



- (51) International Patent Classification:  
G06K 9/00 (2006.01) H04N 13/02 (2006.01)
- (21) International Application Number:  
PCT/US2021/031514
- (22) International Filing Date:  
10 May 2021 (10.05.2021)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
63/023,089 11 May 2020 (11.05.2020) US
- (71) Applicant: **MAGIC LEAP, INC.** [US/US]; 7500 West Sunrise Boulevard, Plantation, FL 33322 (US).
- (72) Inventors: **ZHOU, Lipu**; 455 Brahms Way, Apt. 315, Sunnyvale, CA 94087 (US). **STEINBRUECKER, Frank Thomas**; 1335 Montecito Avenue, Apt. 26, Mountain View, CA 94043 (US). **SWAMINATHAN, Ashwin**; 4327 Table Mountain Road, Dublin, CA 94568 (US). **JU, Hui**;

665 N. San Marcos Dr., Mountain House, CA 95391 (US). **KOPPEL, Daniel Esteban**; 1431 Kingman Avenue, San Jose, CA 95128 (US). **ZAMPOGIANNIS, Konstantinos**; 1 Vista Montana, Apt. 4312, San Jose, CA 95134 (US). **MEHTA, Pooja Piyush**; 1624 Hope Drive, Apt. 631, Santa Clara, CA 95054 (US). **BALAKUMAR, Vinayram**; 2125 Bryant Street, Apt. 305, San Francisco, CA 94110 (US).

(74) Agent: **LI, Lin** et al.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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,

(54) Title: COMPUTATIONALLY EFFICIENT METHOD FOR COMPUTING A COMPOSITE REPRESENTATION OF A 3D ENVIRONMENT

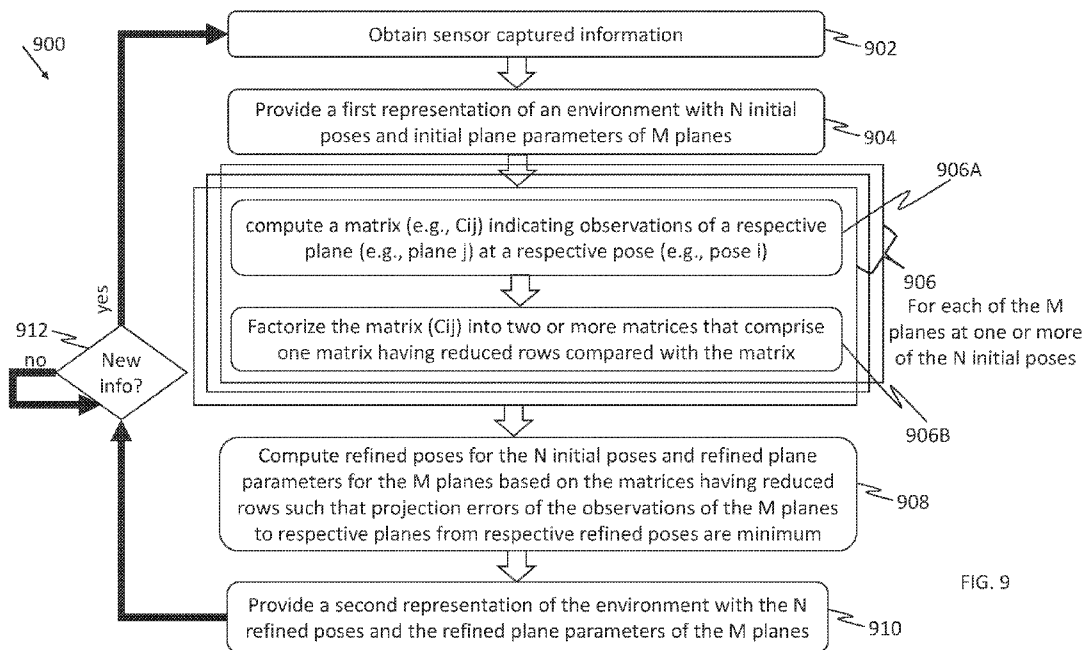


FIG. 9

(57) Abstract: Methods and apparatus for providing a representation of an environment, for example, in an XR system, and any suitable computer vision and robotics applications. A representation of an environment may include one or more planar features. The representation of the environment may be provided by jointly optimizing plane parameters of the planar features and sensor poses that the planar features are observed at. The joint optimization may be based on a reduced matrix and a reduced residual vector in lieu of the Jacobian matrix and the original residual vector.



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))*

# COMPUTATIONALLY EFFICIENT METHOD FOR COMPUTING A COMPOSITE REPRESENTATION OF A 3D ENVIRONMENT

## RELATED APPLICATIONS

[0001] This application claims priority to and the benefit of U.S. Provisional Patent Application Serial No. 63/023,089, filed May 11, 2020 and entitled “COMPUTATIONALLY EFFICIENT METHOD FOR COMPUTING A COMPOSITE REPRESENTATION OF A 3D ENVIRONMENT,” which is hereby incorporated herein by reference in its entirety.

## TECHNICAL FIELD

[0002] This application relates generally to providing representations of environments, for example, in cross reality (XR) systems, autonomous vehicles, or other computer vision systems that include movable sensors.

## BACKGROUND

[0003] Systems that use sensors to acquire information about a 3D environment are used in multiple contexts, such as cross reality (XR) systems or autonomous vehicles. These systems may employ sensors, such as cameras, to acquire information about the 3D environment. The sensors may provide many observations of the environment that may be integrated into a representation of the environment. For example, as an autonomous vehicle moves, its sensors may acquire images of the 3D environment from different poses. Each image may provide another observation of the environment, providing more information about a portion of the environment previously image or information about a new portion of the environment.

[0004] Such a system may assemble the information acquired over multiple observations into a composite representation of the 3D environment. New images may be added to the composite representation of the environment, filling in information based on the pose from which that image was acquired. Initially, the pose of an image may be estimated, such as based on outputs of internal sensors that indicate motion of the sensors acquiring the images or correlation of features relative to features in prior images that have already been incorporated into the

composite representation of the environment. Over time, however, errors in these estimation techniques may accumulate.

**[0005]** To compensate for accumulated errors, an adjustment of the composite representation may be performed from time to time. Such an adjustment may entail adjusting the poses of the images that have been combined into the composite representation such that features in different images representing the same objects in the 3D environment better align. In some scenarios, the features that are aligned may be feature points. In other scenarios, the features may be planes that are identified based on the contents of the images. In such a scenario, the adjustment may entail adjusting both the relative pose of each image and the relative position of planes detected based on the composite representation.

**[0006]** Such an adjustment may provide an accurate representation of the 3D environment that may be used in any of multiple ways. In an XR system, for example, computers may control human user interfaces to create a cross reality (XR) environment in which some or all of the XR environment, as perceived by the user, is generated by the computer. These XR environments may be virtual reality (VR), augmented reality (AR), and mixed reality (MR) environments, in which some or all of an XR environment may be generated by computers using, in part, data that describes the environment. This data may describe, for example, virtual objects that may be rendered in a way that users sense or perceive as a part of a physical world such that users can interact with the virtual objects. The user may experience these virtual objects as a result of the data being rendered and presented through a user interface device, such as, for example, a head-mounted display device. The data may be displayed to the user to see, or may control audio that is played for the user to hear, or may control a tactile (or haptic) interface, enabling the user to experience touch sensations that the user senses or perceives as feeling the virtual object.

**[0007]** XR systems may be useful for many applications, spanning the fields of scientific visualization, medical training, engineering design and prototyping, tele-manipulation and tele-presence, and personal entertainment. AR and MR, in contrast to VR, include one or more virtual objects in relation to real objects of the physical world. The experience of virtual objects interacting with real objects greatly enhances the user's enjoyment in using the XR system, and also opens the door for a variety of applications that present realistic and readily understandable information about how the physical world might be altered.

**[0008]** To realistically render virtual content, an XR system may build a representation of the physical world around a user of the system. This representation, for example, may be constructed by processing images acquired with sensors on a wearable device that forms a part of

the XR system. In such a system, a user might perform an initialization routine by looking around a room or other physical environment in which the user intends to use the XR system until the system acquires sufficient information to construct a representation of that environment. As the system operates and the user moves around the environment or to other environments, the sensors on the wearable devices might acquire additional information to expand or update the representation of the physical world.

#### BRIEF SUMMARY

**[0009]** Aspects of the present application relate to methods and apparatus for providing representations of environments, for example, in cross reality (XR) systems, autonomous vehicles, or other computer vision systems that include movable sensors. Techniques as described herein may be used together, separately, or in any suitable combination.

**[0010]** Some embodiments relate to a method of operating a computing system to generate a representation of an environment. The method comprises obtaining sensor captured information, the sensor captured information comprising a first number of images; providing an initial representation of the environment, the initial representation comprising the first number of initial poses and initial parameters of a second number of planar features based at least in part on the first number of images; for each of the second number of planar features at each pose corresponding to an image comprising one or more observations of the planar feature, computing a matrix indicating the one or more observations of the planar feature, and factorizing the matrix into two or more matrices, the two or more matrices comprising one matrix having reduced rows compared with the matrix; and computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having reduced rows. The representation of the environment comprises the first number of refined poses and the refined parameters of the second number of planar features.

**[0011]** Some embodiments relate to a method of operating a computing system to generate a representation of an environment. The method comprises obtaining sensor captured information, the sensor captured information comprising a first number of images; providing an initial representation of the environment, the initial representation comprising the first number of initial poses and initial parameters of a second number of planar features based at least in part on the first number of images; for each of the second number of planar features at each pose corresponding to an image comprising one or more observations of the planar feature, computing

a matrix having a third number of rows, the third number being smaller than the number of the one or more observations of the planar feature; and computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having the third number of rows. The representation of the environment comprises the first number of refined poses and the refined parameters of the second number of planar features.

**[0012]** The foregoing summary is provided by way of illustration and is not intended to be limiting.

### BRIEF DESCRIPTION OF THE DRAWINGS

**[0013]** The accompanying drawings are not intended to be drawn to scale. In the drawings, each identical or nearly identical component that is illustrated in various figures is represented by a like numeral. For purposes of clarity, not every component may be labeled in every drawing. In the drawings:

**[0014]** Figure 1 is a sketch illustrating an example of a simplified augmented reality (AR) scene, according to some embodiments.

**[0015]** Figure 2 is a sketch of an exemplary simplified AR scene, showing exemplary use cases of an XR system, according to some embodiments.

**[0016]** Figure 3 is a schematic diagram illustrating data flow for a single user in an AR system configured to provide an experience to the user of AR content interacting with a physical world, according to some embodiments.

**[0017]** Figure 4 is a schematic diagram illustrating an exemplary AR display system, displaying virtual content for a single user, according to some embodiments.

**[0018]** Figure 5A is a schematic diagram illustrating a user wearing an AR display system rendering AR content as the user moves through a physical world environment, according to some embodiments.

**[0019]** Figure 5B is a schematic diagram illustrating a viewing optics assembly and attendant components, according to some embodiments.

**[0020]** Figure 6A is a schematic diagram illustrating an AR system using a world reconstruction system, according to some embodiments.

[0021] Figure 6B is a schematic diagram illustrating components of an AR system that maintain a model of a passable world, according to some embodiments.

[0022] Figure 7 is a schematic illustration of a tracking map formed by a device traversing a path through a physical world, according to some embodiments.

[0023] Figure 8A is a schematic illustrating observations of various planes at different poses, according to some embodiments.

[0024] Figure 8B is a schematic illustrating geometric entities involved in computing refined poses and plane parameters of a representation of an environment comprising the planes of Figure 8A observed at the poses of Figure 8A, according to some embodiments.

[0025] Figure 9 is a flow chart illustrating a method of providing a representation of an environment, according to some embodiments.

[0026] Figure 10 is a schematic diagram illustrating a portion of a representation of an environment computed based on sensor captured information using the method of Figure 9, according to some embodiments.

#### DETAILED DESCRIPTION

[0027] Described herein are methods and apparatus for providing a representation of an environment, for example, in an XR system, and any suitable computer vision and robotics applications. The inventors have recognized and appreciated methods and apparatus that provide representations of potentially complex environments, such as rooms with many objects in them, with reduced time and reduced computational cost and memory usage. In some embodiments, an accurate representation of an environment may be provided by computational techniques that reduce the amount of information processed in adjusting a composite representation of a 3D environment to reduce errors accumulated in building the composite representation from multiple observations over time. These techniques may be applied in systems that represent the environment by planes, detected in a representation of the environment constructed by combining images of the environment acquired at multiple poses. Such techniques may be based on a reduced matrix and a reduced residual vector in lieu of a Jacobian matrix and an original residual vector.

[0028] Computationally less burdensome processing and or lower memory usage may be used

in a system to provide for an accurate representation of the 3D environment with lower latency, lower power consumption, less heat generation, lower weight, or other advantages. An XR system is an example of a system that may be improved with such a technique. With low computations complexity, for example, a user device for the XR system may be formed by programming on a smartphone.

**[0029]** In some embodiments, to provide realistic XR experiences to multiple users, an XR system must know the users' physical surroundings in order to correctly correlate locations of virtual objects in relation to real objects. An XR system may build a representation of an environment, in which virtual objects may be displayed. The representation of the environment may be created from information collected with sensors that are part of XR devices of the XR system. An XR device may be a head mounted device with integrated displays and sensors, a handheld mobile device (e.g., smartphone, smartwatch, tablet computing device, etc.), or other device having sensors. However, techniques as described herein may be used on other types of devices having sensors, such as autonomous machines with sensors. Such devices may include one or more types of sensors including, for example, image sensors, LiDAR cameras, RGBD cameras, infrared cameras, structured light sensors, ultrasound sensors, or coherent light sensors. These devices may include one or more components that aid in collecting sensor information, such as infrared emitters, ultrasound emitters, or structured light emitters. Alternatively or additionally, such devices may include sensors that aid in determining pose, such as gyroscopic sensors, accelerometers, magnetometers, altimeters, proximity sensors, and GPS sensors.

**[0030]** The representation of the environment, in some embodiments, may be a local map of the environment surrounding an XR device, and created by the XR device by integrating information from one or more images collected as the device operates. In some embodiments, the coordinate system of a local map may be tied to the position and/or orientation of the device when the device first initiates scanning the environment (e.g. starts a new session). That position and/or orientation of the device may change from session to session.

**[0031]** A local map may include sparse information, representing the environment based on a subset of features detected in sensor captured information used in forming the map. Additionally or alternatively, a local map may include dense information, representing the environment with information of surfaces in the environment such as meshes.

**[0032]** The subset of features in a map may include one or more types of features. In some embodiments, the subset of features may include point features such as table corners, which may be detected based on visual information. In some embodiments, the system may alternatively or

additionally form a representation of a 3D environment based on higher level features, such as planar features such as planes. Planes, for example, may be detected by processing depth information. A device may store planar features instead of or in addition to feature points, but in some embodiments, storing planar features in lieu of corresponding point features may reduce the size of the map. For example, a plane may be stored in a map represented with a plane normal and a signed distance to the origin of the coordinate system of the map. In contrast, the same structure might be stored as multiple feature points.

**[0033]** In addition to potentially supplying input to create maps, sensor captured information may be used to track devices' motions in an environment. The tracking may enable an XR system to localize XR devices by estimating respective devices' poses relative to the frame of reference established by a map. Localizing an XR device may require making comparisons to find a match between a set of measurements extracted from images captured by the XR device and a set of features stored in an existing map.

**[0034]** This co-dependence between the creation of a map and the localization of a device constitutes significant challenges. Even with the simplification of using planes to represent surfaces in the 3D environment, substantial processing may be required to accurately create the map and localize the device simultaneously. The processing must be accomplished quickly as devices move in the environment. On the other hand, an XR device might have limited computational resources so that the device can move with reasonable flexibility at a reasonable speed in the environment.

**[0035]** That processing may include jointly optimizing planar feature parameters and sensor poses used in forming the map of the 3D environment in which planes are detected. Conventional techniques for jointly optimizing planar feature parameters and poses may incur high computational cost and memory consumption. Because a sensor can record different observations of a planar feature at different poses, joint optimization according to a conventional approach may require solution of a very large nonlinear least-squares problem for even for a small-scale workspace.

**[0036]** As used herein, "optimizing" (and similar terms) need not result in a perfect or theoretically best solution. Rather, optimization may result from processing that reduces a measure of error in a solution. Processing to reduce error may be performed until a solution with sufficiently low error is reached. Such processing may be performed iteratively, and iterations may be performed until an exit criteria, indicative a sufficiently low error, is detected. The exit criteria may be an indication that processing has converged to a solution which might be detected

by a percentage reduction in error from iteration to iteration that is below a threshold and/or may be a predetermined number of iterations and/or other criteria that indicate a solution with low error has been identified.

**[0037]** Techniques for efficient map optimization are described herein, with an XR system being used as an example of a system that may optimize a map with these techniques. In some embodiments, information about each plane in a map is collected in each of multiple images acquired from a respective pose. An observation of each plane may be made at each pose at which an image used in creating the 3D representation was captured. This information that indicates the one or more observations of the planes may be formatted as a matrix. In a conventional approach to optimizing the map, this matrix may be a Jacobean and optimization may be based on computations performed using this Jacobean matrix.

**[0038]** In accordance with some embodiments, optimization may be performed based on factorized matrices, which may be mathematically related to a matrix as used in the conventional approach, but which may have reduced rows compared with the matrix. The XR system may compute refined parameters of the planar features and refined poses based on one of the factorized matrices having reduced rows, which reduces computational cost and memory storage. The refined parameters of the planar features and refined poses may be computed such that projection errors of the observations of the planar features to respective planar features from respective refined poses are minimum.

**[0039]** Techniques as described herein may be used together or separately with many types of devices and for many types of environments, including wearable or portable or autonomous devices with limited computational resources. In some embodiments, the techniques may be implemented by one or more services that form a portion of an XR system.

**[0040]** *Exemplary System*

**[0041]** Figures 1 and 2 illustrate scenes with virtual content displayed in conjunction with a portion of the physical world. For purposes of illustration, an AR system is used as an example of an XR system. Figures 3-6B illustrate an exemplary AR system, including one or more processors, memory, sensors and user interfaces that may operate according to the techniques described herein.

**[0042]** Referring to **Figure 1**, an outdoor AR scene 354 is depicted in which a user of an AR technology sees a physical world park-like setting 356, featuring people, trees, buildings in the background, and a concrete platform 358. In addition to these items, the user of the AR

technology also perceives that they "see" a robot statue 357 standing upon the physical world concrete platform 358, and a cartoon-like avatar character 352 flying by which seems to be a personification of a bumble bee, even though these elements (e.g., the avatar character 352, and the robot statue 357) do not exist in the physical world. Due to the extreme complexity of the human visual perception and nervous system, it is challenging to produce an AR technology that facilitates a comfortable, natural-feeling, rich presentation of virtual image elements amongst other virtual or physical world imagery elements.

