



(51) International Patent Classification:

G06F 3/01 (2006.01) G06N 3/02 (2006.01)
G06F 3/0346 (2013.01)

(21) International Application Number:

PCT/US2021/065026

(22) International Filing Date:

22 December 2021 (22.12.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/131,242 28 December 2020 (28.12.2020) US
17/555,126 17 December 2021 (17.12.2021) US

(71) Applicant: **META PLATFORMS TECHNOLOGIES, LLC** [US/US]; 1601 Willow Road, Menlo Park, California 94025 (US).

(72) Inventors: **SENKAL, Doruk**; 1601 Willow Road, Menlo Park, California 94025 (US). **SHEN, Sheng**; 1601 Willow Road, Menlo Park, California 94025 (US).

(74) Agent: **COLBY, Steven et al.**; Rimon Law, 420 West Main Street, Suite 101B, Boise, Idaho 83702 (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, 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).

(54) Title: CONTROLLER POSITION TRACKING USING INERTIAL MEASUREMENT UNITS AND MACHINE LEARNING

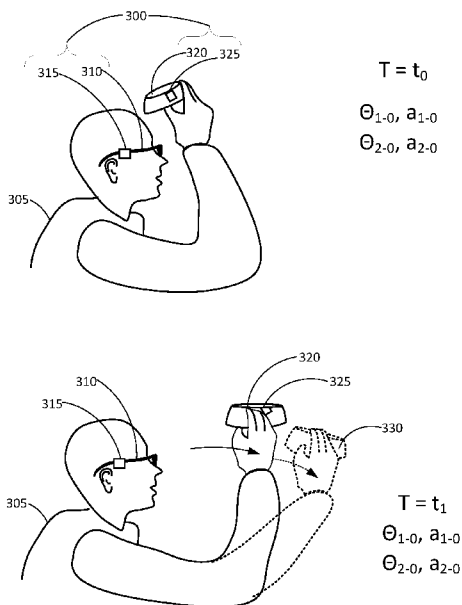


FIG. 3

(57) Abstract: A system having at least a first component and a second component positioned at different locations on a user's body (e.g., on the user's head and held on the user's hand). Each component includes at least one inertial measurement unit (IMU) configured generate measurements indicating acceleration and angular rate data. The generated measurements of the IMUs are used with ground truth information indicating the positions of the first and second component to generate a set of training data to train a neural network configured to predict a relative position between the first and second components based on IMU measurements received over a predetermined time period. Because the neural network is trained based upon movements of a human user, the neural network model takes into account physiological constraints of the user in determining how the set of potential positions of the different components may change over time, reducing potential error.

WO 2022/146858 A1

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

1 **CONTROLLER POSITION TRACKING USING INERTIAL**
2 **MEASUREMENT UNITS AND MACHINE LEARNING**

3 **FIELD OF THE INVENTION**

4 **[0001]** This disclosure relates generally to artificial reality systems, and more specifically
5 to position tracking for use in for artificial reality systems.

6 **BACKGROUND**

7 **[0002]** An artificial reality system, such as a virtual reality (VR), augmented reality (AR),
8 and/or mixed reality (MR) system, may perform various functions based on a user's
9 movements and/or positions of the user's body. For example, in a VR game, a user may be
10 able to control the game by moving their head (e.g., to view different regions of a displayed
11 virtual world) and hands (e.g., to cause an in-game character to perform certain actions). As
12 the user of the artificial reality system moves their body, the system may determine positions
13 of various portions of the user's body over time, and use the determined positional
14 information to control the artificial reality environment.

15 **[0003]** Because position tracking using IMUs is performed by continually integrating the
16 received measurements (e.g., integrating acceleration with respect to time to calculate
17 velocity, and again to calculate position), IMU-based position tracking may suffer from
18 accumulated error, as any measurement errors are accumulated over time, causing the
19 difference between the determined position and the body's actual position to continually
20 increase. This is also known as "drift error" or "drift", as the determined position may begin
21 to drift farther away from the actual position over time.

22 **SUMMARY**

23 **[0004]** Embodiments are directed to a system having different components (e.g., headset,
24 handheld controller, etc.) positioned at different areas of a user's body. Each component
25 comprises at least one inertial measurement unit (IMU) configured generate measurements
26 indicating acceleration and angular rate data of the component. The generated measurements
27 of the IMUs are used as inputs to a trained neural network to predict a relative position
28 between the different components of the system. The neural network is trained based upon
29 training data comprising IMU measurement data generated by a user having the components
30 positioned on their body moving their body, and ground truth information indicating absolute
31 positions of the different components. Because the neural network model is trained based
32 upon movements of a human user, the neural network model is able to take into account the
33 physiological constraints of the user in determining how the set of potential positions of the
34 different components may change over time, reducing potential error that may be caused by
35 drift.

1 [0005] In accordance with a first aspect of the present disclosure, there is provided a
2 system comprising a first component and a second component. The first component is
3 configured to be positioned at a first portion of a user's body, and comprises at least a first
4 IMU configured to generate first measurements indicating motion of the first component,
5 while the second component is configured to be positioned at a second portion of the user's
6 body, and comprises at least a second IMU configured to generate second measurements
7 indicating motion of the second component, wherein a set of potential positions of the first
8 component and the second component is determined by physiological constraints of the first
9 portion and the second portion of the user's body. The system further comprises a controller
10 configured to receive first measurements from the first IMU and the second measurements
11 from the second IMU, and to predict a position of the first component relative to the second
12 component using a trained neural network model and the set of potential positions.

13 [0006] In some embodiments, the first component is a head-mounted device (HMD)
14 configured to be worn on the user's head, and the second component is a handheld device
15 configured to be held in a hand of the user.

16 [0007] In some embodiments, the position of the first component relative to the second
17 component predicted by the trained neural network corresponds to an expected position of the
18 first component relative to the second component corresponding to a future time.

19 [0008] In some embodiments, the controller is configured to receive ground truth
20 information for training the neural network model from a camera system configured to
21 capture images of the first component and the second component.

22 [0009] In some embodiments, the ground truth information comprises position
23 information of the first component and the second component corresponding to a time after a
24 time period during which measurements from the first IMU and the second IMU are received.

25 [0010] In some embodiments, the controller is further configured to periodically receive
26 information indicating positions of the first and second components from an additional
27 sensor, and to periodically use the received information to correct an error of the predicted
28 position.

29 [0011] In some embodiments, the trained neural network model is configured to: receive
30 a set of first measurements from the first IMU and a set of second measurements from the
31 second IMU corresponding to a predetermined period of time; and predict a motion of the
32 first component relative to the second component over the predetermined period of time;
33 wherein the controller is further configured to predict the position of the first component
34 relative to the second component by updating an initial position based upon the motion
35 predicted by the trained neural network model.

1 [0012] In some embodiments, the trained neural network is configured to: receive a set of
2 first measurements from the first IMU and a set of second measurements from the second
3 IMU corresponding to a predetermined period of time; and predict an expected motion of the
4 first component relative to the second component over a future time period following the
5 predetermined period of time; wherein the controller is further configured to predict the
6 position of the first component relative to the second component by updating an initial
7 position based upon the motion predicted by the trained neural network model, and wherein
8 the position corresponds to an expected future position of the first component relative to the
9 second component.

10 [0013] In some embodiments, the trained neural network is configured to: receive
11 velocity information indicating a current velocity of the first component relative to the second
12 component; extrapolate an expected future position estimate using the received velocity
13 information corresponding to a future point in time; predict an expected offset from the
14 expected future position estimate corresponding to the future point in time; and apply the
15 expected offset to the expected future position estimate to predict the expected future position
16 of the first component relative to the second component at the future point in time.

17 [0014] In some embodiments, the trained neural network model is a long short-term
18 memory (LSTM) network which when unrolled matches a length of a set of first
19 measurements from the first IMU and a set of second measurements from the second IMU
20 corresponding to a predetermined period of time.

21 [0015] In some embodiments, the trained neural network model is a convolutional neural
22 network (CNN) model.

23 [0016] In some embodiments, the controller is further configured to: track positions of
24 the first component and second component based upon received information indicating
25 positions of the first and second components from an additional sensor; compare the
26 predicted position of the first component relative to the second component predicted using
27 the trained neural network model to the positions of the first and second components received
28 from the additional sensor; determine whether the additional sensor has lost tracking, based
29 upon the comparison; and responsive to determining that the additional sensor has lost
30 tracking, adjust the tracked positions of the first component and second component based on
31 the predicted position.

32 [0017] In some embodiments, measurements from the first and second IMUs are received
33 over a predetermined period of time, and the predicted position of the first component
34 relative to the second component is determined by predicting a motion of the first component
35 relative to the second component over the predetermined period of time, and updating an

1 initial position based upon the predicted motion. In some embodiments, the position of the
2 first component relative to the second component predicted by the trained neural network
3 corresponds to an expected position of the first component relative to the second component
4 corresponding to a future time, and is determined based upon an expected motion of the first
5 component relative to the second component over a future time period following the
6 predetermined period of time.

7 **[0018]** In accordance with a second aspect of the present disclosure, there is provided a
8 method comprising: receiving, from a first inertial measurement unit (IMU), first
9 measurements indicating motion of a first component, wherein the first component is
10 positioned at a first portions of a user's body; receiving, from a second IMU, second
11 measurements indicating motion of a second component positioned at a second portion of the
12 user's body, wherein a set of potential positions of the first component and the second
13 component is determined by physiological constraints of the first portion and the second
14 portion of the user's body; and predicting a position of the first component relative to the
15 second component using a trained neural network model trained using the set of potential
16 positions.

17 **[0019]** In some embodiments, the first component is a head-mounted device (HMD)
18 configured to be worn on the user's head, and the second component is a handheld device
19 configured to be held in a hand of the user.

20 **[0020]** In some embodiments, the position of the first component relative to the second
21 component predicted by the trained neural network corresponds to an expected position of the
22 first component relative to the second component corresponding to a future time.

23 **[0021]** In some embodiments, the method further comprises, at the trained neural network
24 model: receiving a set of first measurements from the first IMU and a set of second
25 measurements from the second IMU corresponding to a predetermined period of time,
26 predicting a motion of the first component relative to the second component over the
27 predetermined period of time; and predicting the position of the first component relative to
28 the second component by updating an initial position based upon the motion predicted by the
29 trained neural network model.

30 **[0022]** In some embodiments, the method further comprises, at the trained neural network
31 model: receive a set of first measurements from the first IMU and a set of second
32 measurements from the second IMU corresponding to a predetermined period of time,
33 predicting an expected motion of the first component relative to the second component over a
34 future time period following the predetermined period of time; and predicting the position of
35 the first component relative to the second component by updating an initial position based

1 upon the motion predicted by the trained neural network model, wherein the position
2 corresponds to an expected future position of the first component relative to the second
3 component.

4 **[0023]** In some embodiments, the method further comprises tracking positions of the first
5 component and second component based upon received information indicating positions of
6 the first and second components from an additional sensor; comparing the predicted position
7 of the first component relative to the second component predicted using the trained neural
8 network model to the positions of the first and second components received from the
9 additional sensor; determining whether the additional sensor has lost tracking, based upon the
10 comparison; and responsive to determining that the additional sensor has lost tracking,
11 adjusting the tracked positions of the first component and second component based on the
12 predicted position.

13 **[0024]** In accordance with a third aspect of the present disclosure, there is provided a
14 computer readable non-transitory storage medium, storing instructions for: receiving, from a
15 first inertial measurement unit (IMU), first measurements indicating motion of a first
16 component, wherein the first component is positioned at a first portions of a user's body;
17 receiving, from a second IMU, second measurements indicating motion of a second
18 component positioned at a second portion of the user's body, wherein a set of potential
19 positions of the first component and the second component is determined by physiological
20 constraints of the first portion and the second portion of the user's body; and predicting a
21 position of the first component relative to the second component using a trained neural
22 network model and the set of potential positions.

23 **[0025]** In some embodiments, the first component is a head-mounted device (HMD)
24 configured to be worn on the user's head, and the second component is a handheld device
25 configured to be held in a hand of the user.

26 **[0026]** In some embodiments, the system corresponds to an artificial reality system,
27 where the first component corresponds to a headset or head-mounted device of the artificial
28 reality system, and the second component corresponds to a device of the artificial reality
29 system positioned on a hand or arm of the user, such as a handheld controller.

30 **[0027]** It will be appreciated that any features described herein as being suitable for
31 incorporation into one or more aspects or embodiments of the present disclosure are intended
32 to be generalizable across any and all aspects and embodiments of the present disclosure.
33 Other aspects of the present disclosure can be understood by those skilled in the art in light of
34 the description, the claims, and the drawings of the present disclosure. The foregoing general

1 description and the following detailed description are exemplary and explanatory only and
2 are not restrictive of the claims.

3 **BRIEF DESCRIPTION OF THE DRAWINGS**

4 **[0028]** FIG. 1A is a high-level diagram of a system in which a position and orientation of
5 the device is tracked using one or more IMUs, in accordance with some embodiments.

6 **[0029]** FIG. 1B is a perspective view of a headset implemented as a head-mounted
7 display, in accordance with one or more embodiments.

8 **[0030]** FIG. 2 is a side view of a handheld controller, in accordance with one or more
9 embodiments.

10 **[0031]** FIG. 3 is a diagram of a user using an artificial reality system configured to
11 determine positional information using IMUs, in accordance with some embodiments.

12 **[0032]** FIG. 4 is a diagram of a neural network model that may be trained to predict
13 relative positions of components of an artificial reality system, in accordance with some
14 embodiments.

15 **[0033]** FIG. 5 is a diagram of another neural network model that may be trained to predict
16 relative positions of components of an artificial reality system, in accordance with some
17 embodiments.

18 **[0034]** FIG. 6A is a flowchart illustrating a process for training a neural network model to
19 predict relative positions of components using IMUs, in accordance with one or more
20 embodiments.

21 **[0035]** FIG. 6B is a flowchart illustrating a process for predicting relative positions of
22 components using IMUs and a trained neural network model, in accordance with some
23 embodiments.

24 **[0036]** FIG. 7 is a system that includes a headset and controller device, in accordance
25 with one or more embodiments.

26 **[0037]** The figures depict various embodiments for purposes of illustration only. One
27 skilled in the art will readily recognize from the following discussion that alternative
28 embodiments of the structures and methods illustrated herein may be employed without
29 departing from the principles described herein.

30 **DETAILED DESCRIPTION**

31 **[0038]** Inertial measurement units (IMUs) are used in many position tracking applications
32 to determine a position of a body. An IMU comprises at least an accelerometer and a
33 gyroscope, and is configured to generate measurements indicating an acceleration (e.g., using
34 the accelerometer) and an angular rate (e.g., using the gyroscope) of the body over time.
35 These measurements may be used to determine changes in position or orientation of the body

1 from an initial position and orientation, e.g., by integrating the angular rate to determine an
2 angular offset from the initial angular position, and by integrating the measured acceleration
3 to determine velocity, which in turn can be integrated to determine a spatial offset from the
4 initial spatial position, in a process known as “dead reckoning” or “inertial navigation.”

5 **[0039]** However, using IMUs for positional tracking may result in accumulated error.
6 Because position is tracked by continually integrating acceleration with respect to time to
7 calculate velocity and position (as well as integrating angular rate with respect to time to
8 calculate angular position), any measurement errors are accumulated over time, causing the
9 difference between the determined position and the body’s actual position to continually
10 increase. This is also known as “drift error” or “drift”, as the determined position may begin
11 to drift farther away from the actual position over time. For a constant angular rate and
12 acceleration error, angular error will increase linearly (due to being determined by integrating
13 angular rate), while spatial position error increases quadratically (due to being determined by
14 integrating acceleration twice).

15 **[0040]** In some systems, in order to compensate for drift error, IMUs are used in
16 conjunction with a second sensor system configured to determine a position of the body, in
17 order to correct for the drift errors of the IMU. The second sensor system may comprise a
18 positional tracking system such as a visual odometry system, a global positioning system
19 (GPS), and/or the like. In some embodiments, because IMU drift error accumulates with
20 time, in order to prevent the drift error of the IMU-based positioned tracking from exceeding
21 a threshold amount, the second sensor system may be configured to generate positional data
22 at least at a threshold frequency for correcting the drift error of the IMUs. For example,
23 where the second sensor system comprises one or more cameras for capturing images of the
24 tracked device and/or surrounding environment for position determination, one or more
25 cameras may be configured to capture images at a relatively high frame rate, so that drift
26 error can be corrected for at a predetermined frequency to maintain performance.

27 **[0041]** Embodiments described herein are directed to a tracking system that tracks
28 portions of a user’s body using IMUs and predicts positional information using a machine
29 learning model or neural network model, in order to reduce or eliminate the need for a second
30 sensor system for correcting drift error. In some embodiments, the tracking system
31 comprises a first IMU on a first component and a second IMU on a second component. The
32 first and second components are positioned at different portions of the user’s body. For
33 example, in some embodiments, the first component comprises a headset or head-mounted
34 device (HMD), while the second component comprises a handheld device, such as controller
35 held in the user’s hand. The tracking system receives measurement data from the first and

1 second IMUs and uses a trained machine learning model or neural network model to
2 determine a relative position of the first component to the second component. For ease of
3 discussion, further discussion below will refer to the user of a neural network model for
4 predicting the relative positions of the first and second components, although it is understood
5 that in other embodiments, a trained machine learning model may be used.

