the method 1000 may start in response to input from a user, the occurrence of the end of the method 700, or the appearance of parts 110.

At 1004, the machine-vision based system 100 captures an image of a location where one or more of the parts 110 may be present. For example, the control system 106 may cause one of the image capture devices 114, 214, 314a, 314b, 414 to capture an image of all or a portion of the parts 110.

At 1006, the machine-vision based system 100 identifies an object region of the captured image based on reference two-dimensional information or data, for example based on at least one of the reference two-dimensional models of the object created during the training mode or time. For example, the control system 106 may employ any one or more of various two-dimensional machine-vision techniques to recognize objects in the image based on the features stored as the reference two-dimensional information or data or reference two-dimensional models. Such techniques may include one or more of correlation based pattern matching, blob analysis, and/or geometric pattern matching, to name a few. Identification of an objection typically means that the object (e.g., part 110) in the object region has a similar pose relative to the sensor (e.g., image capture device 114) as the pose of the training object that produced the particular reference two-dimensional model.

At 1008, the machine-vision based system 100 identifies a corresponding one of the reference three-dimensional information or data, for example one of the reference three-dimensional models of the training object created during training. For example, the control system 106 may rely on the relationship stored 716 at of the method 700. Such may be stored as a relationship in a database, for example as a lookup table. Such may be stored as a logical connection between elements of a record or as a relationship between records of a data structure. Other approaches to storing and retrieving

or otherwise identifying the relationship will be apparent to those of skill in the computing arts.

At 1010, the machine-vision based system 100 determines a three-dimensional pose of the object (e.g., part 110) based on the reference
5 three-dimensional information or data, for example the reference three-dimensional model identified at 1008.

Optionally, at 1012 the machine-vision based system 100 determines if additional images or portions thereof will be processed. Control returns to 1004 if additional images or portions thereof will be processed.
10 Otherwise control passes to 1014, where the method 1000 terminates. Alternatively, the method 1000 may pass control directly from determining the three-dimensional pose estimation at 1010 to terminating at 1014.

Figure 11 shows a method 1100 of identifying object regions in an image, according to one illustrated embodiment, where the identified object
15 region is a region of the image that contains a representation of at least part of an object. The method 1100 may be suitable for performing the act 1006 of the method 1000 (Figure 10).

The method 1100 starts at 1102, for example called as part of executing act 1006 of the method 1000.

20 At 1104, the machine-vision based system 100 extracts two-dimensional information from a first region of a captured image.

At 1106, the machine-vision based system 100 compares the two-dimensional information or data extracted from the first region of the image to reference two-dimensional information or data (e.g., representing features)
25 such as reference two-dimensional models of the object. Figure 12 shows one method 1200 illustrating some of the types of elements that may be compared as part of the comparison 1106. In particular, the representations of edges, points, and/or image patches in a computer digital two-dimensional representation of the first region of the captured image may be compared with
30 edges, points, and/or image patches in the reference two-dimensional model at 1202. Figure 13 shows one method 1300 of illustrating some of the particular types of comparisons that may be performed on various types of elements. In

particular, the machine-vision based system 100 may computationally perform correlation based pattern matching, blob analysis and/or geometric pattern matching at 1302.

At 1108, the machine-vision based system 100 determines based
5 on the comparison whether the two-dimensional information, data or models of the first region of the captured image match the reference two-dimensional information, data or models within a defined tolerance. If so, an object region containing a representation of at least a portion of an object has been found, and control passes to 1116 where the method 1100 terminates. If not, an
10 object region has not been found and control passes to 1110.

At 1110, the machine-vision based system 100 determines whether there are further portions of the captured image to be analyzed to find object regions. If there are not further portions of the captured image to be analyzed, then the machine-vision based system 100 has determined that the
15 captured image does not include representations of the trained object. The machine-vision based system 100 provides a suitable indication of the lack of objects in the captured image at 1114 and terminates at 1116. If there are further portions of the captured image to be analyzed, control passes to 1112.

At 1112, the machine-vision based system 100 identifies a portion
20 of the captured image that has not been previously analyzed, and returns control to 1104 to repeat the process. The various acts of the method 1100 may be repeated until an object region is located or until it is determined that the captured image does not contain a representation of the object or a time out condition occurs.

25 Figure 14 shows a method 1400 of determining a three-dimensional pose estimation, according to one illustrated embodiment.

The method 1400 may be suitable for performing the act 1010 of method 1000 (Figure 10).

The method 1400 starts at 1402, for example called as part of
30 executing act 1010 of the method 1000.

At 1404, the machine-vision based system 100 extracts three-dimensional information from an object region, for example an object region

identified at 1006 of the method 1000 (Figure 10). For example, the machine-vision system 100 may determine the three-dimensional coordinates for some or all of the points in the object region.

At 1406, the machine-vision based system 100 forms a runtime
5 three-dimensional representation or model of the object region of the image. The runtime three-dimensional representation or model may, for example, take the form of a three-dimensional point cloud of the object region of the image for all or some of the points in the object region.

At 1408, the machine-vision based system 100 performs
10 registration between the reference three-dimensional model of the object region and the runtime three-dimensional representation or model of the object region of the image. Figure 15 shows a method 1500 of performing registration according to one illustrated embodiment. The method 1500 may be suitable for performing the registration 1408 of method 1400 (Figure 14). In particular, at
15 1502, the machine-vision based system 100 executes an error minimization algorithm to minimize an error between a reference three-dimensional model identified at 1008 of method 1000 (Figure 10) and runtime three-dimensional model, for example by executing an iterative closest point algorithm.

In some embodiments, at least an approximate correspondence
20 may be drawn between points in each of the reference three-dimensional models being compared. The correspondence may, for example, be based on a location where the runtime two-dimensional model is found and the stored relationship between the runtime two-dimensional model and the reference two-dimensional model. Additionally or alternatively, the approximate pose
25 determined as a result of identifying an object region (e.g., 1006 of method 1000) may be used to initialize the comparison or registration process.

At 1410, the machine-vision based system 100 determines
whether the registration is successful. If the registration is successful, the three-dimensional pose estimation has been found and control passes to 1418
30 where the method 1400 terminates. In some embodiments, the machine-vision based system 100 may provide a suitable indication regarding the found three-

dimensional pose estimation before terminating at 1418. If the registration is unsuccessful, control passes to 1412.

At 1412, the machine-vision based system 100 determines whether there are further objects regions to be analyzed and/or whether a number of iterations or amount of time is below a defined limit. If there are no further object regions to be analyzed and/or if the number of iterations or amount of time is not below a defined limit control passes to 1414. At 1414, the machine-vision based system 100 provides a suitable indication that a three-dimensional pose estimation was not found, and the method 1400 terminates at 1418.