**[0043]** Such an AR scene may be achieved with a system that builds maps of the physical world based on tracking information, enable users to place AR content in the physical world, determine locations in the maps of the physical world where AR content are placed, preserve the AR scenes such that the placed AR content can be reloaded to display in the physical world during, for example, a different AR experience session, and enable multiple users to share an AR experience. The system may build and update a digital representation of the physical world surfaces around the user. This representation may be used to render virtual content so as to appear fully or partially occluded by physical objects between the user and the rendered location of the virtual content, to place virtual objects, in physics based interactions, and for virtual character path planning and navigation, or for other operations in which information about the physical world is used.

**[0044]** **Figure 2** depicts another example of an indoor AR scene 400, showing exemplary use cases of an XR system, according to some embodiments. The exemplary scene 400 is a living room having walls, a bookshelf on one side of a wall, a floor lamp at a corner of the room, a floor, a sofa, and coffee table on the floor. In addition to these physical items, the user of the AR technology also perceives virtual objects such as images on the wall behind the sofa, birds flying through the door, a deer peeking out from the book shelf, and a decoration in the form of a windmill placed on the coffee table.

**[0045]** For the images on the wall, the AR technology requires information about not only surfaces of the wall but also objects and surfaces in the room such as lamp shape, which are occluding the images to render the virtual objects correctly. For the flying birds, the AR technology requires information about all the objects and surfaces around the room for rendering the birds with realistic physics to avoid the objects and surfaces or bounce off them if the birds collide. For the deer, the AR technology requires information about the surfaces such as the floor or coffee table to compute where to place the deer. For the windmill, the system may identify that is an object separate from the table and may determine that it is movable, whereas

corners of shelves or corners of the wall may be determined to be stationary. Such a distinction may be used in determinations as to which portions of the scene are used or updated in each of various operations.

**[0046]** The virtual objects may be placed in a previous AR experience session. When new AR experience sessions start in the living room, the AR technology requires the virtual objects being accurately displayed at the locations previously placed and realistically visible from different viewpoints. For example, the windmill should be displayed as standing on the books rather than drifting above the table at a different location without the books. Such drifting may happen if the locations of the users of the new AR experience sessions are not accurately localized in the living room. As another example, if a user is viewing the windmill from a viewpoint different from the viewpoint when the windmill was placed, the AR technology requires corresponding sides of the windmill being displayed.

**[0047]** A scene may be presented to the user via a system that includes multiple components, including a user interface that can stimulate one or more user senses, such as sight, sound, and/or touch. In addition, the system may include one or more sensors that may measure parameters of the physical portions of the scene, including position and/or motion of the user within the physical portions of the scene. Further, the system may include one or more computing devices, with associated computer hardware, such as memory. These components may be integrated into a single device or may be distributed across multiple interconnected devices. In some embodiments, some or all of these components may be integrated into a wearable device.

**[0048]** **Figure 3** depicts an AR system 502 configured to provide an experience of AR contents interacting with a physical world 506, according to some embodiments. The AR system 502 may include a display 508. In the illustrated embodiment, the display 508 may be worn by the user as part of a headset such that a user may wear the display over their eyes like a pair of goggles or glasses. At least a portion of the display may be transparent such that a user may observe a see-through reality 510. The see-through reality 510 may correspond to portions of the physical world 506 that are within a present viewpoint of the AR system 502, which may correspond to the viewpoint of the user in the case that the user is wearing a headset incorporating both the display and sensors of the AR system to acquire information about the physical world.

**[0049]** AR contents may also be presented on the display 508, overlaid on the see-through reality 510. To provide accurate interactions between AR contents and the see-through reality 510 on the display 508, the AR system 502 may include sensors 522 configured to capture

information about the physical world 506.

**[0050]** The sensors 522 may include one or more depth sensors that output depth maps 512. Each depth map 512 may have multiple pixels, each of which may represent a distance to a surface in the physical world 506 in a particular direction relative to the depth sensor. Raw depth data may come from a depth sensor to create a depth map. Such depth maps may be updated as fast as the depth sensor can form a new image, which may be hundreds or thousands of times per second. However, that data may be noisy and incomplete, and have holes shown as black pixels on the illustrated depth map.

**[0051]** The sensors 522 may include other sensors, such as image sensors. The image sensors may acquire information such as monocular or stereoscopic information, which may be processed to represent the physical world in other ways. In some embodiments, the system may include other sensors such as, for example, one or more of the following: LiDAR cameras, RGBD cameras, infrared camera sensors, visible spectrum camera sensors, structured light emitters and/or sensors, infrared light emitters, coherent light emitters and/or sensors, gyroscopic sensors, accelerometers, magnetometers, altimeters, proximity sensors, GPS sensors, ultrasonic emitters and detectors, and haptic interfaces. The sensor data may be processed in world reconstruction component 516 to create a mesh, representing connected portions of objects in the physical world. Metadata about such objects, including for example, color and surface texture, may similarly be acquired with the sensors and stored as part of the world reconstruction. Metadata about a location of a device including the system may be determined or inferred based on the sensor data. For example magnetometers, altimeters, GPS sensors, and the like may be used to determine or infer the location of the device. The location of the device may be of varying degrees of granularity. For example, the accuracy of the determined location of the device may vary from coarse (e.g., accurate to 10 meter diameter sphere), to fine (e.g., accurate to a 3 meter diameter sphere), to extra fine (e.g., accurate to a 1 meter diameter sphere), to ultra fine (e.g., accurate to a 0.5 meter diameter sphere), etc.

**[0052]** The system may also acquire information about the headpose (or “pose”) of the user with respect to the physical world. In some embodiments, a headpose tracking component of the system may be used to compute headposes in real time. The headpose tracking component may represent a headpose of a user in a coordinate frame with six degrees of freedom including, for example, translation in three perpendicular axes (e.g., forward/backward, up/down, left/right) and rotation about the three perpendicular axes (e.g., pitch, yaw, and roll). In some embodiments, sensors 522 may include inertial measurement units that may be used to compute and/or

determine a headpose 514. A headpose 514 for a depth map may indicate a present viewpoint of a sensor capturing the depth map with six degrees of freedom, for example, but the headpose 514 may be used for other purposes, such as to relate image information to a particular portion of the physical world or to relate the position of the display worn on the user's head to the physical world.

**[0053]** In some embodiments, the headpose information may be derived in other ways than from an IMU, such as from analyzing objects in an image. For example, the headpose tracking component may compute relative position and orientation of an AR device to physical objects based on visual information captured by cameras and inertial information captured by IMUs. The headpose tracking component may then compute a headpose of the AR device by, for example, comparing the computed relative position and orientation of the AR device to the physical objects with features of the physical objects. In some embodiments, that comparison may be made by identifying features in images captured with one or more of the sensors 522 that are stable over time such that changes of the position of these features in images captured over time can be associated with a change in headpose of the user.

**[0054]** Techniques for operating an XR system may provide XR scenes for a more immersive user experience. In such a system, an XR device may estimate headpose at a frequency of 1 kHz, with low usage of computational resources. Such a device may be configured with, for example, four video graphic array (VGA) cameras operating at 30 Hz, one inertial measurement unit (IMU) operating at 1 kHz, compute power of a single advanced RISC machine (ARM) core, memory less than 1 GB, and network bandwidth less than 100 Mbps. Techniques as described herein may be employed to reduce processing required to generate and maintain maps and estimate headpose as well as to provide and consume data with low computational overhead. The XR system may calculate its pose based on the matched visual features. U.S. Patent Application Publication No. US 2019/0188474 describes hybrid tracking and is hereby incorporated herein by reference in its entirety.

**[0055]** In some embodiments, the AR device may construct a map from the feature points recognized in successive images in a series of image frames captured as a user moves throughout the physical world with the AR device. Though each image frame may be taken from a different pose as the user moves, the system may adjust the orientation of the features of each successive image frame to match the orientation of the initial image frame by matching features of the successive image frames to previously captured image frames. Translations of the successive image frames so that points representing the same features will match corresponding feature

points from previously collected image frames, can be used to align each successive image frame to match the orientation of previously processed image frames. The frames in the resulting map may have a common orientation established when the first image frame was added to the map. This map, with sets of feature points in a common frame of reference, may be used to determine the user's pose within the physical world by matching features from current image frames to the map. In some embodiments, this map may be called a tracking map.

**[0056]** Alternatively or additionally, a map of a 3D environment around a user may be constructed by identifying planes or other surfaces based on image information. The position of these surfaces from image to image, may be correlated to create the representation. Techniques to efficiently optimize such a map, as described herein, may be used to form that map. Such a map may be used for positioning virtual objects with respect to the physical world, or for different or additional functions. For example, the 3D representation based on panes may be used for headpose tracking.

**[0057]** In addition to enabling tracking of the user's pose within the environment, this map may enable other components of the system, such as world reconstruction component 516, to determine the location of physical objects with respect to the user. The world reconstruction component 516 may receive the depth maps 512 and headposes 514, and any other data from the sensors, and integrate that data into a reconstruction 518. The reconstruction 518 may be more complete and less noisy than the sensor data. The world reconstruction component 516 may update the reconstruction 518 using spatial and temporal averaging of the sensor data from multiple viewpoints over time.

**[0058]** The reconstruction 518 may include representations of the physical world in one or more data formats including, for example, voxels, meshes, planes, etc. The different formats may represent alternative representations of the same portions of the physical world or may represent different portions of the physical world. In the illustrated example, on the left side of the reconstruction 518, portions of the physical world are presented as a global surface; on the right side of the reconstruction 518, portions of the physical world are presented as meshes.

**[0059]** In some embodiments, the map maintained by headpose component 514 may be sparse relative to other maps that might be maintained of the physical world. Rather than providing information about locations, and possibly other characteristics, of surfaces, the sparse map may indicate locations of interest points and/or structures, such as corners or edges. In some embodiments, the map may include image frames as captured by the sensors 522. These frames may be reduced to features, which may represent the interest points and/or structures. In

conjunction with each frame, information about a pose of a user from which the frame was acquired may also be stored as part of the map. In some embodiments, every image acquired by the sensor may or may not be stored. In some embodiments, the system may process images as they are collected by sensors and select subsets of the image frames for further computation. The selection may be based on one or more criteria that limits the addition of information yet ensures that the map contains useful information. The system may add a new image frame to the map, for example, based on overlap with a prior image frame already added to the map or based on the image frame containing a sufficient number of features determined as likely to represent stationary objects. In some embodiments, the selected image frames, or groups of features from selected image frames may serve as key frames for the map, which are used to provide spatial information.

**[0060]** In some embodiments, the amount of data that is processed when constructing maps may be reduced, such as by constructing sparse maps with a collection of mapped points and keyframes and/or dividing the maps into blocks to enable updates by blocks. A mapped point may be associated with a point of interest in the environment. A keyframe may include selected information from camera-captured data. U.S. Patent Application Publication No. US 2020/0034624 describes determining and/or evaluating localization maps and is hereby incorporated herein by reference in its entirety.

**[0061]** The AR system 502 may integrate sensor data over time from multiple viewpoints of a physical world. The poses of the sensors (e.g., position and orientation) may be tracked as a device including the sensors is moved. As the sensor's frame pose is known and how it relates to the other poses, each of these multiple viewpoints of the physical world may be fused together into a single, combined reconstruction of the physical world, which may serve as an abstract layer for the map and provide spatial information. The reconstruction may be more complete and less noisy than the original sensor data by using spatial and temporal averaging (i.e. averaging data from multiple viewpoints over time), or any other suitable method.

**[0062]** In the illustrated embodiment in FIG. 3, a map represents the portion of the physical world in which a user of a single, wearable device is present. In that scenario, headpose associated with frames in the map may be represented as a local headpose, indicating orientation relative to an initial orientation for a single device at the start of a session. For example, the headpose may be tracked relative to an initial headpose when the device was turned on or otherwise operated to scan an environment to build a representation of that environment.

**[0063]** In combination with content characterizing that portion of the physical world, the map

may include metadata. The metadata, for example, may indicate time of capture of the sensor information used to form the map. Metadata alternatively or additionally may indicate location of the sensors at the time of capture of information used to form the map. Location may be expressed directly, such as with information from a GPS chip, or indirectly, such as with a wireless (e.g., Wi-Fi) signature indicating strength of signals received from one or more wireless access points while the sensor data was being collected and/or with identifiers, such as BSSID's, of wireless access points to which the user device connected while the sensor data was collected.

**[0064]** The reconstruction 518 may be used for AR functions, such as producing a surface representation of the physical world for occlusion processing or physics-based processing. This surface representation may change as the user moves or objects in the physical world change. Aspects of the reconstruction 518 may be used, for example, by a component 520 that produces a changing global surface representation in world coordinates, which may be used by other components.

**[0065]** The AR content may be generated based on this information, such as by AR applications 504. An AR application 504 may be a game program, for example, that performs one or more functions based on information about the physical world, such as visual occlusion, physics-based interactions, and environment reasoning. It may perform these functions by querying data in different formats from the reconstruction 518 produced by the world reconstruction component 516. In some embodiments, component 520 may be configured to output updates when a representation in a region of interest of the physical world changes. That region of interest, for example, may be set to approximate a portion of the physical world in the vicinity of the user of the system, such as the portion within the view field of the user, or is projected (predicted/determined) to come within the view field of the user.

**[0066]** The AR applications 504 may use this information to generate and update the AR contents. The virtual portion of the AR contents may be presented on the display 508 in combination with the see-through reality 510, creating a realistic user experience.

**[0067]** In some embodiments, an AR experience may be provided to a user through an XR device, which may be a wearable display device, which may be part of a system that may include remote processing and or remote data storage and/or, in some embodiments, other wearable display devices worn by other users. **Figure 4** illustrates an example of system 580 (hereinafter referred to as "system 580") including a single wearable device for simplicity of illustration. The system 580 includes a head mounted display device 562 (hereinafter referred to as "display device 562"), and various mechanical and electronic modules and systems to support the

functioning of the display device 562. The display device 562 may be coupled to a frame 564, which is wearable by a display system user or viewer 560 (hereinafter referred to as "user 560") and configured to position the display device 562 in front of the eyes of the user 560. According to various embodiments, the display device 562 may be a sequential display. The display device 562 may be monocular or binocular. In some embodiments, the display device 562 may be an example of the display 508 in Figure 3.

**[0068]** In some embodiments, a speaker 566 is coupled to the frame 564 and positioned proximate an ear canal of the user 560. In some embodiments, another speaker, not shown, is positioned adjacent another ear canal of the user 560 to provide for stereo/shapeable sound control. The display device 562 is operatively coupled, such as by a wired lead or wireless connectivity 568, to a local data processing module 570 which may be mounted in a variety of configurations, such as fixedly attached to the frame 564, fixedly attached to a helmet or hat worn by the user 560, embedded in headphones, or otherwise removably attached to the user 560 (e.g., in a backpack-style configuration, in a belt-coupling style configuration).

**[0069]** The local data processing module 570 may include a processor, as well as digital memory, such as non-volatile memory (e.g., flash memory), both of which may be utilized to assist in the processing, caching, and storage of data. The data include data a) captured from sensors (which may be, e.g., operatively coupled to the frame 564) or otherwise attached to the user 560, such as image capture devices (such as cameras), microphones, inertial measurement units, accelerometers, compasses, GPS units, radio devices, and/or gyros; and/or b) acquired and/or processed using remote processing module 572 and/or remote data repository 574, possibly for passage to the display device 562 after such processing or retrieval.

**[0070]** In some embodiments, the wearable device may communicate with remote components. The local data processing module 570 may be operatively coupled by communication links 576, 578, such as via a wired or wireless communication links, to the remote processing module 572 and remote data repository 574, respectively, such that these remote modules 572, 574 are operatively coupled to each other and available as resources to the local data processing module 570. In further embodiments, in addition or as alternative to remote data repository 574, the wearable device can access cloud based remote data repositories, and/or services. In some embodiments, the headpose tracking component described above may be at least partially implemented in the local data processing module 570. In some embodiments, the world reconstruction component 516 in Figure 3 may be at least partially implemented in the local data processing module 570. For example, the local data processing module 570 may be configured

to execute computer executable instructions to generate the map and/or the physical world representations based at least in part on at least a portion of the data.

**[0071]** In some embodiments, processing may be distributed across local and remote processors. For example, local processing may be used to construct a map on a user device (e.g. tracking map) based on sensor data collected with sensors on that user's device. Such a map may be used by applications on that user's device. Additionally, previously created maps (e.g., canonical maps) may be stored in remote data repository 574. Where a suitable stored or persistent map is available, it may be used instead of or in addition to the tracking map created locally on the device. In some embodiments, a tracking map may be localized to the stored map, such that a correspondence is established between a tracking map, which might be oriented relative to a position of the wearable device at the time a user turned the system on, and the canonical map, which may be oriented relative to one or more persistent features. In some embodiments, the persistent map might be loaded on the user device to allow the user device to render virtual content without a delay associated with scanning a location to build a tracking map of the user's full environment from sensor data acquired during the scan. In some embodiments, the user device may access a remote persistent map (e.g., stored on a cloud) without the need to download the persistent map on the user device.

**[0072]** In some embodiments, spatial information may be communicated from the wearable device to remote services, such as a cloud service that is configured to localize a device to stored maps maintained on the cloud service. According to one embodiment, the localization processing can take place in the cloud matching the device location to existing maps, such as canonical maps, and return transforms that link virtual content to the wearable device location. In such embodiments, the system can avoid communicating maps from remote resources to the wearable device. Other embodiments can be configured for both device-based and cloud-based localization, for example, to enable functionality where network connectivity is not available or a user opts not to enable cloud-based localization.

**[0073]** Alternatively or additionally, the tracking map may be merged with previously stored maps to extend or improve the quality of those maps. The processing to determine whether a suitable previously created environment map is available and/or to merge a tracking map with one or more stored environment maps may be done in local data processing module 570 or remote processing module 572.

**[0074]** In some embodiments, the local data processing module 570 may include one or more processors (e.g., a graphics processing unit (GPU)) configured to analyze and process data and/or

image information. In some embodiments, the local data processing module 570 may include a single processor (e.g., a single-core or multi-core ARM processor), which would limit the local data processing module 570's compute budget but enable a more miniature device. In some embodiments, the world reconstruction component 516 may use a compute budget less than a single Advanced RISC Machine (ARM) core to generate physical world representations in real-time on a non-predefined space such that the remaining compute budget of the single ARM core can be accessed for other uses such as, for example, extracting meshes.

**[0075]** Processing as described herein to optimize a map of a 3D environment may be performed in any processor of the system. However, the reduced computation and reduced memory required for optimization as described herein may enable such operations to be performed quickly and with low latency on local processors that are part of a wearable device.

**[0076]** In some embodiments, the remote data repository 574 may include a digital data storage facility, which may be available through the Internet or other networking configuration in a "cloud" resource configuration. In some embodiments, all data is stored and all computations are performed in the local data processing module 570, allowing fully autonomous use from a remote module. In some embodiments, all data is stored and all or most computations are performed in the remote data repository 574, allowing for a smaller device. A world reconstruction, for example, may be stored in whole or in part in this repository 574.