6 **[0042]** The trained neural network model may take into account physiological constraints
7 of the human in predicting the relative position of the first and second components. For
8 example, where the first component is positioned on the user's head, and the second
9 component is positioned on the user's hand, the ways in which the position of the second
10 component can change relative to the first component is constrained by human physiology.
11 The neural network model receives measurements from the first and second IMUs, and is
12 trained to predict the relative position of the first component to the second component given
13 the IMU measurements received over a predetermined time period (e.g., past 1 second, past
14 10 seconds, etc.). Because the neural network model determines a current relative position
15 based on a limited set of measurements from the first and second IMUs, and is trained based
16 on movements of the human body, the drift error of the IMUs can be limited, reducing or
17 eliminating a need to correct the position determined using the IMUs with a second sensor
18 system. In some embodiments, the neural network model determines a current position based
19 on a determined relative motion from a known starting position. However, even though drift
20 error may accumulate when the neural network model is run in an iterative fashion, because
21 the model takes into account human physiological constraints, the amount of drift error in
22 comparison to conventional dead reckoning techniques may be greatly reduced. In some
23 embodiments, the neural network model is used to determine the positions of the first and/or
24 second components when at least one of the first or second component is unable to be tracked
25 using vision-based tracking (e.g., if the user moves a handheld controller device outside the
26 view of the vision tracking cameras, such as behind the head) for a period of time (e.g., 10-20
27 seconds), but is expected to return to a position where vision-based tracking can be resumed.
28 During this "blind" period, the neural network model can be used to determine the relative
29 positions of the components with greater accuracy and/or less accumulation of drift error.

30 **[0043]** In some embodiments, the neural network is configured to predict expected
31 positions of the first and/or second components a certain period of time into the future, e.g.,
32 an expected future position of the second component relative to the first component, with
33 limited change in accuracy. Predicting a future position of the first and/or second component
34 may be performed for latency compensation purposes, e.g., to absorb latency induced by the
35 entire tracking system (IMU data acquisition, wireless communication, processing, rendering

1 etc.). For example, in a system in which a virtual object corresponding to a controller is
2 rendered to the user in an augmented or virtual reality environment, the virtual object may be
3 rendered at a position corresponding to where the controller will be expected to be,
4 compensating for an expected level of latency or lag (e.g., 50ms in the future, to compensate
5 for an expected 50ms of latency). A neural network model based approach can produce
6 better future predictions compared to more rudimentary approaches that rely on extrapolating
7 position changes based on last measured velocity/acceleration, as it can use the knowledge of
8 trajectory over a past time period and how controller has moved in the past after such a
9 trajectory, e.g., based upon patterns of movement of a human user, which are limited by
10 physiological constraints of the human body. In some embodiments, because a latency level
11 of the system may change over time, the neural network is configured to output, instead of
12 relative position information for a specific future point in time, one or more coefficients of a
13 function (e.g., a polynomial function) corresponding to an expected path over which the
14 second component is expected to travel relative to the first component over a future time
15 period. Thus, the same neural network model may be used to determine expected future
16 positions for different amounts of latency.

17 **[0044]** Embodiments of the invention may include or be implemented in conjunction with
18 an artificial reality system. Artificial reality is a form of reality that has been adjusted in
19 some manner before presentation to a user, which may include, e.g., a virtual reality (VR), an
20 augmented reality (AR), a mixed reality (MR), a hybrid reality, or some combination and/or
21 derivatives thereof. Artificial reality content may include completely generated content or
22 generated content combined with captured (e.g., real-world) content. The artificial reality
23 content may include video, audio, haptic feedback, or some combination thereof, any of
24 which may be presented in a single channel or in multiple channels (such as stereo video that
25 produces a three-dimensional effect to the viewer). Additionally, in some embodiments,
26 artificial reality may also be associated with applications, products, accessories, services, or
27 some combination thereof, that are used to create content in an artificial reality and/or are
28 otherwise used in an artificial reality. The artificial reality system that provides the artificial
29 reality content may be implemented on various platforms, including a wearable device (e.g.,
30 headset) connected to a host computer system, a standalone wearable device (e.g., headset), a
31 mobile device or computing system, or any other hardware platform capable of providing
32 artificial reality content to one or more viewers.

33 **[0045]** FIG. 1A is a high-level diagram of a system in which a position of the device is
34 tracked using one or more IMUs, in accordance with some embodiments. The system 100
35 comprises a plurality of devices, including a headset 102 and an additional device 104. Each

1 of the plurality of devices may be configured to be positioned at respective portion of a body
2 of a user. For example, in some embodiments, the headset 102 is worn on the user's head,
3 while the device 104 may be worn at a different position on the user's body (e.g., strapped to
4 the user's chest, leg, or arm, etc.). In other embodiments, the device 104 may be configured
5 to be held in a hand of the user.

6 **[0046]** The headset 102 may be implemented as an eyewear device, in accordance with
7 one or more embodiments. In some embodiments, the headset 102 is a near eye display
8 (NED). In general, the headset 102 may be worn on the face of a user such that content (e.g.,
9 media content) is presented using a display assembly and/or an audio system. However, the
10 headset 102 may also be used such that media content is presented to a user in a different
11 manner. Examples of media content presented by the headset 102 include one or more
12 images, video, audio, or some combination thereof. The headset 102 includes a frame, and
13 may include, among other components, a display assembly including one or more display
14 elements 120, a depth camera assembly (DCA), an audio system, and an IMU 190. While
15 FIG. 1A illustrates the components of the headset 102 in example locations on the headset
16 102, the components may be located elsewhere on the headset 102, on a peripheral device
17 paired with the headset 102, or some combination thereof. Similarly, there may be more or
18 fewer components on the headset 102 than what is shown in FIG. 1B.

19 **[0047]** The frame 110 holds the other components of the headset 102. The frame 110
20 includes a front part that holds the one or more display elements 120 and end pieces (e.g.,
21 temples) to attach to a head of the user. The front part of the frame 110 bridges the top of a
22 nose of the user. The length of the end pieces may be adjustable (e.g., adjustable temple
23 length) to fit different users. The end pieces may also include a portion that curls behind the
24 ear of the user (e.g., temple tip, ear piece).

25 **[0048]** The one or more display elements 120 provide light to a user wearing the headset
26 102. As illustrated the headset includes a display element 120 for each eye of a user. In
27 some embodiments, a display element 120 generates image light that is provided to an
28 eyebox of the headset 102. The eyebox is a location in space that an eye of user occupies
29 while wearing the headset 102. For example, a display element 120 may be a waveguide
30 display. A waveguide display includes a light source (e.g., a two-dimensional source, one or
31 more line sources, one or more point sources, etc.) and one or more waveguides. Light from
32 the light source is in-coupled into the one or more waveguides which outputs the light in a
33 manner such that there is pupil replication in an eyebox of the headset 102. In-coupling
34 and/or outcoupling of light from the one or more waveguides may be done using one or more
35 diffraction gratings. In some embodiments, the waveguide display includes a scanning

1 element (e.g., waveguide, mirror, etc.) that scans light from the light source as it is in-coupled
2 into the one or more waveguides. Note that in some embodiments, one or both of the display
3 elements 120 are opaque and do not transmit light from a local area around the headset 102.

4 The local area is the area surrounding the headset 102. For example, the local area may be a
5 room that a user wearing the headset 102 is inside, or the user wearing the headset 102 may
6 be outside and the local area is an outside area. In this context, the headset 102 generates VR
7 content. Alternatively, in some embodiments, one or both of the display elements 120 are at
8 least partially transparent, such that light from the local area may be combined with light
9 from the one or more display elements to produce AR and/or MR content.

10 **[0049]** In some embodiments, a display element 120 does not generate image light, and
11 instead is a lens that transmits light from the local area to the eyebox. For example, one or
12 both of the display elements 120 may be a lens without correction (non-prescription) or a
13 prescription lens (e.g., single vision, bifocal and trifocal, or progressive) to help correct for
14 defects in a user's eyesight. In some embodiments, the display element 120 may be polarized
15 and/or tinted to protect the user's eyes from the sun.

16 **[0050]** In some embodiments, the display element 120 may include an additional optics
17 block (not shown). The optics block may include one or more optical elements (e.g., lens,
18 Fresnel lens, etc.) that direct light from the display element 120 to the eyebox. The optics
19 block may, e.g., correct for aberrations in some or all of the image content, magnify some or
20 all of the image, or some combination thereof.

21 **[0051]** The DCA determines depth information for a portion of a local area surrounding
22 the headset 102. The DCA includes one or more imaging devices 130 and a DCA controller
23 (not shown in FIG. 1A), and may also include an illuminator 140. In some embodiments, the
24 illuminator 140 illuminates a portion of the local area with light. The light may be, e.g.,
25 structured light (e.g., dot pattern, bars, etc.) in the infrared (IR), IR flash for time-of-flight,
26 etc. In some embodiments, the one or more imaging devices 130 capture images of the
27 portion of the local area that include the light from the illuminator 140. As illustrated, FIG.
28 1A shows a single illuminator 140 and two imaging devices 130. In alternate embodiments,
29 there is no illuminator 140 and at least two imaging devices 130.

30 **[0052]** The DCA controller computes depth information for the portion of the local area
31 using the captured images and one or more depth determination techniques. The depth
32 determination technique may be, e.g., direct time-of-flight (ToF) depth sensing, indirect ToF
33 depth sensing, structured light, passive stereo analysis, active stereo analysis (uses texture
34 added to the scene by light from the illuminator 140), some other technique to determine
35 depth of a scene, or some combination thereof.

1 [0053] The audio system provides audio content. The audio system includes a transducer
2 array, a sensor array, and an audio controller 150. However, in other embodiments, the audio
3 system may include different and/or additional components. Similarly, in some cases,
4 functionality described with reference to the components of the audio system can be
5 distributed among the components in a different manner than is described here. For example,
6 some or all of the functions of the controller may be performed by a remote server.

7 [0054] The transducer array presents sound to user. The transducer array includes a
8 plurality of transducers. A transducer may be a speaker 160 or a tissue transducer 170 (e.g., a
9 bone conduction transducer or a cartilage conduction transducer). Although the speakers 160
10 are shown exterior to the frame 110, the speakers 160 may be enclosed in the frame 110. In
11 some embodiments, instead of individual speakers for each ear, the headset 102 includes a
12 speaker array comprising multiple speakers integrated into the frame 110 to improve
13 directionality of presented audio content. The tissue transducer 170 couples to the head of
14 the user and directly vibrates tissue (e.g., bone or cartilage) of the user to generate sound.
15 The number and/or locations of transducers may be different from what is shown in FIG. 1A.

16 [0055] The sensor array detects sounds within the local area of the headset 102. The
17 sensor array includes a plurality of acoustic sensors 180. An acoustic sensor 180 captures
18 sounds emitted from one or more sound sources in the local area (e.g., a room). Each
19 acoustic sensor is configured to detect sound and convert the detected sound into an
20 electronic format (analog or digital). The acoustic sensors 180 may be acoustic wave sensors,
21 microphones, sound transducers, or similar sensors that are suitable for detecting sounds.

22 [0056] In some embodiments, one or more acoustic sensors 180 may be placed in an ear
23 canal of each ear (e.g., acting as binaural microphones). In some embodiments, the acoustic
24 sensors 180 may be placed on an exterior surface of the headset 102, placed on an interior
25 surface of the headset 102, separate from the headset 102 (e.g., part of some other
26 component), or some combination thereof. The number and/or locations of acoustic sensors
27 180 may be different from what is shown in FIG. 1A. For example, the number of acoustic
28 detection locations may be increased to increase the amount of audio information collected
29 and the sensitivity and/or accuracy of the information. The acoustic detection locations may
30 be oriented such that the microphone is able to detect sounds in a wide range of directions
31 surrounding the user wearing the headset 102.

32 [0057] The audio controller 150 processes information from the sensor array that
33 describes sounds detected by the sensor array. The audio controller 150 may comprise a
34 processor and a computer-readable storage medium. The audio controller 150 may be
35 configured to generate direction of arrival (DOA) estimates, generate acoustic transfer

1 functions (e.g., array transfer functions and/or head-related transfer functions), track the
2 location of sound sources, form beams in the direction of sound sources, classify sound
3 sources, generate sound filters for the speakers 160, or some combination thereof.

4 **[0058]** Each of the headset 102 and the additional device 104 of the system 100 may
5 comprise at least one inertial measurement unit (IMU). For example, as shown in FIG. 1A,
6 the device 104 comprises an IMU 105, and the headset comprises an IMU 190. Each of the
7 IMUs 105 and 190 includes at least one or more accelerometers and one or more gyroscopes
8 configured to measure an acceleration and angular rate caused by movement of the device
9 104 or headset 102. In some embodiments, the device 104 or the headset 102 comprises a
10 controller configured to determine positional information based upon measurements
11 generated by the IMUs (e.g., using a neural network model configured to predict a relative
12 position of the device 104 to the headset 102, discussed in greater detail below). In other
13 embodiments, the neural network model is maintained on a separate console (not shown), and
14 device 104 and the headset 102 are configured to transmit IMU measurements to the separate
15 console which determines the positional information based on the received information using
16 the neural network model.

17 **[0059]** As shown in FIG. 1A, the IMU 190 of the headset 102 may be located on a
18 portion of the frame 110 of the headset 102, allowing for the one or more accelerometers and
19 one or more gyroscopes of the IMU to measure an acceleration and angular rate caused by
20 movement of the headset 102.

21 **[0060]** In some embodiments, the headset 102 and/or the device 104 may further
22 comprise other types of sensors, such as one or more magnetometers, another suitable type of
23 sensor that detects motion, a type of sensor used for error correction of the IMU, or some
24 combination thereof. For example, in some embodiments, the headset 102 and/or the device
25 104 may include a pressure sensor or altimeter to measure a height of the headset and/or
26 device, a gyroscopic sensor or a magnetometer to measure an orientation of the headset
27 and/or device, an ultrasound rangefinder, an ultra-wideband sensor, and/or the like.

28 **[0061]** In some embodiments, the headset 102 may provide for simultaneous localization
29 and mapping (SLAM) for a position of the headset 102 and updating of a model of the local
30 area. For example, the headset 102 may include a passive camera assembly (PCA) that
31 generates color image data. The PCA may include one or more RGB cameras that capture
32 images of some or all of the local area. In some embodiments, some or all of the imaging
33 devices 130 of the DCA may also function as the PCA. The images captured by the PCA and
34 the depth information determined by the DCA may be used to determine parameters of the
35 local area, generate a model of the local area, update a model of the local area, or some

1 combination thereof. Furthermore, the IMU 190 may be used to track the position (e.g.,
2 location and pose) of the headset 102 within the room. Additional details regarding the
3 components of the headset 102 are discussed below in connection with FIG 6.

4 **[0062]** FIG. 1B is a perspective view of a headset 106 implemented as a HMD, in
5 accordance with one or more embodiments. In embodiments that describe an AR system
6 and/or a MR system, portions of a front side of the HMD are at least partially transparent in
7 the visible band (~380 nm to 750 nm), and portions of the HMD that are between the front
8 side of the HMD and an eye of the user are at least partially transparent (e.g., a partially
9 transparent electronic display). The HMD includes a front rigid body 115 and a band 175.
10 The headset 106 includes many of the same components described above with reference to
11 FIG. 1A, but modified to integrate with the HMD form factor. For example, the HMD
12 includes a display assembly, a DCA, an audio system, and an IMU 190. FIG. 1B shows the
13 illuminator 140, a plurality of the speakers 160, a plurality of the imaging devices 130, a
14 plurality of acoustic sensors 180, and the IMU 190. The speakers 160 may be located in
15 various locations, such as coupled to the band 175 (as shown), coupled to front rigid body
16 115, or may be configured to be inserted within the ear canal of a user.

17 **[0063]** In some embodiments, a user of an artificial reality system, in addition to wearing
18 a head-mounted device such as the headset 102 of FIG. 1A or the headset 106 of FIG. 1B,
19 may also be associated with one or more additional devices located at different positions on
20 the user's body. For example, in some embodiments, the user may use a controller device to
21 perform various actions within the artificial reality environment. In some embodiments, the
22 controller device comprises a handheld controller held in the user's hand.

23 **[0064]** FIG. 2 is a side view of a handheld controller 200, in accordance with one or more
24 embodiments. The handheld controller 200 may be an embodiment of the device 104
25 illustrated in FIG. 1A, and is a controller device configured to be held in a hand of the user.
26 The handheld controller 200 includes one or more interface elements configured to receive
27 input from a user of the handheld controller 200, such as a button assembly 210, a trigger
28 assembly 220, or a combination thereof. The user may use the interface elements to perform
29 action requests and receive responses from a console of the system. An action request is a
30 request to perform a particular action. For example, an action request may be an instruction
31 to start or end capture of image or video data, or an instruction to perform a particular action
32 within an application. The handheld controller 200 additionally comprises a position sensor
33 corresponding to an IMU 260. The IMU 260 of the handheld controller 200 may be
34 correspond to the IMU 105 illustrated in FIG. 1A. In other embodiments, the handheld
35 controller 100 contains additional components than those listed, e.g., the handheld controller

1 200 may have additional interface elements, multiple IMUs 260, and/or other types of
2 sensors. In the embodiment of FIG. 2, the handheld controller 200 is configured to operate
3 with one hand of the user; however, in other embodiments, a second handheld controller with
4 mirror symmetry in relation to the handheld controller 200 may be configured to operate with
5 the other hand of the user.