If there are further object regions to be analyzed and/or if the number of iterations or amount of time is below a defined limit control passes to 1416. At 1416, the machine-vision based system 100 may return to find another object region of the image to analyze or process, for example returning to 1006 of method 1000 (Figure 10), the method 1400 terminating at 1418.

In the above-described various embodiments, the image capture device 114 was mounted on a member 116c of the sensor robotic system 116. In alternative embodiments, the image capture device 114 may be mounted on a portion of the robotic system 104 or mounted on a non-machine-vision based system, such as a track system, chain/pulley system or other suitable system. In other embodiments, a moveable mirror or the like may be adjustable to provide different views for a fixed image capture device 114.

In the above-described various embodiments, a plurality of images are successively captured as the image capture device 114 is moved until the pose of an object is determined. The process may end upon the robotic system 104 successfully manipulating one or more parts 110. In an alternative embodiment, the process of successively capturing a plurality of images, and the associated analysis of the image data, determination of three-dimensional pose estimates, and driving of the robotic system 104 continues until a time period expires, referred to as a cycle time or the like. The cycle time limits the amount of time that an embodiment may search for an object region of interest. In such situations, it is desirable to end the process, move

the image capture device to the start position (or a different start position), and begin the process anew. That is, upon expiration of the cycle time, the process starts over or otherwise resets.

In other embodiments, if the three-dimensional pose estimation
5 for one or more objects of interest are determined before expiration of the cycle time, the process of capturing images and analyzing captured image information continues so that other objects of interest are identified and/or their respective three-dimensional pose estimates determined. Then, after the current object of interest is engaged, the next object of interest has already
10 been identified and/or its respective three-dimensional pose estimate determined before the start of the next cycle time. Or, the identified next object of interest may be directly engaged without the start of a new cycle time.

In the above-described various embodiments, the control system
106 (Figure 1) may employ a microprocessor, a digital signal processor (DSP),
15 an application specific integrated circuit (ASIC) and/or a drive board or circuitry, along with any associated memory, such as random access memory (RAM), read only memory (ROM), electrically erasable read only memory (EEPROM), or other memory device storing instructions to control operation.

The above description of illustrated embodiments, including what
20 is described in the Abstract, is not intended to be exhaustive or to limit the invention to the precise forms disclosed. Although specific embodiments of and examples are described herein for illustrative purposes, various equivalent modifications can be made without departing from the spirit and scope of the invention, as will be recognized by those skilled in the relevant art. The
25 teachings provided herein of the invention can be applied to other object recognition systems, not necessarily the exemplary machine-vision based system embodiments generally described above.

The foregoing detailed description has set forth various
embodiments of the devices and/or processes via the use of block diagrams,
30 schematics, and examples. Insofar as such block diagrams, schematics, and examples contain one or more functions and/or operations, it will be understood by those skilled in the art that each function and/or operation within such block

diagrams, flowcharts, or examples can be implemented, individually and/or collectively, by a wide range of hardware, software, firmware, or virtually any combination thereof. In one embodiment, the present subject matter may be implemented via Application Specific Integrated Circuits (ASICs). However, 5 those skilled in the art will recognize that the embodiments disclosed herein, in whole or in part, can be equivalently implemented in standard integrated circuits, as one or more computer programs running on one or more computers (*e.g.*, as one or more programs running on one or more computer systems), as one or more programs running on one or more controllers (*e.g.*, 10 microcontrollers) as one or more programs running on one or more processors (*e.g.*, microprocessors), as firmware, or as virtually any combination thereof, and that designing the circuitry and/or writing the code for the software and or firmware would be well within the skill of one of ordinary skill in the art in light of this disclosure.

15 For convenience, the various communications paths are illustrated as hardwire connections. However, one or more of the various paths may employ other communication media, such as, but not limited to, radio frequency (RF) media, optical media, fiber optic media, or any other suitable communication media.

20 In addition, those skilled in the art will appreciate that the control mechanisms taught herein are capable of being distributed as a program product in a variety of forms, and that an illustrative embodiment applies equally regardless of the particular type of signal bearing media used to actually carry out the distribution. Examples of signal bearing media include, but are not 25 limited to, the following: recordable type media such as floppy disks, hard disk drives, CD ROMs, digital tape, and computer memory; and transmission type media such as digital and analog communication links using TDM or IP based communication links (*e.g.*, packet links).

30 These and other changes can be made to the present systems and methods in light of the above-detailed description. In general, in the following claims, the terms used should not be construed to limit the invention to the specific embodiments disclosed in the specification and the claims, but should be

construed to include all power systems and methods that read in accordance with the claims. Accordingly, the invention is not limited by the disclosure, but instead its scope is to be determined entirely by the following claims.

CLAIMS

1. A method of object pose estimation using machine-vision, comprising:
 - identifying an object region of an image on which pose estimation is being performed based on a correspondence between at least a portion of a representation of an object in the image and at least a corresponding one of a plurality of reference two-dimensional models of the object, the object region being a portion of the image that contains the representation of at least a portion of the object; and
 - determining a three-dimensional pose of the object based on at least one of a plurality of reference three-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known.
2. The method of claim 1 wherein identifying an object region of an image includes comparing each of a number of features in a digital representation of the image with a number of features in the corresponding one of the reference two-dimensional models.
3. The method of claim 2 wherein identifying an object region of an image includes comparing each of the number of features in the digital representation of the image to a number of features in successive ones of the reference two-dimensional models until a match is found within a defined tolerance or no match is found among all of the reference two-dimensional models of a set of the reference two-dimensional models.
4. The method of claim 3 wherein comparing a number of features in digital representation of the image to a number of features in successive ones of the reference two-dimensional models includes comparing

a representation of at least one of an edge, a point, or an image patch in the digital representation of the image to a representation at least one of an edge, a point, or an image patch in the reference two-dimensional model.

5. The method of claim 1 wherein identifying an object region of an image includes computationally performing at least one of correlation based pattern matching, blob analysis, or geometric pattern matching.

6. The method of claim 1, further comprising:
identifying the at least one of the plurality of reference three-dimensional models of the object based on the at least corresponding one of the reference two-dimensional models of the object.

7. The method of claim 6 wherein identifying at least one of the plurality of reference three-dimensional models of the object based on the at least corresponding one of the reference two-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known includes identifying at least one of the reference three-dimensional models based on a stored relationship between the corresponding one of the reference two-dimensional models and the at least one of the reference three-dimensional models.