**[0077]** In embodiments in which data is stored remotely, and accessible over a network, data may be shared by multiple users of an augmented reality system. For example, user devices may upload their tracking maps to augment a database of environment maps. In some embodiments, the tracking map upload occurs at the end of a user session with a wearable device. In some embodiments, the tracking map uploads may occur continuously, semi-continuously, intermittently, at a pre-defined time, after a pre-defined period from the previous upload, or when triggered by an event. A tracking map uploaded by any user device may be used to expand or improve a previously stored map, whether based on data from that user device or any other user device. Likewise, a persistent map downloaded to a user device may be based on data from that user device or any other user device. In this way, high quality environment maps may be readily available to users to improve their experiences with the AR system.

**[0078]** In further embodiments, persistent map downloads can be limited and/or avoided based on localization executed on remote resources (e.g., in the cloud). In such configurations, a wearable device or other XR device communicates to the cloud service feature information coupled with pose information (e.g., positioning information for the device at the time the

features represented in the feature information were sensed) . One or more components of the cloud service may match the feature information to respective stored maps (e.g., canonical maps) and generates transforms between a tracking map maintained by the XR device and the coordinate system of the canonical map. Each XR device that has its tracking map localized with respect to the canonical map may accurately render virtual content in locations specified with respect to the canonical map based on its own tracking.

**[0079]** In some embodiments, the local data processing module 570 is operatively coupled to a battery 582. In some embodiments, the battery 582 is a removable power source, such as over the counter batteries. In other embodiments, the battery 582 is a lithium-ion battery. In some embodiments, the battery 582 includes both an internal lithium-ion battery chargeable by the user 560 during non-operation times of the system 580 and removable batteries such that the user 560 may operate the system 580 for longer periods of time without having to be tethered to a power source to charge the lithium-ion battery or having to shut the system 580 off to replace batteries.

**[0080]** **Figure 5A** illustrates a user 530 wearing an AR display system rendering AR content as the user 530 moves through a physical world environment 532 (hereinafter referred to as "environment 532"). The information captured by the AR system along the movement path of the user may be processed into one or more tracking maps. The user 530 positions the AR display system at positions 534, and the AR display system records ambient information of a passable world (e.g., a digital representation of the real objects in the physical world that can be stored and updated with changes to the real objects in the physical world) relative to the positions 534. That information may be stored as poses in combination with images, features, directional audio inputs, or other desired data. The positions 534 are aggregated to data inputs 536, for example, as part of a tracking map, and processed at least by a passable world module 538, which may be implemented, for example, by processing on a remote processing module 572 of Figure 4. In some embodiments, the passable world module 538 may include the headpose component 514 and the world reconstruction component 516, such that the processed information may indicate the location of objects in the physical world in combination with other information about physical objects used in rendering virtual content.

**[0081]** The passable world module 538 determines, at least in part, where and how AR content 540 can be placed in the physical world as determined from the data inputs 536. The AR content is "placed" in the physical world by presenting via the user interface both a representation of the physical world and the AR content, with the AR content rendered as if it were interacting with objects in the physical world and the objects in the physical world presented as if the AR content

were, when appropriate, obscuring the user's view of those objects. In some embodiments, the AR content may be placed by appropriately selecting portions of a fixed element 542 (e.g., a table) from a reconstruction (e.g., the reconstruction 518) to determine the shape and position of the AR content 540. As an example, the fixed element may be a table and the virtual content may be positioned such that it appears to be on that table. In some embodiments, the AR content may be placed within structures in a field of view 544, which may be a present field of view or an estimated future field of view. In some embodiments, the AR content may be persisted relative to a model 546 of the physical world (e.g. a mesh).

**[0082]** As depicted, the fixed element 542 serves as a proxy (e.g. digital copy) for any fixed element within the physical world which may be stored in the passable world module 538 so that the user 530 can perceive content on the fixed element 542 without the system having to map to the fixed element 542 each time the user 530 sees it. The fixed element 542 may, therefore, be a mesh model from a previous modeling session or determined from a separate user but nonetheless stored by the passable world module 538 for future reference by a plurality of users. Therefore, the passable world module 538 may recognize the environment 532 from a previously mapped environment and display AR content without a device of the user 530 mapping all or part of the environment 532 first, saving computation process and cycles and avoiding latency of any rendered AR content.

**[0083]** The mesh model 546 of the physical world may be created by the AR display system and appropriate surfaces and metrics for interacting and displaying the AR content 540 can be stored by the passable world module 538 for future retrieval by the user 530 or other users without the need to completely or partially recreate the model. In some embodiments, the data inputs 536 are inputs such as geolocation, user identification, and current activity to indicate to the passable world module 538 which fixed element 542 of one or more fixed elements are available, which AR content 540 has last been placed on the fixed element 542, and whether to display that same content (such AR content being "persistent" content regardless of user viewing a particular passable world model).

**[0084]** Even in embodiments in which objects are considered to be fixed (e.g. a kitchen table), the passable world module 538 may update those objects in a model of the physical world from time to time to account for the possibility of changes in the physical world. The model of fixed objects may be updated with a very low frequency. Other objects in the physical world may be moving or otherwise not regarded as fixed (e.g. kitchen chairs). To render an AR scene with a realistic feel, the AR system may update the position of these non-fixed objects with a much

higher frequency than is used to update fixed objects. To enable accurate tracking of all of the objects in the physical world, an AR system may draw information from multiple sensors, including one or more image sensors.

**[0085]** Figure 5B is a schematic illustration of a viewing optics assembly 548 and attendant components. In some embodiments, two eye tracking cameras 550, directed toward user eyes 549, detect metrics of the user eyes 549, such as eye shape, eyelid occlusion, pupil direction and glint on the user eyes 549.

**[0086]** In some embodiments, one of the sensors may be a depth sensor 551, such as a time of flight sensor, emitting signals to the world and detecting reflections of those signals from nearby objects to determine distance to given objects. A depth sensor, for example, may quickly determine whether objects have entered the field of view of the user, either as a result of motion of those objects or a change of pose of the user. However, information about the position of objects in the field of view of the user may alternatively or additionally be collected with other sensors. Depth information, for example, may be obtained from stereoscopic visual image sensors or plenoptic sensors.

**[0087]** In some embodiments, world cameras 552 record a greater-than-peripheral view to map and/or otherwise create a model of the environment 532 and detect inputs that may affect AR content. In some embodiments, the world camera 552 and/or camera 553 may be grayscale and/or color image sensors, which may output grayscale and/or color image frames at fixed time intervals. Camera 553 may further capture physical world images within a field of view of the user at a specific time. Pixels of a frame-based image sensor may be sampled repetitively even if their values are unchanged. Each of the world cameras 552, the camera 553 and the depth sensor 551 have respective fields of view of 554, 555, and 556 to collect data from and record a physical world scene, such as the physical world environment 532 depicted in Figure 34A.

**[0088]** Inertial measurement units 557 may determine movement and orientation of the viewing optics assembly 548. In some embodiments, inertial measurement units 557 may provide an output indicating a direction of gravity. In some embodiments, each component is operatively coupled to at least one other component. For example, the depth sensor 551 is operatively coupled to the eye tracking cameras 550 as a confirmation of measured accommodation against actual distance the user eyes 549 are looking at.

**[0089]** It should be appreciated that a viewing optics assembly 548 may include some of the components illustrated in Figure 34B and may include components instead of or in addition to

the components illustrated. In some embodiments, for example, a viewing optics assembly 548 may include two world camera 552 instead of four. Alternatively or additionally, cameras 552 and 553 need not capture a visible light image of their full field of view. A viewing optics assembly 548 may include other types of components. In some embodiments, a viewing optics assembly 548 may include one or more dynamic vision sensor (DVS), whose pixels may respond asynchronously to relative changes in light intensity exceeding a threshold.

**[0090]** In some embodiments, a viewing optics assembly 548 may not include the depth sensor 551 based on time of flight information. In some embodiments, for example, a viewing optics assembly 548 may include one or more plenoptic cameras, whose pixels may capture light intensity and an angle of the incoming light, from which depth information can be determined. For example, a plenoptic camera may include an image sensor overlaid with a transmissive diffraction mask (TDM). Alternatively or additionally, a plenoptic camera may include an image sensor containing angle-sensitive pixels and/or phase-detection auto-focus pixels (PDAF) and/or micro-lens array (MLA). Such a sensor may serve as a source of depth information instead of or in addition to depth sensor 551.

**[0091]** It also should be appreciated that the configuration of the components in Figure 5B is provided as an example. A viewing optics assembly 548 may include components with any suitable configuration, which may be set to provide the user with the largest field of view practical for a particular set of components. For example, if a viewing optics assembly 548 has one world camera 552, the world camera may be placed in a center region of the viewing optics assembly instead of at a side.

**[0092]** Information from the sensors in viewing optics assembly 548 may be coupled to one or more of processors in the system. The processors may generate data that may be rendered so as to cause the user to perceive virtual content interacting with objects in the physical world. That rendering may be implemented in any suitable way, including generating image data that depicts both physical and virtual objects. In other embodiments, physical and virtual content may be depicted in one scene by modulating the opacity of a display device that a user looks through at the physical world. The opacity may be controlled so as to create the appearance of the virtual object and also to block the user from seeing objects in the physical world that are occluded by the virtual objects. In some embodiments, the image data may only include virtual content that may be modified such that the virtual content is perceived by a user as realistically interacting with the physical world (e.g. clip content to account for occlusions), when viewed through the user interface.

**[0093]** The location on the viewing optics assembly 548 at which content is displayed to create the impression of an object at a particular location may depend on the physics of the viewing optics assembly. Additionally, the pose of the user's head with respect to the physical world and the direction in which the user's eyes are looking may impact where in the physical world content displayed at a particular location on the viewing optics assembly content will appear. Sensors as described above may collect this information, and or supply information from which this information may be calculated, such that a processor receiving sensor inputs may compute where objects should be rendered on the viewing optics assembly 548 to create a desired appearance for the user.

**[0094]** Regardless of how content is presented to a user, a model of the physical world may be used so that characteristics of the virtual objects, which can be impacted by physical objects, including the shape, position, motion, and visibility of the virtual object, can be correctly computed. In some embodiments, the model may include the reconstruction of a physical world, for example, the reconstruction 518.

**[0095]** That model may be created from data collected from sensors on a wearable device of the user. Though, in some embodiments, the model may be created from data collected by multiple users, which may be aggregated in a computing device remote from all of the users (and which may be "in the cloud").

**[0096]** The model may be created, at least in part, by a world reconstruction system such as, for example, the world reconstruction component 516 of Figure 3 depicted in more detail in **Figure 6A**. The world reconstruction component 516 may include a perception module 660 that may generate, update, and store representations for a portion of the physical world. In some embodiments, the perception module 660 may represent the portion of the physical world within a reconstruction range of the sensors as multiple voxels. Each voxel may correspond to a 3D cube of a predetermined volume in the physical world, and include surface information, indicating whether there is a surface in the volume represented by the voxel. Voxels may be assigned values indicating whether their corresponding volumes have been determined to include surfaces of physical objects, determined to be empty or have not yet been measured with a sensor and so their value is unknown. It should be appreciated that values indicating that voxels that are determined to be empty or unknown need not be explicitly stored, as the values of voxels may be stored in computer memory in any suitable way, including storing no information for voxels that are determined to be empty or unknown.

**[0097]** In addition to generating information for a persisted world representation, the

perception module 660 may identify and output indications of changes in a region around a user of an AR system. Indications of such changes may trigger updates to volumetric data stored as part of the persisted world, or trigger other functions, such as triggering components 604 that generate AR content to update the AR content.

**[0098]** In some embodiments, the perception module 660 may identify changes based on a signed distance function (SDF) model. The perception module 660 may be configured to receive sensor data such as, for example, depth maps 660a and headposes 660b, and then fuse the sensor data into a SDF model 660c. Depth maps 660a may provide SDF information directly, and images may be processed to arrive at SDF information. The SDF information represents distance from the sensors used to capture that information. As those sensors may be part of a wearable unit, the SDF information may represent the physical world from the perspective of the wearable unit and therefore the perspective of the user. The headposes 660b may enable the SDF information to be related to a voxel in the physical world.

**[0099]** In some embodiments, the perception module 660 may generate, update, and store representations for the portion of the physical world that is within a perception range. The perception range may be determined based, at least in part, on a sensor's reconstruction range, which may be determined based, at least in part, on the limits of a sensor's observation range. As a specific example, an active depth sensor that operates using active IR pulses may operate reliably over a range of distances, creating the observation range of the sensor, which may be from a few centimeters or tens of centimeters to a few meters.

**[00100]** The world reconstruction component 516 may include additional modules that may interact with the perception module 660. In some embodiments, a persisted world module 662 may receive representations for the physical world based on data acquired by the perception module 660. The persisted world module 662 also may include various formats of representations of the physical world. For example, volumetric metadata 662b such as voxels may be stored as well as meshes 662c and planes 662d. In some embodiments, other information, such as depth maps could be saved.

**[00101]** In some embodiments, representations of the physical world, such as those illustrated in FIG. 6A may provide relatively dense information about the physical world in comparison to sparse maps, such as a tracking map based on feature points as described above.

**[00102]** In some embodiments, the perception module 660 may include modules that generate representations for the physical world in various formats including, for example, meshes 660d,

planes and semantics 660e. The representations for the physical world may be stored across local and remote storage mediums. The representations for the physical world may be described in different coordinate frames depending on, for example, the location of the storage medium. For example, a representation for the physical world stored in the device may be described in a coordinate frame local to the device. The representation for the physical world may have a counterpart stored in a cloud. The counterpart in the cloud may be described in a coordinate frame shared by all devices in an XR system.

**[00103]** In some embodiments, these modules may generate representations based on data within the perception range of one or more sensors at the time the representation is generated as well as data captured at prior times and information in the persisted world module 662. In some embodiments, these components may operate on depth information captured with a depth sensor. However, the AR system may include vision sensors and may generate such representations by analyzing monocular or binocular vision information.

**[00104]** In some embodiments, these modules may operate on regions of the physical world. Those modules may be triggered to update a subregion of the physical world, when the perception module 660 detects a change in the physical world in that subregion. Such a change, for example, may be detected by detecting a new surface in the SDF model 660c or other criteria, such as changing the value of a sufficient number of voxels representing the subregion.

**[00105]** The world reconstruction component 516 may include components 664 that may receive representations of the physical world from the perception module 660. Information about the physical world may be pulled by these components according to, for example, a use request from an application. In some embodiments, information may be pushed to the use components, such as via an indication of a change in a pre-identified region or a change of the physical world representation within the perception range. The components 664, may include, for example, game programs and other components that perform processing for visual occlusion, physics-based interactions, and environment reasoning.

**[00106]** Responding to the queries from the components 664, the perception module 660 may send representations for the physical world in one or more formats. For example, when the component 664 indicates that the use is for visual occlusion or physics-based interactions, the perception module 660 may send a representation of surfaces. When the component 664 indicates that the use is for environmental reasoning, the perception module 660 may send meshes, planes and semantics of the physical world.

**[00107]** In some embodiments, the perception module 660 may include components that format information to provide the component 664. An example of such a component may be raycasting component 660f. A use component (e.g., component 664), for example, may query for information about the physical world from a particular point of view. Raycasting component 660f may select from one or more representations of the physical world data within a field of view from that point of view.

**[00108]** As should be appreciated from the foregoing description, the perception module 660, or another component of an AR system, may process data to create 3D representations of portions of the physical world. Data to be processed may be reduced by culling parts of a 3D reconstruction volume based at least in part on a camera frustum and/or depth image, extracting and persisting plane data, capturing, persisting, and updating 3D reconstruction data in blocks that allow local update while maintaining neighbor consistency, providing occlusion data to applications generating such scenes, where the occlusion data is derived from a combination of one or more depth data sources, and/or performing a multi-stage mesh simplification. The reconstruction may contain data of different levels of sophistication including, for example, raw data such as live depth data, fused volumetric data such as voxels, and computed data such as meshes.

**[00109]** In some embodiments, components of a passable world model may be distributed, with some portions executing locally on an XR device and some portions executing remotely, such as on a network connected server, or otherwise in the cloud. The allocation of the processing and storage of information between the local XR device and the cloud may impact functionality and user experience of an XR system. For example, reducing processing on a local device by allocating processing to the cloud may enable longer battery life and reduce heat generated on the local device. But, allocating too much processing to the cloud may create undesirable latency that causes an unacceptable user experience.

**[00110]** **Figure 6B** depicts a distributed component architecture 600 configured for spatial computing, according to some embodiments. The distributed component architecture 600 may include a passable world component 602 (e.g., PW 538 in FIG. 5A), a Lumin OS 604, API's 606, SDK 608, and Application 610. The Lumin OS 604 may include a Linux-based kernel with custom drivers compatible with an XR device. The API's 606 may include application programming interfaces that grant XR applications (e.g., Applications 610) access to the spatial computing features of an XR device. The SDK 608 may include a software development kit that allows the creation of XR applications.

**[00111]** One or more components in the architecture 600 may create and maintain a model of a passable world. In this example sensor data is collected on a local device. Processing of that sensor data may be performed in part locally on the XR device and partially in the cloud. PW 538 may include environment maps created based, at least in part, on data captured by AR devices worn by multiple users. During sessions of an AR experience, individual AR devices (such as wearable devices described above in connection with Figure 4 may create tracking maps, which is one type of map.

**[00112]** In some embodiments, the device may include components that construct both sparse maps and dense maps. A tracking map may serve as a sparse map and may include headposes of the AR device scanning an environment as well as information about objects detected within that environment at each headpose. Those headposes may be maintained locally for each device. For example, the headpose on each device may be relative to an initial headpose when the device was turned on for its session. As a result, each tracking map may be local to the device creating it and may have its own frame of reference defined by its own local coordinate system. In some embodiments, however, the tracking map on each device may be formed such that one coordinate of its local coordinate system is aligned with the direction of gravity as measured by its sensors, such as inertial measurement unit 557.

**[00113]** The dense map may include surface information, which may be represented by a mesh or depth information. Alternatively or additionally, a dense map may include higher level information derived from surface or depth information, such as the location and/or characteristics of planes and/or other objects.

**[00114]** Creation of the dense maps may be independent of the creation of sparse maps, in some embodiments. The creation of dense maps and sparse maps, for example, may be performed in separate processing pipelines within an AR system. Separating processing, for example, may enable generation or processing of different types of maps to be performed at different rates. Sparse maps, for example, may be refreshed at a faster rate than dense maps. In some embodiments, however, the processing of dense and sparse maps may be related, even if performed in different pipelines. Changes in the physical world revealed in a sparse map, for example, may trigger updates of a dense map, or *vice versa*. Further, even if independently created, the maps might be used together. For example, a coordinate system derived from a sparse map may be used to define position and/or orientation of objects in a dense map.

**[00115]** The sparse map and/or dense map may be persisted for re-use by the same device and/or sharing with other devices. Such persistence may be achieved by storing information in

the cloud. The AR device may send the tracking map to a cloud to, for example, merge with environment maps selected from persisted maps previously stored in the cloud. In some embodiments, the selected persisted maps may be sent from the cloud to the AR device for merging. In some embodiments, the persisted maps may be oriented with respect to one or more persistent coordinate frames. Such maps may serve as canonical maps, as they can be used by any of multiple devices. In some embodiments, a model of a passable world may comprise or be created from one or more canonical maps. Devices, even though they perform some operations based on a coordinate frame local to the device, may nonetheless use the canonical map by determining a transformation between their coordinate frame local to the device and the canonical map.