6 **[0065]** The button assembly 210 contains one or more buttons that each receive input
7 when at least partially depressed. Each button of the button assembly 210 translates along
8 one or more axes. In some embodiments, a button translates along a single axis. Translation
9 of a button along the one or more axes may correspond to various inputs. In one
10 embodiment, a button only receives an input when the button is depressed by a threshold
11 distance corresponding to a threshold translation of the button along an axis. In some
12 embodiments, a button of the button assembly 210 may comprise a capacitive sensor capable
13 of detecting a proximity of the user's skin to the button. The button assembly 210 may
14 comprise a multitude of buttons each configured to receive one or more inputs. In one
15 embodiment, the button assembly comprises one or more buttons configured as a directional
16 pad (D-pad) with each button of the D-pad corresponding to a directional input.

17 **[0066]** The trigger assembly 220 contains one or more triggers that each receive input
18 when depressed. Each trigger of the trigger assembly 220 is fixed at a point providing
19 rotational movement of the trigger about the fixed point. In one embodiment, a trigger only
20 receives an input when the trigger is rotated more than some threshold degree. In another
21 embodiment, a trigger may detect a range of angular rotation of the trigger. In some
22 embodiments, each trigger is a capacitive sensor capable of detecting a proximity of the
23 user's skin to the trigger.

24 **[0067]** The IMU 260 is configured to generate one or more measurement signals in
25 response to motion of the headset 102. The IMU 260 may correspond to the IMU 105
26 illustrated in FIG. 1A, and be located on a structure of the handheld controller 200, allowing
27 for the one or more accelerometers and one or more gyroscopes of the IMU to measure an
28 acceleration and angular rate caused by movement of the handheld controller 200. In some
29 embodiments, the IMU 260 comprises a controller configured to determine positional
30 information based upon measurements generated by the IMU. In other embodiments, the
31 IMU 260 is configured to transmit sensor information (e.g., IMU measurements) to a separate
32 device or console which determines the positional information of the handheld controller 200
33 based on the received information.

34 **[0068]** FIG. 3 is a diagram of a user using an artificial reality system configured to
35 determine positional information using IMUs, in accordance with some embodiments. As

1 shown in FIG. 3, a user 305 of an artificial reality system 300 may have a plurality of
2 components positioned at different portions of their body, including a first component
3 corresponding to an HMD 310 worn on the user's head, and a second component
4 corresponding to a handheld controller 320 held in the user's hand. The first
5 component/HMD 310 and second component/handheld controller 320 may correspond to
6 devices of the system 100 as illustrated in FIG. 1A, e.g., the HMD 310 corresponding to the
7 headset 102 or 106 of FIG. 1A or 1B, and the handheld controller 320 corresponding to the
8 device 104 of FIG. 1A or the handheld controller 200 of FIG. 2.

9 **[0069]** The HMD 310 comprises a first IMU 315, and the handheld controller 320
10 comprises a second IMU 325. It is understood that although FIG. 3 illustrates the user 305
11 having components positioned at certain parts of the user's body (e.g., head and hand), in
12 other embodiments, the first and second components may correspond to other types of
13 components of the artificial reality system positioned at other portions of the user's body.
14 For example, in some embodiments, the user 305 may have a component containing an IMU
15 strapped to the user's chest in addition to or instead of a component worn on the user's head
16 or held in the user's hand. In some embodiments, the second component 320 may correspond
17 to a watch or other component worn on the user's wrist instead of held in the user's hand. In
18 some embodiments, at least one of the first or second component may correspond to a
19 component strapped to the user's thigh or leg, e.g., an ankle bracelet.

20 **[0070]** In addition, in some embodiments, the artificial reality system 300 may comprise
21 an external console (not shown) separate from the components positioned on the user's body.
22 For example, in some embodiments the console is configured to receive data from the HMD
23 310 and/or handheld controller 320 (e.g., IMU measurements, action requests, etc.), and to
24 provide content to the HMD 310 and/or handheld controller 320 responsive to the received
25 data (e.g., visual or audio content to the HMD 310, haptic feedback to the handheld controller
26 320, etc.).

27 **[0071]** When components of the artificial reality system are attached to portions of the
28 user's body (e.g., worn on the user's head and/or held in the user's hand as shown in FIG. 3),
29 the movement and relative positions of the components are constrained by human
30 physiology. For example, the possible positions of the HMD 310 are limited by the possible
31 positions of the head of the user. For example, the user has a limited range of motion of their
32 head which is constrained by, e.g., the range of motion of their neck relative to their torso. In
33 addition, movement of the handheld controller 320 relative to the HMD 310 is constrained by
34 the range of motion of the user's arm. For example, the maximum distance to which the
35 handheld controller 320 may move from the HMD 310 is constrained based upon a length of

1 the user's arm and neck. In addition, paths on which the handheld controller 320 is able to
2 move relative to the HMD 310 will be limited based upon the degrees of freedom in which
3 the user's arm can move. For example, when a user raises or lowers their hand by bending
4 their forearm, the angular orientation of the handheld controller 320 held in their hand may
5 change due to the rotation of the user's forearm, instead of remaining constant throughout the
6 movement.

7 **[0072]** The first and second IMUs 315 and 325 measure angular rate and acceleration
8 information of the HMD 310 and the handheld controller 320 over time, as the user moves
9 their body when using the artificial reality system. For example, FIG. 3 illustrates that at a
10 time t_0 , the user 305 may be holding their arm out to the side. At the time t_0 , the first IMU
11 315 may measure an initial angular rate θ_{1-0} and an initial acceleration a_{1-0} , while the second
12 IMU 325 measures an initial angular rate θ_{2-0} and an initial acceleration a_{2-0} (the first
13 subscript corresponding to IMU, e.g., the first or second IMU, and the second subscript
14 corresponding to time). At a later time (e.g., time t_1), the user 305 has moved their arm
15 downwards, causing a change in the position and orientation of the handheld controller 320
16 relative to the HMD 310. At the time t_1 , the first IMU 315 may measure an angular rate θ_{1-1}
17 and an acceleration a_{1-1} , while the second IMU 325 measures an angular rate θ_{2-1} and an
18 acceleration a_{2-1} . In some embodiments, the accelerometer and gyroscope of each IMU
19 generates acceleration and angular rate data, which is sampled by the IMU to generate the
20 IMU measurements (e.g., 100 samples per second). As discussed above, these rapidly-
21 generated IMU measurements may be integrated over time to estimate the position and
22 angular orientation of the HMD 310 and handheld controller 325. However, any errors in the
23 IMU measurements may build up over time, causing "drift error."

24 **[0073]** In some embodiments, to limit the effects of drift error, the position of the first
25 and second components (e.g., handheld controller 320 to the HMD 310) can be determined
26 using a model constructed based upon the user's body, e.g., the user's arm. In other
27 embodiments involving components positioned at other portions of the user's body, the
28 model may be constructed based upon other portions of the user's body, such as the user's
29 neck, torso, head, leg, etc. Because the model is based upon the structure of the user's body,
30 physiological constraints of the user's body are considered when calculating the positions and
31 orientation of the HMD and handheld controller, limiting the amount of drift error that can be
32 accumulated. In some embodiments, the model may correspond to a default model, or may
33 be customized from a default model to match a specific user. However, due to the large
34 number of joints and degrees of freedom within the human body (e.g., within the human

1 arm), manually constructing a model of portions the user's body and determining positions
2 using the model may be impractical from a computational standpoint due to the number of
3 potential calculations involved.

4 **[0074]** In some embodiments, a neural network model is trained to predict a relative
5 position of the handheld controller 320 to the HMD 310 based on received IMU
6 measurements. The neural network model is configured to receive measurements taken by
7 the first and second IMUs over time, and, based upon the IMU measurements received over a
8 predetermined time period (e.g., past 1 second, past 10 seconds, etc.), predict a relative
9 position of the handheld controller 320 relative to the HMD 310. For example, in some
10 embodiments, the neural network model uses a set of IMU measurements including a first set
11 of measurements (θ_{1-0}, a_{1-0}) , (θ_{2-0}, a_{2-0}) corresponding to time t_0 , a second set of
12 measurements (θ_{1-1}, a_{1-1}) , (θ_{2-1}, a_{2-1}) corresponding to time t_1 , and any intervening
13 measurements, and predicts a motion of the handheld controller 320 relative to the HMD 310
14 between times t_0 and t_1 . For example, in some embodiments, the neural network model may
15 access a stream of IMU measurement data received from the HMD 310 and handheld
16 controller 320, and, responsive to a signal to generate a prediction and/or at predetermined
17 time intervals, uses a set of IMU measurements from the stream corresponding to a previous
18 time period (e.g., between t_0 and t_1) to predict a motion of the handheld controller 320
19 relative to the HMD 310 over the time period. The predicted motion may be used to update a
20 previously determined initial position to generate an updated position. In some embodiments,
21 the neural network model may predict a set of potential relative positions between the
22 handheld controller 320 and the HMD 310 based on the IMU measurements corresponding to
23 the time period, which may be narrowed down to a specific position based upon the
24 previously-determined initial position. In some embodiments, the neural network model may
25 further receive initial position information (e.g., initial position, orientation and velocity
26 information) as an input.

27 **[0075]** In some embodiments, a duration of measurements (e.g., time between t_0 and t_1)
28 used by the neural network model to generate relative position data may be selected based
29 upon the needs of the particular application, such as a period of time over which the neural
30 network model is expected to be used determine the relative position of the first/second
31 components (e.g., a maximum expected duration in which the controller moves within a
32 "blind zone" during which the position of the controller cannot be determined using visual
33 tracking techniques). In some embodiments, the system maintains multiple neural network
34 models trained based upon different sequence lengths, which dynamically selects a model
35 during runtime based upon application needs.

1 [0076] In some embodiments, the neural network model is further trained to determine
2 changes in relative angular position (relative orientation). For example, the model may be
3 trained against relative orientation changes of the controller with respect to the headset. In
4 some embodiments, orientation can be represented using Euler Angles (3 parameters or 3 x 1)
5 or Quaternion notation (4 parameters or 4 x 1). In some embodiment, the neural network
6 model is trained to predict position and orientation simultaneously, e.g., the network is
7 trained against a position + orientation ground truth (6 x 1 or 7 x 1 parameters).

8 [0077] In some embodiments, a frequency at which the neural network model predicts
9 relative motion/position of the first and second components, and a time period analyzed by
10 the neural network model to generate each position are configurable. For example, in some
11 embodiments, the neural network model may be instructed to generate relative
12 motion/position predictions more frequently if a previously predicted relative motion exceeds
13 a threshold amount, in order to generate more granular predictions. In some embodiments,
14 the length of a time period analyzed by the neural network model may be based upon a
15 frequency of prediction, or be independent from the prediction frequency (e.g., the neural
16 network model may be configured to generate a prediction each second, each prediction
17 based on a previous 10 seconds of IMU data of the first and second components).

18 [0078] In some embodiments, the neural network model is maintained on an external
19 console (not shown) separate from the HMD 310 and the handheld controller 320. For
20 example, each of the first and second IMUs 315 and 325 may generate measurement samples
21 over time, and transmit the generated measurements to the external console to be fed to the
22 neural network model maintained on the external console. The neural network model
23 receives the generated IMU measurements, and predicts a relative motion of the HMD 310
24 and handheld controller 320, from which the relative positions of the HMD 310 and the
25 handheld controller 320 can be determined. In other embodiments, the neural network model
26 may be maintained on the HMD 310 or the handheld controller 320.

27 [0079] In some embodiments, the console may use the determined positional information
28 to generate content to be presented to the user 305 (e.g., visual content to be displayed to the
29 user through a display of the HMD, audio content to be provided to the user through an audio
30 system of the HMD, haptic feedback to be provided through the handheld controller, etc.).
31 For example, in some embodiments, the artificial reality system may analyze the relative
32 position of the second component 320 to the first component 310 to identify one or more
33 movements or gestures corresponding to action requests, and cause the requested actions to
34 be performed in the artificial reality system.

1 **[0080]** In some embodiments, the neural network model is initially trained using training
2 data obtained by instructing a training user (e.g., the user 305) having first and second
3 components positioned on their body (e.g., wearing the HMD 310 and holding the handheld
4 controller 320) to perform a variety of movements over time (e.g., moving their arms and/or
5 head to cause changes in the relative positions of the HMD 310 and handheld controller 320),
6 during which measurements from the first and second IMUs are received. The IMU
7 measurements are used to train the neural network model in conjunction with ground truth
8 information indicating positions of the first and second components generated in parallel with
9 the IMU measurements (e.g., using a second sensor system). Once the neural network model
10 is trained, the system may be used by an end user, during which the trained neural network
11 model may be used to predict relative positions of the first and second components.

12 **[0081]** The training may take place in an environment where accurate ground truth
13 information can be obtained in parallel with the measurements of the IMUs of the first and
14 second components. For example, in some embodiments, the training user may be located
15 within a room having a camera system comprising one or more cameras configured to capture
16 images of the first and second components positioned at different portions of the user's body.
17 As the training user moves their body, the cameras capture images of first and second
18 components, and uses the captured images to determine positional information of the first and
19 second components. In some embodiments, each component may comprise one or more
20 locators located at specific positions of the component. Each locator may be a light emitting
21 diode (LED), a corner cube reflector, a reflective marker, a type of light source that contrasts
22 with an environment in which the artificial reality system operates, or some combination
23 thereof. In embodiments where the locators are active (i.e., an LED or other type of light
24 emitting device), the locators may emit light in the visible band (~380 nm to 750 nm), in the
25 infrared (IR) band (~750 nm to 1 mm), in the ultraviolet band (10 nm to 380 nm), some other
26 portion of the electromagnetic spectrum, or some combination thereof. In embodiments
27 where the locators are passive, the locators may reflect light in the visible band, in the
28 infrared (IR) band, in the ultraviolet band, some other portion of the electromagnetic
29 spectrum, or some combination thereof. As the one or more cameras capture images of the
30 first and second components, the camera system may analyze the captured images and
31 identify locations of the locators within the captured images to determine a position and/or
32 orientation of each component. In some embodiments, the training user may be prompted to
33 perform certain types of motions with their head and/or arms, or to perform certain actions
34 within an application (e.g., for a gaming system, the training user may be instructed to play
35 the game with the device).

1 [0082] In other embodiments, the ground truth information may be generated by a
2 different sensor system, such as a DCA implemented as part of each of the first and second
3 components and configured to determine a position for each component, from which the
4 relative positions of the first component and the second component can be determined. For
5 example, in embodiments where the first component 310 corresponds to the headset 102 of
6 FIG. 1A, the DCA (e.g., comprising the illuminator 140 and imaging devices 130) may be
7 used to determine positional information for the first component 310.

8 [0083] By comparing the IMU measurement information received over time with the
9 ground truth information determined using another sensor system (e.g., camera system or
10 other type of sensor system), the neural network model is trained to predict the relative
11 position of the handheld controller 320 relative to the HMD 310 for a human user. Because
12 the set of potential positions of the first and second components in the training data are
13 subject to physiological constraints of the user's body, the neural network is trained to predict
14 the relative position of the first and second components subject to the physical structure of
15 the human body. For example, the amount of error of the relative position between the HMD
16 310 and the handheld controller 320 may be limited or capped due to the physical constraints
17 of the human body. In addition, because the joints of the human body have certain ranges of
18 motion, the user may move the handheld controller 320 relative to the HMD 310 more
19 naturally along certain trajectories, which may be recognized by the neural network during
20 the training process.

21 [0084] The neural network model limits an amount of error when determining the relative
22 positions of the first and second components. As such, in embodiments where a second
23 sensor system is used to periodically correct the position determined using the IMUs for drift
24 error, position determination by the second sensor system may be performed with reduced
25 frequency and/or less precision. For example, in embodiments that use a camera system to
26 determine absolute positions of the first/second components, the rate of position
27 determination may be reduced. In addition, the second sensor system may be configured to
28 operate at lower power and/or lower precision (e.g., the camera system may capture lower
29 resolution images for determining position and/or comprise fewer cameras compared to that
30 used for training the neural network model), while the error of the relative positions as
31 determining using the IMUs is maintained to be below a threshold amount. For example,
32 when predicting relative positions of the first/second components using a neural network
33 model, a camera system may be operated at a frame rate that is several times lower in
34 comparison to systems that simply integrate received IMU measurements to perform
35 positional determination. In some embodiments, a frequency at which positional information

1 from the second sensor system is used to correct for error of positions based upon the IMUs
2 is configured to not exceed a predetermined amount, e.g., from 30-60Hz to 1Hz or lower.

3 **[0085]** In some embodiments, the neural network model may allow for the relative
4 positions of the HMD 310 and handheld controller 320 to be determined entirely using the
5 first and second IMUs 315 and 325 (e.g., without the need for a second sensor system).
6 Because the neural network model limits the amount of potential error of the relative position
7 between the first and second components (e.g., due to the physical constraints of the human
8 body), determination of relative positions between the components can be performed based
9 upon only the IMU measurements while maintaining at least a threshold level of accuracy,
10 without suffering from the problem of ever-increasing drift error.

