



(86) Date de dépôt PCT/PCT Filing Date: 2010/03/02
(87) Date publication PCT/PCT Publication Date: 2010/09/23
(45) Date de délivrance/Issue Date: 2017/09/12
(85) Entrée phase nationale/National Entry: 2011/08/18
(86) N° demande PCT/PCT Application No.: US 2010/025932
(87) N° publication PCT/PCT Publication No.: 2010/107577
(30) Priorité/Priority: 2009/03/20 (US12/408,589)

(51) Cl.Int./Int.Cl. *G06F 3/01* (2006.01),
A63F 13/21 (2014.01), *A63F 13/428* (2014.01)

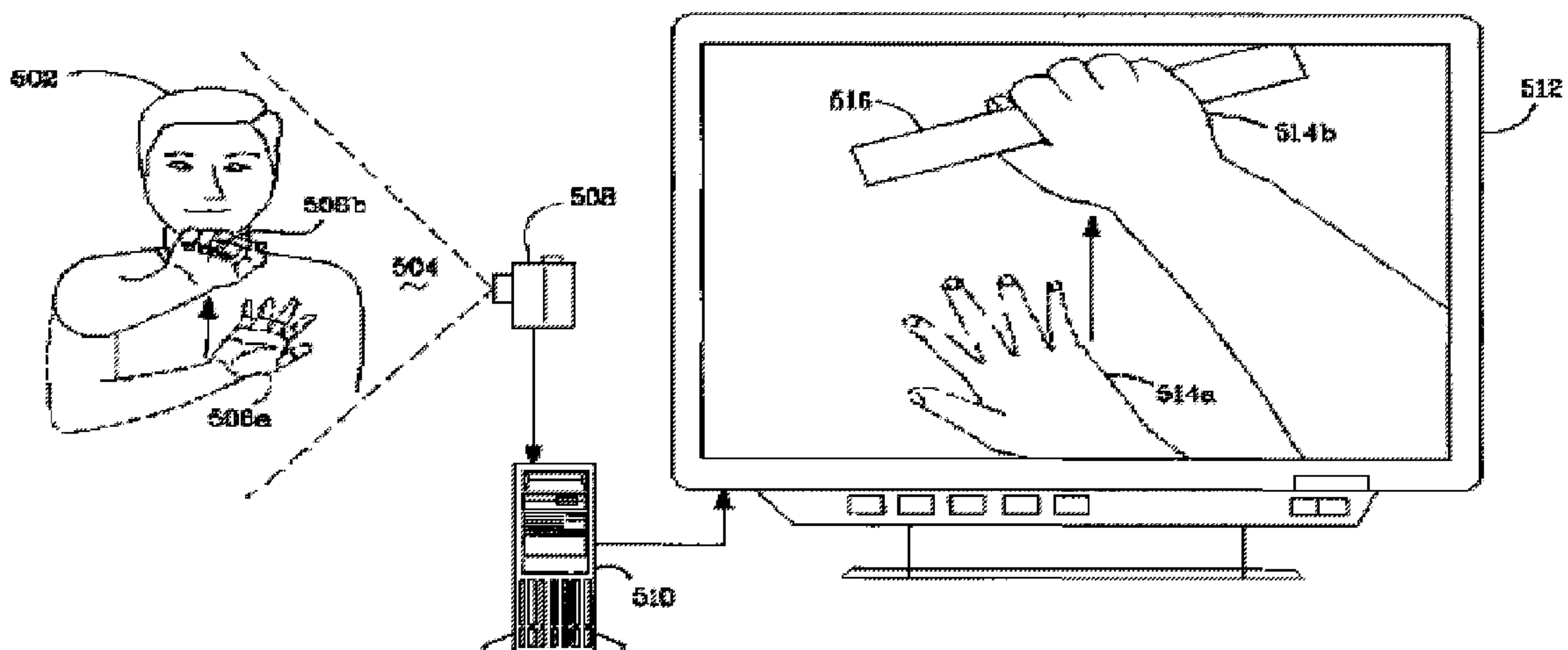
(72) Inventeurs/Inventors:
LATTA, STEPHEN G., US;
GEISNER, KEVIN, US;
MARKOVIC, RELJA, US;
BENNETT, DARREN ALEXANDER, US;
TOMLIN, ARTHUR CHARLES, US

(73) Propriétaire/Owner:
MICROSOFT TECHNOLOGY LICENSING, LLC, US

(74) Agent: SMART & BIGGAR

(54) Titre : MANIPULATION D'OBJETS VIRTUELS

(54) Title: VIRTUAL OBJECT MANIPULATION



(57) Abrégé/Abstract:

Systems, methods and computer readable media are disclosed for manipulating virtual objects. A user may utilize a controller, such as his hand, in physical space to associate with a cursor in a virtual environment. As the user manipulates the controller in physical space, this is captured by a depth camera. The image data from the depth camera is parsed to determine how the controller is manipulated, and a corresponding manipulation of the cursor is performed in virtual space. Where the cursor interacts with a virtual object in the virtual space, that virtual object is manipulated by the cursor.

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

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
23 September 2010 (23.09.2010)(10) International Publication Number
WO 2010/107577 A3

(51) International Patent Classification:

G06F 3/03 (2006.01) *G06F 3/048* (2006.01)

(21) International Application Number:

PCT/US2010/025932

(22) International Filing Date:

2 March 2010 (02.03.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

12/408,589 20 March 2009 (20.03.2009) US

(71) Applicant (for all designated States except US): **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).(72) Inventors: **LATTA, Stephen, G**; One Microsoft Way, Redmond, Washington 98052-6399 (US). **GEISNER, Kevin**; One Microsoft Way, Redmond, Washington 98052-6399 (US). **MARKOVIC, Relja**; One Microsoft Way, Redmond, Washington 98052-6399 (US). **BENNETT, Darren, Alexander**; One Microsoft Way, Redmond, Washington 98052-6399 (US). **TOMLIN, Arthur, Charles**; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))

(88) Date of publication of the international search report:

6 January 2011

(54) Title: VIRTUAL OBJECT MANIPULATION

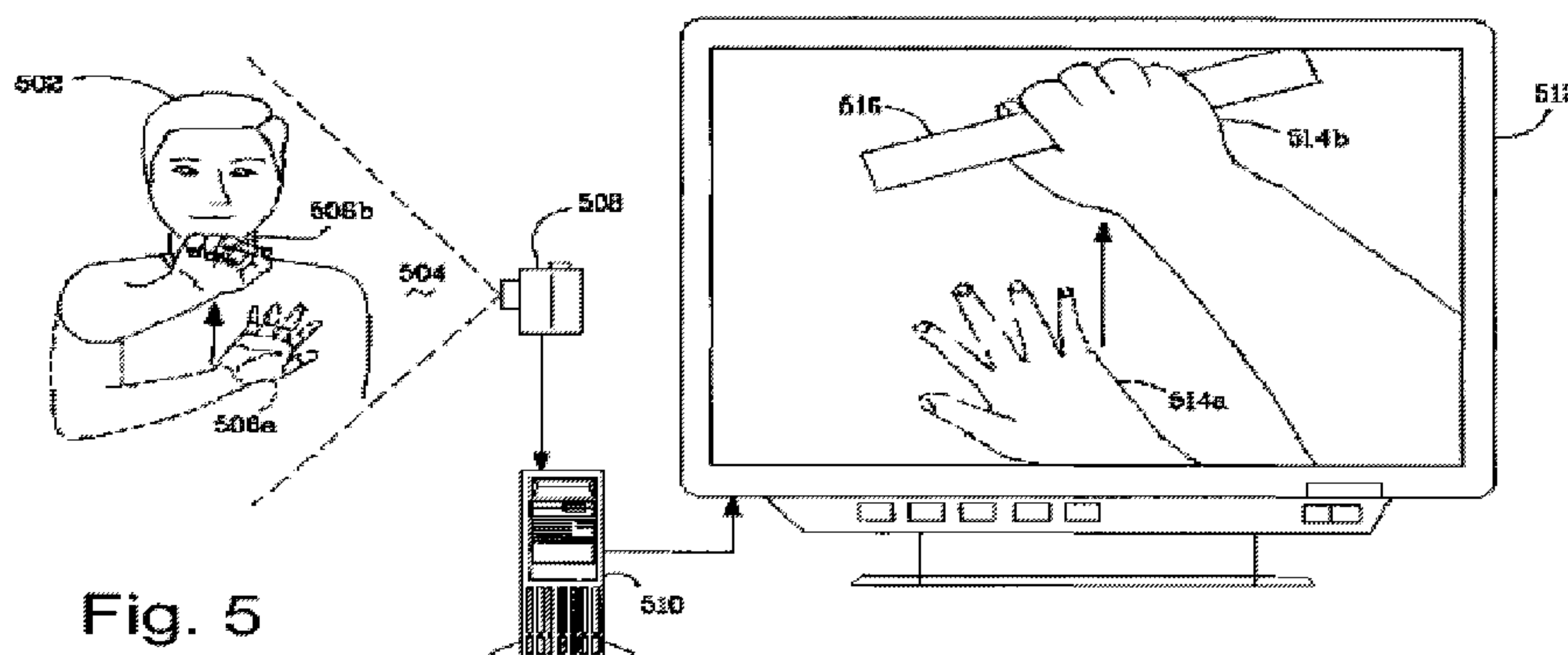


Fig. 5

(57) Abstract: Systems, methods and computer readable media are disclosed for manipulating virtual objects. A user may utilize a controller, such as his hand, in physical space to associate with a cursor in a virtual environment. As the user manipulates the controller in physical space, this is captured by a depth camera. The image data from the depth camera is parsed to determine how the controller is manipulated, and a corresponding manipulation of the cursor is performed in virtual space. Where the cursor interacts with a virtual object in the virtual space, that virtual object is manipulated by the cursor.

51050-145

VIRTUAL OBJECT MANIPULATION BACKGROUND OF THE INVENTION

[0001] Many computing applications, such as computer games, present to a user virtual environments or worlds. These environments often contain virtual objects that the user may interact with. Typically, such interaction occurs through input devices such as controllers, remotes, keyboards, mice, or the like. Unfortunately, such methods of input do not correlate well with the action being performed, such as pressing a button on a controller to open a door.

SUMMARY OF THE INVENTION

[0002] It would therefore be an improvement to enable the user to interact with virtual objects in a virtual environment using a mapping between a physical object that the user controls, such as his hand, that corresponds to a cursor in the virtual environment. Disclosed herein are methods, systems and computer readable media that perform such virtual object manipulation.

[0003] In an exemplary embodiment, a physical controller of the user is associated with a cursor in the virtual environment. User controller input is received, and from that, corresponding cursor manipulation is determined. Where the cursor manipulation corresponds to the manipulation of a virtual object, that manipulation on the virtual object is performed, and the result is displayed on a display device.

[0003a] According to one embodiment of the present invention, there is provided a method for manipulating a first virtual object in a virtual space, the first virtual object being associated with a physical object in a first physical space, the physical object being separate from a first user, comprising: receiving first depth image data indicative of the first physical space; generating the first virtual object in the virtual space based on the physical object as represented in the first depth image data; in response to receiving second depth image data indicative of the first physical space where the physical object is being moved by the first user, displaying a corresponding moving of the first virtual object in the virtual space; and in response to receiving third depth image data indicative of a second user

51050-145

in a second physical space, the second physical space differing from the first physical space in which the physical object is located, and in response to determining that the third depth image data is indicative of the second user attempting to move a second virtual object in the virtual space and the first virtual object in the virtual space without the second user having moved the physical object in the first physical space, determining to display movement of the second
5 virtual object and not to display movement of the first virtual object.

[0003b] According to another embodiment of the present invention, there is provided a system for interacting with a first virtual object in a virtual environment, the first virtual object being associated with a physical object in a first physical space, the physical
10 object being separate from a first user, comprising: a memory bearing instructions that, upon execution by a processor, cause the system at least to: generate the first virtual object based in the virtual environment on the physical object as represented in first depth image data; in response to receiving second depth image data indicative of the first physical space where the physical object is being moved by the first user, display a corresponding moving of the first
15 virtual object in the virtual environment; and in response to receiving third depth image data indicative of a second user in a second physical space, the second physical space differing from the first physical space in which the physical object is located, and in response to determining that the third depth image data is indicative of the second user attempting to move a second virtual object in the virtual environment and the first virtual object in the
20 virtual environment without the second user having moved the physical object in the first physical space, determine to display movement of the second virtual object and not to display movement of the first virtual object.