8. The method of claim 7 wherein determining a three-dimensional pose of the object based on at least one of a plurality of reference three-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known includes performing a registration between the at least one of the

reference three-dimensional models and a digital runtime three-dimensional representation of the object region of the image.

9. The method of claim 8 wherein the digital runtime three-dimensional representation of the object region of the image is a runtime three-dimensional model of the object region of the image and wherein performing a registration between the at least one of the reference three-dimensional models and a digital runtime three-dimensional representation of the object region of the image includes executing an error minimization algorithm to minimize error between the at least one of the reference three-dimensional models and the runtime three-dimensional model of the object region of the image.

10. The method of claim 9 wherein executing an error minimization algorithm includes executing an iterative closest point algorithm.

11. The method of claim 8, further comprising:
extracting three-dimensional information from the object region of the image.

12. The method of claim 11, further comprising:
forming the runtime three-dimensional model of the object region of the image from the runtime three-dimensional information extracted from the object region of the image.

13. The method of claim 8, further comprising:
providing an indication that the three-dimensional pose of the object has not been found if an outcome of the registration is unsuccessful.

14. The method of claim 1 wherein determining a three-dimensional pose of the object based on at least one of a plurality of reference three-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between

the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known includes performing a registration between a set of dense three-dimensional data stored as the at least one reference three-dimensional model and a set of dense three-dimensional data stored as the runtime three-dimensional representation of the object region of the image.

15. The method of claim 1, further comprising:
capturing the image.

16. The method of claim 15 wherein the capturing an image, the identifying an object region of the image and the determining a three-dimensional pose of the object all occur during a runtime mode that follows a training mode.

17. The method of claim 16, further comprising:
acquiring the reference two-dimensional models during the training mode.

18. The method of claim 17 wherein acquiring the reference two-dimensional models during the training mode includes at least one of accessing an existing computer model of the object or sensing data from a representative object.

19. The method of claim 17, further comprising:
acquiring the reference three-dimensional models during the training mode.

20. The method of claim 19 wherein acquiring the reference three-dimensional models during the training mode includes acquiring information using at least one of a dense stereo sensor system, a laser triangulation system, a laser time of flight system or an ultrasound transducer.

21. The method of claim 19 wherein acquiring the reference three-dimensional models during the training mode includes identifying a portion of the image that contains a digital representation of at least part of the object.

22. The method of claim 21 wherein identifying a portion of the image that contains a digital representation of at least part of the object is performed either manually or automatically.

23. A computer readable medium that stores instructions for causing a computer to perform object pose estimation using machine-vision, by:
identifying an object region of an image based on a correspondence between at least a portion of a representation of an object in the object region of the image and at least a corresponding one of a plurality of reference two-dimensional models of the object, the object region being a portion of the image that contains the representation of at least a portion of the object; and

determining a three-dimensional pose of the object based on at least one of a plurality of reference three-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known.

24. The computer readable medium of claim 23 wherein the instructions cause the computer to perform object pose estimation using machine-vision, further by:

extracting two-dimensional information from the object region of the image during a runtime to form a runtime two-dimensional digital model of the object region of the image.

25. The computer readable medium of claim 24 wherein identifying an object region of an image includes comparing each of the number of features in the runtime two-dimensional digital model of the object region of the image to a number of features in successive ones of the reference two-dimensional models until a match within a defined tolerance is found or no match is found among all of the reference two-dimensional models of a set of the reference two-dimensional models.

26. The computer readable medium of claim 24 wherein the instructions cause the computer to perform object pose estimation using machine-vision, further by:

extracting three-dimensional information from the object region of the image during the runtime to form a runtime three-dimensional model of the object region of the image.

27. The computer readable medium of claim 26 wherein the instructions cause the computer to perform object pose estimation using machine-vision, further by:

identifying the at least one of the plurality of reference three-dimensional models of the object based on a stored relationship between the corresponding one of the reference two-dimensional models and the at least one of the reference three-dimensional models.

28. The computer readable medium of claim 26 wherein determining a three-dimensional pose of the object based on at least one of a plurality of reference three-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known includes performing a registration between the at least one of the reference three-dimensional models and the runtime three-dimensional model of the object region of the image.

29. The computer readable medium of claim 23 wherein the identifying an object region of the image and the determining a three-dimensional pose of the object all occur during a runtime mode following a training mode.

30. The computer readable medium of claim 29 wherein the instructions cause the computer to perform object pose estimation using machine-vision, further by:

acquiring the reference two-dimensional models during the training mode; and

acquiring the reference three-dimensional models during the training mode.

31. A system to perform three-dimensional pose estimation, the system comprising:

at least one sensor;

at least one processor; and

at least one memory storing processor executable instructions that cause the at least one processor to segment an image captured by the at least one sensor into a number of object regions based at least in part on a correspondence between at least a portion of a representation of an object in the object region of the image and at least a corresponding one of a plurality of reference two-dimensional models of the object and to cause the at least one processor to determine a three-dimensional pose of the object based on at least one of a plurality of reference three-dimensional models of the object that is related to the corresponding one of the plurality of reference two-dimensional models of the object and a runtime three-dimensional representation of the object region where a point-to-point relationship between the reference three-dimensional models of the object and the runtime three-dimensional representation of the object region is not necessarily previously known.

32. The system of claim 31 wherein the instructions further cause at least one processor to identify the at least one of the plurality of reference three-dimensional models of the object based on a set of stored relationships between a plurality of reference two-dimensional models and the plurality of reference three-dimensional models of the object.

33. The system of claim 31 wherein the at least one sensor includes at least one of an imager mounted for movement, a stereo pair of cameras, and a laser.

34. The system of claim 31 wherein the at least one sensor includes at least one imager mounted for movement with respect to the object.

35. The system of claim 31 wherein the at least one sensor includes at least one imager and at least one of a laser or a set of structured lighting.

36. The system of claim 31 wherein the at least one sensor includes at least one stereo pair of cameras.

37. The system of claim 31 wherein the instructions further cause the at least one processor to provide drive signals to drive a robotic member based on the determined three-dimensional pose estimation.