11 **[0086]** In some embodiments, the neural network model may be used to compensate for
12 loss of tracking in a second sensor system. For example, in systems that use a camera system
13 to determine positions of the first and second component based on captured images (e.g., by
14 detecting the positions of locators on each component in the captured images), tracking can
15 be lost if the first or second component (or the locators thereon) are obscured from a camera
16 of the camera system (e.g., due to being covered by the user or by other objects). In some
17 cases, loss of tracking may cause objects in the artificial environment the depend upon the
18 determined positions of tracked component to fly away (e.g., due to the determined position
19 going to infinity due to lost tracking), which may be abrupt and jarring to the user. The
20 neural network model may be used to predict the relative positions of the first and second
21 components when tracking is lost. For example, the system may compare the relative
22 positions predicted using the neural network model with positions determined by the second
23 sensor system to assess whether the positions or motions of the first and second components
24 determined by the second sensor system reflect realistic motion (e.g., do not “fly away” or go
25 to infinity, do not move at an unrealistic speed), and, if not, display content to the user based
26 upon the relative positions determined by the neural network model and IMUs instead of the
27 second sensor system until tracking can be reestablished, potentially improving user
28 experience.

29 **[0087]** In some embodiments, the neural network model may determine the relative
30 positions of the HMD 310 and handheld controller 320 using inputs from the first and second
31 IMUs 315 and 325, as well as one or more additional sensors. For example, as discussed
32 above, additional sensors may include a sensor to measure a height of the HMD or handheld
33 controller (e.g., an altimeter), or an orientation of the HMD or handheld controller (e.g., a
34 magnetometer or gyroscopic sensor), etc. In some embodiments, the additional sensor data
35 may be used by the neural network to further refine the relative position information

1 determined based on the IMU measurements, allowing for more accurate position
2 determinations without the use of the second sensor system (e.g., camera system).

3 **[0088]** In some embodiments, in addition to or instead of predicting relative position
4 changes for the last time step in the sequence (e.g., corresponding to a current relative
5 position), the neural network model is trained to predict a relative position a certain period of
6 time into the future, e.g., for latency-compensation purposes. For example, as illustrated in
7 FIG. 3, the neural network may be trained to predict an expected location 330 of the handheld
8 controller 320 relative to the HMD 310 a set amount of time into the future (e.g., a set
9 amount of time after t_1). In some embodiments, the neural network model outputs data
10 indicating a position offset of the expected future location 330 from an initial position of the
11 handheld controller 320 at a beginning of the analyzed time period (e.g., t_0), or from a current
12 position of the handheld controller 320 (e.g., position at t_1).

13 **[0089]** In some embodiments, the neural network model is trained to predict the expected
14 future position of the handheld controller 320 corresponding to an amount of time based upon
15 an expected amount of latency is known. In this way, the predicted future position of the
16 controller may be used to eliminate “perceived” latency or lag (e.g., by rendering a virtual
17 object at the expected future position, performing some control function based on the
18 expected future position, etc.). The neural network is able to take into account the
19 physiological constraints of the user, and is trained based upon patterns of human motion,
20 allowing for the expected future position to be predicted with greater accuracy in comparison
21 to simply extrapolating the expected future position based upon a last measured
22 velocity/acceleration.

23 **[0090]** In some embodiments, a neural network trained for future prediction requires no
24 architectural change in comparison to a neural network trained to estimate a current position
25 (e.g., neural network model 400 illustrated in FIG. 4). Instead, when training the neural
26 network, the ground truth position data corresponding to x ms after a last time step of the
27 input data may be used, instead of or in addition to current position ground truth.

28 **[0091]** FIG. 4 is a diagram of a neural network model that may be trained to predict
29 relative positions of components of an artificial reality system, in accordance with some
30 embodiments. The neural network model 400 may be stored and maintained on a console in
31 communication with the first and second components containing the first and second IMUs
32 (e.g., IMUs 315 of 325 of FIG. 3), on the first component or the second component (e.g., a
33 headset or HMD, and/or a handheld controller), or some combination thereof. The neural
34 network model 400 comprises a plurality of layers, including an input layer 405, at least one

1 intermediate layer (e.g., intermediate layers 410 and 415), a dense layer 420, and output pose
2 estimate layer 425.

3 **[0092]** The input layer 405 is configured to receive IMU measurement data from at least
4 a first IMU and a second IMU. In some embodiments, the first and second IMUs may
5 correspond to a headset or HMD and a handheld controller respectively (e.g., IMU 315 of the
6 HMD 310, and IMU 325 of the handheld controller 320, as illustrated in FIG. 3). In some
7 embodiments, the IMUs corresponds to a plurality of channels of data (e.g., 12 channels).
8 For example, in some embodiments, each IMU has 6 channels (e.g., 3 for gyro, and 3 for
9 accel X/Y/Z), and the system comprises two IMUs (e.g., one on the headset, and one on the
10 handheld controller). In other embodiments, a different number of channels may be used
11 (e.g., if only accelerometer data is used, then only 6 channels would be needed). In other
12 embodiments, the system may comprise additional IMUs, and the input layer is configured to
13 receive a greater number of channels. For example, the headset or controller may comprise
14 additional IMUs (e.g., increasing the number of channels to 18 or more). In some
15 embodiments, there may be separate left and right controllers, each having a respective IMU.
16 In some embodiments, there can be additional sensors at different locations, e.g., elbow,
17 chest, leg IMU.

18 **[0093]** In some embodiments, the input layer 405 may also receive additional inputs from
19 other sensors. For example, in addition to raw IMU measurements generated by the first and
20 second IMUs, the input layer 405 may receive additional parameters such as an initial
21 position (if known), one or more velocity parameters (e.g., obtained by integrating IMU
22 acceleration data over a time period, or from one or more additional sensors), orientation
23 information (e.g., generated by fusing accelerometer and gyroscopic information to obtain an
24 orientation vector with respect to gravity), etc.

25 **[0094]** In some embodiments, the input layer is configured to receive batches of input
26 sequences corresponding to IMU measurements generated over a time period of up to a
27 predetermined length. For example, in some embodiments, the IMUs are configured to
28 generate, for each second of data, 100 samples. The input layer 405 may contain sufficient
29 channels to receive up to a certain number of IMU samples corresponding to a predetermined
30 time period, such as 10 seconds, resulting in input layer 405 receiving 1000 samples at a time
31 (e.g., 12×1000 channels).

32 **[0095]** The neural network model 400 further comprises a plurality of intermediate
33 layers. For example, FIG. 4 illustrates a first intermediate layer 410 and a second
34 intermediate layer 415. In some embodiments, each intermediate layer comprises a plurality
35 of long short-term memory (LSTM) units configured to, when unrolled, match a length of an

1 IMU input sequence received at the input layer 405. For example, in an embodiment where
2 the input layer is configured to receive up to 1000 samples at a time, each intermediate layer
3 represents a number of LSTM cells (e.g., 192 LSTM cells) each unrolled 1000 times. In
4 some embodiments, the number of LSTM cells is selected based upon a complexity of the
5 problem (e.g., to achieve a balance between computational efficiency and performance). In
6 addition, although FIG. 4 illustrates the intermediate layers of the neural network model 400
7 as comprising LSTM cells, it is understood that in other embodiments, other types of neural
8 network nodes, such as Recurrent Neural Network (RNN) units, Gated Recurrent Units
9 (GRU), etc. may be used.

10 **[0096]** In some embodiments, the plurality of intermediate layers includes a first
11 intermediate layer 410 configured to return intermediate sequences, and a last intermediate
12 layer (e.g., second intermediate layer 415) configured to return final values corresponding to
13 a last unrolling step for each LSTM (e.g., 192 x 1).

14 **[0097]** In some embodiments, the neural network model 400 further comprises a dense
15 layer 420. The dense layer 420 is configured to receive an output of the final intermediate
16 layer (e.g., second intermediate layer 415) corresponding to a last unrolling step value for
17 each LSTM, and maps the received values to the pose estimate layer 425.

18 **[0098]** The neural network model 400 further comprises a pose estimate layer 425
19 configured to receive an output of the dense layer 420 and generate position change (e.g., x,
20 y, z position change) indicating a relative motion of the second component relative to the first
21 component over the analyzed time period (e.g., 10 seconds). For example, in some
22 embodiments, the pose estimate layer 425 returns a 3 x 1 vector of 3 numbers, either a
23 floating point vector or integer vector corresponding to X, Y and Z position delta in 3-D
24 space, e.g., [0, 1, -2.5], would mean 0 cm in X, 1 cm, in Y and 2.5 cm in -Z direction. It is
25 understood that in other embodiments, different representations of position and/or orientation
26 can be used, e.g., cartesian, cylindrical or spherical coordinate systems for delta position or,
27 Euler angles, axis-angle representation or Quaternion notation for relative orientation.

28 **[0099]** In some embodiments, the position information generated by the pose estimate
29 layer 425 indicates a position change of the second component (e.g., hand controller) relative
30 to the first component (e.g., HMD), in a coordinate system of the first component (e.g., a
31 head coordinate system or HMD coordinate system). In some embodiments, determining
32 relative position in terms of the user's head coordinate system instead of a room coordinate
33 system or global coordinate system may improve the accuracy of the neural network model in
34 predicting the relative positions, as the user's head coordinate system is able to more
35 accurately account for the physiological constraints of the user's body. For example, when a

1 user performs a particular arm motion, the physiological constraints of the user's body may
2 be the same regardless of which direction the user is facing when performing the motion. In
3 some embodiments, the neural network model may receive input information corresponding
4 to a sensor (e.g., IMU) on the user's body or torso, and predict the position of the hand
5 controller using a body or torso coordinate system instead of a head coordinate system. In
6 some embodiments, the position of the hand controller may be predicted relative to a
7 modified head coordinate system that is constrained along one or more axes. For example, in
8 some embodiments, the user's modified head coordinate system is constrained such that the
9 z-axis remains aligned with a gravity vector, even if the user tilts their head (e.g., where the
10 gravity vector is determined using a gyroscopic sensor on the HMD worn by the user).

11 **[00100]** Although FIG. 4 illustrates the neural network model using intermediate LSTM
12 layers, it is understood that in other embodiments, other types of NN architectures, such as
13 Gated Recurrent Unit Network (GRU), Recurrent Neural Network (RNN), Convolutional
14 Neural Network (CNN) Architectures and their combinations, may be used.

15 **[00101]** FIG. 5 is a diagram of another neural network model that may be trained to predict
16 relative positions of components of an artificial reality system, in accordance with some
17 embodiments. In contrast to the LSTM-based neural network illustrated in FIG. 4, the neural
18 network 500 illustrated in FIG. 5 is based on a CNN architecture. In some embodiments, the
19 neural network model may be trained to perform position tracking (e.g., predicting relative
20 position changes for the last time step in the sequence) and/or, as discussed above, future
21 position prediction/latency compensation.

22 **[00102]** Similar to the neural network model 400, the neural network model 500 may be
23 stored and maintained on a console in communication with the first and second components
24 containing the first and second IMUs (e.g., IMUs 315 of 325 of FIG. 3), on the first
25 component or the second component (e.g., a headset or HMD, and/or a handheld controller),
26 or some combination thereof. The neural network model 500 comprises a plurality of layers,
27 including an input layer 505, decimation layer 510, CNN layers 515, dense layers 520, and
28 one or more output layers 525. In some embodiments, the neural network model may further
29 comprise a feed forward path 530.

30 **[00103]** The input layer 505 is configured to receive batches of input sequences
31 corresponding to IMU measurements generated over a time period of up to a predetermined
32 length. For example, in some embodiments, the input layer 505 is configured to receive sets
33 of input sequences corresponding to a set number of time steps (e.g., 128 time steps at a
34 time). Each input sequence corresponding to a particular time step may comprise a plurality
35 of features, such as a time stamp, IMU measurement data (e.g., linear acceleration

1 parameters, angular acceleration parameters) from at least a first IMU and a second IMU
2 (e.g., corresponding to a headset or HMD and a handheld controller, respectively). In
3 addition, in some embodiments, similar to as discussed above in relation to FIG. 4, the input
4 layer 505 may also receive additional inputs from other sensors. For example, the plurality of
5 features included in the input sequence may, in addition to include IMU measurements
6 generated by the first and second IMUs, include one or more position parameters (e.g., based
7 upon a previous position determination, or generated using a second sensor system, such as a
8 camera system, if available), one or more orientation parameters (e.g., based upon a previous
9 orientation determination, or generated using a separate sensor, such as a gyroscopic sensor,
10 magnetometer, etc.), one or more velocity parameters (e.g., generated by integrating IMU
11 acceleration data over a previous time period, or generated using a separate sensor), or some
12 combination thereof. In some embodiments, the input layer 505 may compute one or more
13 additional parameters not received as part of the input sequence (e.g., integrating a plurality
14 of previously received IMU acceleration parameters to determine a velocity parameter).

15 **[00104]** In some embodiments, the neural network model 500 performs decimation on the
16 received input sequences at a decimation layer 510. For example, in some embodiments,
17 input sequence data may be received at a first rate (e.g., at 500 samples per second, or
18 500Hz). However, the neural network model 500 may be able to produce results of a
19 sufficient accuracy using samples produced at a second, slower rate (e.g., between 50 and
20 100Hz). As such, the neural network 500 may use the decimation layer to decimate the
21 received samples by a specified amount, e.g., by averaging sets of consecutive sample to
22 generate a smaller number of aggregated samples. For example, in some embodiments, the
23 decimation layer 510 is configured to average sets of 8 consecutive samples or 16
24 consecutive samples to generate a smaller number of samples, e.g., decimating a received set
25 of samples corresponding to 128 time steps to aggregated samples corresponding to 16 time
26 steps or 8 time steps.

27 **[00105]** The CNN layers 515 correspond to a multi-layer CNN model. For example, in
28 some embodiments, the CNN layers 515 may correspond to a 7-layer CNN, where each layer
29 includes batch normalization, and a rectified linear unit (ReLU), and generates one or more
30 features from the received set of input sequences, e.g., encoded information relating to
31 motion of the first and/or second IMUs in a tensor format.

32 **[00106]** The dense layers 520 may comprise a first dense layer corresponding to a non-
33 linear activation function with ReLU and dropout, and a second fully-connected dense layer
34 with no ReLU or dropout. In some embodiments, the dense layers 520 are configured to
35 receive an output of a final CNN layer, and map the received values to the output layer 525.

1 **[00107]** The output layer 525 receives an output of the dense layers 520 and generates
2 pose information, which may include position change information (e.g., x, y, z position
3 change), orientation information, or a combination thereof. In embodiments where the neural
4 network is performing position tracking, the position information may indicate a distance
5 travelled within the analyzed time window (e.g., between t_0 and t_1). On the other hand, if
6 performing future position prediction or latency compensation, the position change
7 information indicates a distance the second component is expected to travel during a future
8 time period (e.g., 50ms after t_1). For example, in some embodiments, the output layer 525
9 returns a 3 x 1 vector of 3 numbers, corresponding to X, Y, and Z position delta in 3-D space,
10 indicating a distance and direction that the second component is expected to travel during the
11 future time period. In some embodiments, the output layer 525 may return a 6 x 1 vector or a
12 7 x 1 vector corresponding to position and orientation. In other embodiments, other formats
13 for expressing position and/or orientation information may be used (e.g., a 12 x 1 vector or 3
14 x 4 array corresponding to 3 x 1 position information and 9 x 1 rotation matrix expressing
15 orientation information). In some embodiments, the output vector may indicate a coordinate
16 in a head coordinate system of the user.

17 **[00108]** In some embodiments, the output layer 525 comprises a first encoder layer that
18 outputs a set of coefficients for a position determination function, and a second output layer
19 that outputs the final position information. The position determination function is a function
20 indicating an expected path of the second component over a future time period. For example,
21 in some embodiments the position determination function is a cubic polynomial function of a
22 time that outputs a three-dimensional pose vector, having coefficients corresponding to
23 bias/position, velocity, acceleration, and jerk. In some embodiments, the polynomial function
24 may be divided into three polynomial functions corresponding to X, Y, and Z coordinate
25 axes. As such, the encoder layer may output a set 12 coefficients (e.g., corresponding to
26 bias/position, velocity, acceleration, and jerk coefficients for each function). In some
27 embodiments, the output coefficients may also include coefficients for a function for
28 determining orientation information.

29 **[00109]** In some embodiments, the output layer 525 determines coefficients for a position
30 determination function to allow for position information to be determined for a range of
31 future time steps. This is because the latency values of a system may not be known
32 beforehand, and may change based on how the system is being used (e.g., based on a
33 particular application that the user is using). For example, if the user is using the first and
34 second components to navigate a menu that is not very graphics intensive, latency may be
35 low. On the other hand, if the user is using the first and second components to play a

1 graphically-intensive game, latency may be higher. In some embodiments, the neural
2 network 500 is trained using information corresponding to a range of latency values (e.g.,
3 between 10 and 100 ms of latency).

4 **[00110]** The second output layer receives the coefficients generated by the encoder layer,
5 and determines position information for one or more time steps using the position
6 determination function. In some embodiments, the second output layer receives a specific
7 latency value, and determines position information corresponding to a time step
8 corresponding to the received latency value. In other embodiments, the second output layer
9 uses the position determination function to generate a plurality of positions corresponding to
10 a plurality of time steps, e.g., to be used by another processing element to select a specific
11 position based upon a specific latency value.

12 **[00111]** In some embodiments, the neural network model may, in addition to outputting
13 relative position information of the second component, also be trained to output orientation
14 information. In other embodiments, orientation information may instead be determined using
15 one or more additional sensors, such as by fusing accelerometer and gyroscopic information
16 obtain an orientation vector with respect to gravity, using a magnetometer to determine
17 magnetic north, and/or the like.