[0003c] According to still another embodiment of the present invention, there is provided a computer-readable storage medium having stored thereon computer executable
25 instructions that upon execution on a computer, cause the computer to perform operations comprising: receiving first depth image data indicative of a first physical space; generating a first virtual object in a virtual space based on a physical object as represented in the first depth image data; in response to receiving second depth image data indicative of the first physical space where the physical object is being moved by a first user, displaying a corresponding
30 moving of the first virtual object in the virtual space; and in response to receiving third depth

51050-145

image data indicative of a second user in a second physical space, the second physical space differing from the first physical space in which the physical object is located, and in response to determining that the third depth image data is indicative of the second user attempting to move a second virtual object in the virtual space and the first virtual object in the virtual space
5 without the second user having moved the physical object in the first physical space, determining to display movement of the second virtual object and not to display movement of the first virtual object.

[0004] It can be appreciated by one of skill in the art that one or more various aspects of the disclosure may include but are not limited to circuitry and/or programming for
10 effecting the herein-referenced aspects of the present disclosure; the circuitry and/or programming can be virtually any combination of hardware, software, and/or firmware configured to effect the herein-referenced aspects depending upon the design choices of the system designer.

[0005] The foregoing is a summary and thus contains, by necessity,
15 simplifications, generalizations and omissions of detail. Those skilled in the art will appreciate that the summary is illustrative only and is not intended to be in any way limiting.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] The systems, methods, and computer readable media for virtual object manipulation in accordance with this specification are further described with reference to the
20 accompanying drawings in which:

[0007] FIGs. 1A and 1B illustrate an example embodiment of a target recognition, analysis, and tracking system with a user playing a game.

[0008] FIG. 2 illustrates an example embodiment of a capture device that may be used in a target recognition, analysis, and tracking system.

5 [0009] FIG. 3A illustrates an example embodiment of a computing environment that may be used to interpret one or more gestures in a target recognition, analysis, and tracking system.

[0010] FIG. 3B illustrates another example embodiment of a computing environment that may be used to interpret one or more gestures in a target recognition,
10 analysis, and tracking system.

[0011] FIG. 4A illustrates a skeletal mapping of a user that has been generated from the target recognition, analysis, and tracking system of FIG. 2.

[0012] FIG. 5 illustrates a system that captures user controller manipulation and maps it to cursor manipulation of a virtual object.

15 [0013] FIG. 6 illustrates exemplary operational procedures for virtual object manipulation.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0014] As will be described herein, a user may control an application executing on a computing environment such as a game console, a computer, or the like by
20 performing one or more gestures. According to one embodiment, the gestures may be received by, for example, a capture device. For example, the capture device may capture a depth image of a scene. In one embodiment, the capture device may determine whether one or more targets or objects in the scene corresponds to a human target such as the user. To determine whether a target or object in the scene corresponds a human target, each of
25 the targets may be flood filled and compared to a pattern of a human body model. Each target or object that matches the human body model may then be scanned to generate a skeletal model associated therewith. The skeletal model may then be provided to the computing environment such that the computing environment may track the skeletal model, render an avatar associated with the skeletal model, and may determine which
30 controls to perform in an application executing on the computer environment based on, for example, gestures of the user that have been recognized from the skeletal model. A gesture recognizer engine, the architecture of which is described more fully below, is used to determine when a particular gesture has been made by the user.

[0015] FIGs. 1A and 1B illustrate an example embodiment of a configuration of a target recognition, analysis, and tracking system 10 with a user 18 playing a boxing game. In an example embodiment, the target recognition, analysis, and tracking system 10 may be used to recognize, analyze, and/or track a human target such as the user 18.

5 [0016] As shown in FIG. 1A, the target recognition, analysis, and tracking system 10 may include a computing environment 12. The computing environment 12 may be a computer, a gaming system or console, or the like. According to an example embodiment, the computing environment 12 may include hardware components and/or software components such that the computing environment 12 may be used to execute
10 applications such as gaming applications, non-gaming applications, or the like.

[0017] As shown in FIG. 1A, the target recognition, analysis, and tracking system 10 may further include a capture device 20. The capture device 20 may be, for example, a camera that may be used to visually monitor one or more users, such as the user 18, such that gestures performed by the one or more users may be captured, analyzed,
15 and tracked to perform one or more controls or actions within an application, as will be described in more detail below.

[0018] According to one embodiment, the target recognition, analysis, and tracking system 10 may be connected to an audiovisual device 16 such as a television, a monitor, a high-definition television (HDTV), or the like that may provide game or
20 application visuals and/or audio to a user such as the user 18. For example, the computing environment 12 may include a video adapter such as a graphics card and/or an audio adapter such as a sound card that may provide audiovisual signals associated with the game application, non-game application, or the like. The audiovisual device 16 may receive the audiovisual signals from the computing environment 12 and may then output
25 the game or application visuals and/or audio associated with the audiovisual signals to the user 18. According to one embodiment, the audiovisual device 16 may be connected to the computing environment 12 via, for example, an S-Video cable, a coaxial cable, an HDMI cable, a DVI cable, a VGA cable, or the like.

[0019] As shown in FIGs. 1A and 1B, the target recognition, analysis, and
30 tracking system 10 may be used to recognize, analyze, and/or track a human target such as the user 18. For example, the user 18 may be tracked using the capture device 20 such that the movements of user 18 may be interpreted as controls that may be used to affect the application being executed by computer environment 12. Thus, according to one embodiment, the user 18 may move his or her body to control the application.

[0020] As shown in FIGs. 1A and 1B, in an example embodiment, the application executing on the computing environment 12 may be a boxing game that the user 18 may be playing. For example, the computing environment 12 may use the audiovisual device 16 to provide a visual representation of a boxing opponent 22 to the user 18. The computing environment 12 may also use the audiovisual device 16 to provide a visual representation of a player avatar 24 that the user 18 may control with his or her movements. For example, as shown in FIG. 1B, the user 18 may throw a punch in physical space to cause the player avatar 24 to throw a punch in game space. Thus, according to an example embodiment, the computer environment 12 and the capture device 20 of the target recognition, analysis, and tracking system 10 may be used to recognize and analyze the punch of the user 18 in physical space such that the punch may be interpreted as a game control of the player avatar 24 in game space.

[0021] Other movements by the user 18 may also be interpreted as other controls or actions, such as controls to bob, weave, shuffle, block, jab, or throw a variety of different power punches. Furthermore, some movements may be interpreted as controls that may correspond to actions other than controlling the player avatar 24. For example, the player may use movements to end, pause, or save a game, select a level, view high scores, communicate with a friend, etc.

[0022] In example embodiments, the human target such as the user 18 may have an object. In such embodiments, the user of an electronic game may be holding the object such that the motions of the player and the object may be used to adjust and/or control parameters of the game. For example, the motion of a player holding a racket may be tracked and utilized for controlling an on-screen racket in an electronic sports game. In another example embodiment, the motion of a player holding an object may be tracked and utilized for controlling an on-screen weapon in an electronic combat game.