**[00116]** A canonical map may originate as a tracking map (TM) (e.g., TM 1102 in Figure 31A), which may be promoted to a canonical map. The canonical map may be persisted such that devices that access the canonical map may, once determining a transformation between their local coordinate system and a coordinate system of the canonical map, use the information in the canonical map to determine locations of objects represented in the canonical map in the physical world around the device. In some embodiments, a TM may be a headpose sparse map created by an XR device. In some embodiments, the canonical map may be created when an XR device sends one or more TMs to a cloud server for merging with additional TMs captured by the XR device at a different time or by other XR devices.

**[00117]** In embodiments in which tracking maps are formed on local devices with one coordinate of a local coordinate frame aligned with gravity, this orientation with respect to gravity may be preserved upon creation of a canonical map. For example, when a tracking map that is submitted for merging does not overlap with any previously stored map, that tracking map may be promoted to a canonical map. Other tracking maps, which may also have an orientation relative to gravity, may be subsequently merged with that canonical map. The merging may be done so as to ensure that the resulting canonical map retains its orientation relative to gravity. Two maps, for example, may not be merged, regardless of correspondence of feature points in those maps, if coordinates of each map aligned with gravity do not align with each other with a sufficiently close tolerance.

**[00118]** The canonical maps, or other maps, may provide information about the portions of the physical world represented by the data processed to create respective maps. **Figure 7** depicts an exemplary tracking map 700, according to some embodiments. The tracking map 700 may provide a floor plan 706 of physical objects in a corresponding physical world, represented by

points 702. In some embodiments, a map point 702 may represent a feature of a physical object that may include multiple features. For example, each corner of a table may be a feature that is represented by a point on a map. The features may be derived from processing images, such as may be acquired with the sensors of a wearable device in an augmented reality system. The features, for example, may be derived by processing an image frame output by a sensor to identify features based on large gradients in the image or other suitable criteria. Further processing may limit the number of features in each frame. For example, processing may select features that likely represent persistent objects. One or more heuristics may be applied for this selection.

**[00119]** The tracking map 700 may include data on points 702 collected by a device. For each image frame with data points included in a tracking map, a pose may be stored. The pose may represent the orientation from which the image frame was captured, such that the feature points within each image frame may be spatially correlated. The pose may be determined by positioning information, such as may be derived from the sensors, such as an IMU sensor, on the wearable device. Alternatively or additionally, the pose may be determined from matching image frames to other image frames that depict overlapping portions of the physical world. By finding such positional correlation, which may be accomplished by matching subsets of features points in two frames, the relative pose between the two frames may be computed. A relative pose may be adequate for a tracking map, as the map may be relative to a coordinate system local to a device established based on the initial pose of the device when construction of the tracking map was initiated.

**[00120]** Not all of the feature points and image frames collected by a device may be retained as part of the tracking map, as much of the information collected with the sensors is likely to be redundant. Rather, only certain frames may be added to the map. Those frames may be selected based on one or more criteria, such as degree of overlap with image frames already in the map, the number of new features they contain or a quality metric for the features in the frame. Image frames not added to the tracking map may be discarded or may be used to revise the location of features. As a further alternative, all or most of the image frames, represented as a set of features may be retained, but a subset of those frames may be designated as key frames, which are used for further processing.

**[00121]** The key frames may be processed to produce keyrigs 704. The key frames may be processed to produce three dimensional sets of feature points and saved as keyrigs 704. Such processing may entail, for example, comparing image frames derived simultaneously from two

cameras to stereoscopically determine the 3D position of feature points. Metadata may be associated with these keyframes and/or keyrigs, such as poses.

**[00122]** The environment maps may have any of multiple formats depending on, for example, the storage locations of an environment map including, for example, local storage of AR devices and remote storage. For example, a map in remote storage may have higher resolution than a map in local storage on a wearable device where memory is limited. To send a higher resolution map from remote storage to local storage, the map may be down sampled or otherwise converted to an appropriate format, such as by reducing the number of poses per area of the physical world stored in the map and/or the number of feature points stored for each pose. In some embodiments, a slice or portion of a high resolution map from remote storage may be sent to local storage, where the slice or portion is not down sampled.

**[00123]** A database of environment maps may be updated as new tracking maps are created. To determine which of a potentially very large number of environment maps in a database is to be updated, updating may include efficiently selecting one or more environment maps stored in the database relevant to the new tracking map. The selected one or more environment maps may be ranked by relevance and one or more of the highest ranking maps may be selected for processing to merge higher ranked selected environment maps with the new tracking map to create one or more updated environment maps. When a new tracking map represents a portion of the physical world for which there is no preexisting environment map to update, that tracking map may be stored in the database as a new environment map.

**[00124]** *Techniques For Efficient Processing of Maps With Planes*

**[00125]** FIG. 7 illustrates a map based on feature points. Some systems may create a map based on recognized surfaces, such as planes, rather than individual points. In some embodiments, rather than containing map points 702, the map may contain planes. For example, such a map may contain a plane to represent a table top, rather than a collection of feature points representing corners of the table and possibly points on the surface.

**[00126]** The processing of creating/updating such a map may include computing features of planes (such as position and orientation) in images acquired from various poses. The determination of the poses may, however, depend on the estimated features of planes previously detected. Therefore, creating and updating a map may entail jointly optimizing estimated features describing the planes and sensor poses from which images, used to estimate the features of the planes, were captured. The joint optimization may be referred to as planar bundle adjustment.

[00127] The inventors have recognized and appreciated that conventional approaches to planar bundle adjustment that applies to a map with one or more planar features can result in a large-scale nonlinear least-square problem, which incurs high computational cost and memory consumption. Unlike bundle adjustment conventionally performed for maps represented by feature points in which one recording of a sensor can yield a single observation of a point feature, a planar feature may be an infinite object, and one image from a sensor can provide multiple observations of a planar feature. Each observation of the planar feature may put a constraint on plane parameters of the planar feature and pose parameters of the poses, resulting in a large-scale nonlinear least-square problem even for a small workplace.

[00128] The inventors have recognized and appreciated a planar bundle adjustment that reduces the computational cost and memory usage. **Figure 8A** is a schematic illustrating observations of various planes at different poses, according to some embodiments. In the illustrated example, there are  $M$  planes and  $N$  sensor poses. Denote the rotation and translation of the  $i$ th pose as  $R_i \in SO_3$  and  $t_i \in \mathbb{R}^3$ . Suppose the  $j$ th plane has the parameters as  $\pi_j = [n_j; d_j]$ , where  $n$  is the plane normal with  $\|n\|_2 = 1$  and  $d$  is the negative distance from the coordinate system origin to the plane. The measurements of the  $j$ th plane at the  $i$ th pose are a set of  $K_{ij}$  points defined as:

$$\mathbb{P}_{ij} = \{\mathbf{p}_{ijk}\}_{k=1}^{K_{ij}} \quad (6)$$

Each  $\mathbf{p}_{ijk} \in \mathbb{P}_{ij}$  provides one constraint on the  $i$ th pose and  $j$ th plane, which is illustrated in **Figure 8B**. The residual  $\delta_{ijk}$  is the signed distance from  $\mathbf{p}_{ijk}$  to plane  $\pi_j$  which can be written as

$$\delta_{ijk} = \mathbf{n}_j \cdot (\mathbf{R}_i \mathbf{p}_{ijk} + \mathbf{t}_i) + d_j, \quad (7)$$

[00129] Unlike  $t_i$ , the rotation  $R_i$  and plane parameters  $\pi_j$  have extra constraints. For instance,  $R_i$  can be parameterized by quaternion, angle-axis or Euler angle. The plane parameters  $\pi_j$  can be represented by homogeneous coordinates, closest point parameterization or the minimal parameterization based on quaternion. Regardless of a special parameterization, this example defines  $\theta_i \rightarrow R(\theta_i)$  and  $\omega_j \rightarrow \pi(\omega_j)$  to present arbitrary parameterization for rotation and plane parameters.  $\theta_i$  and  $t_i$  are related to the sensor pose, which may be combined as  $\rho_i = [\theta_i; t_i]$ .  $\rho_i$  may have 6 or 7 unknowns, changed with respect to the parameterization of the rotation matrix (7 for the quaternion and 6 for minimal representations of the rotation, such as the angle-axis parameterization).  $\omega_j$  may have 3 or 4 unknowns (3 for minimal representations of a plane and

4 for the homogeneous coordinates of a plane). Using these notation, the residual  $\delta_{ijk}$  is a function of  $\rho_i$  and  $\omega_j$ .

[00130] The planar bundle adjustment may be the problem of jointly refining all  $\rho_i (i \neq 1)$  and  $\omega_j$  by minimizing the following nonlinear least-squares problem

$$\min_{\substack{\rho_i, \omega_j \\ i \neq 1}} \sum_i \sum_j \sum_k \delta_{ijk}^2 (\rho_i, \omega_j). \tag{8}$$

Here the first pose  $\rho_1$  is fixed during the optimization to anchor the coordinate system rigidly.

[00131] The Jacobian Matrix of the planar bundle adjustment may be computed, and may be provided to an algorithm that solves least-squares problem (e.g., the Levenberg-Marquardt (LM) algorithm, and the Gaussian-Newton algorithm). The observation of the  $j$ th plane at the  $i$ th pose is a point set  $P_{ij}$ . The Jacobian matrix  $J_{ij}$  may be derived from the point set  $P_{ij}$ . The Jacobian matrix  $J$  may be the stack of all  $J_{ij}$ . Assume there are  $K_{ij}$  points in  $P_{ij}$ . The  $J_{ij} (i \neq 1)$  has the form as

$$J_{ij} = \begin{matrix} & \begin{matrix} \text{\textit{i}th pose} & & \text{\textit{j}th plane} \end{matrix} \\ \begin{matrix} 0 & \dots & \frac{\partial \delta_{ijk}}{\partial \rho_i} & \dots & 0 & \dots & \frac{\partial \delta_{ijk}}{\partial \omega_j} & \dots & 0 \\ 0 & \dots & \frac{\partial \delta_{ijk}}{\partial \rho_i} & \dots & 0 & \dots & \frac{\partial \delta_{ijk}}{\partial \omega_j} & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\partial \delta_{ijk}}{\partial \rho_i} & \dots & 0 & \dots & \frac{\partial \delta_{ijk}}{\partial \omega_j} & \dots & 0 \end{matrix} \\ \underbrace{\hspace{10em}}_{N-1 \text{ pose}} & & \underbrace{\hspace{10em}}_{M \text{ plane}} \end{matrix} \tag{9}$$

[00132] The partial derivatives of  $\delta_{ijk}$ ,  $\frac{\partial \delta_{ijk}}{\partial \rho_i}$  and  $\frac{\partial \delta_{ijk}}{\partial \omega_j}$ , may be computed based on the form of the residual  $\delta_{ijk}$  in (7). It may be defined that

$$\mathbf{R}_i = \begin{bmatrix} R_i^{1,1} & R_i^{1,2} & R_i^{1,3} \\ R_i^{2,1} & R_i^{2,2} & R_i^{2,3} \\ R_i^{3,1} & R_i^{3,2} & R_i^{3,3} \end{bmatrix}, \mathbf{t}_i = \begin{bmatrix} t_i^1 \\ t_i^2 \\ t_i^3 \end{bmatrix}, \mathbf{n}_j = \begin{bmatrix} n_j^1 \\ n_j^2 \\ n_j^3 \end{bmatrix}, \mathbf{p}_{ijk} = \begin{bmatrix} x_{ijk} \\ y_{ijk} \\ z_{ijk} \end{bmatrix}. \tag{10}$$

It is noted that the elements of  $\mathbf{R}_i$  defined above are functions of  $\theta_i$ , and  $d_j$  and the elements of  $\mathbf{n}_j$  are functions of  $\omega_j$ .  $p_{ijk}$  may represent the  $k$ th measurement of the  $j$ th plane at the  $i$ th pose.  $x_{ijk}$ ,  $y_{ijk}$ , and  $z_{ijk}$  may represent  $p_{ijk}$ 's position in a coordinate frame that may be a sensor coordinate

frame or a device coordinate frame or a canonical coordinate frame. The residual  $\delta_{ijk}$  can be computed by substituting (10) into (7) and expanding it. This results

$$\begin{aligned} \delta_{ijk} = & x_{ijk}R_i^{1,1}n_j^1 + y_{ijk}R_i^{1,2}n_j^1 + z_{ijk}R_i^{1,3}n_j^1 + \\ & x_{ijk}R_i^{2,1}n_j^2 + y_{ijk}R_i^{2,2}n_j^2 + z_{ijk}R_i^{2,3}n_j^2 + \\ & x_{ijk}R_i^{3,1}n_j^3 + y_{ijk}R_i^{3,2}n_j^3 + z_{ijk}R_i^{3,3}n_j^3 + \\ & n_j^1t_i^1 + n_j^2t_i^2 + n_j^3t_i^3 + d_j. \end{aligned} \quad (11)$$

(11) may be rewritten as

$$\delta_{ijk} = \mathbf{c}_{ijk} \cdot \mathbf{v}_{ij}, \quad (12)$$

where  $\mathbf{c}_{ijk}$  and  $\mathbf{v}_{ij}$  are 13-dimensional vectors as

$$\begin{aligned} \mathbf{c}_{ijk} = & [x_{ijk}, y_{ijk}, z_{ijk}, x_{ijk}, y_{ijk}, z_{ijk}, x_{ijk}, y_{ijk}, z_{ijk}, 1, 1, 1, 1]^T, \\ \mathbf{v}_{ij} = & [R_i^{1,1}n_j^1, R_i^{1,2}n_j^2, R_i^{1,3}n_j^3, R_i^{2,1}n_j^1, R_i^{2,2}n_j^2, R_i^{2,3}n_j^3, R_i^{3,1}n_j^1, \\ & R_i^{3,2}n_j^2, R_i^{3,3}n_j^3, n_j^1t_i^1, n_j^2t_i^2, n_j^3t_i^3, d_j]^T. \end{aligned} \quad (13)$$

The elements in  $\mathbf{c}_{ijk}$  are from the observation  $\mathbf{p}_{ijk}$  or 1, which are constants. On the other hand, the elements in  $\mathbf{v}_{ij}$  are functions of  $\rho_i$  and  $\omega_j$ .

[00133] The partial derivatives of  $\delta_{ijk}$  can be computed by assume that  $\rho_i$  has  $n_\rho$  unknowns and  $\omega_j$  has  $n_\omega$  unknowns. It may be defined

$$\zeta_{ij} = \left. \begin{array}{l} \left[ \rho_i \right] \\ \left[ \omega_j \right] \end{array} \right\} \begin{array}{l} n_\rho \text{ unknowns} \\ n_\omega \text{ unknowns} \end{array} \quad (14)$$

Suppose  $\zeta_{ij}^d$  is the  $d$ th element of  $\zeta_{ij}$ . According to (12), the partial derivative of  $\delta_{ijk}$  with respect to  $\zeta_{ij}^d$  has the form as

$$\frac{\partial \delta_{ijk}}{\partial \zeta_{ij}^d} = \mathbf{c}_{ijk} \cdot \frac{\partial \mathbf{v}_{ij}}{\partial \zeta_{ij}^d}, \quad (15)$$

where  $\frac{\partial \mathbf{v}_{ij}}{\partial \zeta_{ij}^d}$  is a 13-dimensional vector whose elements are the partial derivatives of the elements of  $\mathbf{v}_{ij}$  with respect to  $\zeta_{ij}^d$ . According to (15),  $\frac{\partial \delta_{ijk}}{\partial \rho_i}$  has the form as

$$\begin{aligned} \frac{\partial \delta_{ijk}}{\partial \rho_i} &= \left[ \frac{\partial \delta_{ijk}}{\partial \zeta_{ij}^1}, \dots, \frac{\partial \delta_{ijk}}{\partial \zeta_{ij}^{n_p}} \right] \\ &= \left[ c_{ijk} \cdot \frac{\partial \mathbf{v}_{ij}}{\partial \zeta_{ij}^1}, \dots, c_{ijk} \cdot \frac{\partial \mathbf{v}_{ij}}{\partial \zeta_{ij}^{n_p}} \right] \\ &= c_{ijk}^T \underbrace{\left[ \frac{\partial \mathbf{v}_{ij}}{\partial \zeta_{ij}^1}, \dots, \frac{\partial \mathbf{v}_{ij}}{\partial \zeta_{ij}^{n_p}} \right]}_{\mathbf{V}_{\rho_i}} \\ &= c_{ijk}^T \mathbf{V}_{\rho_i}, \end{aligned} \tag{16}$$

$\mathbf{V}_{\rho_i}$  may be a  $13 \times 6$  or  $13 \times 7$  matrix ( $13 \times 7$  for quaternion and  $13 \times 6$  for minimal representations of the rotation matrix). Similarly,  $\frac{\partial \delta_{ijk}}{\partial \omega_j}$  has the form as

$$\begin{aligned} \frac{\partial \delta_{ijk}}{\partial \omega_j} &= c_{ijk}^T \underbrace{\left[ \frac{\partial \mathbf{v}_{ij}}{\partial \zeta_{ij}^{n_p+1}}, \dots, \frac{\partial \mathbf{v}_{ij}}{\partial \zeta_{ij}^{n_p+n_\omega}} \right]}_{\mathbf{V}_{\omega_j}} \\ &= c_{ijk}^T \mathbf{V}_{\omega_j} \end{aligned} \tag{17}$$

$\mathbf{V}_{\omega_j}$  may be a  $13 \times 3$  or  $13 \times 4$  matrix ( $13 \times 3$  for minimal representations of a plane and  $13 \times 4$  for homogeneous coordinates of a plane).