18 **[00112]** In some embodiments, the neural network 500 includes a velocity feed-forward
19 path 530. As discussed above, a simple way of predicting a future relative position of a
20 component (e.g., controller) is to extrapolate a current velocity of the component multiplied
21 by an expected time value (e.g., latency value). In some embodiments, the neural network
22 model 500 receives current velocity information as part of an input sequence at the input
23 layer 505, or determines a current velocity using received acceleration information, and uses
24 the velocity information to determine an extrapolated relative position to be used as a
25 baseline for predicting an expected future position of the component. In some embodiments,
26 the neural network 500 may further take into account a current acceleration when
27 extrapolating the expected relative position of the component. In some embodiments, the
28 neural network feeds forward the velocity (and/or acceleration) information via the velocity
29 feed forward path 530, to be extrapolated at a later layer (e.g., by the output layer 525).

30 **[00113]** In some embodiments, the neural network 500 is trained to use the extrapolated
31 relative position as a baseline value when predicting an expected future relative position. In
32 some embodiments, the neural network 500 is trained to predict an offset of the future
33 relative position from the extrapolated relative position. For example, the output layer 525
34 may determine an output of the position determination function that corresponds to an offset
35 from an extrapolated position, where the extrapolated position is received via the velocity

1 feed forward path 530 or determined based on velocity information received via the feed
2 forward path 530. The output layer 525 applies the predicted offset to the extrapolated
3 relative position to determine a final relative position prediction.

4 **[00114]** FIGs. 6A and 6B illustrate flowcharts of processes for training and using a neural
5 network model for predicting relative position of components of an artificial reality system
6 using IMUs, in accordance with one or more embodiments. The processes shown in FIGs.
7 6A and 6B may be performed by components of an artificial reality system (e.g., artificial
8 reality system 300 illustrated in FIG. 3). Other entities may perform some or all of the steps
9 in FIGs. 6A and 6B in other embodiments. Embodiments may include different and/or
10 additional steps, or perform the steps in different orders. In some embodiments, the
11 processes of FIGs. 6A and 6B are performed by components of different artificial reality
12 systems. For example, the training process of FIG. 6A may be performed using a first
13 artificial reality system to train the neural network model, which may then be loaded onto a
14 second artificial reality system to be used by an end user (e.g., based on the process of FIG.
15 6B). In some embodiments, the neural network model for the artificial reality system is pre-
16 trained. For example, the neural network model may be initially trained as a default model,
17 and downloaded to each a plurality of artificial reality systems.

18 **[00115]** FIG. 6A illustrates a flowchart of a process for training the neural network model
19 for predicting relative positions between components using IMUs, in accordance with some
20 embodiments. The artificial reality system receives 610, over a time period, positional
21 information from a first IMU of a first component of the artificial reality system positioned at
22 a first position on a user's body, and a second IMU of a second component of the artificial
23 reality system positioned at a second position on the user's body. In some embodiments, the
24 first component corresponds to an HMD worn on the head of the user, while the second
25 component corresponds to a handheld controller held in a hand of the user. The received
26 positional information may comprise a plurality of IMU measurement samples generated by
27 the first and second IMUs over the time period, each indicating an angular rate and an
28 acceleration of the first or second IMU. In some embodiments, the artificial reality system
29 may further receive additional positional information for one or more additional sensors.

30 **[00116]** In some embodiments, the first and second components are positioned on a body
31 of a training user, who may be instructed to move their body during the time period, such that
32 the received positional information corresponds to different positions of the first and second
33 components over the time period. For example, where the first and second components
34 correspond to an HMD and a handheld controller, the user may turn their head and move
35 their arm in various directions. In some embodiments, the artificial reality system may

1 prompt the user to move their body in certain ways in order to achieve adequate coverage of
2 different positions in the training data. In some embodiments, the user may perform a variety
3 of different motions expected to be performed when controlling the artificial reality system.
4 For example, in systems in which the user may perform predetermined gestures using the first
5 and second components to generate action requests, the user may be prompted to perform
6 such gestures to ensure they are reflected in the training data. In some embodiments, the
7 system may prompt the user to perform additional types of motions that may not be directly
8 related to controlling the artificial reality system, in order to capture a more diverse set of
9 possible movements.

10 **[00117]** The artificial reality system receives 620, over the time period, positional
11 information of the first and second components generated using a ground truth sensor system.
12 In some embodiments, training of the neural network model may occur in a controlled
13 environment where a ground truth sensor system is used to determine absolute positions of
14 the first and second components in order to generate ground truth information for use in
15 training the model. In some embodiments, the ground truth sensor system comprises one or
16 more tracking cameras to capture images of the first and second components positioned at
17 different positions on the user's body. In some embodiments, the first and second
18 components may comprise one or more locators at specific positions on the first and second
19 components, which may be used to determine a position of the first and second components
20 using the captured images. In other embodiments, other types of sensor systems usable to
21 determine positional information of the first and second components, e.g., a DCA
22 implemented as part of the first and/or second component, one or more magnetic sensors,
23 and/or the like, may be used. The positional information generated by the second sensor
24 system may indicate absolute positions of the first and second components over the time
25 period, which, combined with the IMU measurements from the first and second IMUs
26 indicating angular rate and acceleration of the first and second IMUs over the time period,
27 may form a training data set for training the neural network model.

28 **[00118]** The artificial reality system trains 630 the neural network model based on the
29 received IMU measurements and ground truth information indicating the positions of the first
30 and second components. The neural network model uses IMU data corresponding a
31 predetermined time period (e.g., 1 second, 10 seconds, etc.), and predicts a relative motion of
32 the second component relative to the first component over the predetermined time period.
33 The predicted relative motion is used to determine a relative position of the first and second
34 components, which is compared to the ground truth information indicating positions of the
35 first and second components to generate feedback data for training the neural network model.

1 [00119] In addition, as discussed above, in some embodiments, the neural network model
2 may be trained to predict a certain period of time into the future, e.g., for latency
3 compensation purposes. In such cases, the predicted relative position of the first and second
4 components is compared to ground truth data corresponding to position measured a set
5 amount of time (e.g., between 10 and 100 ms) after the end of the predetermined time period,
6 instead of at the end of the predetermined time period.

7 [00120] In some embodiments, the training method of FIG. 6A may be performed with
8 over multiple time periods and/or with multiple training users, to generate a more diverse set
9 of training data, such that the neural network model is applicable to a general population of
10 potential users. In addition, in some embodiments, an end user of the artificial reality system
11 may perform additional training on the neural network model, in order to calibrate the neural
12 network model to more accurately reflect the body of the specific end user.

13 [00121] In some embodiments where the neural network model is configured to receive
14 input sequences containing additional data beyond raw IMU measures (e.g., position
15 parameters, velocity parameters, and/or other parameters received from one or more
16 additional sensors), the neural network model may perform regularization during training to
17 prevent the model from placing too much weight on specific parameters. For example, in
18 some embodiments, if a weight value associated with a specific parameter becomes too large
19 during training relative to weight values of other parameters, the weight value may be
20 reduced or penalized, to ensure that the neural network model does not over rely on any
21 single parameter.

22 [00122] FIG. 6B illustrates a flowchart of a method for position determination of
23 components of an artificial reality system using IMUs, in accordance with some
24 embodiments. The artificial reality system receives 650, as the user uses the artificial reality
25 system, measurement data from the first and second IMUs indicating angular rate and
26 acceleration information of the first and second IMUs. The first IMU is located on a first
27 component of the artificial reality system positioned at a first position on a user's body (e.g.,
28 an HMD worn on the user's hand), while the second IMU is located on a second component
29 of the artificial reality system positioned at a second position on the user's body (e.g., a
30 handheld controller held in a hand of the user). In some embodiments, the artificial reality
31 system may further receive additional measurement data for one or more additional sensors.

32 [00123] The artificial reality system uses 660 a trained neural network model (e.g., a
33 neural network model trained using the process described in relation to FIG. 6A) to predict a
34 relative position of the first and second components. In some embodiments, the neural
35 network model uses as input IMU data corresponding to a predetermined time period (e.g., a

1 previous 1 second, previous 10 seconds, etc.) to predict a motion of the second component
2 relative to the first component during the time period. The predicted motion may be used to
3 update an initial position and determine an updated relative position of the second component
4 to the first component. In some embodiments, the initial position may be determined using a
5 visual tracking system, or based upon a prompt to the user to initially orient the first and
6 second components in a certain way (e.g., “touch your head with the controller to start”). In
7 some embodiments, the system may use a past history of motion as determined by the neural
8 network model as an indication of absolute position, or to deduce an initial position. For
9 example, responsive to a determination that the controller moved 5 feet up, deducing that the
10 controller was being held down, and it was moved up such that the arm is extended upwards.

11 **[00124]** In some embodiments, the neural network model may predict a relative position of
12 the first and second components a certain period of time into the future, e.g., to predict future
13 position/orientation with approximately 10-100 ms into the future, based on an expected level
14 of latency. This is essentially a way to do pose extrapolation in time into the future, and
15 absorb latency induced by the system, e.g., eliminating “perceived” latency or lag by
16 predicting where the controller will be some period of time into the future. As discussed
17 above, training the neural network model for future prediction requires no architectural
18 change, but instead uses ground truth position data from x ms into the future for training
19 instead of current position ground truth.

20 **[00125]** The artificial reality system optionally periodically adjusts 670 the determined
21 relative position of the second component to the first component based on measurements
22 from a second sensor system. In some embodiments, the second sensor system may comprise
23 a tracking camera system, visual odometry system, magnetic sensors, or other type of sensor
24 system usable to determine absolute position information of the first and second
25 components. The second sensor system may be different from the ground truth sensor system
26 used to generate the ground truth data for training the model described in relation to FIG. 6A.
27 For example, in some embodiments, training of the neural network model may take place in a
28 controlled environment prior to the artificial reality system being used by an end user (e.g., in
29 a controller room having multiple high-resolution cameras), while the second sensor system
30 may comprise one or more sensors located on the first or second components, or can be set up
31 by the end-user of the artificial reality system.

32 **[00126]** In some embodiments, because the use of the neural network model to predict
33 relative position and movement of the first and second components may limit an amount of
34 error that is able to accumulate in determining the relative position between the first and
35 second components, the second sensor system may be used to adjust the determined relative

1 position less frequently in comparison to systems in which the positions of the components is
2 determined by integrating the receiving IMU measurements, or be configured to run in a
3 lower power or lower resolution mode. For example, in embodiments where the second
4 sensor system comprises a camera system configured to capture images of the first and
5 second components from which positional information of the first and second components
6 can be determined, the camera system may comprise fewer cameras than the ground truth
7 sensor system used for the initial training of the neural network model (e.g., as described in
8 relation to FIG. 6A) and/or lower resolution cameras.

9 **[00127]** In addition, in some embodiments the artificial reality system may update the
10 neural network model based upon measurements and/or positional information obtained using
11 the second sensor system. For example, in some embodiments, the trained neural network
12 model may initially correspond to a baseline or default model. As a specific end user uses the
13 artificial reality system, the neural network model may be updated based on absolute position
14 information determined using the second sensor system, to further train the neural network
15 model based on the body of the specific user.

16 **[00128]** FIG. 7 illustrates an artificial reality system 700, in accordance with one or more
17 embodiments. In some embodiments, the artificial reality system 700 comprises a headset
18 705 which may correspond to the headset 102 of FIG. 1A or the HMD 106 of FIG. 1B, and a
19 controller device 710 which may correspond to the handheld controller 200 of FIG. 2. The
20 artificial reality system 700 may operate in a virtual reality environment, an augmented
21 reality environment, a mixed reality environment, or some combination thereof. The headset
22 705 and the controller device 710 may be coupled to a console 715. While FIG. 7 shows an
23 example system 700 including one headset 705 and one controller device 710, in other
24 embodiments any number of these components may be included in the system 700. For
25 example, in some embodiments a user may wear the headset 705 while also operating two
26 different controller devices 710 (e.g., one in each hand). In some embodiments, there may be
27 multiple headsets each having an associated controller device 710, with each headset and
28 controller device 710 communicating with the console 715. In alternative configurations,
29 different and/or additional components may be included in the system 700 (e.g., an audio
30 system for providing audio content to the user). Additionally, functionality described in
31 conjunction with one or more of the components shown in FIG. 7 may be distributed among
32 the components in a different manner than described in conjunction with FIG. 7 in some
33 embodiments. For example, some or all of the functionality of the console 715 may be
34 provided by the headset 705.

1 [00129] The headset 705 includes the display assembly 730, an optics block 735, one or
2 more position sensors 740, and a DCA 745. Some embodiments of headset 705 have
3 different components than those described in conjunction with FIG. 7. Additionally, the
4 functionality provided by various components described in conjunction with FIG. 7 may be
5 differently distributed among the components of the headset 705 in other embodiments, or be
6 captured in separate assemblies remote from the headset 705.

7 [00130] The display assembly 730 displays content to the user in accordance with data
8 received from the console 715. The display assembly 730 displays the content using one or
9 more display elements (e.g., the display elements 120 illustrated in FIG. 1A). A display
10 element may be, e.g., an electronic display. In various embodiments, the display assembly
11 730 comprises a single display element or multiple display elements (e.g., a display for each
12 eye of a user). Examples of an electronic display include: a liquid crystal display (LCD), an
13 organic light emitting diode (OLED) display, an active-matrix organic light-emitting diode
14 display (AMOLED), a waveguide display, some other display, or some combination thereof.
15 Note in some embodiments, the display element 730 may also include some or all of the
16 functionality of the optics block 735.

17 [00131] The optics block 735 may magnify image light received from the electronic
18 display, corrects optical errors associated with the image light, and presents the corrected
19 image light to one or both eyeboxes of the headset 705. In various embodiments, the optics
20 block 735 includes one or more optical elements. Example optical elements included in the
21 optics block 735 include: an aperture, a Fresnel lens, a convex lens, a concave lens, a filter, a
22 reflecting surface, or any other suitable optical element that affects image light. Moreover,
23 the optics block 735 may include combinations of different optical elements. In some
24 embodiments, one or more of the optical elements in the optics block 735 may have one or
25 more coatings, such as partially reflective or anti-reflective coatings.

26 [00132] Magnification and focusing of the image light by the optics block 735 allows the
27 electronic display to be physically smaller, weigh less, and consume less power than larger
28 displays. Additionally, magnification may increase the field of view of the content presented
29 by the electronic display. For example, the field of view of the displayed content is such that
30 the displayed content is presented using almost all (e.g., approximately 110 degrees
31 diagonal), and in some cases, all of the user's field of view. Additionally, in some
32 embodiments, the amount of magnification may be adjusted by adding or removing optical
33 elements.

34 [00133] In some embodiments, the optics block 735 may be designed to correct one or
35 more types of optical error. Examples of optical error include barrel or pincushion distortion,

1 longitudinal chromatic aberrations, or transverse chromatic aberrations. Other types of
2 optical errors may further include spherical aberrations, chromatic aberrations, or errors due
3 to the lens field curvature, astigmatisms, or any other type of optical error. In some
4 embodiments, content provided to the electronic display for display is pre-distorted, and the
5 optics block 735 corrects the distortion when it receives image light from the electronic
6 display generated based on the content.

7 **[00134]** The IMU 740 may correspond to the IMU 190 illustrated in FIGs. 1A and 1B, and
8 comprises at least one accelerometer to measure translational motion (forward/back,
9 up/down, left/right) and at least one gyroscope to measure rotational motion (e.g., pitch, yaw,
10 roll). In some embodiments, the IMU 740 rapidly samples the measurement signals of the
11 accelerometer and gyroscope to generate measurements usable to calculate a position of the
12 headset 705. For example, the generated measurements may be used as input to a trained
13 neural network model (e.g., maintained by a tracking module 760) to predict a position of a
14 reference point on the headset 705 (e.g., a relative position to another component on the
15 user's body, such as the controller device 710). The reference point is a point that may be
16 used to describe the position of the headset 705. While the reference point may generally be
17 defined as a point in space, however, in practice the reference point is defined as a point
18 within the headset 705.

19 **[00135]** The DCA 745 generates depth information for a portion of the local area. The
20 DCA includes one or more imaging devices and a DCA controller. The DCA 745 may also
21 include an illuminator. Operation and structure of the DCA 745 is described above with
22 regard to FIG. 1A.

23 **[00136]** In some embodiments, the headset 705 further comprises a second position sensor
24 750 configured to generate data indicating a position of the headset 705. In some
25 embodiments, the second position sensor 750 is configured to determine an absolute position
26 of the headset 705, e.g., a position of the headset 705 relative to a static reference point in the
27 surrounding environment (e.g., corresponding to the console 715, a fixed point in a room,
28 etc.). In some embodiments, the second position sensor 750 may be implemented as part of
29 the DCA 745, in which the position of the headset 705 may be determined based upon the
30 generated depth information. In other embodiments, instead of a second position sensor 750
31 included as part of the headset 705, the second sensor system may comprise a separate
32 camera system configured to capture images of the headset 705 to determine the position
33 and/or orientation of the headset 705. For example, in some embodiments, the headset 705
34 may include one or more locators, which may be used by the determine a position of the
35 headset 705 using images captured by the camera system.

1 **[00137]** The controller device 710 may correspond to a handheld device such as a
2 handheld controller (e.g., the handheld controller 200 of FIG. 2), and is a device that allows a
3 user to send action requests and receive responses from the console 715. An action request is
4 a request to perform a particular action. For example, an action request may be an instruction
5 to start or end capture of image or video data, or an instruction to perform a particular action
6 within an application. In some embodiments, the user sends action requests using an I/O
7 interface 770 of the controller device 710, which may include one or more input elements,
8 such as one or more buttons, triggers, joysticks, etc. In addition, in some embodiments, an
9 action request may be sent based upon a movement or position of the controller device 715
10 (e.g., determined based upon the IMU 775 of the controller device 710). An action request is
11 communicated to the console 715, which performs an action corresponding to the action
12 request. In some embodiments, the I/O interface 770 may provide haptic feedback to the user
13 in accordance with instructions received from the console 715. For example, haptic feedback
14 is provided when an action request is received, or the console 715 communicates instructions
15 to the controller device 710 causing the I/O interface 770 to generate haptic feedback when
16 the console 715 performs an action.