[0023] According to other example embodiments, the target recognition, analysis, and tracking system 10 may further be used to interpret target movements as operating system and/or application controls that are outside the realm of games. For example, virtually any controllable aspect of an operating system and/or application may be controlled by movements of the target such as the user 18.

[0024] FIG. 2 illustrates an example embodiment of the capture device 20 that may be used in the target recognition, analysis, and tracking system 10. According to an example embodiment, the capture device 20 may be configured to capture video with depth information including a depth image that may include depth values via any suitable

technique including, for example, time-of-flight, structured light, stereo image, or the like. According to one embodiment, the capture device 20 may organize the calculated depth information into “Z layers,” or layers that may be perpendicular to a Z axis extending from the depth camera along its line of sight.

5 **[0025]** As shown in FIG. 2, the capture device 20 may include an image camera component 22. According to an example embodiment, the image camera component 22 may be a depth camera that may capture the depth image of a scene. The depth image may include a two-dimensional (2-D) pixel area of the captured scene where each pixel in the 2-D pixel area may represent a length in, for example, centimeters, millimeters, or the
10 like of an object in the captured scene from the camera.

[0026] As shown in FIG. 2, according to an example embodiment, the image camera component 22 may include an IR light component 24, a three-dimensional (3-D) camera 26, and an RGB camera 28 that may be used to capture the depth image of a scene. For example, in time-of-flight analysis, the IR light component 24 of the capture device 20
15 may emit an infrared light onto the scene and may then use sensors (not shown) to detect the backscattered light from the surface of one or more targets and objects in the scene using, for example, the 3-D camera 26 and/or the RGB camera 28. In some embodiments, pulsed infrared light may be used such that the time between an outgoing light pulse and a corresponding incoming light pulse may be measured and used to determine a physical
20 distance from the capture device 20 to a particular location on the targets or objects in the scene. Additionally, in other example embodiments, the phase of the outgoing light wave may be compared to the phase of the incoming light wave to determine a phase shift. The phase shift may then be used to determine a physical distance from the capture device to a particular location on the targets or objects.

25 **[0027]** According to another example embodiment, time-of-flight analysis may be used to indirectly determine a physical distance from the capture device 20 to a particular location on the targets or objects by analyzing the intensity of the reflected beam of light over time via various techniques including, for example, shuttered light pulse imaging.

30 **[0028]** In another example embodiment, the capture device 20 may use a structured light to capture depth information. In such an analysis, patterned light (i.e., light displayed as a known pattern such as grid pattern or a stripe pattern) may be projected onto the scene via, for example, the IR light component 24. Upon striking the surface of one or more targets or objects in the scene, the pattern may become deformed in response.

Such a deformation of the pattern may be captured by, for example, the 3-D camera 26 and/or the RGB camera 28 and may then be analyzed to determine a physical distance from the capture device to a particular location on the targets or objects.

5 [0029] According to another embodiment, the capture device 20 may include two or more physically separated cameras that may view a scene from different angles, to obtain visual stereo data that may be resolved to generate depth information

[0030] The capture device 20 may further include a microphone 30. The microphone 30 may include a transducer or sensor that may receive and convert sound into an electrical signal. According to one embodiment, an array of microphones 30 may
10 be used to reduce feedback or audio distortion as compared to a single microphone that receives sound. Additionally, the microphone 30 may be used to receive audio signals that may also be provided by the user to control applications such as game applications, non-game applications, or the like that may be executed by the computing environment 12.

[0031] In an example embodiment, the capture device 20 may further include a
15 processor 32 that may be in operative communication with the image camera component 22. The processor 32 may include a standardized processor, a specialized processor, a microprocessor, or the like that may execute instructions that may include instructions for receiving the depth image, determining whether a suitable target may be included in the depth image, converting the suitable target into a skeletal representation or model of the
20 target, or any other suitable instruction.

[0032] The capture device 20 may further include a memory component 34 that may store the instructions that may be executed by the processor 32, images or frames of images captured by the 3-D camera or RGB camera, or any other suitable information, images, or the like. According to an example embodiment, the memory component 34
25 may include random access memory (RAM), read only memory (ROM), cache, Flash memory, a hard disk, or any other suitable storage component. As shown in FIG. 2, in one embodiment, the memory component 34 may be a separate component in communication with the image capture component 22 and the processor 32. According to another embodiment, the memory component 34 may be integrated into the processor 32 and/or
30 the image capture component 22.

[0033] As shown in FIG. 2, the capture device 20 may be in communication with the computing environment 12 via a communication link 36. The communication link 36 may be a wired connection including, for example, a USB connection, a Firewire connection, an Ethernet cable connection, or the like and/or a wireless connection such as

a wireless 802.11b, g, a, or n connection. According to one embodiment, the computing environment 12 may provide a clock to the capture device 20 that may be used to determine when to capture, for example, a scene via the communication link 36.

[0034] Additionally, the capture device 20 may provide the depth information and images captured by, for example, the 3-D camera 26 and/or the RGB camera 28, and a skeletal model that may be generated by the capture device 20 to the computing environment 12 via the communication link 36. The computing environment 12 may then use the skeletal model, depth information, and captured images to, for example, recognize user gestures and in response control an application such as a game or word processor. For example, as shown, in FIG. 2, the computing environment 12 may include a gestures recognizer engine 190. The gestures recognizer engine 190 may include a collection of gesture filters, each comprising information concerning a gesture that may be performed by the skeletal model (as the user moves). The data captured by the cameras 26, 28 and device 20 in the form of the skeletal model and movements associated with it may be compared to the gesture filters in the gesture recognizer engine 190 to identify when a user (as represented by the skeletal model) has performed one or more gestures. Those gestures may be associated with various controls of an application. Thus, the computing environment 12 may use the gesture recognizer engine 190 to interpret movements of the skeletal model and to control an application based on the movements.

[0035] FIG. 3A illustrates an example embodiment of a computing environment that may be used to interpret one or more gestures in a target recognition, analysis, and tracking system. The computing environment such as the computing environment 12 described above with respect to FIGs. 1A-2 may be a multimedia console 100, such as a gaming console. As shown in FIG. 3A, the multimedia console 100 has a central processing unit (CPU) 101 having a level 1 cache 102, a level 2 cache 104, and a flash ROM (Read Only Memory) 106. The level 1 cache 102 and a level 2 cache 104 temporarily store data and hence reduce the number of memory access cycles, thereby improving processing speed and throughput. The CPU 101 may be provided having more than one core, and thus, additional level 1 and level 2 caches 102 and 104. The flash ROM 106 may store executable code that is loaded during an initial phase of a boot process when the multimedia console 100 is powered ON.