[00134] The Jacobian matrix  $\mathbf{J}_{ij}$  may be computed based on

$$\mathbf{C}_{ij} = \begin{bmatrix} c_{ij1}^T \\ c_{ij2}^T \\ \vdots \\ c_{ijK_{ij}}^T \end{bmatrix}, \tag{18}$$

where the  $k$ th row  $\mathbf{c}_{ijk}$  is defined in (13).  $\mathbf{C}_{ij}$  is a matrix of size  $K_{ij} \times 13$ . The Jacobian matrix  $\mathbf{J}_{ij}$  may be computed by substituting (16) and (17) into (9) and using the definition of  $\mathbf{C}_{ij}$  in (18). This results

$$\begin{aligned} \mathbf{J}_{ij} &= \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{c}_{ij1}^T \mathbf{V}_{\rho_i} & \cdots & \mathbf{0} & \cdots & \mathbf{c}_{ij1}^T \mathbf{V}_{\omega_j} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{c}_{ij2}^T \mathbf{V}_{\rho_i} & \cdots & \mathbf{0} & \cdots & \mathbf{c}_{ij2}^T \mathbf{V}_{\omega_j} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{c}_{ijK_{ij}}^T \mathbf{V}_{\rho_i} & \cdots & \mathbf{0} & \cdots & \mathbf{c}_{ijK_{ij}}^T \mathbf{V}_{\omega_j} & \cdots & \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0} & \cdots & \underbrace{\mathbf{C}_{ij}^T \mathbf{V}_{\rho_i}}_{\mathbf{C}_{ij} \mathbf{V}_{\rho_i}} & \cdots & \mathbf{0} & \cdots & \underbrace{\mathbf{C}_{ij}^T \mathbf{V}_{\omega_j}}_{\mathbf{C}_{ij} \mathbf{V}_{\omega_j}} & \cdots & \mathbf{0} \end{bmatrix}, \end{aligned} \quad (19)$$

[00135] Suppose  $\mathbf{P}_{ij}$  is the measurements from the  $j$ th plane  $\omega_j$  at the first pose  $\rho_1$ . As  $\rho_1$  may be fixed during the optimization, the Jacobian matrix  $\mathbf{J}_{ij}$  derived from  $\mathbf{P}_{ij}$  may have the form as

$$\mathbf{J}_{ij} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{C}_{ij} \mathbf{V}_{\omega_j} & \cdots & \mathbf{0} \end{bmatrix} \quad (20)$$

[00136]  $\mathbf{C}_{ij}$  can be written in the form as

$$\mathbf{C}_{ij} = \mathbf{Q}_{ij} \mathbf{M}_{ij}, \quad (21)$$

where  $\mathbf{M}_{ij}$  is of the size  $4 \times 13$  and  $\mathbf{Q}_{ij}^T \mathbf{Q}_{ij} = \mathbf{I}_4$ , where  $\mathbf{I}_4$  is the  $4 \times 4$  identity matrix. As shown in the definition of  $\mathbf{c}_{ijk}$  in (13),  $x_{ijk}$ ,  $y_{ijk}$ ,  $z_{ijk}$  and 1 are duplicated several times to form  $\mathbf{c}_{ijk}$ . Therefore, there are only 4 individual columns among the 13 columns of  $\mathbf{C}_{ij}$ , which contains the constant 1 and the  $x$ ,  $y$ ,  $z$  coordinates of points within  $\mathbf{P}_{ij}$ . It is denoted as

$$\begin{aligned} \mathbf{E}_{ij} &= \begin{bmatrix} x_{ij1} & y_{ij1} & z_{ij1} & 1 \\ x_{ij2} & y_{ij2} & z_{ij2} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{ijK_{ij}} & y_{ijK_{ij}} & z_{ijK_{ij}} & 1 \end{bmatrix} \\ &= \begin{bmatrix} \underbrace{x_{ij}}_{\mathbf{x}_{ij}} & \underbrace{y_{ij}}_{\mathbf{y}_{ij}} & \underbrace{z_{ij}}_{\mathbf{z}_{ij}} & \underbrace{\mathbf{1}}_{\mathbf{1}} \end{bmatrix} \end{aligned} \quad (22)$$

The 13 columns in the  $\mathbf{C}_{ij}$  are copies of the 4 columns in  $\mathbf{E}_{ij}$ .  $\mathbf{E}_{ij}$  can be factorized as

$$\mathbf{E}_{ij} = \mathbf{Q}_{ij} \mathbf{U}_{ij} \quad (23)$$

where  $Q_{ij}^T Q_{ij} = I_4$  and  $U_{ij}$  is an upper triangular matrix.  $Q_{ij}$  is of the size  $K_{ij} \times 4$  and  $U_{ij}$  is of the size  $4 \times 4$ . The thin QR decomposition may be used since the number of points  $K_{ij}$  is generally much larger than 4. The thin QR decomposition can reduce computational time.  $U_{ij}$  can be partitioned by its columns as

$$U_{ij} = \begin{bmatrix} \mathbf{u}_{ij}^1 & \mathbf{u}_{ij}^2 & \mathbf{u}_{ij}^3 & \mathbf{u}_{ij}^4 \end{bmatrix}. \quad (24)$$

Substituting (24) into (22) results

$$E_{ij} = Q_{ij} \begin{bmatrix} \mathbf{u}_{ij}^1 & \mathbf{u}_{ij}^2 & \mathbf{u}_{ij}^3 & \mathbf{u}_{ij}^4 \end{bmatrix} \quad (25)$$

Comparing (22) and (23) results

$$\begin{aligned} \mathbf{x}_{ij} &= Q_{ij} \mathbf{u}_{ij}^1, & \mathbf{y}_{ij} &= Q_{ij} \mathbf{u}_{ij}^2, \\ \mathbf{z}_{ij} &= Q_{ij} \mathbf{u}_{ij}^3, & \mathbf{1} &= Q_{ij} \mathbf{u}_{ij}^4 \end{aligned}$$

As the columns of  $C_{ij}$  are a copy of the columns of  $E_{ij}$ , according to the form of  $c_{ijk}$  in (13) and the definition of  $E_{ij}$  in (22),  $C_{ij}$  can be written as

$$C_{ij} = [\mathbf{x}_{ij}, \mathbf{y}_{ij}, \mathbf{z}_{ij}, \mathbf{x}_{ij}, \mathbf{y}_{ij}, \mathbf{z}_{ij}, \mathbf{x}_{ij}, \mathbf{y}_{ij}, \mathbf{z}_{ij}, \mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1}] \quad (27)$$

Substituting (26) into (27) results

$$\begin{aligned} C_{ij} &= Q_{ij} \underbrace{\begin{bmatrix} \mathbf{u}_{ij}^1, \mathbf{u}_{ij}^2, \mathbf{u}_{ij}^3, \mathbf{u}_{ij}^1, \mathbf{u}_{ij}^2, \mathbf{u}_{ij}^3, \mathbf{u}_{ij}^1, \mathbf{u}_{ij}^2, \mathbf{u}_{ij}^3, \mathbf{u}_{ij}^4, \mathbf{u}_{ij}^4, \mathbf{u}_{ij}^4, \mathbf{u}_{ij}^4 \end{bmatrix}}_{M_{ij}} \\ &= Q_{ij} M_{ij} \end{aligned} \quad (28)$$

**[00137]** The factorization of  $C_{ij}$  can be used to significantly reduce the computational cost. Although the thin QR decomposition is described in the example, other factorization methods may be used including, for example, singular value decomposition (SVD).

[00138] A reduced Jacobian matrix  $J'_{ij}$  of  $J_{ij}$  may be computed based on  $C_{ij}$ , which may be factorized as  $C_{ij} = Q_{ij}M_{ij}$  as described above.

$$J'_{ij} = \begin{cases} \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{M}_{ij}\mathbf{V}_{\omega_j} & \cdots & \mathbf{0} \end{bmatrix} & i = 1 \\ \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{M}_{ij}\mathbf{V}_{\rho_j} & \cdots & \mathbf{0} & \cdots & \mathbf{M}_{ij}\mathbf{V}_{\omega_j} & \cdots & \mathbf{0} \end{bmatrix} & i \neq 1 \end{cases} \quad (29)$$

[00139] The reduced Jacobian matrix  $J'_{ij}$  has fewer rows than the Jacobian matrix  $J_{ij}$ , because  $M_{ij}$  is a much smaller matrix than  $C_{ij}$ .  $C_{ij}$  has the size  $K_{ij} \times 13$ . In contrast,  $M_{ij}$  has the size  $4 \times 13$ . Generally,  $K_{ij}$  is much larger than 4.

[00140]  $J_{ij}$  and  $J'_{ij}$  may be stacked to form the Jacobian matrix  $J$  and the reduced Jacobian matrix  $J'$  for the cost function (8), as

$$J = \begin{bmatrix} \vdots \\ J_{ij} \\ \vdots \end{bmatrix}, \quad J' = \begin{bmatrix} \vdots \\ J'_{ij} \\ \vdots \end{bmatrix}, \quad (30)$$

[00141]  $J'$  can replace  $J$  to calculate  $J^T J$  in the algorithm that solves least-squares problem. For the planar bundle adjustment,  $J^T J = J'^T J'$ .  $J$  and  $J'$  are block vectors in terms of  $J_{ij}$  and  $J'_{ij}$  as defined in (30). According to block matrix multiplication,

$$J^T J = \sum_{i,j} J_{ij}^T J_{ij}, \quad J'^T J' = \sum_{i,j} J'_{ij}{}^T J'_{ij} \quad (31)$$

For  $i \neq 1$ , using the expression in (19),  $J'^T J_{ij}$  has the form

$$J'^T J_{ij} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \mathbf{0} & \cdots & \mathbf{V}_{\rho_j}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{V}_{\rho_j} & \cdots & \mathbf{V}_{\rho_j}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{V}_{\omega_j} & \cdots & \mathbf{0} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \mathbf{0} & \cdots & \mathbf{V}_{\omega_j}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{V}_{\rho_j} & \cdots & \mathbf{V}_{\omega_j}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{V}_{\omega_j} & \cdots & \mathbf{0} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \quad (32)$$

Similarly, using the expression in (29),  $J'_{ij}{}^T J'_{ij}$  has the form

$$\mathbf{J}_{ij}^T \mathbf{J}_{ij} = \begin{bmatrix} 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & \mathbf{V}_{\rho_i}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{V}_{\rho_i} & \dots & \mathbf{V}_{\rho_i}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{V}_{\omega_j} & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & \mathbf{V}_{\omega_j}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{V}_{\rho_i} & \dots & \mathbf{V}_{\omega_j}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{V}_{\omega_j} & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 \end{bmatrix} \quad (33)$$

Substituting (21) into  $\mathbf{V}_{\rho_i}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{V}_{\rho_i}$  and using the fact  $\mathbf{Q}_{ij}^T \mathbf{Q}_{ij} = \mathbf{I}_4$ ,

$$\begin{aligned} \mathbf{V}_{\rho_i}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{V}_{\rho_i} &= \mathbf{V}_{\rho_i}^T \mathbf{M}_{ij}^T (\mathbf{Q}_{ij}^T \mathbf{Q}_{ij}) \mathbf{M}_{ij} \mathbf{V}_{\rho_i} \\ &= \mathbf{V}_{\rho_i}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{V}_{\rho_i} \end{aligned} \quad (34)$$

Similarly,

$$\begin{aligned} \mathbf{V}_{\rho_i}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{V}_{\omega_j} &= \mathbf{V}_{\rho_i}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{V}_{\omega_j} \\ \mathbf{V}_{\omega_j}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{V}_{\rho_i} &= \mathbf{V}_{\omega_j}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{V}_{\rho_i} \\ \mathbf{V}_{\omega_j}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{V}_{\omega_j} &= \mathbf{V}_{\omega_j}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{V}_{\omega_j} \end{aligned} \quad (35)$$

For  $i = 1$ , according to (20), the only non-zero term for  $\mathbf{J}_{1j}^T \mathbf{J}_{1j}$  is  $\mathbf{V}_{\omega_j}^T \mathbf{C}_{1j}^T \mathbf{C}_{1j} \mathbf{V}_{\omega_j}$ . On the other hand, according to (29),  $\mathbf{J}_{1j}^T \mathbf{J}_{1j}$  has only one corresponding non-zero term  $\mathbf{V}_{\omega_j}^T \mathbf{M}_{1j}^T \mathbf{M}_{1j} \mathbf{V}_{\omega_j}$ . Similar to the derivation in (34), we have

$$\mathbf{V}_{\omega_j}^T \mathbf{C}_{1j}^T \mathbf{C}_{1j} \mathbf{V}_{\omega_j} = \mathbf{V}_{\omega_j}^T \mathbf{M}_{1j}^T \mathbf{M}_{1j} \mathbf{V}_{\omega_j}. \quad (36)$$

In summary, using (34), (35) and (36),  $\mathbf{J}_{ij}^T \mathbf{J}_{ij} = \mathbf{J}_{ij}^T \mathbf{J}_{ij}$ . According to (31), consequently  $\mathbf{J}^T \mathbf{J} = \mathbf{J}^T \mathbf{J}$ .

**[00142]** The residual vector for the  $K_{ij}$  points in  $P_{ij}$  can be defined as  $\delta_{ij} = [\delta_{ij1}, \delta_{ij2}, \dots, \delta_{ijK_{ij}}]^T$ . According to (12),  $\delta_{ij}$  can be written as

$$\boldsymbol{\delta}_{ij} = \mathbf{C}_{ij} \mathbf{v}_{ij}. \quad (37)$$

A reduced residual vector  $\delta'_{ij}$  of  $\delta_{ij}$  can be defined as

$$\boldsymbol{\delta}_{ij}^r = \mathbf{M}_{ij} \mathbf{v}_{ij} \tag{38}$$

Stacking all  $\delta_{ij}$  and  $\delta_{ij}^r$ , the residual vector  $\delta$  and the reduced residual vector  $\delta^r$  have the forms as

$$\boldsymbol{\delta} = \begin{bmatrix} \vdots \\ \boldsymbol{\delta}_{ij} \\ \vdots \end{bmatrix}, \quad \boldsymbol{\delta}^r = \begin{bmatrix} \vdots \\ \boldsymbol{\delta}_{ij}^r \\ \vdots \end{bmatrix}. \tag{39}$$

The reduced residual vector  $\delta^r$  can replace the residual vector  $\delta$  in the algorithm that solves least-squares problem. For the planar bundle adjustment,  $\mathbf{J}^T \boldsymbol{\delta} = \mathbf{J}^{rT} \boldsymbol{\delta}^r$ .  $\mathbf{J}$ ,  $\mathbf{J}^r$ ,  $\boldsymbol{\delta}$  and  $\boldsymbol{\delta}^r$  are block vectors with elements  $J_{ij}$ ,  $J_{ij}^r$ ,  $\delta_{ij}$  and  $\delta_{ij}^r$  as defined in (30) and (39), respectively. Applying the block matrix multiplication,

$$\mathbf{J}^T \boldsymbol{\delta} = \sum_{ij} \mathbf{J}_{ij}^T \boldsymbol{\delta}_{ij}, \quad \mathbf{J}^{rT} \boldsymbol{\delta}^r = \sum_{ij} \mathbf{J}_{ij}^{rT} \boldsymbol{\delta}_{ij}^r \tag{40}$$

For  $i \neq 1$ , using the expression of  $J_{ij}$  in (19) and  $J_{ij}^r$  in (29), and the expression of  $\delta_{ij}$  in (37) and  $\delta_{ij}^r$  in (38),  $\mathbf{J}_{ij}^T \boldsymbol{\delta}_{ij}$  and  $\mathbf{J}_{ij}^{rT} \boldsymbol{\delta}_{ij}^r$  have the forms as

$$\mathbf{J}_{ij}^T \boldsymbol{\delta}_{ij} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{v}_{\rho_{ij}}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{v}_{ij} \\ \vdots \\ \mathbf{v}_{\omega_{ij}}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{v}_{ij} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{J}_{ij}^{rT} \boldsymbol{\delta}_{ij}^r = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{v}_{\rho_{ij}}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{v}_{ij} \\ \vdots \\ \mathbf{v}_{\omega_{ij}}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{v}_{ij} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{41}$$

Substituting (21) into  $\mathbf{v}_{\rho_{ij}}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{v}_{ij}$  and using the fact  $\mathbf{Q}_{ij}^T \mathbf{Q}_{ij} = \mathbf{I}_4$ ,

$$\begin{aligned} \mathbf{v}_{\rho_{ij}}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{v}_{ij} &= \mathbf{v}_{\rho_{ij}}^T \mathbf{M}_{ij}^T (\mathbf{Q}_{ij}^T \mathbf{Q}_{ij}) \mathbf{M}_{ij} \mathbf{v}_{ij} \\ &= \mathbf{v}_{\rho_{ij}}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{v}_{ij} \end{aligned} \tag{42}$$

Similarly,

$$\mathbf{V}_{\omega_{ij}}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{v}_{ij} = \mathbf{V}_{\omega_{ij}}^T \mathbf{M}_{ij}^T \mathbf{M}_{ij} \mathbf{v}_{ij} \tag{43}$$

For  $i = 1$ , substituting (20) and (37) into  $\mathbf{J}_{1j}^T \delta_{1j}$  and applying the block matrix multiplication, the only non-zero term of  $\mathbf{J}_{1j}^T \delta_{1j}$  is  $\mathbf{V}_{\omega_{1j}}^T \mathbf{C}_{1j}^T \mathbf{C}_{1j} \mathbf{v}_{1j}$ . On the other hand, substituting (29) and (38) into  $\mathbf{J}'_{1j}{}^T \delta'_{1j}$ ,  $\mathbf{J}'_{1j}{}^T \delta'_{1j}$  only has one non-zero term  $\mathbf{V}'_{\omega_{1j}}{}^T \mathbf{M}'_{1j}{}^T \mathbf{M}'_{1j} \mathbf{v}'_{1j}$ . Similar to the derivation in (42),

$$\mathbf{V}_{\omega_{ij}}^T \mathbf{C}_{ij}^T \mathbf{C}_{ij} \mathbf{v}_{ij} = \mathbf{V}'_{\omega_{ij}}{}^T \mathbf{M}'_{ij}{}^T \mathbf{M}'_{ij} \mathbf{v}'_{ij} \tag{44}$$

In summary, from (42), (43) and (44),  $\mathbf{J}_{ij}^T \delta_{ij} = \mathbf{J}'_{ij}{}^T \delta'_{ij}$ . According to (40),  $\mathbf{J}^T \delta = \mathbf{J}'^T \delta'$ .

**[00143]** It should be appreciated that, in some embodiments a system using a reduced Jacobean matrix may compute the reduced Jacobean matrix directly from sensor data or other information, without forming the Jacobean matrix or repeating the computations to arrive at the form of the reduced matrix described herein.

**[00144]** For the planar bundle adjustment, the reduced Jacobian matrix  $\mathbf{J}'$  and the reduced residual vector  $\delta'$  can replace  $\mathbf{J}$  and  $\delta$  in (4) to compute the step in the algorithm that solves least-squares problem, and each block  $\mathbf{J}'_{ij}$  and  $\delta'_{ij}$  in  $\mathbf{J}'$  and  $\delta'$  have 4 rows. The algorithm that solves least-squares problem uses (4) to calculate the step for each iteration. As  $\mathbf{J}^T \mathbf{J} = \mathbf{J}'^T \mathbf{J}'$ ,  $\mathbf{J}^T \delta = \mathbf{J}'^T \delta'$ ,  $(\mathbf{J}^T \mathbf{J} + \lambda \mathbf{I}) \xi = \mathbf{J}^T \delta$  is equivalent to  $(\mathbf{J}'^T \mathbf{J}' + \lambda \mathbf{I}) \xi = \mathbf{J}'^T \delta'$ . Thus,  $\mathbf{J}'$  and  $\delta'$  can replace  $\mathbf{J}$  and  $\delta$  for computing the step in the algorithm that solves least-squares problem. According to the definition of  $\mathbf{J}'_{ij}$  in (29) and  $\delta'_{ij}$  in (38), the number of rows of  $\mathbf{J}'_{ij}$  and  $\delta'_{ij}$  is the same as the number of rows of  $\mathbf{M}_{ij}$ , which has 4 rows. Consequently,  $\mathbf{J}'$  and  $\delta'$  have 4 rows. Therefore, no matter how many points are in  $P_{ij}$ , the reduced  $\mathbf{J}'_{ij}$  and  $\delta'_{ij}$  have at most 4 rows. This significantly reduces the computational cost in the algorithm that solves least-squares problem. The additional cost here is to calculate  $\mathbf{C}_{ij} = \mathbf{Q}_{ij} \mathbf{M}_{ij}$ . As  $\mathbf{C}_{ij}$  keeps constant during the iteration, it is computed for a single time before the iteration.