1/10

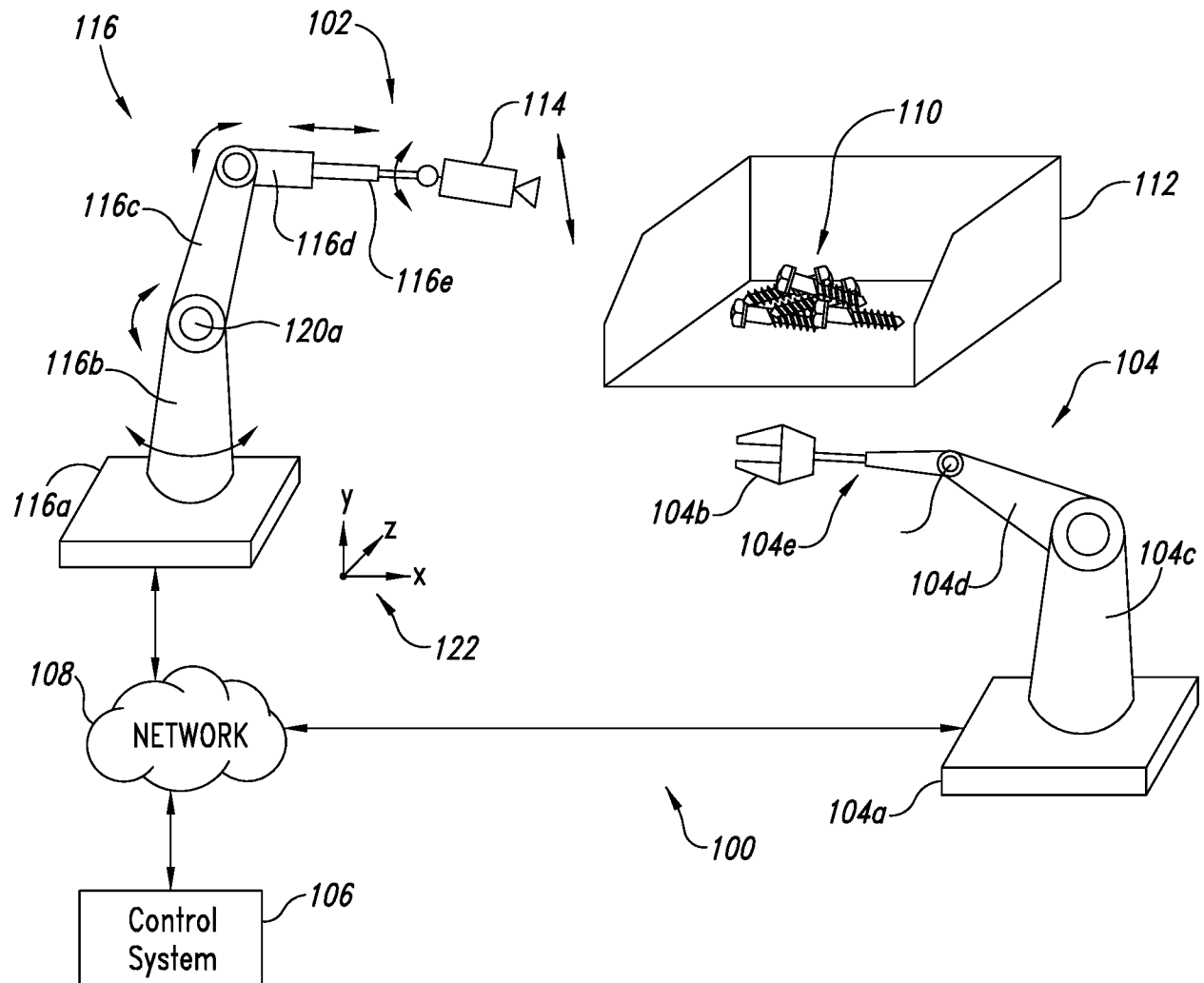


FIG. 1

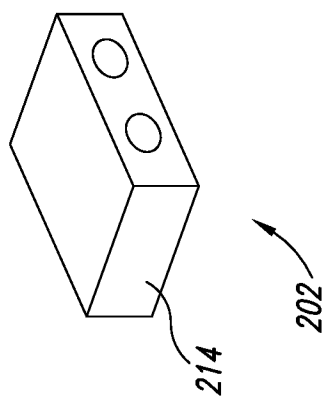
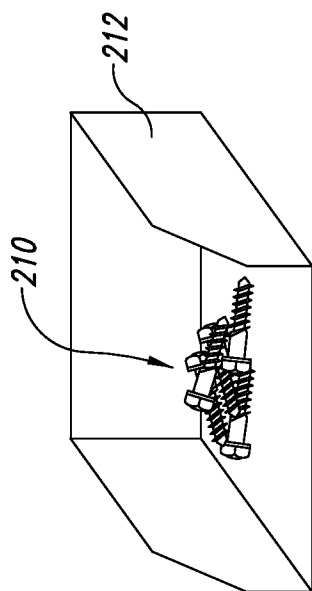