[0036] A graphics processing unit (GPU) 108 and a video encoder/video codec (coder/decoder) 114 form a video processing pipeline for high speed and high resolution graphics processing. Data is carried from the graphics processing unit 108 to the video

encoder/video codec 114 via a bus. The video processing pipeline outputs data to an A/V (audio/video) port 140 for transmission to a television or other display. A memory controller 110 is connected to the GPU 108 to facilitate processor access to various types of memory 112, such as, but not limited to, a RAM (Random Access Memory).

5 **[0037]** The multimedia console 100 includes an I/O controller 120, a system management controller 122, an audio processing unit 123, a network interface controller 124, a first USB host controller 126, a second USB controller 128 and a front panel I/O subassembly 130 that are preferably implemented on a module 118. The USB controllers 126 and 128 serve as hosts for peripheral controllers 142(1)-142(2), a wireless adapter 10 148, and an external memory device 146 (e.g., flash memory, external CD/DVD ROM drive, removable media, etc.). The network interface 124 and/or wireless adapter 148 provide access to a network (e.g., the Internet, home network, etc.) and may be any of a wide variety of various wired or wireless adapter components including an Ethernet card, a modem, a Bluetooth module, a cable modem, and the like.

15 **[0038]** System memory 143 is provided to store application data that is loaded during the boot process. A media drive 144 is provided and may comprise a DVD/CD drive, hard drive, or other removable media drive, etc. The media drive 144 may be internal or external to the multimedia console 100. Application data may be accessed via the media drive 144 for execution, playback, etc. by the multimedia console 100. The 20 media drive 144 is connected to the I/O controller 120 via a bus, such as a Serial ATA bus or other high speed connection (e.g., IEEE 1394).

[0039] The system management controller 122 provides a variety of service functions related to assuring availability of the multimedia console 100. The audio processing unit 123 and an audio codec 132 form a corresponding audio processing 25 pipeline with high fidelity and stereo processing. Audio data is carried between the audio processing unit 123 and the audio codec 132 via a communication link. The audio processing pipeline outputs data to the A/V port 140 for reproduction by an external audio player or device having audio capabilities.

[0040] The front panel I/O subassembly 130 supports the functionality of the 30 power button 150 and the eject button 152, as well as any LEDs (light emitting diodes) or other indicators exposed on the outer surface of the multimedia console 100. A system power supply module 136 provides power to the components of the multimedia console 100. A fan 138 cools the circuitry within the multimedia console 100.

[0041] The CPU 101, GPU 108, memory controller 110, and various other components within the multimedia console 100 are interconnected via one or more buses, including serial and parallel buses, a memory bus, a peripheral bus, and a processor or local bus using any of a variety of bus architectures. By way of example, such architectures can include a Peripheral Component Interconnects (PCI) bus, PCI-Express bus, etc.

[0042] When the multimedia console 100 is powered ON, application data may be loaded from the system memory 143 into memory 112 and/or caches 102, 104 and executed on the CPU 101. The application may present a graphical user interface that provides a consistent user experience when navigating to different media types available on the multimedia console 100. In operation, applications and/or other media contained within the media drive 144 may be launched or played from the media drive 144 to provide additional functionalities to the multimedia console 100.

[0043] The multimedia console 100 may be operated as a standalone system by simply connecting the system to a television or other display. In this standalone mode, the multimedia console 100 allows one or more users to interact with the system, watch movies, or listen to music. However, with the integration of broadband connectivity made available through the network interface 124 or the wireless adapter 148, the multimedia console 100 may further be operated as a participant in a larger network community.

[0044] When the multimedia console 100 is powered ON, a set amount of hardware resources are reserved for system use by the multimedia console operating system. These resources may include a reservation of memory (e.g., 16MB), CPU and GPU cycles (e.g., 5%), networking bandwidth (e.g., 8 kbs), etc. Because these resources are reserved at system boot time, the reserved resources do not exist from the application's view.

[0045] In particular, the memory reservation preferably is large enough to contain the launch kernel, concurrent system applications and drivers. The CPU reservation is preferably constant such that if the reserved CPU usage is not used by the system applications, an idle thread will consume any unused cycles.

[0046] With regard to the GPU reservation, lightweight messages generated by the system applications (e.g., popups) are displayed by using a GPU interrupt to schedule code to render popup into an overlay. The amount of memory required for an overlay depends on the overlay area size and the overlay preferably scales with screen resolution. Where a full user interface is used by the concurrent system application, it is preferable to

use a resolution independent of application resolution. A scaler may be used to set this resolution such that the need to change frequency and cause a TV resynch is eliminated.

[0047] After the multimedia console 100 boots and system resources are reserved, concurrent system applications execute to provide system functionalities. The system functionalities are encapsulated in a set of system applications that execute within the reserved system resources described above. The operating system kernel identifies threads that are system application threads versus gaming application threads. The system applications are preferably scheduled to run on the CPU 101 at predetermined times and intervals in order to provide a consistent system resource view to the application. The scheduling is to minimize cache disruption for the gaming application running on the console.

[0048] When a concurrent system application requires audio, audio processing is scheduled asynchronously to the gaming application due to time sensitivity. A multimedia console application manager (described below) controls the gaming application audio level (e.g., mute, attenuate) when system applications are active.

[0049] Input devices (e.g., controllers 142(1) and 142(2)) are shared by gaming applications and system applications. The input devices are not reserved resources, but are to be switched between system applications and the gaming application such that each will have a focus of the device. The application manager preferably controls the switching of input stream, without knowledge the gaming application's knowledge and a driver maintains state information regarding focus switches. The cameras 26, 28 and capture device 20 may define additional input devices for the console 100.