**[00145]** In some embodiments, a planar bundle adjustment may obtain initial predictions of  $N$  poses and  $M$  plane parameters, and the measurements  $\{P_{ij}\}$ . The planar bundle adjustment may compute a matrix block  $c_{ijk}$  for each  $p_{ijk} \in P_{ij}$  as (11), and stacking the matrix blocks to  $\mathbf{C}_{ij}$  as (16). The planar bundle adjustment may factorize  $\mathbf{C}_{ij}$  into an orthogonal matrix  $\mathbf{Q}_{ij}$  and an upper

triangular matrix  $M_{ij}$  as (19). The upper triangular matrix  $M_{ij}$  may be provided to an algorithm that solves least-squares problem (e.g., the Levenberg-Marquardt (LM) algorithm, and the Gaussian-Newton algorithm) to compute the reduced Jacobian matrix block  $J_{ij}^r$  in (27) and the reduced residual block  $\delta_{ij}^r$  in (36), which may be stacked to form the reduced Jacobian matrix  $J^r$  as (28) and the reduced residual vector  $\delta^r$  as (37). The algorithm that solves least-squares problem may compute refined poses and plane parameters until converging.

**[00146]** In some embodiments, planar features may be combined in 3D reconstruction with other features such as point features. In the combined cost function derived from multiple features, the Jacobian matrix from the plane cost would have the form as  $J$  in (28), and the residual vector also has the form as  $\delta$  in (37). Therefore, the reduced Jacobian matrix  $J^r$  in (28) and the reduced residual vector  $\delta^r$  in (37) may replace the Jacobian matrix and the original residual vector in a bundle adjustment with hybrid features.

**[00147]** Representations of environments may be computed with planar bundle adjustment. **Figure 9** is a flow chart illustrating a method 900 of providing a representation of an environment, according to some embodiments. **Figure 10** is a schematic diagram illustrating a portion 1000 of a representation of an environment computed based on sensor captured information using the method 900, according to some embodiments.

**[00148]** The method 900 may start by obtaining (Act 902) information captured by one or more sensors at respective poses. In some embodiments, the obtained information may be visual images and/or depth images. In some embodiments, an image may be a keyframe that may be a combination of multiple images. The example of FIG. 10 shows three sensors *sensor0*, *sensor1*, and *sensor2*. In some embodiments, the sensors may belong to a device, which may have a device coordinate frame with an origin  $O$ . The device coordinate frame may represent the location and the orientation when the device first initiates scanning the environment for this session. Each sensor may have a corresponding sensor coordinate frame. Each sensor may capture one or more images such as images 1002A, 1002B, 1002C at respective poses  $\rho_{i-1}$ ,  $\rho_i$ , and  $\rho_{i+1}$ . In some embodiments, the coordinate frame with origin  $O$  may represent a shared coordinate frame by one or more devices. For example, the sensors may belong to three devices that share the coordinate frame.

**[00149]** The method 900 may include providing (Act 904) a first representation of an environment. The first representation may include initial predictions of feature parameters for features extracted from the sensor captured information and initial predictions of corresponding poses. In the example of FIG. 10, a first plane 1004 is observed in image 1002A and image

1002B. The first plane 1004 may be represented by plane parameters  $\pi_j = [n_j; d_j]$ . Image 1002A and image 1002B observe multiple points  $P_{(i-1)j}$  and  $P_{ij}$  in the plane 1004, respectively. Portions of the first plane 1004 may be observed by both image 1002A and 1002B. Portions of the first plane 1004 may be observed by only image 1002A or only 1002B. A second plane 1006 is observed in image 1002B and image 1002C. The second plane 1006 may be represented by plane parameters  $\pi_{j+1} = [n_{(j+1)}; d_{(j+1)}]$ . Image 1002B and image 1002C observe multiple points  $P_{i(j+1)}$  and  $P_{(i+1)j}$  in the plane 1006, respectively. Portions of the second plane 1006 may be observed by both image 1002B and 1002C. Portions of the second plane 1006 may be observed by only image 1002B or only 1002C. Image 1002B and image 1002C also observe a point feature 1008. The first representation may include initial predictions of the poses  $\rho_{i-1}$ ,  $\rho_i$ , and  $\rho_{i+1}$ , plane parameters for the first plane 1004 and the second plane 1006, and point feature parameters for the point feature 1008.

**[00150]** The initial predictions may be based on the images that first observe the features. For the example of FIG. 10, the initial predictions of the pose  $\rho_{i-1}$  and plane parameters for the first plane 1004 may be based on image 1002A. The initial predictions of the pose  $\rho_i$ , plane parameters for the second plane 1006, and feature parameters for the point feature 1008 may be based on image 1002B. The initial predictions of the pose  $\rho_{i+1}$  may be based on image 1002C.

**[00151]** As the features are observed by subsequent images, the initial predictions may be refined to reduce drift and improve the quality of the presentation. The method 900 may include Act 906 and Act 908 to compute refined poses and refined feature parameters, which may include refined plane parameters and refined point feature parameters. Act 906 may be performed for each of the planes at corresponding poses. In the example of FIG. 10, Act 906 is performed for the first plane 1004 at poses  $\rho_{i-1}$  and  $\rho_i$ , respectively. Act 906 is also performed for the second plane 1006 at poses  $\rho_i$  and  $\rho_{i+1}$ , respectively. Act 906 may include computing (Act 906A) a matrix (e.g.,  $C_{ij}$  as (16)) that indicates observations of a respective plane (e.g., plane  $j$ ) at a respective pose (e.g., pose  $i$ ). Act 906 may include factorizing (Act 906B) the matrix into two or more matrices that comprise one matrix having reduced rows compared with the matrix (e.g., an upper triangular matrix  $M_{ij}$  as (19)).

**[00152]** The method 900 may include providing (Act 910) a second representation of the environment with the refined poses and refined feature parameters. The method 900 may include determining (Act 912) whether new information are observed such that Acts 902 to 910 should be repeated.

**[00153]** Having thus described several aspects of some embodiments, it is to be appreciated

that various alterations, modifications, and improvements will readily occur to those skilled in the art.

**[00154]** As one example, embodiments are described in connection with an augmented (AR) environment. It should be appreciated that some or all of the techniques described herein may be applied in an MR environment or more generally in other XR environments, in VR environments, and any other computer vision and robotics applications.

**[00155]** As another example, embodiments are described in connection with devices, such as wearable devices. It should be appreciated that some or all of the techniques described herein may be implemented via networks (such as cloud), discrete applications, and/or any suitable combinations of devices, networks, and discrete applications.

**[00156]** Such alterations, modifications, and improvements are intended to be part of this disclosure, and are intended to be within the spirit and scope of the disclosure. Further, though advantages of the present disclosure are indicated, it should be appreciated that not every embodiment of the disclosure will include every described advantage. Some embodiments may not implement any features described as advantageous herein and in some instances. Accordingly, the foregoing description and drawings are by way of example only.

**[00157]** The above-described embodiments of the present disclosure can be implemented in any of numerous ways. For example, the embodiments may be implemented using hardware, software or a combination thereof. When implemented in software, the software code can be executed on any suitable processor or collection of processors, whether provided in a single computer or distributed among multiple computers. Such processors may be implemented as integrated circuits, with one or more processors in an integrated circuit component, including commercially available integrated circuit components known in the art by names such as CPU chips, GPU chips, microprocessor, microcontroller, or co-processor. In some embodiments, a processor may be implemented in custom circuitry, such as an ASIC, or semicustom circuitry resulting from configuring a programmable logic device. As yet a further alternative, a processor may be a portion of a larger circuit or semiconductor device, whether commercially available, semi-custom or custom. As a specific example, some commercially available microprocessors have multiple cores such that one or a subset of those cores may constitute a processor. Though, a processor may be implemented using circuitry in any suitable format.

**[00158]** Further, it should be appreciated that a computer may be embodied in any of a number of forms, such as a rack-mounted computer, a desktop computer, a laptop computer, or a

tablet computer. Additionally, a computer may be embedded in a device not generally regarded as a computer but with suitable processing capabilities, including a Personal Digital Assistant (PDA), a smart phone or any other suitable portable or fixed electronic device.

**[00159]** Also, a computer may have one or more input and output devices. These devices can be used, among other things, to present a user interface. Examples of output devices that can be used to provide a user interface include printers or display screens for visual presentation of output and speakers or other sound generating devices for audible presentation of output. Examples of input devices that can be used for a user interface include keyboards, and pointing devices, such as mice, touch pads, and digitizing tablets. As another example, a computer may receive input information through speech recognition or in other audible format. In the embodiment illustrated, the input/output devices are illustrated as physically separate from the computing device. In some embodiments, however, the input and/or output devices may be physically integrated into the same unit as the processor or other elements of the computing device. For example, a keyboard might be implemented as a soft keyboard on a touch screen. In some embodiments, the input/output devices may be entirely disconnected from the computing device, and functionally integrated through a wireless connection.

**[00160]** Such computers may be interconnected by one or more networks in any suitable form, including as a local area network or a wide area network, such as an enterprise network or the Internet. Such networks may be based on any suitable technology and may operate according to any suitable protocol and may include wireless networks, wired networks or fiber optic networks.

**[00161]** Also, the various methods or processes outlined herein may be coded as software that is executable on one or more processors that employ any one of a variety of operating systems or platforms. Additionally, such software may be written using any of a number of suitable programming languages and/or programming or scripting tools, and also may be compiled as executable machine language code or intermediate code that is executed on a framework or virtual machine.

**[00162]** In this respect, the disclosure may be embodied as a computer readable storage medium (or multiple computer readable media) (e.g., a computer memory, one or more floppy discs, compact discs (CD), optical discs, digital video disks (DVD), magnetic tapes, flash memories, circuit configurations in Field Programmable Gate Arrays or other semiconductor devices, or other tangible computer storage medium) encoded with one or more programs that, when executed on one or more computers or other processors, perform methods that implement the various embodiments of the disclosure discussed above. As is apparent from the foregoing

examples, a computer readable storage medium may retain information for a sufficient time to provide computer-executable instructions in a non-transitory form. Such a computer readable storage medium or media can be transportable, such that the program or programs stored thereon can be loaded onto one or more different computers or other processors to implement various aspects of the present disclosure as discussed above. As used herein, the term "computer-readable storage medium" encompasses only a computer-readable medium that can be considered to be a manufacture (i.e., article of manufacture) or a machine. In some embodiments, the disclosure may be embodied as a computer readable medium other than a computer-readable storage medium, such as a propagating signal.

**[00163]** The terms "program" or "software" are used herein in a generic sense to refer to any type of computer code or set of computer-executable instructions that can be employed to program a computer or other processor to implement various aspects of the present disclosure as discussed above. Additionally, it should be appreciated that according to one aspect of this embodiment, one or more computer programs that when executed perform methods of the present disclosure need not reside on a single computer or processor, but may be distributed in a modular fashion amongst a number of different computers or processors to implement various aspects of the present disclosure.

**[00164]** Computer-executable instructions may be in many forms, such as program modules, executed by one or more computers or other devices. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. Typically the functionality of the program modules may be combined or distributed as desired in various embodiments.

**[00165]** Also, data structures may be stored in computer-readable media in any suitable form. For simplicity of illustration, data structures may be shown to have fields that are related through location in the data structure. Such relationships may likewise be achieved by assigning storage for the fields with locations in a computer-readable medium that conveys relationship between the fields. However, any suitable mechanism may be used to establish a relationship between information in fields of a data structure, including through the use of pointers, tags or other mechanisms that establish relationship between data elements.

**[00166]** Various aspects of the present disclosure may be used alone, in combination, or in a variety of arrangements not specifically discussed in the embodiments described in the foregoing and is therefore not limited in its application to the details and arrangement of components set forth in the foregoing description or illustrated in the drawings. For example, aspects described

in one embodiment may be combined in any manner with aspects described in other embodiments.

**[00167]** Also, the disclosure may be embodied as a method, of which an example has been provided. The acts performed as part of the method may be ordered in any suitable way. Accordingly, embodiments may be constructed in which acts are performed in an order different than illustrated, which may include performing some acts simultaneously, even though shown as sequential acts in illustrative embodiments.

**[00168]** Use of ordinal terms such as “first,” “second,” “third,” etc., in the claims to modify a claim element does not by itself connote any priority, precedence, or order of one claim element over another or the temporal order in which acts of a method are performed, but are used merely as labels to distinguish one claim element having a certain name from another element having a same name (but for use of the ordinal term) to distinguish the claim elements.

**[00169]** Also, the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of "including," "comprising," or "having," “containing,” “involving,” and variations thereof herein, is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

## CLAIMS

What is claimed is:

1. A method of operating a computing system to generate a representation of an environment, the method comprising:

obtaining sensor captured information, the sensor captured information comprising a first number of images;

providing an initial representation of the environment, the initial representation comprising the first number of initial poses and initial parameters of a second number of planar features based at least in part on the first number of images;

for each of the second number of planar features at each pose corresponding to an image comprising one or more observations of the planar feature,

computing a matrix indicating the one or more observations of the planar feature, and

factorizing the matrix into two or more matrices, the two or more matrices comprising one matrix having reduced rows compared with the matrix; and

computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having reduced rows,

wherein the representation of the environment comprises the first number of refined poses and the refined parameters of the second number of planar features.

2. The method of claim 1, wherein for each of the second number of planar features at each pose corresponding to an image comprising one or more observations of the planar feature, factorizing the matrix into two or more matrices comprises

computing an orthogonal matrix and an upper triangular matrix.

3. The method of claim 2, wherein computing the first number of refined poses and refined parameters of the second number of planar features is based at least in part on the upper triangular matrices.

4. The method of claim 1, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having reduced rows comprises

computing reduced Jacobian matrix blocks based at least in part on the matrices having reduced rows.

5. The method of claim 4, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having reduced rows comprises

stacking the reduced Jacobian matrix blocks to form a reduced Jacobian matrix.

6. The method of claim 5, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having reduced rows comprises

providing the reduced Jacobian matrix to an algorithm that solves least-squares problem to update current estimate.

7. The method of claim 1, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having reduced rows comprises

computing reduced residual blocks based at least in part on the matrices having reduced rows.

8. The method of claim 7, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having reduced rows comprises

stacking the reduced residual blocks to form a reduced residual vector.

9. The method of claim 8, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having reduced rows comprises

providing the reduced residual vector to an algorithm that solves least-squares problem to update current estimate.

10. The method of claim 1, wherein for each of the second number of planar features at each pose corresponding to an image comprising one or more observations of the planar feature, computing the matrix indicating the one or more observations of the planar feature comprises

for each of the one or more observations of the planar feature, computing a matrix block indicating said observation, and

stacking the matrix blocks into the matrix indicating the one or more observations of the planar feature.

11. A method of operating a computing system to generate a representation of an environment, the method comprising:

obtaining sensor captured information, the sensor captured information comprising a first number of images;

providing an initial representation of the environment, the initial representation comprising the first number of initial poses and initial parameters of a second number of planar features based at least in part on the first number of images;

for each of the second number of planar features at each pose corresponding to an image comprising one or more observations of the planar feature,

computing a matrix having a third number of rows, the third number being smaller than the number of the one or more observations of the planar feature; and

computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having the third number of rows,

wherein the representation of the environment comprises the first number of refined poses and the refined parameters of the second number of planar features.

12. The method of claim 11, wherein for each of the second number of planar features at each pose corresponding to an image comprising one or more observations of the planar feature, computing the matrix having the third number of rows comprises

computing a matrix indicating the one or more observations of the planar feature, and

factorizing the matrix into two or more matrices, the two or more matrices comprising the matrix having the third number of rows.

13. The method of claim 12, wherein factorizing the matrix into two or more matrices comprises computing an orthogonal matrix and an upper triangular matrix.

14. The method of claim 11, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having the third number of rows comprises

computing reduced Jacobian matrix blocks based at least in part on the matrices having the third number of rows.

15. The method of claim 14, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having

the third number of rows comprises

stacking the reduced Jacobian matrix blocks to form a reduced Jacobian matrix.

16. The method of claim 15, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having the third number of rows comprises

providing the reduced Jacobian matrix to an algorithm that solves least-squares problem to update current estimate.

17. The method of claim 11, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having the third number of rows comprises

computing reduced residual blocks based at least in part on the matrices having the third number of rows.

18. The method of claim 17, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having the third number of rows comprises

stacking the reduced residual blocks to form a reduced residual vector.

19. The method of claim 18, wherein computing the first number of refined poses and refined parameters of the second number of planar features based at least in part on the matrices having the third number of rows comprises

providing the reduced residual vector to an algorithm that solves least-squares problem to update current estimate.

20. The method of claim 12, wherein for each of the second number of planar features at each pose corresponding to an image comprising one or more observations of the planar feature, computing the matrix indicating the one or more observations of the planar feature comprises

for each of the one or more observations of the planar feature, computing a matrix block indicating said observation, and

stacking the matrix blocks into the matrix indicating the one or more observations of the planar feature.