FIG. 2



2/10

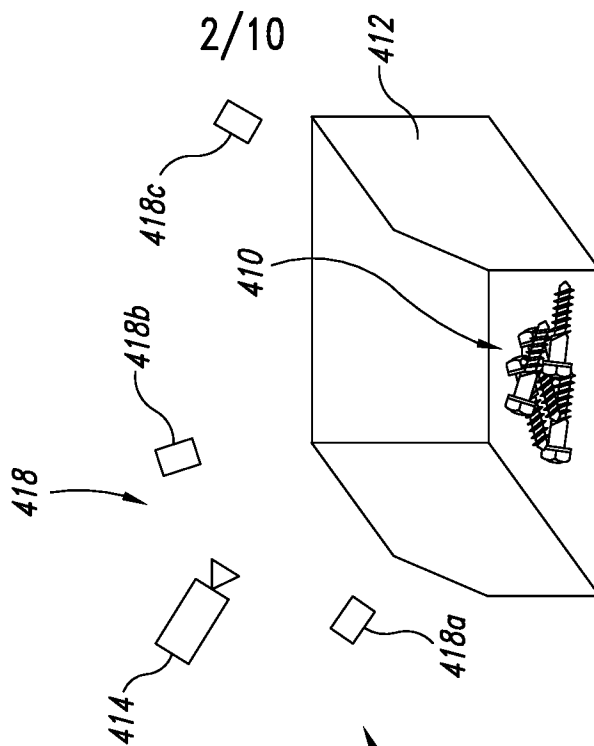


FIG. 4

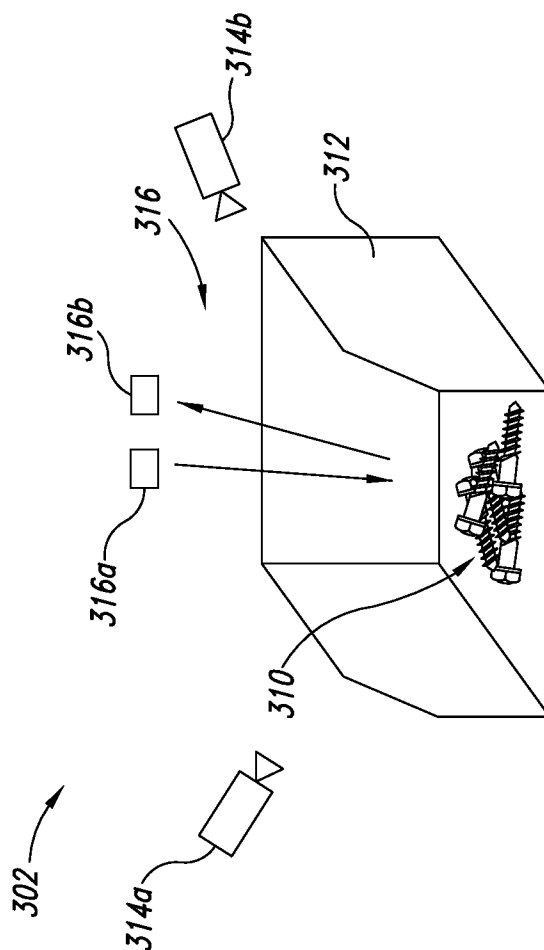


FIG. 3

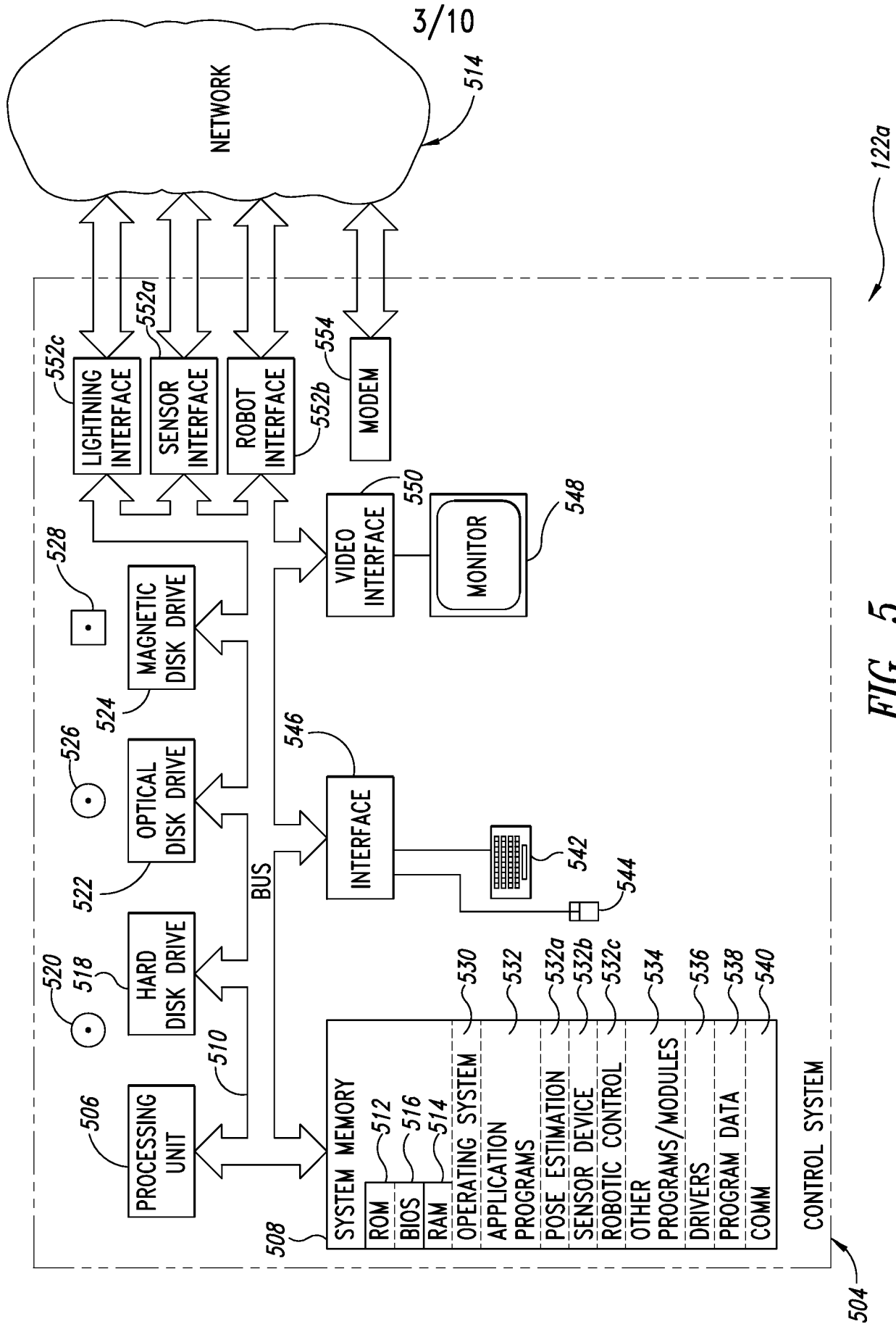
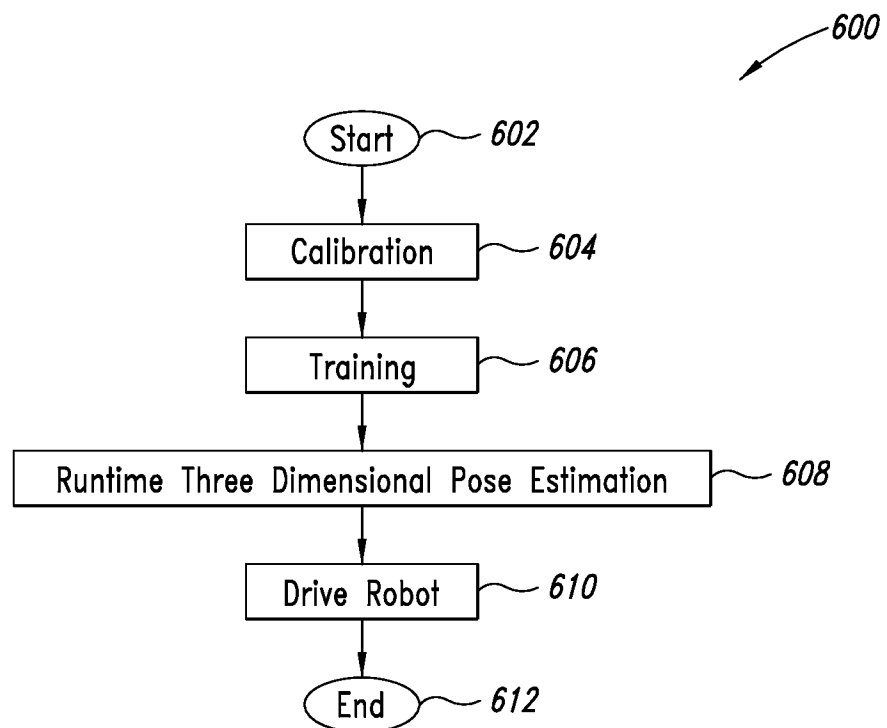