17 **[00138]** The IMU 775 of the controller device 710 may be similar to the IMU 740 of the
18 headset 705, comprising at least one accelerometer to measure translational motion
19 (forward/back, up/down, left/right) and at least one gyroscope to measure rotational motion
20 (e.g., pitch, yaw, roll). The IMU 775 rapidly samples the measurement signals of the
21 accelerometer and gyroscope to generate measurements usable to calculate a position of the
22 controller device 710. For example, the generated measurements may be used as input to a
23 trained neural network model (e.g., maintained by the tracking module 760 of the console
24 715) to predict a position of a reference point on the controller device 710 (e.g., a relative
25 position to another component on the user's body, such as the headset 705). The reference
26 point is a point that may be used to describe the position of the controller device 710. While
27 the reference point may generally be defined as a point in space, however, in practice the
28 reference point is defined as a point within the headset 705.

29 **[00139]** The console 715 provides content to the headset 705 for processing in accordance
30 with information received from the headset 705 and the controller device 710. In the
31 example shown in FIG. 7, the console 715 includes an application store 755, a tracking
32 module 760, and an engine 765. Some embodiments of the console 715 have different
33 modules or components than those described in conjunction with FIG. 7. Similarly, the
34 functions further described below may be distributed among components of the console 715
35 in a different manner than described in conjunction with FIG. 7. In some embodiments, the

1 functionality discussed herein with respect to the console 715 may be implemented in the
2 headset 705, or a remote system.

3 **[00140]** The application store 755 stores one or more applications for execution by the
4 console 715. An application is a group of instructions, that when executed by a processor,
5 generates content for presentation to the user. Content generated by an application may be in
6 response to inputs received from the user via movement of the headset 705 or the controller
7 device 710 (e.g., action requests received at the I/O interface 770). Examples of applications
8 include: gaming applications, conferencing applications, video playback applications, or
9 other suitable applications.

10 **[00141]** The tracking module 760 tracks movements of the headset 705 and the controller
11 device 710 using information from IMUs 740 and 775. The tracking module 760 comprises a
12 trained neural network model configured to receive measurement data captured by the IMUs
13 740 and 775 over a predetermined time period, and predict a relative motion of the controller
14 device 710 relative to the headset 705 over the time period. The tracking module 760 uses
15 the predicted motion to determine a relative position of the controller device 710 to the
16 headset 705. In some embodiments, the trained neural network model may correspond to the
17 neural network model 400 illustrated in FIG. 4, and is trained based on the process illustrated
18 in FIG. 5A. Although FIG. 7 illustrates the tracking module as located on the console 715, it
19 is understood that in some embodiments, at least a portion of the tracking module may be
20 implemented on a different device, such as the headset 705.

21 **[00142]** In some embodiments, the tracking module 760 may further be configured to
22 periodically adjust one or more relative positions determined using the trained neural network
23 model using position information determined based upon the DCA 745, the second position
24 sensor 750, or some combination thereof. For example, the tracking module 760 determines
25 a position of a reference point of the headset 705 in a mapping of a local area based on
26 information from the headset 705. The tracking module 760 may also determine positions of
27 an object or virtual object.

28 **[00143]** Additionally, in some embodiments, the tracking module 760 may use portions of
29 data indicating a position of the headset 705 and/or controller device 710 (e.g., as determined
30 using the trained neural network model), the second position sensor 760, and/or
31 representations of the local area from the DCA 745, to predict a future location of the headset
32 705. The tracking module 760 provides the estimated or predicted future position of the
33 headset 705 or the controller device 710 to the engine 765.

34 **[00144]** The engine 765 executes applications and receives position information,
35 acceleration information, velocity information, predicted future positions, or some

1 combination thereof, of the headset 705 and/or controller device 710 from the tracking
2 module 760. Based on the received information, the engine 765 determines content to
3 provide to the headset 705 for presentation to the user. For example, if the received
4 information indicates that the user has looked to the left, the engine 765 generates content for
5 the headset 705 that mirrors the user's movement in a virtual local area or in a local area
6 augmenting the local area with additional content. Additionally, the engine 765 performs an
7 action within an application executing on the console 715 in response to an action request
8 received from the controller device 710 and provides feedback to the user that the action was
9 performed. The provided feedback may be visual or audible feedback via the headset 705 or
10 haptic feedback via the controller device 710.

11 **[00145]** One or more components of system 700 may contain a privacy module that stores
12 one or more privacy settings for user data elements. The user data elements describe the user
13 or the headset 705. For example, the user data elements may describe a physical
14 characteristic of the user, an action performed by the user, a location of the user of the
15 headset 705, a location of the headset 705, an HRTF for the user, etc. Privacy settings (or
16 "access settings") for a user data element may be stored in any suitable manner, such as, for
17 example, in association with the user data element, in an index on an authorization server, in
18 another suitable manner, or any suitable combination thereof.

19 **[00146]** A privacy setting for a user data element specifies how the user data element (or
20 particular information associated with the user data element) can be accessed, stored, or
21 otherwise used (e.g., viewed, shared, modified, copied, executed, surfaced, or identified). In
22 some embodiments, the privacy settings for a user data element may specify a "blocked list"
23 of entities that may not access certain information associated with the user data element. The
24 privacy settings associated with the user data element may specify any suitable granularity of
25 permitted access or denial of access. For example, some entities may have permission to see
26 that a specific user data element exists, some entities may have permission to view the
27 content of the specific user data element, and some entities may have permission to modify
28 the specific user data element. The privacy settings may allow the user to allow other entities
29 to access or store user data elements for a finite period of time.

30 **[00147]** The privacy settings may allow a user to specify one or more geographic locations
31 from which user data elements can be accessed. Access or denial of access to the user data
32 elements may depend on the geographic location of an entity who is attempting to access the
33 user data elements. For example, the user may allow access to a user data element and
34 specify that the user data element is accessible to an entity only while the user is in a
35 particular location. If the user leaves the particular location, the user data element may no

1 longer be accessible to the entity. As another example, the user may specify that a user data
2 element is accessible only to entities within a threshold distance from the user, such as
3 another user of a headset within the same local area as the user. If the user subsequently
4 changes location, the entity with access to the user data element may lose access, while a new
5 group of entities may gain access as they come within the threshold distance of the user.

6 **[00148]** The system 700 may include one or more authorization/privacy servers for
7 enforcing privacy settings. A request from an entity for a particular user data element may
8 identify the entity associated with the request and the user data element may be sent only to
9 the entity if the authorization server determines that the entity is authorized to access the user
10 data element based on the privacy settings associated with the user data element. If the
11 requesting entity is not authorized to access the user data element, the authorization server
12 may prevent the requested user data element from being retrieved or may prevent the
13 requested user data element from being sent to the entity. Although this disclosure describes
14 enforcing privacy settings in a particular manner, this disclosure contemplates enforcing
15 privacy settings in any suitable manner.

16 **Additional Configuration Information**

17 **[00149]** The foregoing description of the embodiments has been presented for illustration;
18 it is not intended to be exhaustive or to limit the patent rights to the precise forms disclosed.
19 Persons skilled in the relevant art can appreciate that many modifications and variations are
20 possible considering the above disclosure.

21 **[00150]** Some portions of this description describe the embodiments in terms of algorithms
22 and symbolic representations of operations on information. These algorithmic descriptions
23 and representations are commonly used by those skilled in the data processing arts to convey
24 the substance of their work effectively to others skilled in the art. These operations, while
25 described functionally, computationally, or logically, are understood to be implemented by
26 computer programs or equivalent electrical circuits, microcode, or the like. Furthermore, it
27 has also proven convenient at times, to refer to these arrangements of operations as modules,
28 without loss of generality. The described operations and their associated modules may be
29 embodied in software, firmware, hardware, or any combinations thereof.

30 **[00151]** Any of the steps, operations, or processes described herein may be performed or
31 implemented with one or more hardware or software modules, alone or in combination with
32 other devices. In one embodiment, a software module is implemented with a computer
33 program product comprising a computer-readable medium containing computer program
34 code, which can be executed by a computer processor for performing any or all the steps,
35 operations, or processes described.

1 **[00152]** Embodiments may also relate to an apparatus for performing the operations
2 herein. This apparatus may be specially constructed for the required purposes, and/or it may
3 comprise a general-purpose computing device selectively activated or reconfigured by a
4 computer program stored in the computer. Such a computer program may be stored in a
5 non-transitory, tangible computer readable storage medium, or any type of media suitable for
6 storing electronic instructions, which may be coupled to a computer system bus.

7 Furthermore, any computing systems referred to in the specification may include a single
8 processor or may be architectures employing multiple processor designs for increased
9 computing capability.

10 **[00153]** Embodiments may also relate to a product that is produced by a computing
11 process described herein. Such a product may comprise information resulting from a
12 computing process, where the information is stored on a non-transitory, tangible computer
13 readable storage medium and may include any embodiment of a computer program product
14 or other data combination described herein.

15 **[00154]** Finally, the language used in the specification has been principally selected for
16 readability and instructional purposes, and it may not have been selected to delineate or
17 circumscribe the patent rights. It is therefore intended that the scope of the patent rights be
18 limited not by this detailed description, but rather by any claims that issue on an application
19 based hereon. Accordingly, the disclosure of the embodiments is intended to be illustrative,
20 but not limiting, of the scope of the patent rights, which is set forth in the following claims.

1 CLAIMS:

- 2 1. A system comprising:
3 a first component configured to be positioned at a first portion of a user's body,
4 the first component comprising at least a first inertial measurement unit
5 (IMU) configured to generate first measurements indicating motion of the
6 first component;
7 a second component configured to be positioned at a second portion of the user's
8 body, comprising at least a second IMU configured to generate second
9 measurements indicating motion of the second component, wherein a set
10 of potential positions of the first component and the second component is
11 determined by physiological constraints of the first portion and the second
12 portion of the user's body; and
13 a controller configured to:
14 receive first measurements from the first IMU and the second
15 measurements from the second IMU,
16 predict a position of the first component relative to the second
17 component using a trained neural network model trained using
18 the set of potential positions.
19 2. The system of claim 1, wherein the first component is a head-mounted device
20 (HMD) configured to be worn on the user's head, and the second component is a handheld
21 device configured to be held in a hand of the user.
22 3. The system of claim 1 or claim 2, wherein the position of the first component
23 relative to the second component predicted by the trained neural network corresponds to an
24 expected position of the first component relative to the second component corresponding to a
25 future time.
26 4. The system of any preceding claim, wherein the controller is configured to
27 receive ground truth information for training the neural network model from a camera system
28 configured to capture images of the first component and the second component; and
29 preferably wherein the ground truth information comprises position information of the first
30 component and the second component corresponding to a time after a time period during
31 which measurements from the first IMU and the second IMU are received.
32 5. The system of any of claims 1 to 3, wherein the controller is further
33 configured to periodically receive information indicating positions of the first and second
34 components from an additional sensor, and to periodically use the received information to
35 correct an error of the predicted position.

1 6. The system of any preceding claim, wherein the trained neural network model
2 is configured to:

3 receive a set of first measurements from the first IMU and a set of second
4 measurements from the second IMU corresponding to a predetermined
5 period of time; and
6 predict a motion of the first component relative to the second component over the
7 predetermined period of time;

8 wherein the controller is further configured to predict the position of the first
9 component relative to the second component by updating an initial
10 position based upon the motion predicted by the trained neural network
11 model; or preferably wherein the trained neural network is configured to:

12 receive a set of first measurements from the first IMU and a set of second
13 measurements from the second IMU corresponding to a predetermined
14 period of time; and
15 predict an expected motion of the first component relative to the second
16 component over a future time period following the predetermined period
17 of time;

18 wherein the controller is further configured to predict the position of the first
19 component relative to the second component by updating an initial
20 position based upon the motion predicted by the trained neural network
21 model, and wherein the position corresponds to an expected future position
22 of the first component relative to the second component; and preferably
23 wherein the trained neural network is configured to:

24 receive velocity information indicating a current velocity of the first component
25 relative to the second component;

26 extrapolate an expected future position estimate using the received velocity
27 information corresponding to a future point in time;

28 predict an expected offset from the expected future position estimate
29 corresponding to the future point in time; and

30 apply the expected offset to the expected future position estimate to predict the
31 expected future position of the first component relative to the second
32 component at the future point in time.

33 7. The system of any preceding claim, wherein the trained neural network model
34 is a long short-term memory (LSTM) network which when unrolled matches a length of a set
35 of first measurements from the first IMU and a set of second measurements from the second

1 IMU corresponding to a predetermined period of time; or preferably wherein the trained
2 neural network model is a convolutional neural network (CNN) model.

3 8. The system of any of claims 1 to 4 or 6 to 7, wherein the controller is further
4 configured to:

5 track positions of the first component and second component based upon received
6 information indicating positions of the first and second components from
7 an additional sensor;

8 compare the predicted position of the first component relative to the second
9 component predicted using the trained neural network model to the
10 positions of the first and second components received from the additional
11 sensor;

12 determine whether the additional sensor has lost tracking, based upon the
13 comparison; and

14 responsive to determining that the additional sensor has lost tracking, adjust the
15 tracked positions of the first component and second component based on
16 the predicted position.

17 9. A method comprising:

18 receiving, from a first inertial measurement unit (IMU), first measurements
19 indicating motion of a first component, wherein the first component is
20 positioned at a first portions of a user's body;

21 receiving, from a second IMU, second measurements indicating motion of a
22 second component positioned at a second portion of the user's body,
23 wherein a set of potential positions of the first component and the second
24 component is determined by physiological constraints of the first portion
25 and the second portion of the user's body; and

26 predicting a position of the first component relative to the second component
27 using a trained neural network model trained using the set of potential
28 positions.

29 10. The method of claim 9, wherein the first component is a head-mounted device
30 (HMD) configured to be worn on the user's head, and the second component is a handheld
31 device configured to be held in a hand of the user.

32 11. The method of claim 9 or claim 10, wherein the position of the first
33 component relative to the second component predicted by the trained neural network
34 corresponds to an expected position of the first component relative to the second component
35 corresponding to a future time.

- 1 12. The method of any of claims 9 to 11, further comprising:
2 at the trained neural network model:
3 receiving a set of first measurements from the first IMU and a set of
4 second measurements from the second IMU corresponding to a
5 predetermined period of time,
6 predicting a motion of the first component relative to the second
7 component over the predetermined period of time; and
8 predicting the position of the first component relative to the second component by
9 updating an initial position based upon the motion predicted by the trained
10 neural network model; or preferably further comprising:
11 at the trained neural network model:
12 receive a set of first measurements from the first IMU and a set of
13 second measurements from the second IMU corresponding to a
14 predetermined period of time,
15 predicting an expected motion of the first component relative to the
16 second component over a future time period following the
17 predetermined period of time; and
18 predicting the position of the first component relative to the second component by
19 updating an initial position based upon the motion predicted by the trained
20 neural network model, wherein the position corresponds to an expected
21 future position of the first component relative to the second component.
- 22 13. The method of claim 9, further comprising:
23 tracking positions of the first component and second component based upon
24 received information indicating positions of the first and second
25 components from an additional sensor;
26 comparing the predicted position of the first component relative to the second
27 component predicted using the trained neural network model to the
28 positions of the first and second components received from the additional
29 sensor;
30 determining whether the additional sensor has lost tracking, based upon the
31 comparison; and
32 responsive to determining that the additional sensor has lost tracking, adjusting the
33 tracked positions of the first component and second component based on
34 the predicted position.
- 35 14. A computer readable non-transitory storage medium, storing instructions for:

1 receiving, from a first inertial measurement unit (IMU), first measurements
2 indicating motion of a first component, wherein the first component is
3 positioned at a first portions of a user's body;
4 receiving, from a second IMU, second measurements indicating motion of a
5 second component positioned at a second portion of the user's body,
6 wherein a set of potential positions of the first component and the second
7 component is determined by physiological constraints of the first portion
8 and the second portion of the user's body; and
9 predicting a position of the first component relative to the second component
10 using a trained neural network model and the set of potential positions.
11 15. The computer readable non-transitory storage medium of claim 14, wherein
12 the first component is a head-mounted device (HMD) configured to be worn on the user's
13 head, and the second component is a handheld device configured to be held in a hand of the
14 user.