[0050] FIG. 3B illustrates another example embodiment of a computing environment 220 that may be the computing environment 12 shown in FIGs. 1A-2 used to interpret one or more gestures in a target recognition, analysis, and tracking system. The computing system environment 220 is only one example of a suitable computing environment and is not intended to suggest any limitation as to the scope of use or functionality of the presently disclosed subject matter. Neither should the computing environment 220 be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in the exemplary operating environment 220. In some embodiments the various depicted computing elements may include circuitry configured to instantiate specific aspects of the present disclosure. For example, the term circuitry used in the disclosure can include specialized hardware components configured to perform function(s) by firmware or switches. In other examples

embodiments the term circuitry can include a general purpose processing unit, memory, etc., configured by software instructions that embody logic operable to perform function(s). In example embodiments where circuitry includes a combination of hardware and software, an implementer may write source code embodying logic and the source code
5 can be compiled into machine readable code that can be processed by the general purpose processing unit. Since one skilled in the art can appreciate that the state of the art has evolved to a point where there is little difference between hardware, software, or a combination of hardware/software, the selection of hardware versus software to effectuate specific functions is a design choice left to an implementer. More specifically, one of skill
10 in the art can appreciate that a software process can be transformed into an equivalent hardware structure, and a hardware structure can itself be transformed into an equivalent software process. Thus, the selection of a hardware implementation versus a software implementation is one of design choice and left to the implementer.

[0051] In FIG. 3B, the computing environment 220 comprises a computer 241,
15 which typically includes a variety of computer readable media. Computer readable media can be any available media that can be accessed by computer 241 and includes both volatile and nonvolatile media, removable and non-removable media. The system memory 222 includes computer storage media in the form of volatile and/or nonvolatile memory such as read only memory (ROM) 223 and random access memory (RAM) 260.
20 A basic input/output system 224 (BIOS), containing the basic routines that help to transfer information between elements within computer 241, such as during start-up, is typically stored in ROM 223. RAM 260 typically contains data and/or program modules that are immediately accessible to and/or presently being operated on by processing unit 259. By way of example, and not limitation, FIG. 3B illustrates operating system 225, application
25 programs 226, other program modules 227, and program data 228.

[0052] The computer 241 may also include other removable/non-removable, volatile/nonvolatile computer storage media. By way of example only, FIG. 3B illustrates a hard disk drive 238 that reads from or writes to non-removable, nonvolatile magnetic media, a magnetic disk drive 239 that reads from or writes to a removable, nonvolatile
30 magnetic disk 254, and an optical disk drive 240 that reads from or writes to a removable, nonvolatile optical disk 253 such as a CD ROM or other optical media. Other removable/non-removable, volatile/nonvolatile computer storage media that can be used in the exemplary operating environment include, but are not limited to, magnetic tape cassettes, flash memory cards, digital versatile disks, digital video tape, solid state RAM,

solid state ROM, and the like. The hard disk drive 238 is typically connected to the system bus 221 through a non-removable memory interface such as interface 234, and magnetic disk drive 239 and optical disk drive 240 are typically connected to the system bus 221 by a removable memory interface, such as interface 235.

5 **[0053]** The drives and their associated computer storage media discussed above and illustrated in FIG. 3B, provide storage of computer readable instructions, data structures, program modules and other data for the computer 241. In FIG. 3B, for example, hard disk drive 238 is illustrated as storing operating system 258, application programs 257, other program modules 256, and program data 255. Note that these
10 components can either be the same as or different from operating system 225, application programs 226, other program modules 227, and program data 228. Operating system 258, application programs 257, other program modules 256, and program data 255 are given different numbers here to illustrate that, at a minimum, they are different copies. A user may enter commands and information into the computer 241 through input devices such as
15 a keyboard 251 and pointing device 252, commonly referred to as a mouse, trackball or touch pad. Other input devices (not shown) may include a microphone, joystick, game pad, satellite dish, scanner, or the like. These and other input devices are often connected to the processing unit 259 through a user input interface 236 that is coupled to the system bus, but may be connected by other interface and bus structures, such as a parallel port,
20 game port or a universal serial bus (USB). The cameras 26, 28 and capture device 20 may define additional input devices for the console 100. A monitor 242 or other type of display device is also connected to the system bus 221 via an interface, such as a video interface 232. In addition to the monitor, computers may also include other peripheral output devices such as speakers 244 and printer 243, which may be connected through a
25 output peripheral interface 233.

[0054] The computer 241 may operate in a networked environment using logical connections to one or more remote computers, such as a remote computer 246. The remote computer 246 may be a personal computer, a server, a router, a network PC, a peer device or other common network node, and typically includes many or all of the elements
30 described above relative to the computer 241, although only a memory storage device 247 has been illustrated in FIG. 3B. The logical connections depicted in FIG. 3B include a local area network (LAN) 245 and a wide area network (WAN) 249, but may also include other networks. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets and the Internet.

[0055] When used in a LAN networking environment, the computer 241 is connected to the LAN 245 through a network interface or adapter 237. When used in a WAN networking environment, the computer 241 typically includes a modem 250 or other means for establishing communications over the WAN 249, such as the Internet. The
5 modem 250, which may be internal or external, may be connected to the system bus 221 via the user input interface 236, or other appropriate mechanism. In a networked environment, program modules depicted relative to the computer 241, or portions thereof, may be stored in the remote memory storage device. By way of example, and not limitation, FIG. 3B illustrates remote application programs 248 as residing on memory
10 device 247. It will be appreciated that the network connections shown are exemplary and other means of establishing a communications link between the computers may be used.

[0056] FIG. 4A depicts an example skeletal mapping of a user that may be generated from the capture device 20. In this embodiment, a variety of joints and bones are identified: each hand 302, each forearm 304, each elbow 306, each bicep 308, each
15 shoulder 310, each hip 312, each thigh 314, each knee 316, each foreleg 318, each foot 320, the head 322, the torso 324, the top 326 and bottom 328 of the spine, and the waist 330. Where more points are tracked, additional features may be identified, such as the bones and joints of the fingers or toes, or individual features of the face, such as the nose and eyes.

[0057] Through moving his body, a user may create gestures. A gesture
20 comprises a motion or pose by a user that may be captured as image data and parsed for meaning. A gesture may be dynamic, comprising a motion, such as mimicking throwing a ball. A gesture may be a static pose, such as holding one's crossed forearms 304 in front of his torso 324. A gesture may also incorporate props, such as by swinging a mock
25 sword. A gesture may comprise more than one body part, such as clapping the hands 302 together, or a subtler motion, such as pursing one's lips.

[0058] Gestures may be used for input in a general computing context. For instance, various motions of the hands 302 or other body parts may correspond to common system wide tasks such as navigate up or down in a hierarchical list, open a file, close a
30 file, and save a file. Gestures may also be used in a video-game-specific context, depending on the game. For instance, with a driving game, various motions of the hands 302 and feet 320 may correspond to steering a vehicle in a direction, shifting gears, accelerating, and breaking.