FIG. 5

4/10

*FIG. 6*

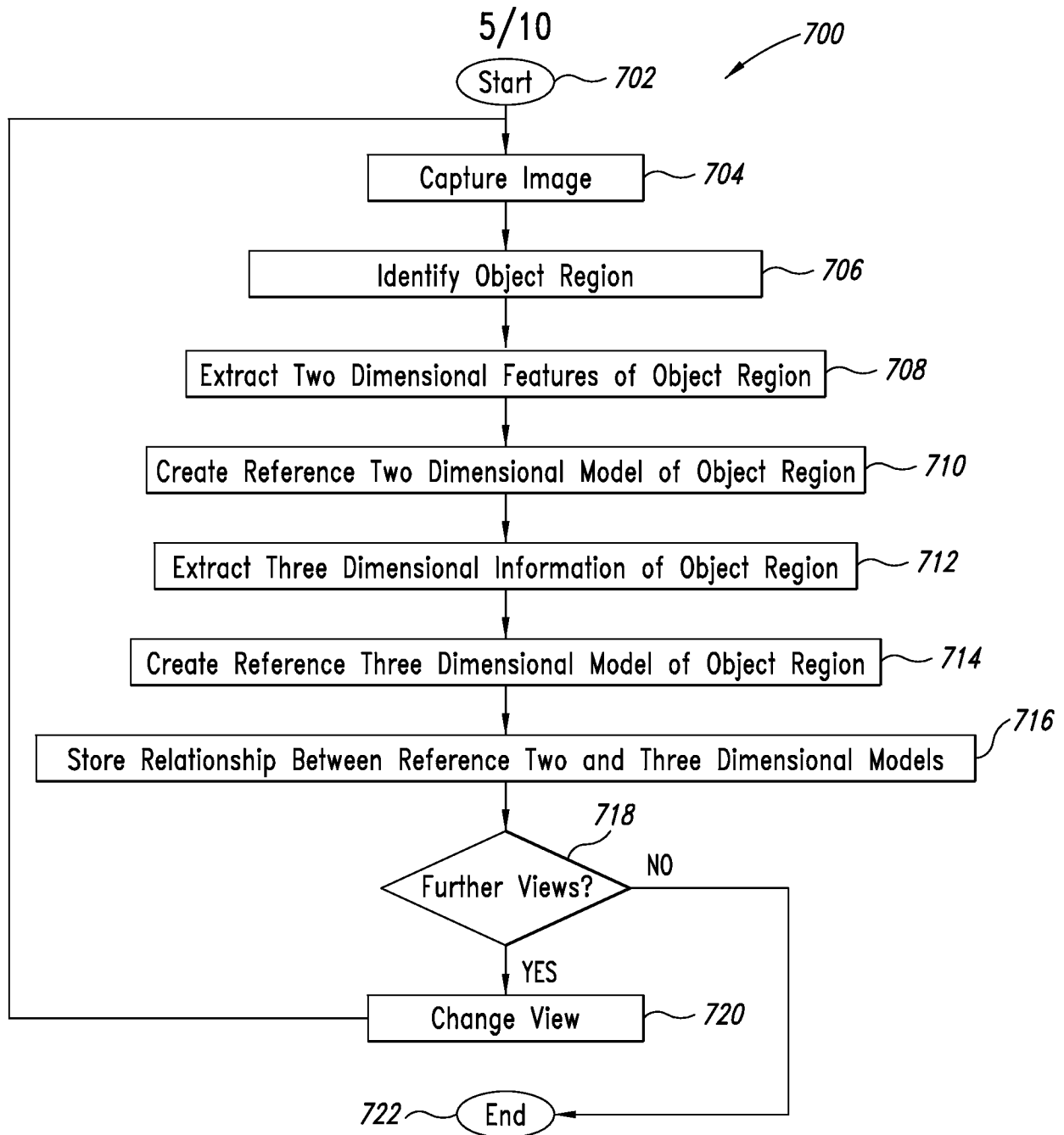


FIG. 7

800

802 Access Existing Computer Model (CAD) Of Object

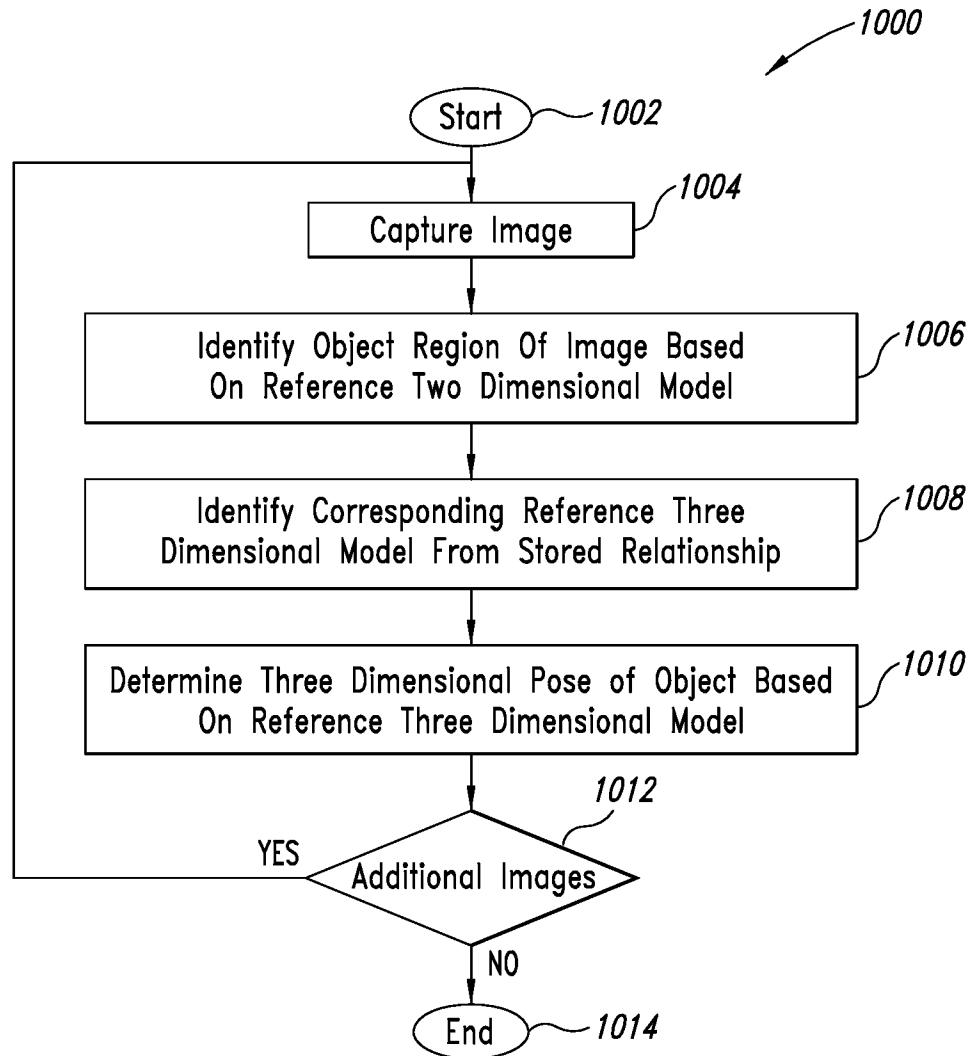
FIG. 8

900