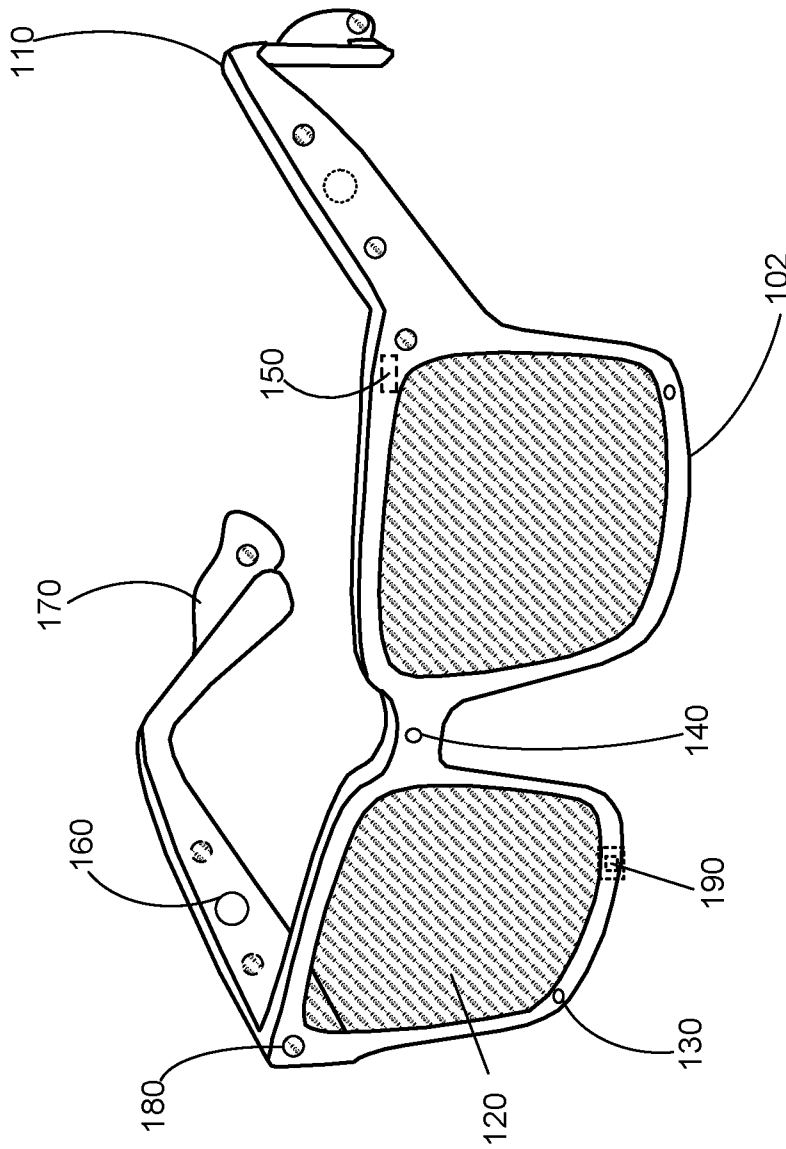
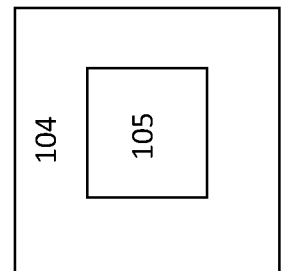