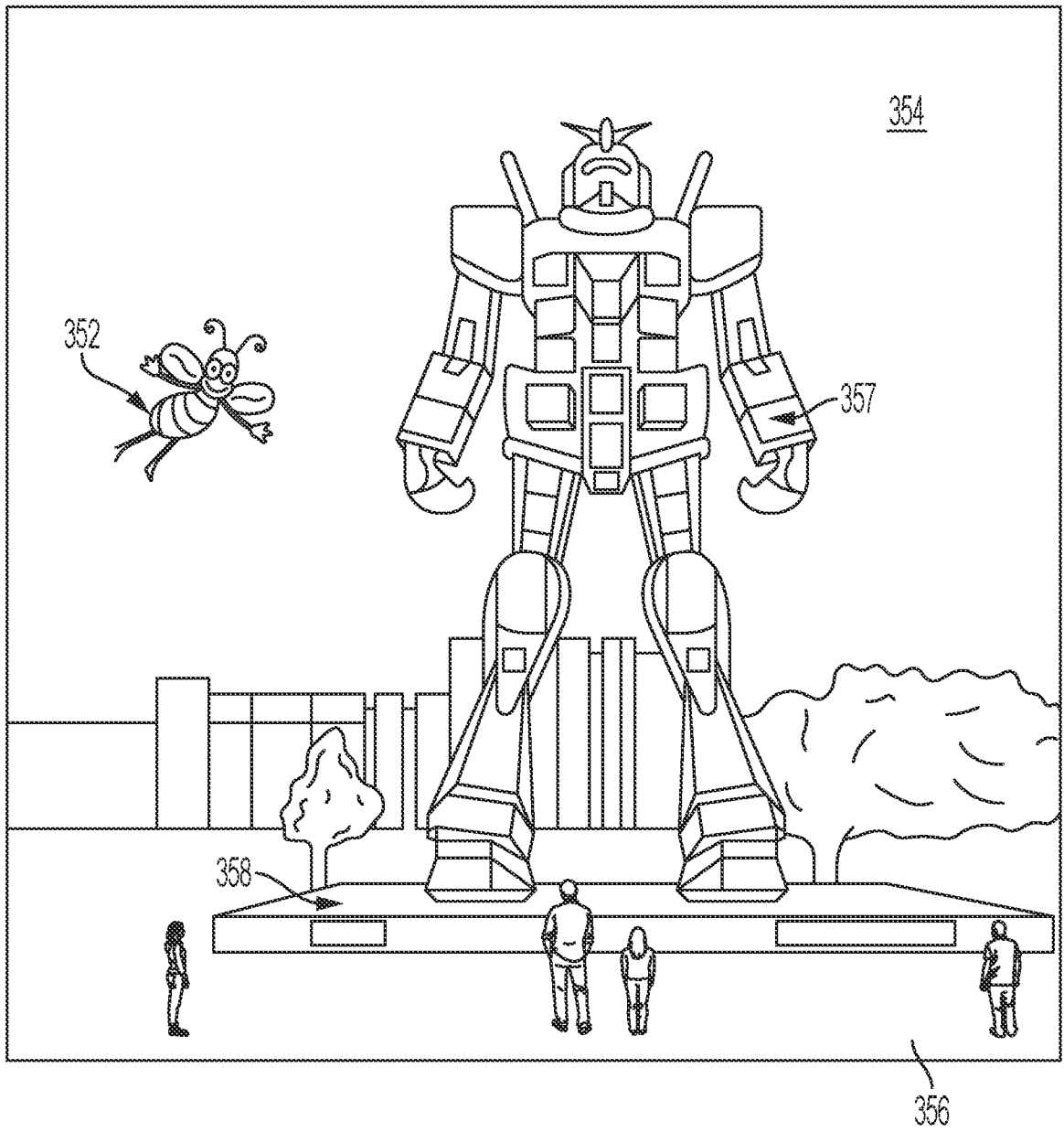


FIG. 1

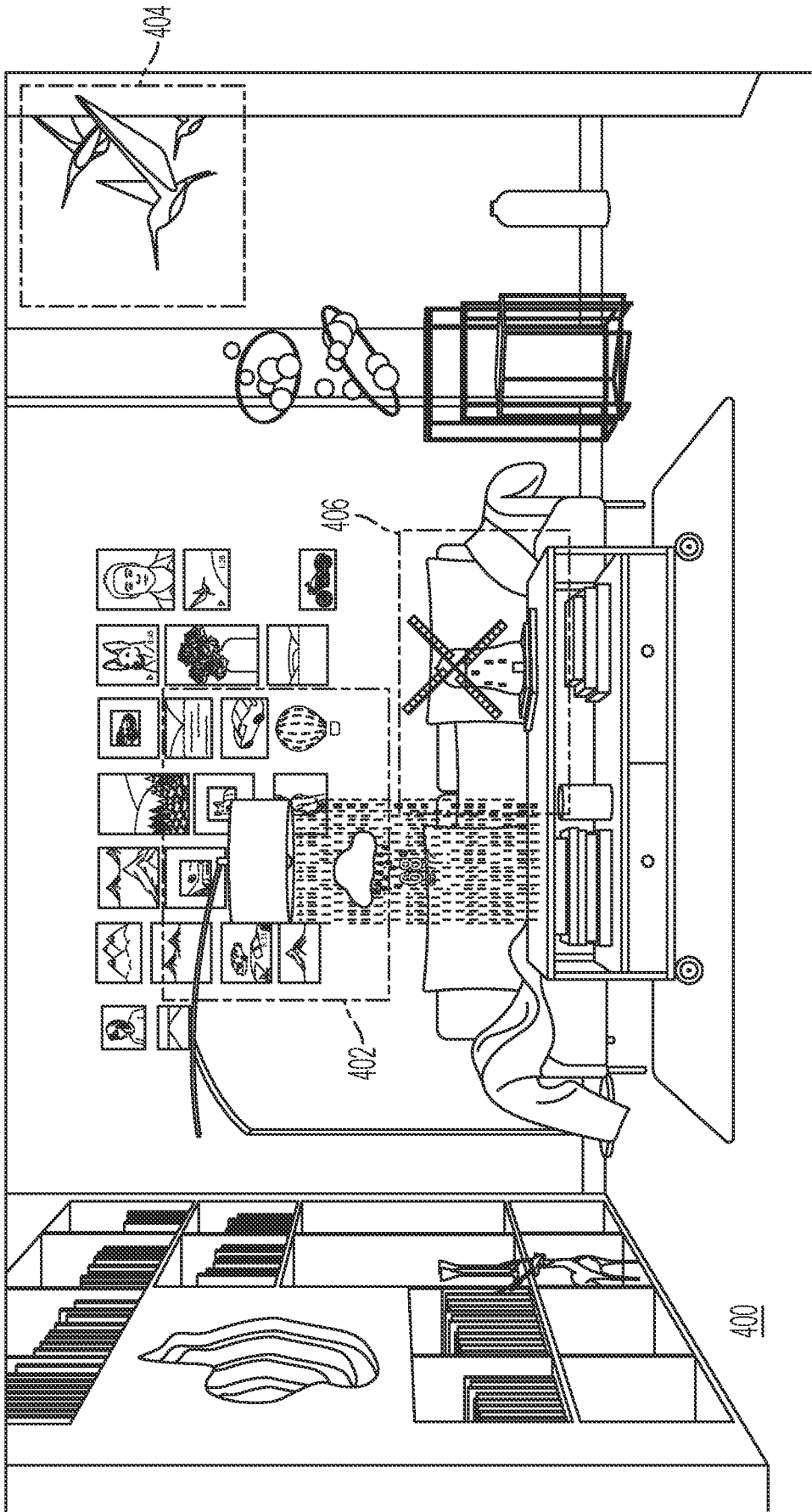


FIG. 2

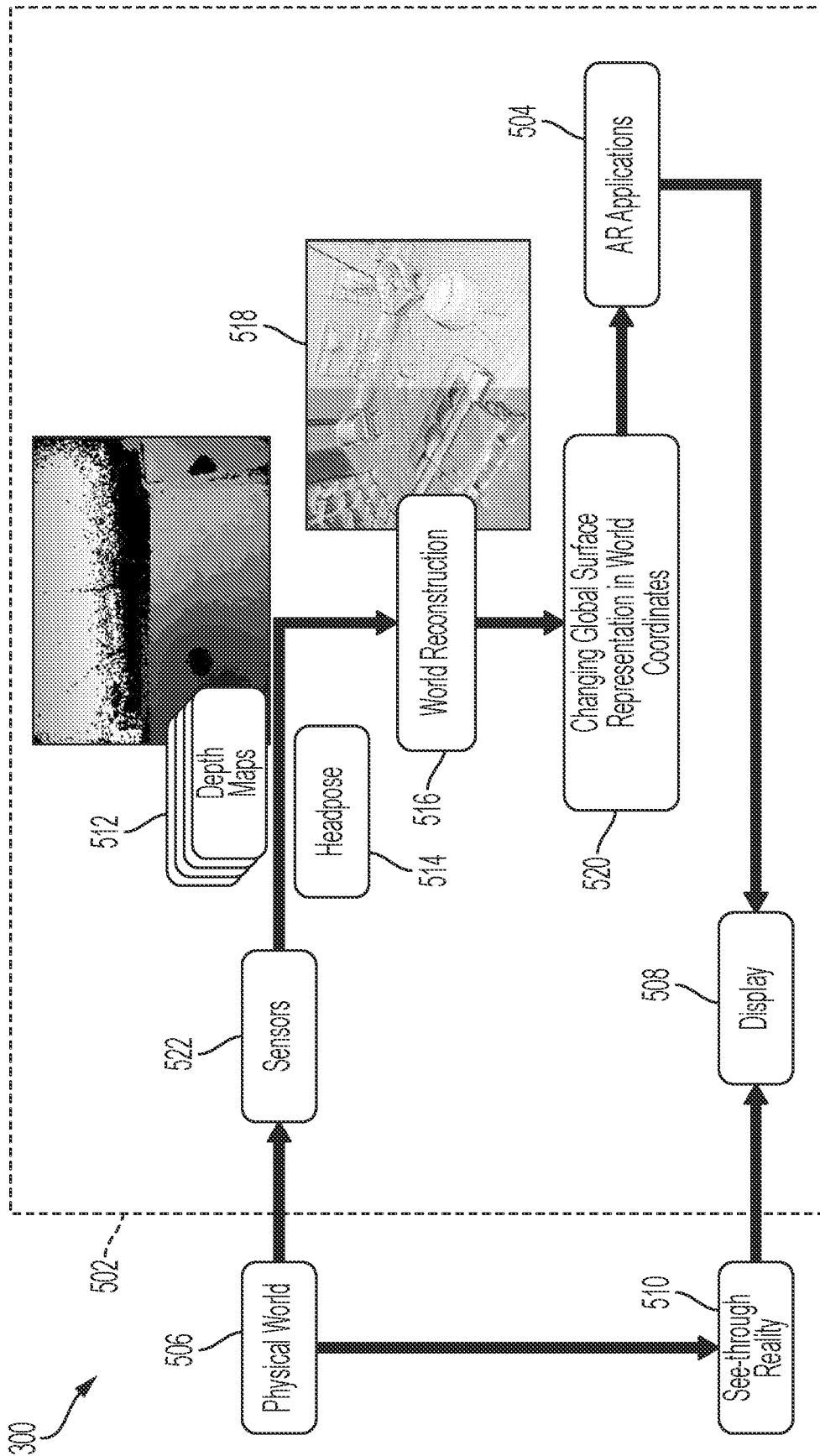


FIG. 3

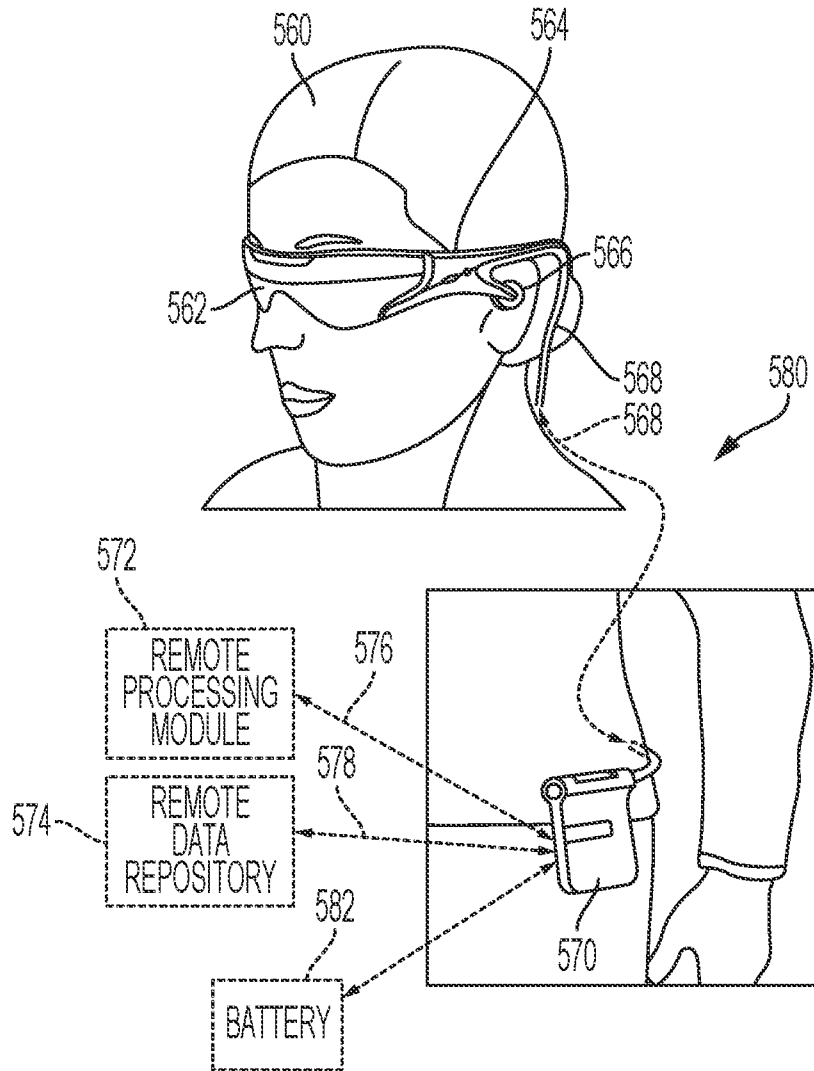


FIG. 4

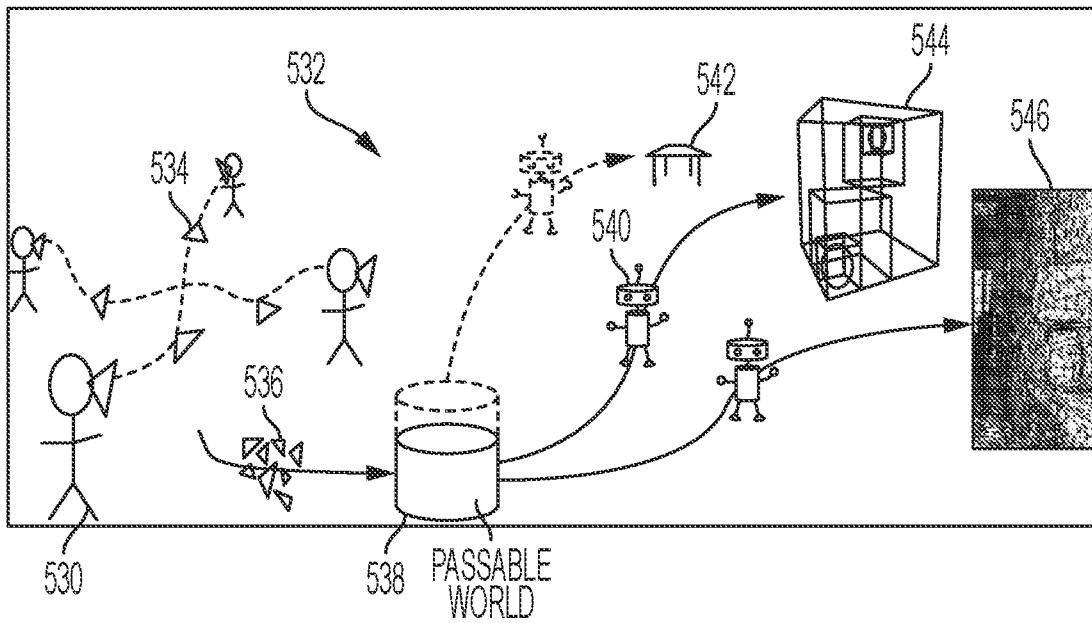


FIG. 5A

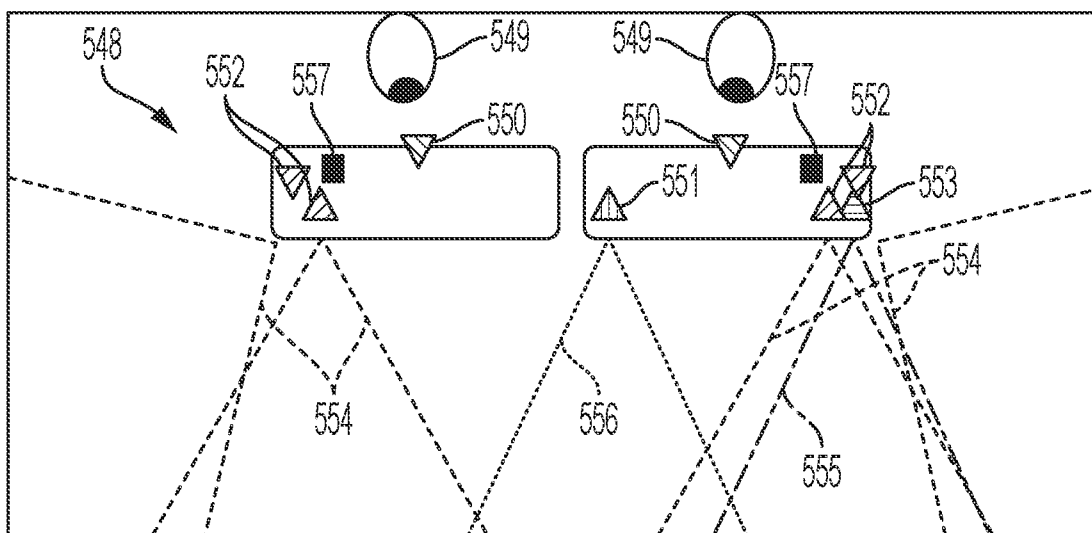


FIG. 5B

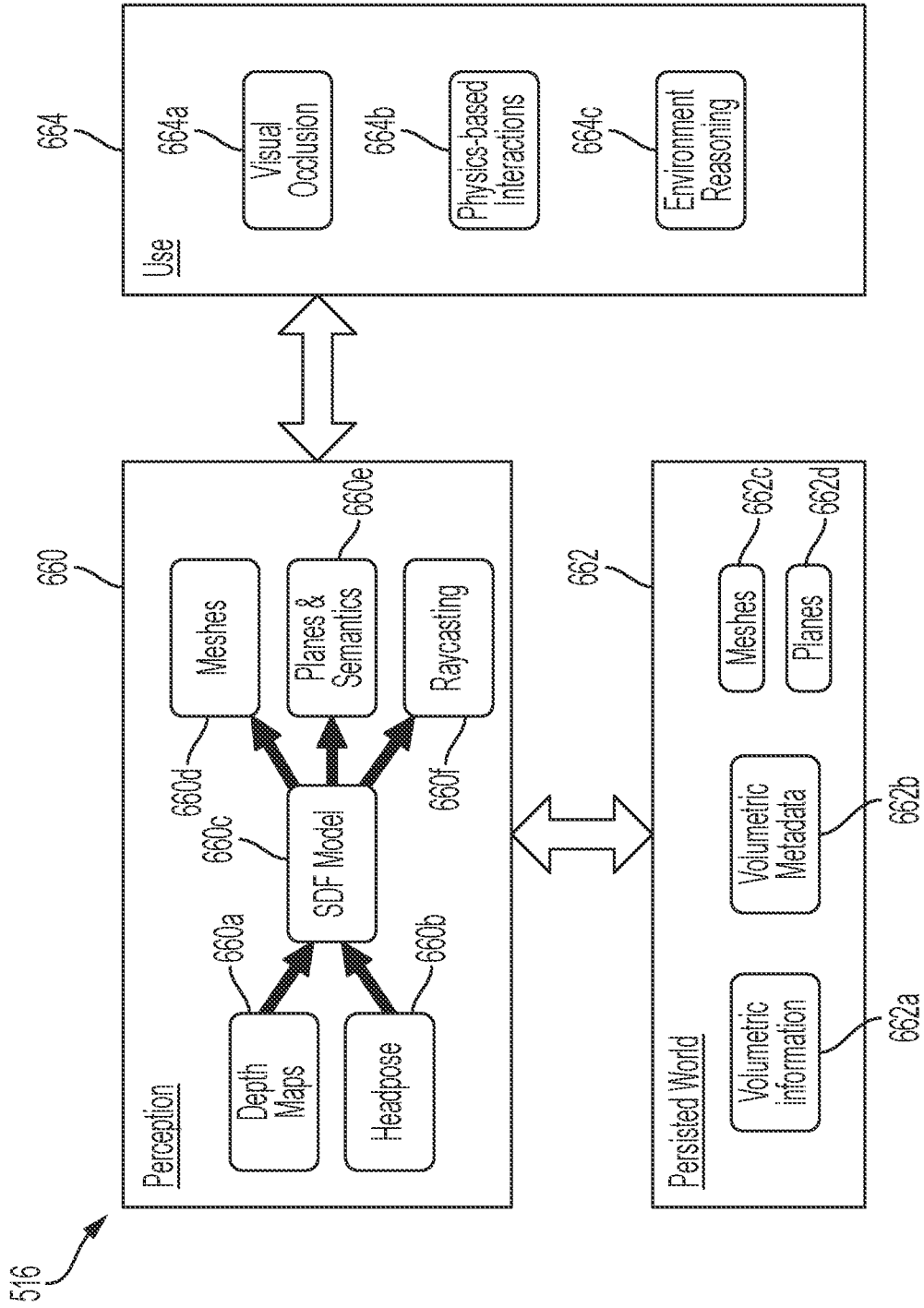


FIG. 6A

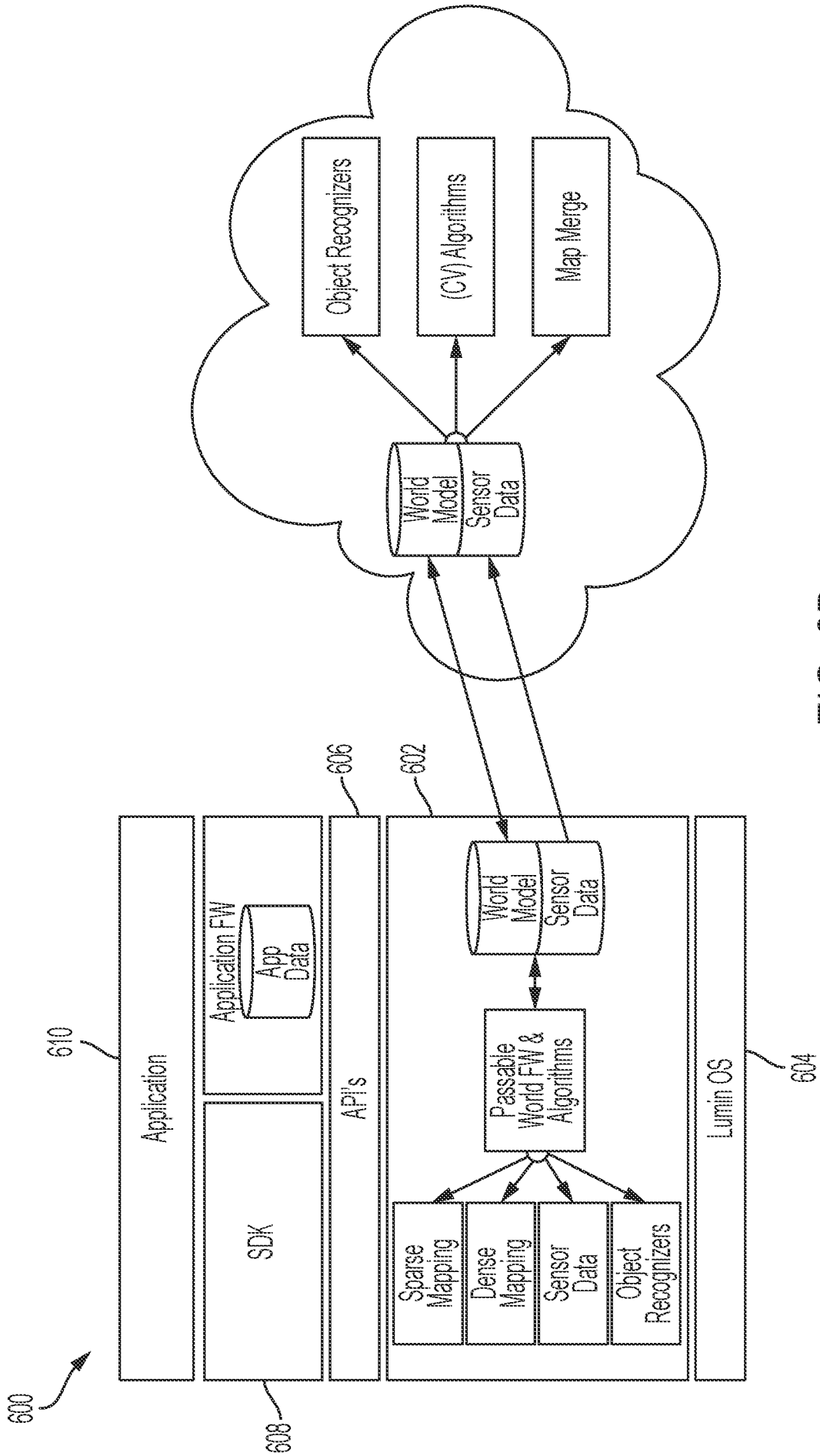


FIG. 6B

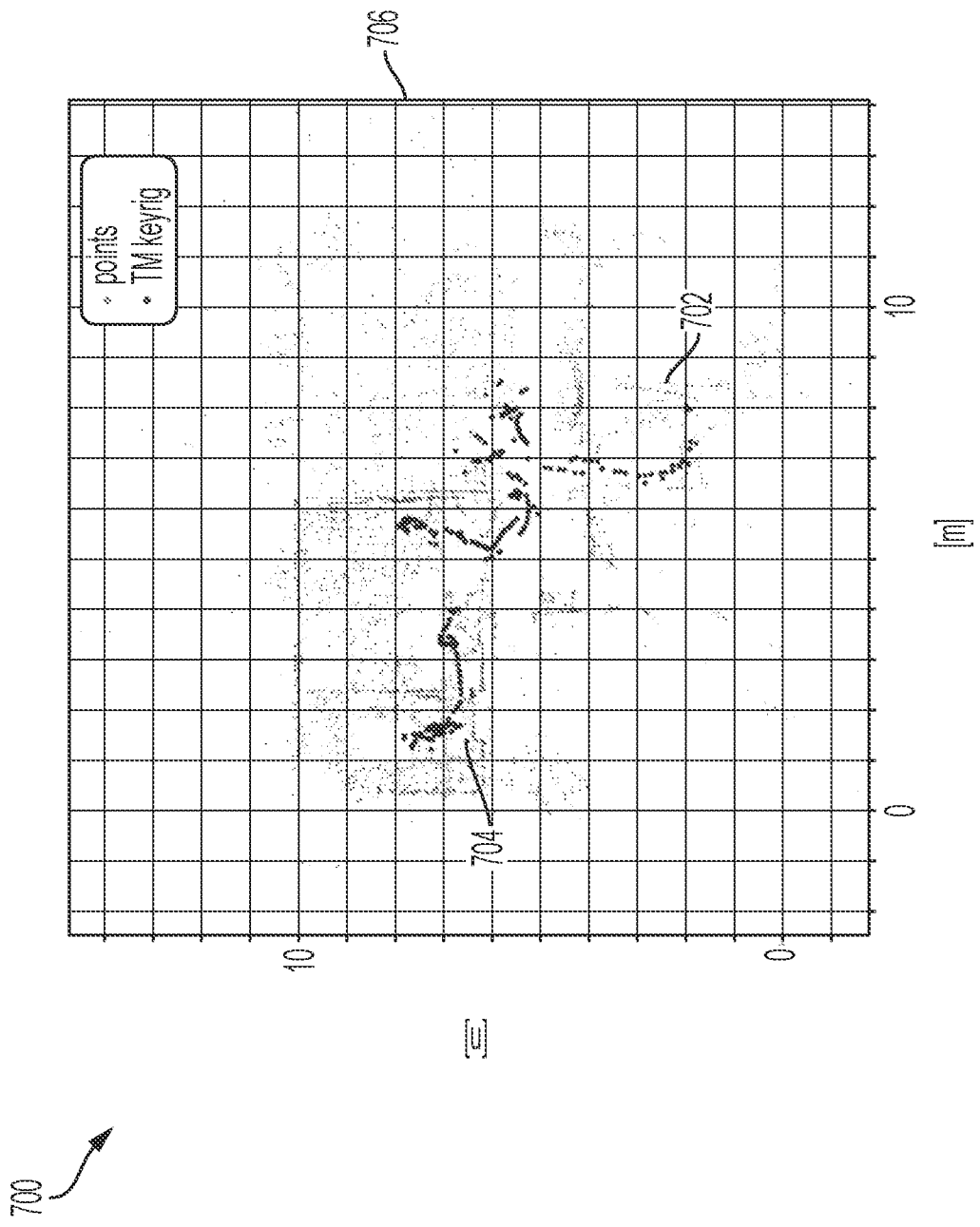


FIG. 7



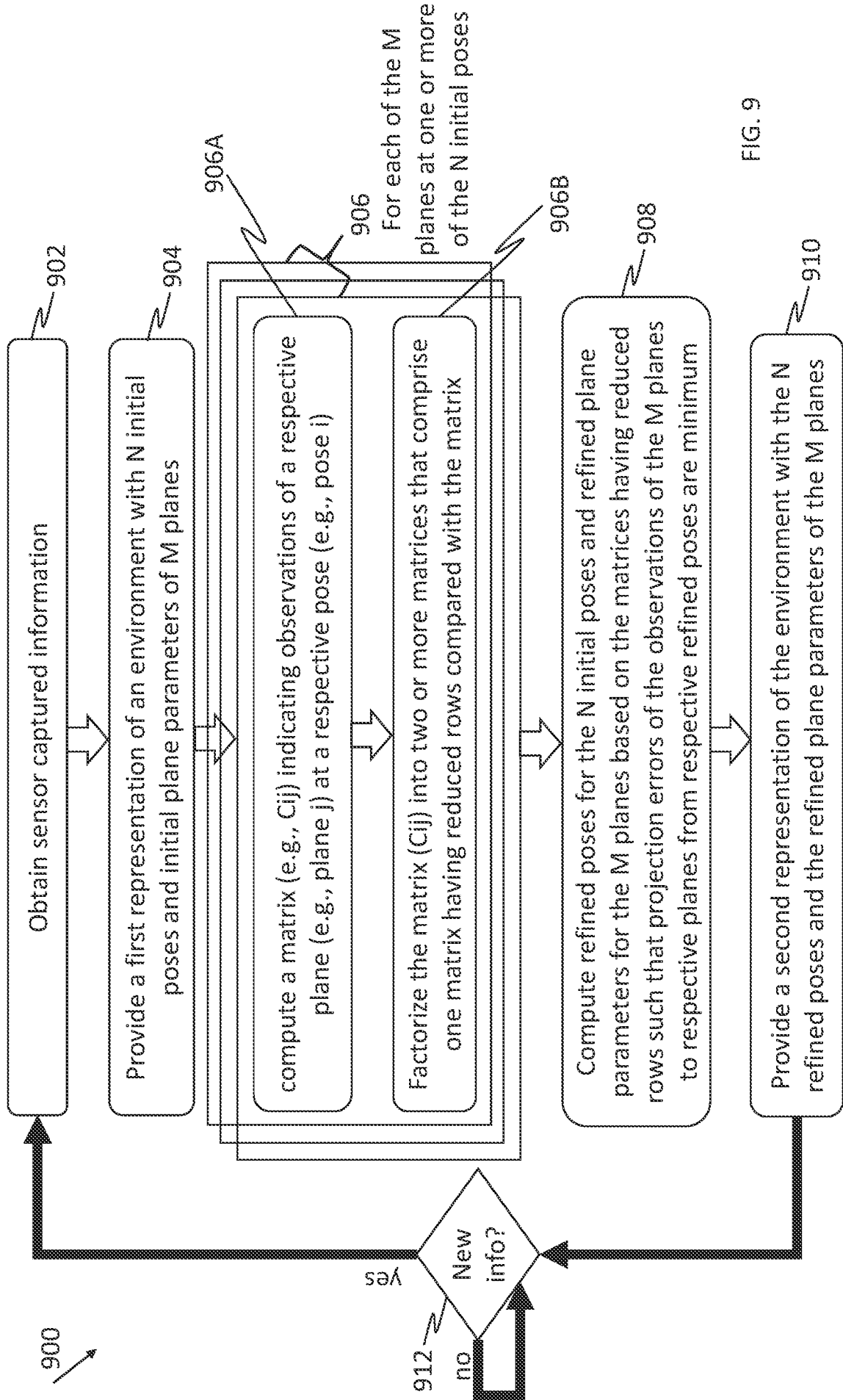


FIG. 9

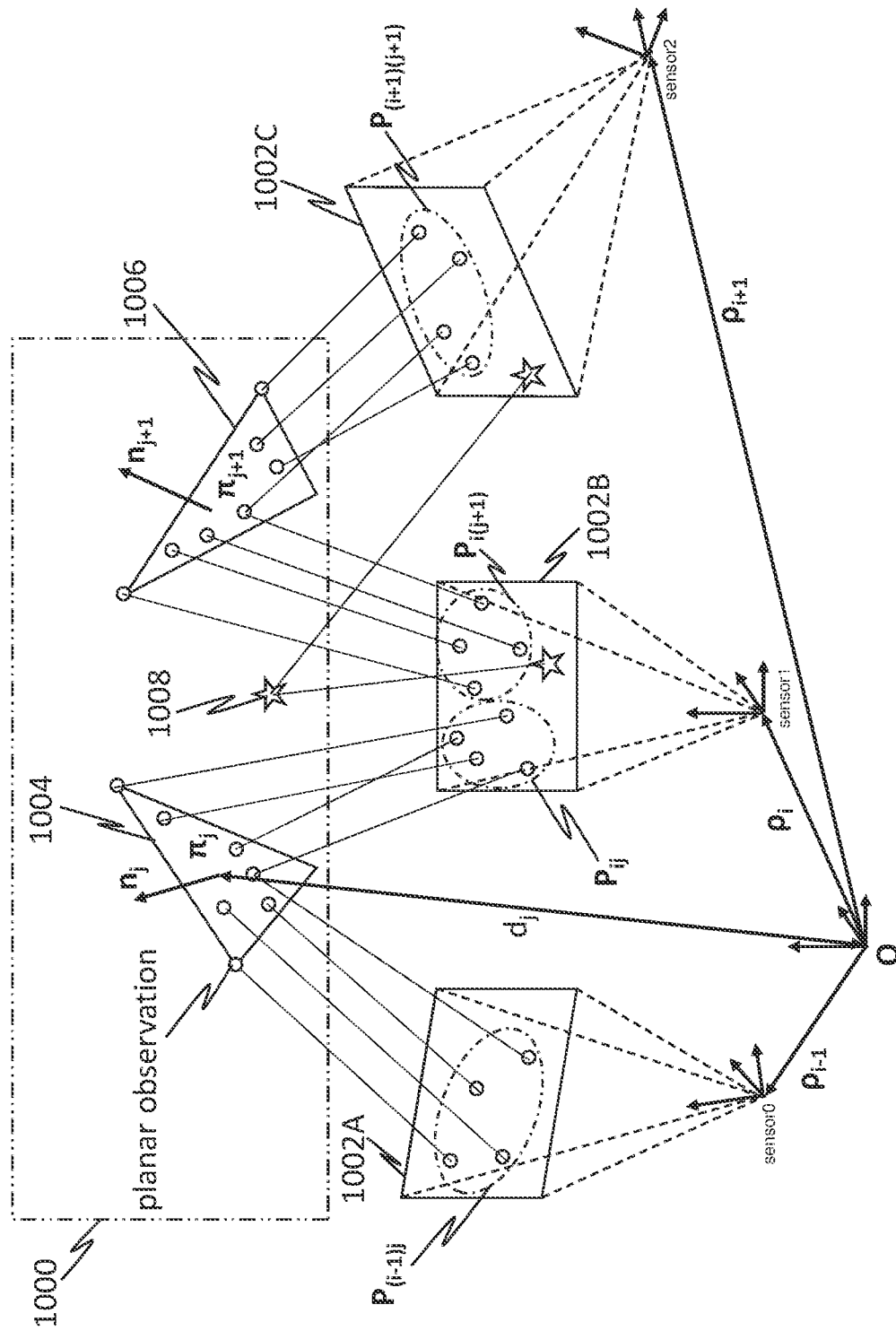


FIG. 10

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/US 21/31514

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC - G06K 9/00, H04N 13/02 (2021.01)

CPC - G06T 17/10, G06T 2207/30244, G01S 5/163, H04N 13/261, G06T 17/00

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)

See Search History document

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

See Search History document

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

See Search History document

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	- Coorg "Pose Imagery and Automated Three-Dimensional Modeling of Urban Environments." In: Massachusetts Institute of Technology. Department of Electrical Engineering and Computer Science, 1998, [online] [retrieved on 13 July 2021 (13.07.2021)] Retrieved from the Internet < URL: <a href="http://hdl.handle.net/1721.1/9630">http://hdl.handle.net/1721.1/9630</a> >, entire document	1-20