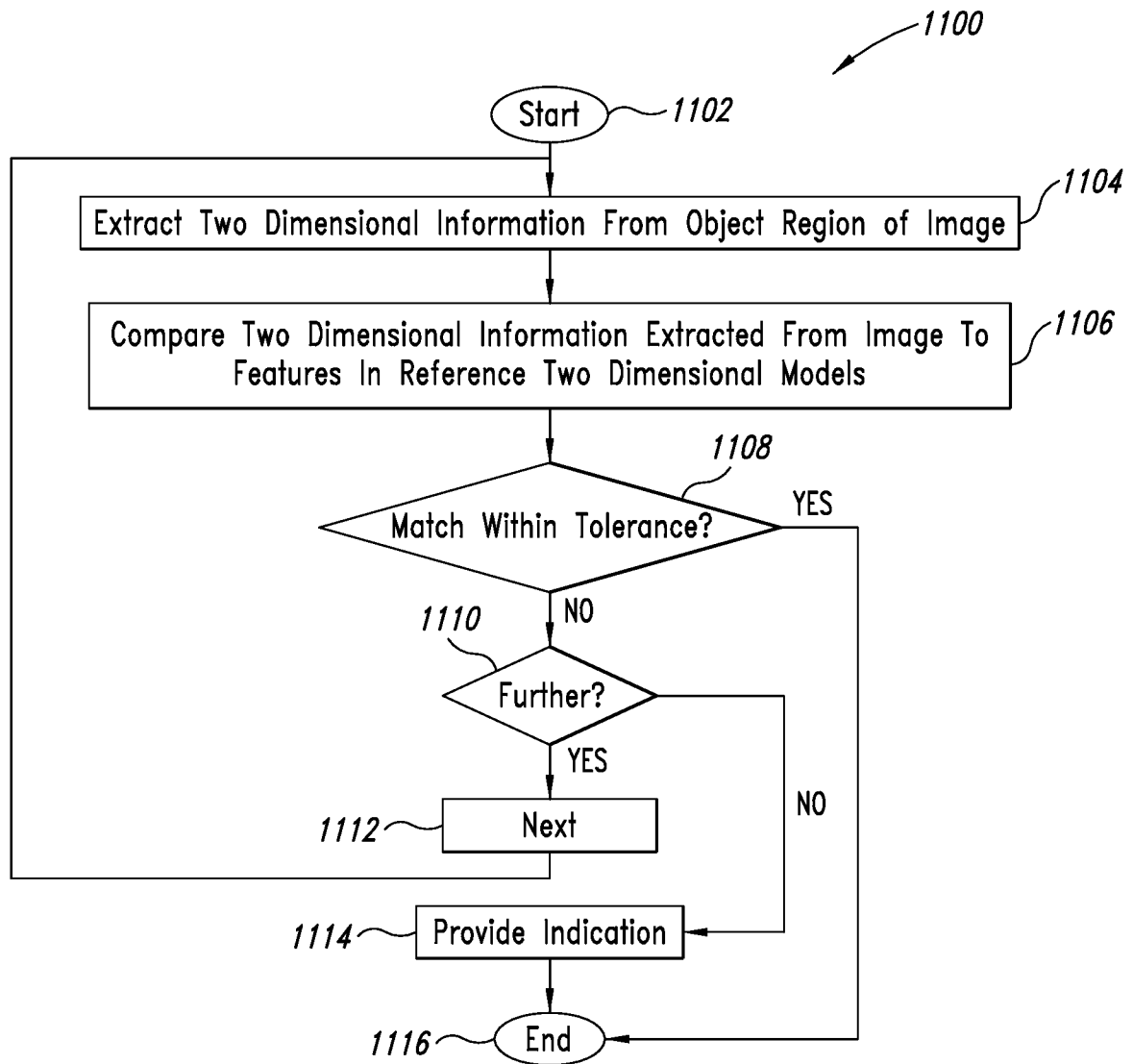
902 Sense Data From Representative or Training Object

FIG. 9

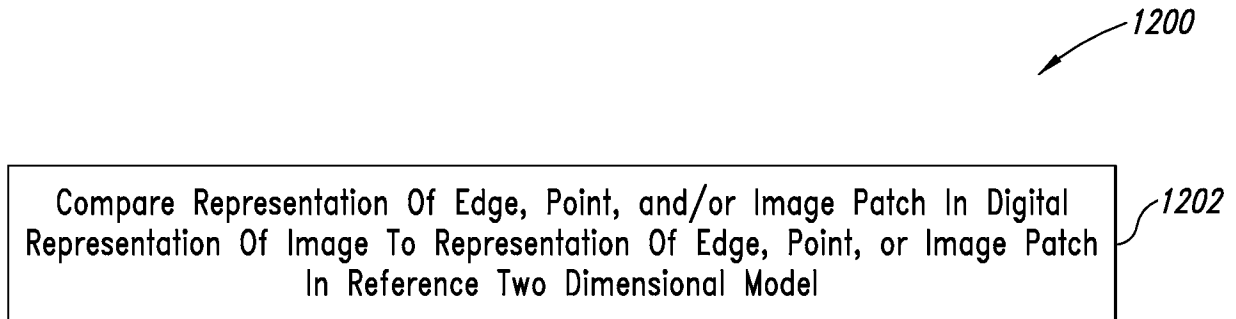
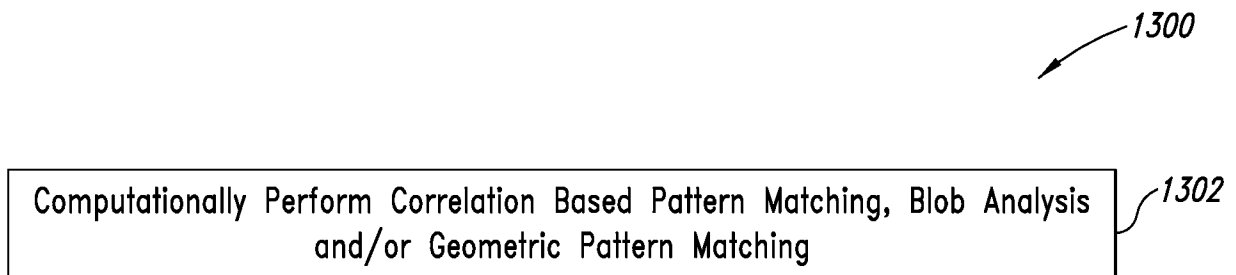
6/10