FIG. 1A

100



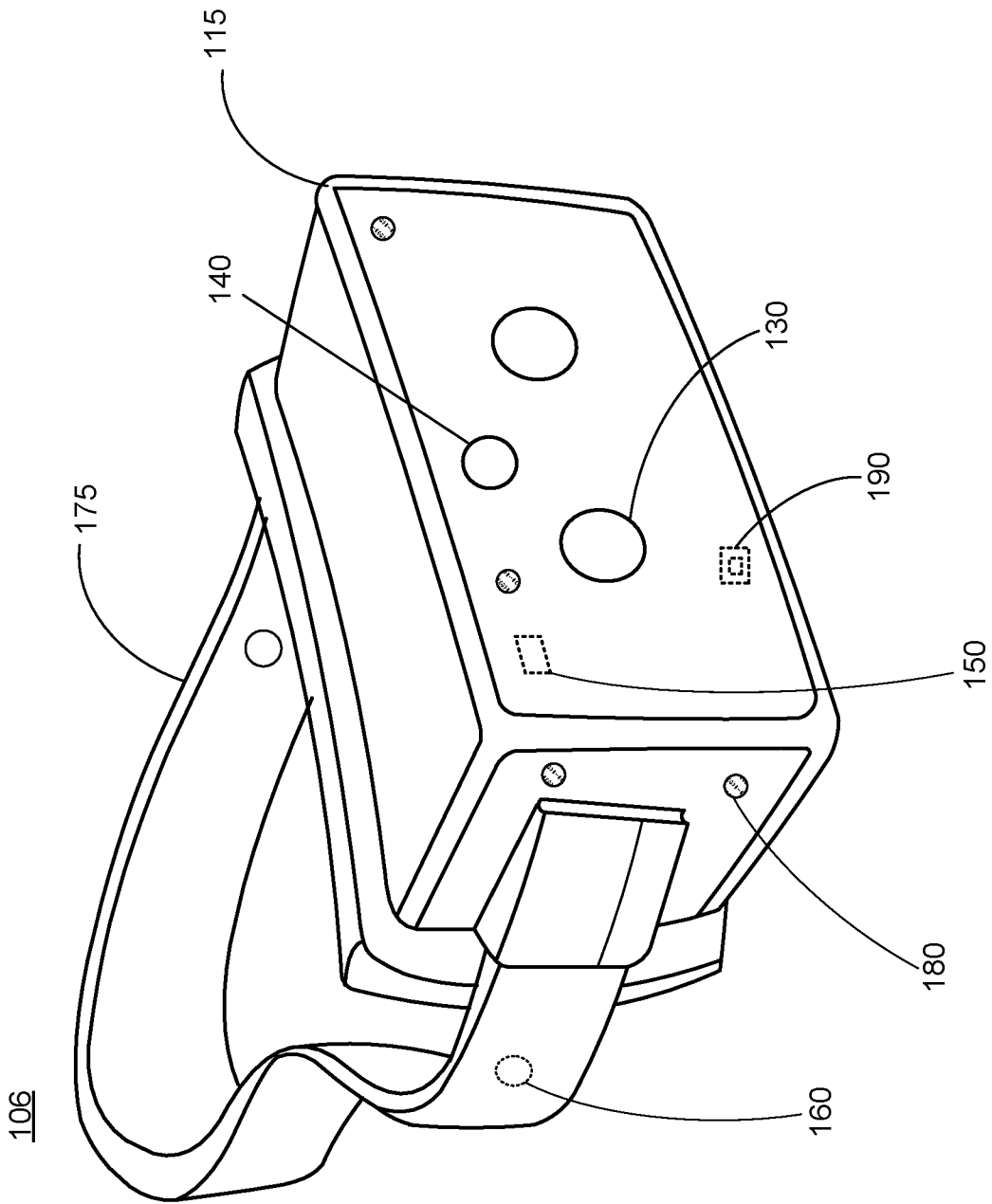


FIG. 1B

3/9

200

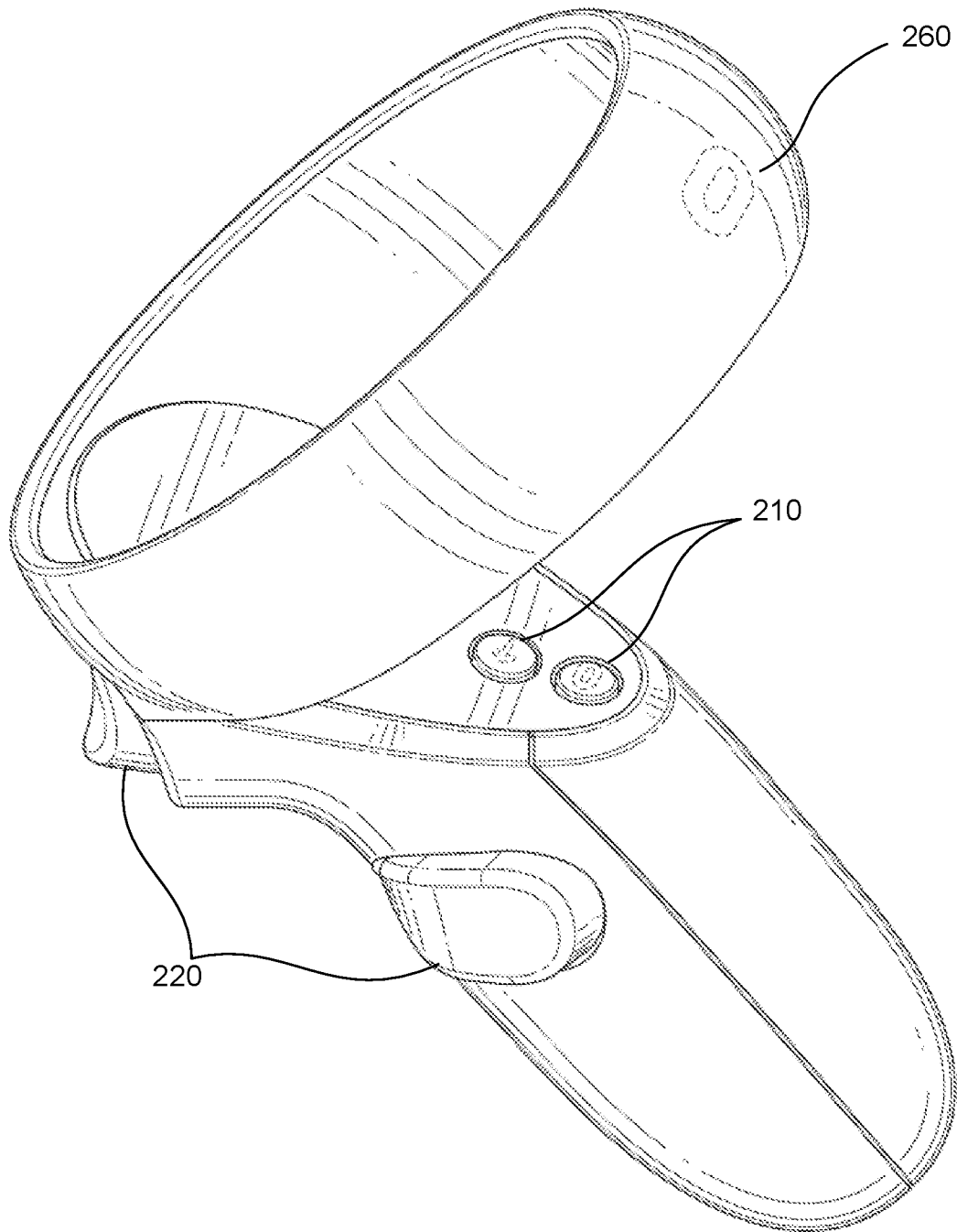


FIG. 2

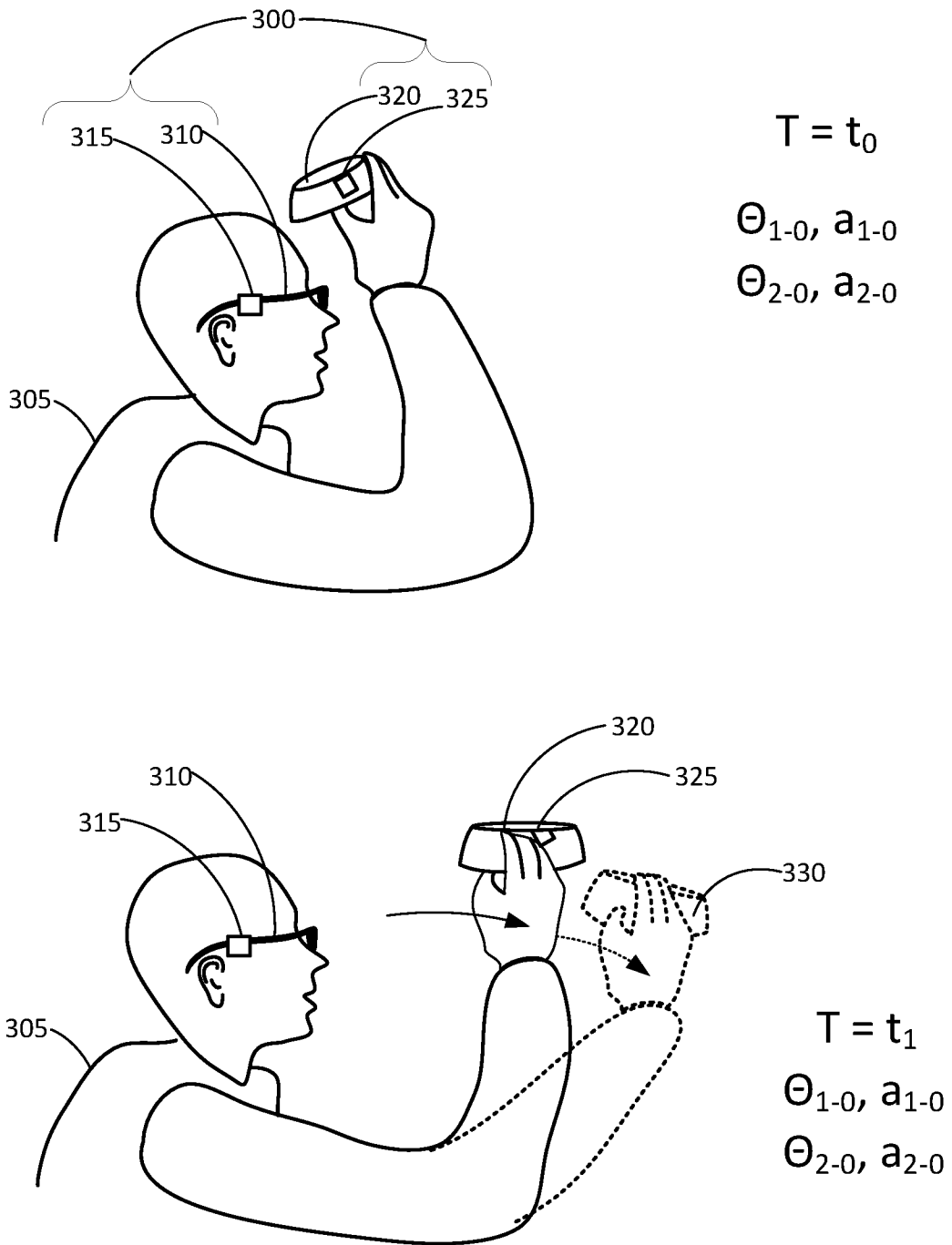


FIG. 3

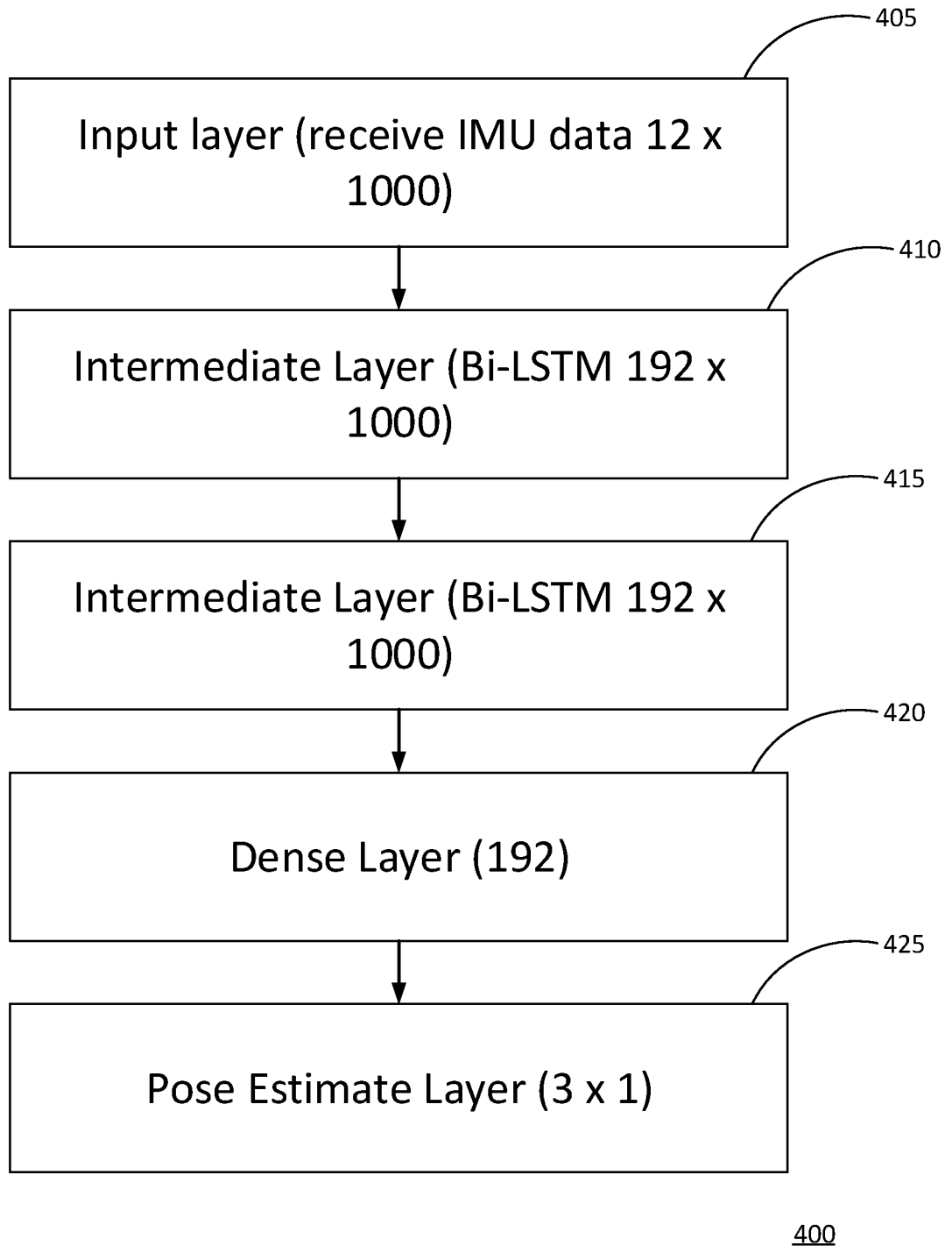


FIG. 4

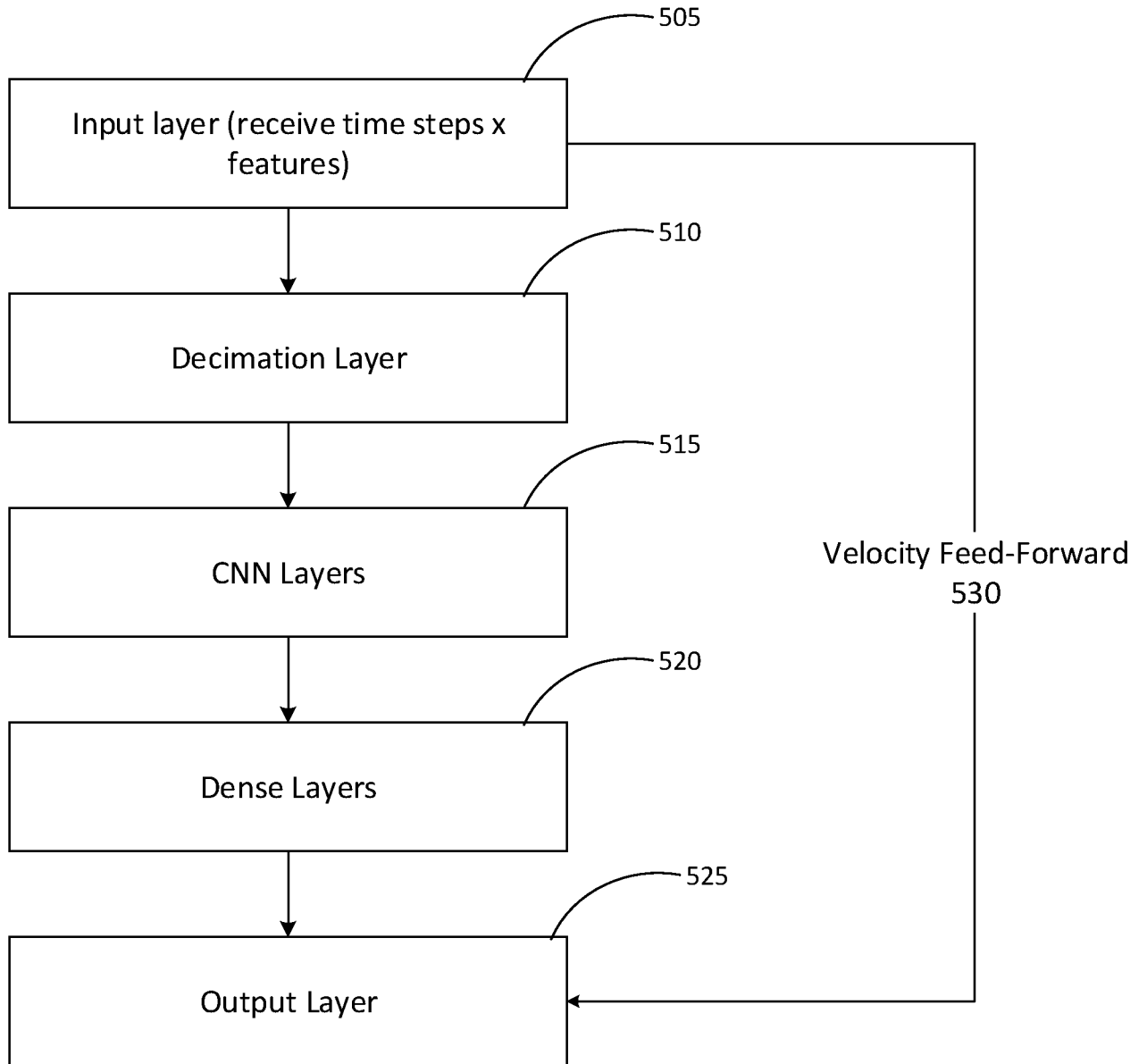


FIG. 5

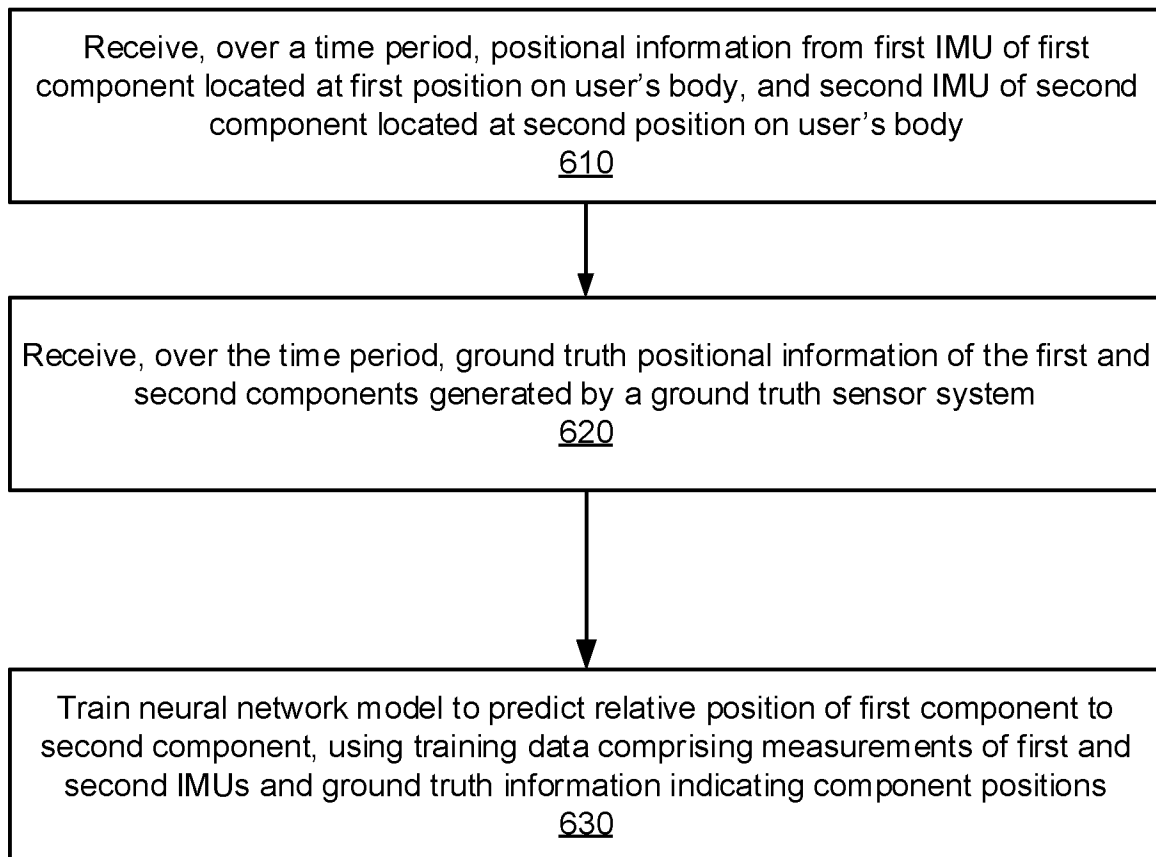


FIG. 6A

8/9

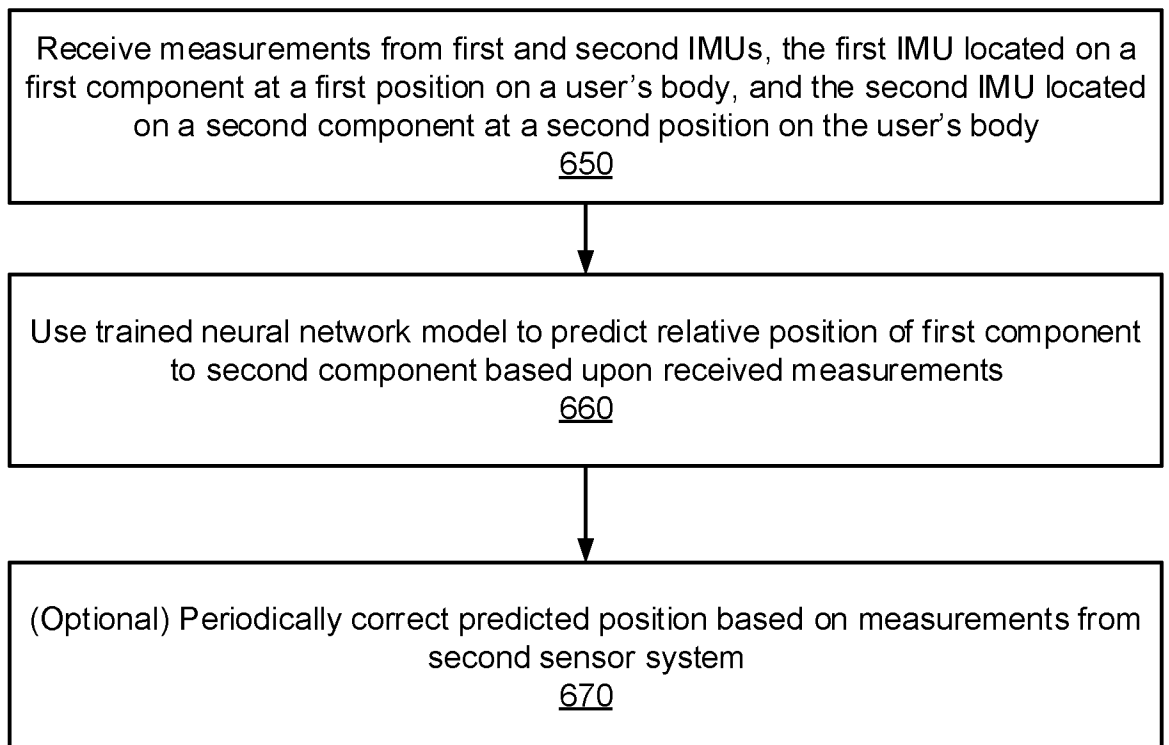


FIG. 6B

700

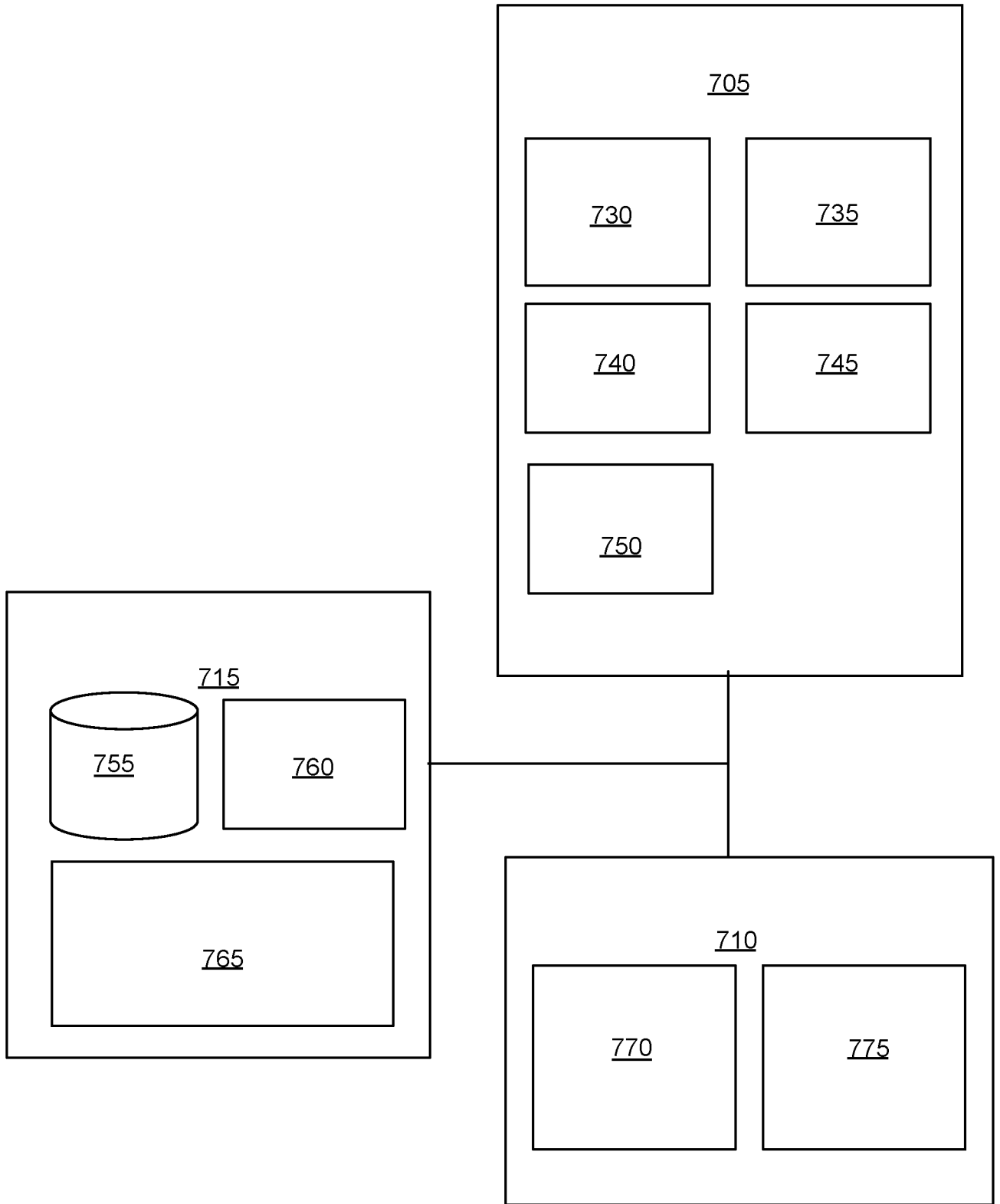


FIG. 7

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2021/065026

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F3/01 G06F3/0346 G06N3/02
ADD.

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

B. FIELDS SEARCHED
 Minimum documentation searched (classification system followed by classification symbols)
G06F G06N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2020/225738 A1 (ERIVANTCEV VIKTOR VLADIMIROVICH [RU] ET AL) 16 July 2020 (2020-07-16)	1, 2, 9, 10, 14, 15
Y	paragraphs [0039], [0057] - [0061], [0068] - [0070], [0083], [0088] - [0096], [0102]; figures 1, 2	3, 4, 6, 11, 12
Y	US 2020/364901 A1 (CHOUDHURI CHIRANJIB [IN] ET AL) 19 November 2020 (2020-11-19) claims 15, 17, 21, 22	3, 4, 6, 11, 12
	----- -/--	

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents :

<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"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</p> <p>"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</p> <p>"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</p> <p>"&" document member of the same patent family</p>
---	---

Date of the actual completion of the international search 17 March 2022	Date of mailing of the international search report 31/05/2022
---	---

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

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2021/065026

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>SONG CHEN ET AL: "Sensor Fusion for Learning-based Tracking of Controller Movement in Virtual Reality", 2019 27TH EUROPEAN SIGNAL PROCESSING CONFERENCE (EUSIPCO), EURASIP, 2 September 2019 (2019-09-02), pages 1-5, XP033660204, DOI: 10.23919/EUSIPCO.2019.8902570 Section I, First paragraph of Section II, Section III.D and Fig. 1</p> <p>-----</p>	3, 4, 6, 11, 12
Y	<p>WANG BOYU ET AL: "Predicting Body Movement and Recognizing Actions: An Integrated Framework for Mutual Benefits", 2018 13TH IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE & GESTURE RECOGNITION (FG 2018), IEEE, 15 May 2018 (2018-05-15), pages 341-348, XP033354718, DOI: 10.1109/FG.2018.00056 Sections I and III and Fig. 2</p> <p>-----</p>	3, 4, 6, 11, 12
Y	<p>KRATZER PHILIPP ET AL: "Prediction of Human Full-Body Movements with Motion Optimization and Recurrent Neural Networks", 2020 IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION (ICRA), IEEE, 31 May 2020 (2020-05-31), pages 1792-1798, XP033826625, DOI: 10.1109/ICRA40945.2020.9197290 Section III.C, Fig. 2</p> <p>-----</p>	3, 4, 6, 11, 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/065026

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims;; it is covered by claims Nos.:
1-4, 6, 9-12, 14, 15

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-4, 6, 9-12, 14, 15

(i) The subject-matter of claim 1 (respectively claims 9 and 14)

(ii) The subject-matter of claims 2, 10, 15

(iii) The subject-matter of claims 4, 6 and 12, insofar as it relates to an estimation for the present time

(iv) The subject-matter of claims 3 and 11

(v) The subject-matter of claims 4, 6 and 12, insofar as it relates to an estimation for a future time

2. claim: 5

(i) The subject-matter of claim 1 (respectively claims 9 and 14)

(vi) The subject-matter of claim 5

3. claim: 7

(i) The subject-matter of claim 1 (respectively claims 9 and 14)

(vii) The subject-matter of claim 7

4. claims: 8, 13

(i) (i) The subject-matter of claim 1 (respectively claims 9 and 14)

(viii) The subject-matter of claims 8 and 13

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2021/065026

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2020225738 A1	16-07-2020	US 10416755 B1	17-09-2019
		US 2019369715 A1	05-12-2019
		US 2020225738 A1	16-07-2020

US 2020364901 A1	19-11-2020	CN 113811920 A	17-12-2021
		EP 3970117 A1	23-03-2022
		US 2020364901 A1	19-11-2020
		WO 2020232174 A1	19-11-2020