Y	- Thrun et al. "The GraphSLAM Algorithm with Applications to Large-Scale Mapping of Urban Structures." In: The International Journal of Robotics Research, Vol. 25, No. 5-6, May-June 2006, [online] [retrieved on 13 July 2021 (13.07.2021)] Retrieved from the Internet < URL: <a href="https://journals.sagepub.com/doi/10.1177/0278364906065387">https://journals.sagepub.com/doi/10.1177/0278364906065387</a> >, entire document	1-20
Y	- Alismail "Direct Pose Estimation and Refinement." In: The Robotics Institute Carnegie Mellon University Pittsburgh, 2016, [online] [retrieved on 13 July 2021 (13.07.2021)] Retrieved from the Internet < URL: <a href="https://www.cs.cmu.edu/~halismail/h_alismail_robotics_2016.pdf">https://www.cs.cmu.edu/~halismail/h_alismail_robotics_2016.pdf</a> >, entire document	1-20
A	US 2001/0043738 A1 (Sawhney et al.), 22 November 2001 (22.11.2001), entire document	1-20
A	US 2013/0335528 A1 (Board of Regents of the University of Texas System), 19 December 2013 (19.12.2013), entire document	1-20

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

"D" document cited by the applicant in the international application

"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

14 July 2021

Date of mailing of the international search report

**AUG 13 2021**

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents  
P.O. Box 1450, Alexandria, Virginia 22313-1450  
Facsimile No. 571-273-8300

Authorized officer

Kari Rodriguez

Telephone No. PCT Helpdesk: 571-272-4300