*FIG. 10*

7/10

*FIG. 11*

8/10

*FIG. 12**FIG. 13*

9/10

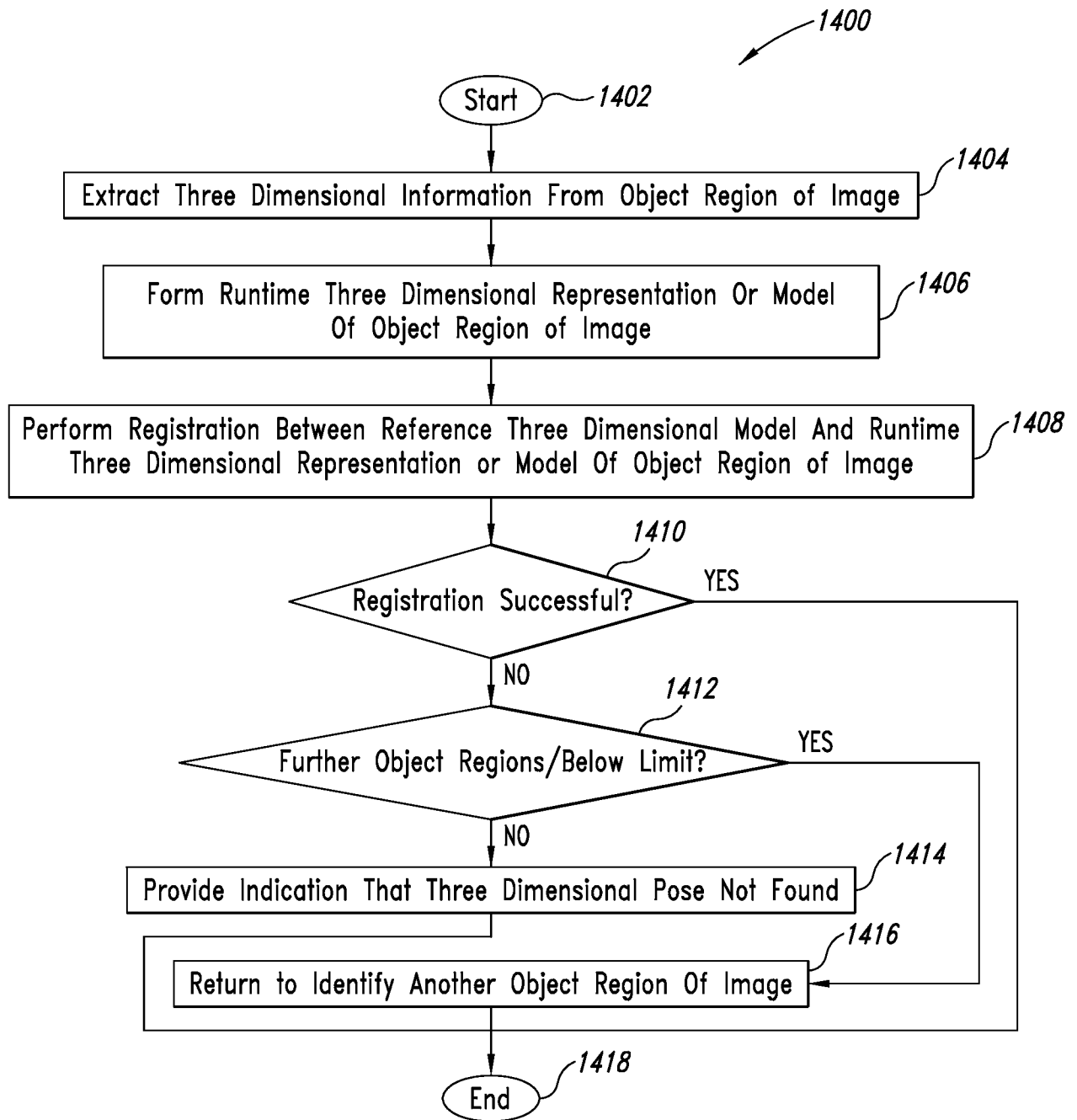
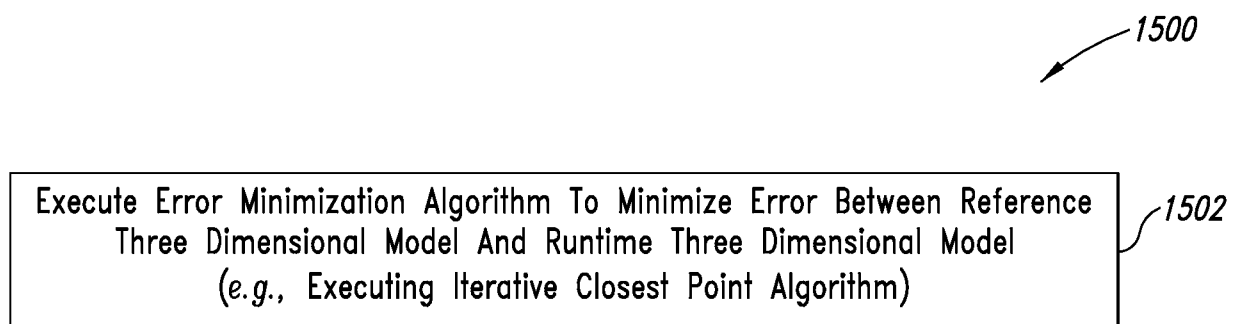


FIG. 14

10/10

*FIG. 15*