[0059] A user may generate a gesture that corresponds to walking or running, by walking or running in place himself. The user may alternately lift and drop each leg 312-320 to mimic walking without moving. The system may parse this gesture by analyzing each hip 312 and each thigh 314. A step may be recognized when one hip-thigh angle (as measured relative to a vertical line, wherein a standing leg has a hip-thigh angle of 0°, and a forward horizontally extended leg has a hip-thigh angle of 90°) exceeds a certain threshold relative to the other thigh. A walk or run may be recognized after some number of consecutive steps by alternating legs. The time between the two most recent steps may be thought of as a period. After some number of periods where that threshold angle is not met, the system may determine that the walk or running gesture has ceased.

[0060] Given a “walk or run” gesture, an application may set values for application-determined parameters associated with this gesture. These parameters may include the above threshold angle, the number of steps required to initiate a walk or run gesture, a number of periods where no step occurs to end the gesture, and a threshold period that determines whether the gesture is a walk or a run. A fast period may correspond to a run, as the user will be moving his legs quickly, and a slower period may correspond to a walk.

[0061] A gesture may be associated with a set of default parameters at first that the application may override with its own parameters. In this scenario, an application is not forced to provide parameters, but may instead use a set of default parameters that allow the gesture to be recognized in the absence of application-defined parameters.

[0062] There are a variety of outputs that may be associated with the gesture. There may be a baseline “yes or no” as to whether a gesture is occurring. There also may be a confidence level, which corresponds to the likelihood that the user’s tracked movement corresponds to the gesture. This could be a linear scale that ranges over floating point numbers between 0 and 1, inclusive. Wherein an application receiving this gesture information cannot accept false-positives as input, it may use only those recognized gestures that have a high confidence level, such as at least .95. Where an application must recognize every instance of the gesture, even at the cost of false-positives, it may use gestures that have at least a much lower confidence level, such as those merely greater than .2. The gesture may have an output for the time between the two most recent steps, and where only a first step has been registered, this may be set to a reserved value, such as -1 (since the time between any two steps must be positive). The

gesture may also have an output for the highest thigh angle reached during the most recent step.

[0063] Another exemplary gesture is a “heel lift jump.” In this, a user may create the gesture by raising his heels off the ground, but keeping his toes planted.

5 Alternatively, the user may jump into the air where his feet 320 leave the ground entirely. The system may parse the skeleton for this gesture by analyzing the angle relation of the shoulders 310, hips 312 and knees 316 to see if they are in a position of alignment equal to standing up straight. Then these points and upper 326 and lower 328 spine points may be monitored for any upward acceleration. A sufficient combination of acceleration may
10 trigger a jump gesture.

[0064] Given this “heel lift jump” gesture, an application may set values for application-determined parameters associated with this gesture. The parameters may include the above acceleration threshold, which determines how fast some combination of the user’s shoulders 310, hips 312 and knees 316 must move upward to trigger the gesture,
15 as well as a maximum angle of alignment between the shoulders 310, hips 312 and knees 316 at which a jump may still be triggered.

[0065] The outputs may comprise a confidence level, as well as the user’s body angle at the time of the jump.

[0066] Setting parameters for a gesture based on the particulars of the application
20 that will receive the gesture is important in accurately identifying gestures. Properly identifying gestures and the intent of a user greatly helps in creating a positive user experience. Where a gesture recognizer system is too sensitive, and even a slight forward motion of the hand 302 is interpreted as a throw, the user may become frustrated because gestures are being recognized where he has no intent to make a gesture, and thus, he lacks
25 control over the system. Where a gesture recognizer system is not sensitive enough, the system may not recognize conscious attempts by the user to make a throwing gesture, frustrating him in a similar manner. At either end of the sensitivity spectrum, the user becomes frustrated because he cannot properly provide input to the system.

[0067] Another parameter to a gesture may be a distance moved. Where a user’s
30 gestures control the actions of an avatar in a virtual environment, that avatar may be arm’s length from a ball. If the user wishes to interact with the ball and grab it, this may require the user to extend his arm 302-310 to full length while making the grab gesture. In this situation, a similar grab gesture where the user only partially extends his arm 302-310 may not achieve the result of interacting with the ball.

[0068] A gesture or a portion thereof may have as a parameter a volume of space in which it must occur. This volume of space may typically be expressed in relation to the body where a gesture comprises body movement. For instance, a football throwing gesture for a right-handed user may be recognized only in the volume of space no lower
5 than the right shoulder 310a, and on the same side of the head 322 as the throwing arm 302a-310a. It may not be necessary to define all bounds of a volume, such as with this throwing gesture, where an outer bound away from the body is left undefined, and the volume extends out indefinitely, or to the edge of scene that is being monitored.

[0069] FIG. 5 illustrates a system that captures user controller manipulation in a
10 physical space and maps it to cursor manipulation of a virtual object in a virtual environment.

[0070] A virtual space may comprise a representation of a three-dimensional space that a user may affect – say by moving an object – through user input. That virtual space may be a completely virtual space that has no correlation to a physical space of the
15 user – such as a representation of a castle or a classroom not found in physical reality. That virtual space may also be based on a physical space that the user has no relation to, such as a physical classroom in Des Moines, Iowa that the user in Redmond, Washington has never seen or been inside.

[0071] The virtual space may comprise a representation of some part of the
20 user's physical space. A depth camera that is capturing the user may also capture the environment that the user is physically in, parse it to determine the boundaries of the space visible by the camera as well as discrete objects in that space, and create virtual representations of all or part of that, which are then presented to the user as a virtual space.

[0072] The virtual space may comprise a representation of some part of a second
25 user's physical space. Where two users are interacting with each other through a virtual environment and by being captured by respective depth cameras, that virtual space may be a representation of that second user's virtual space.

[0073] The virtual space may also be some combination of these spaces. It may be a concatenation of a representation of the user's physical space and a purely virtual
30 space (which may be called a purely virtual sub-space, since the concatenation is a virtual space), such that both are represented as a single virtual space. Likewise, it may be a concatenation of a representation of the user's physical space, a purely virtual space and a representation of the second user's virtual space, a concatenation of a representation of the user's physical space and a representation of the second user's physical space, or a

concatenation of a representation of the second user's physical space and a purely virtual space.

5 [0074] In an embodiment, the user views this virtual space from a first-person perspective or a third-person perspective. In a first person perspective, the virtual space is displayed to the user as if the user were in the space – the display shows what the user would see with his own eyes. In a third-person perspective, the virtual space is displayed to the user with a representation of that user, such as an avatar, in that virtual space, similar to as if the user were being observed in that space by a third person.

10 [0075] In an embodiment, a virtual object comprises an object in the virtual space that may be interacted with. Where the virtual space is purely virtual and represents a classroom, things such as chalk, erasers, blackboards, chairs, desks, globes, windows, doorknobs and doors may be virtual objects. The user may manipulate or interact with these objects in a variety of ways, such as by writing with, lifting, moving, throwing, or breaking in half a piece of chalk.

15 [0076] In an embodiment, the virtual object corresponds to a physical object. This may be a physical object in the possession of the user. For instance, if the user has a chair, that physical chair may be captured by a depth camera and a representation of the chair may be inserted into the virtual environment. Where the user moves the physical chair, the depth camera may capture this, and display a corresponding movement of the
20 virtual chair.

[0077] In an embodiment where the virtual object corresponds to a physical object, there may be situations where one is moved independently of the other. Given the physical chair and virtual chair above, where the user moves the virtual chair, it may be that the corresponding physical chair does not move. For instance, the user may throw a
25 virtual ball at the virtual chair and the resulting collision will knock over the virtual chair. However, since no physical force was imparted on the physical chair, it will remain upright. Likewise, the physical object may be moved without moving its corresponding virtual object. For instance, where the user is in a large room, but the virtual space is comparatively smaller, the user may move the physical chair far to his right. If the virtual
30 chair is already touching a virtual wall in the corresponding direction, it may not move though the physical chair did move.

[0078] In an embodiment, the virtual object corresponds to a physical object of a second user. Where it is a physical chair of the second user that is virtualized, the user will be presented with only a virtual representation of that second user's chair. As the user

interacts with that virtual chair, those interactions (such as moving it) may be reflected in the virtual environment. This may not be reflected in actual movement of the physical chair. In an embodiment, this interaction is displayed to the second user as well, such that the virtual environment displayed to both the user and the second user remains consistent.

5 In an embodiment, this interaction is not displayed to the second user, such that the second user's virtual environment and the second user's physical environment remain consistent.

[0079] In an embodiment, where a virtual object corresponds to a physical object of a second user, the user cannot manipulate this virtual object. That ensures that the virtual environment displayed to both users and the second user's physical environment
10 remain consistent. In an embodiment, the second user may manipulate the virtual object by manipulating the corresponding physical object. As the second user has physical access to the physical object, his manipulation of the physical object may be reflected in both users' virtual environments, so those virtual environments and the second user's physical environment would remain consistent.

15 [0080] A controller comprises something that the user manipulates in physical space to cause a cursor in virtual space to be manipulated in a corresponding fashion. A controller may comprise a physical thing. It may be a body part, such as the user's hand. It likewise may be a physical object, such as a baton or pointer, a disc, or a ball. Where the controller is a body part, a depth camera may be used to provide depth data that is then
20 parsed to determine skeletal data of the user. From this skeletal data and how this skeletal data changes over time, the user's movements of his controller may be determined.

[0081] A controller may comprise an imaginary thing. The user may motion to throw an imaginary ball. This may be detected using user skeletal data as above. For instance, the user may motion to throw an imaginary ball towards the display device that is
25 displaying a virtual object. Where that ball would intersect the display may be calculated and then the way in which this would affect the virtual object – such as through a collision – may be determined and displayed. In an embodiment, when the imaginary ball intersects the display, it may be shown as a virtual object in virtual space and proceed on its flight in that virtual space.

30 [0082] A cursor may comprise a virtual representation of the controller in virtual space that may interact with a virtual object in virtual space. By manipulating one or more controllers in physical space, a user is able to cause the manipulation of one or more corresponding cursors in virtual space, and where those cursors cause a virtual object in

virtual space to be manipulated (such as a virtual block being pushed), that manipulation occurs. Further, a cursor may be invisible in that it is not displayed on the display device.

[0083] A capture device 508 captures a scene 504 in a physical space in which a user 502 is present. The capture device is coupled to a computing device 510 that outputs
5 to a display 512. In an embodiment, capture device 508 comprises capture device 20 of FIG. 2, and computing device 510 comprises computing environment 12 of FIG. 2, multimedia console 100 of FIG. 3A or computing environment 220 of FIG. 3B. The capture device captures this scene and produces depth data. From this depth data, a user controller 506 is determined and monitored for how the user 502 manipulates it. The
10 computing device 510 associates the user controller 506 with a cursor 514 (cursors 514a and 514b depict the cursor at two points in time as it corresponds to controller 506a and 506b at two points in time) in a virtual environment displayed on the display 512.

[0084] The controller 506 may be pre-determined – the system may always find the user's 502 right hand to be the controller 506, and thus always associate the user's 502
15 right hand with the cursor. The controller may also be user-determined. For instance, the computing device 510 may display to the user 502 the image data captured by the capture device 508 with an indication to place the controller in a specific area of the screen. If the user 502 places his left hand in that area, that may be determined to be the controller. The same would apply if the user 502 places his right hand, a foot, or some object such as a
20 baseball bat in that area.

[0085] The controller 506 is then used to manipulate an on-screen cursor 514. The cursor 514 may be a representation of the controller 506. The cursor 514 may be a right hand where the controller 506 is the user's right hand. Likewise, the cursor 514 may be the user's left hand where the controller 506 is the user's left hand. In an embodiment,
25 the cursor 514 and the controller 506 are not directly related in this manner. For instance, the controller 506 may be the user's hand, but the cursor 514 may be a representation of a tennis racket or a stevedore claw. Multiple controllers 506 and multiple cursors 514 may be used within a single virtual environment. For instance each hand of a user 502 may comprise a controller 506, or two users (present in either the same physical environment,
30 or separate physical environments, each having a computing device that communicate together across a communications network) may each have a controller 506, like their right hand, and each of these controllers 506 may map to a cursor 514.

[0086] As the user manipulates the controller 506, this manipulation is captured by the capture device 508 and processed by the computing device 510 to impart a

corresponding manipulation on the cursor 514. For instance, where the user extends his controller 506a hand and closes his fingers 506b, the virtual hand cursor 514a may extend into the virtual scene and the fingers of that virtual hand may close 514b.

[0087] The computing device 510 may keep track of position data for the cursor 514 and at least some virtual object in the scene (like depicted virtual baton 516), such that it can determine when a cursor 514 is touching the virtual object 516 as a result of user manipulation of the controller 506. Where this manipulation of the cursor 514 via the controller 506 causes the virtual object 516 to be manipulated, the computing device 510 may determine the manipulation that has occurred and display the result of this manipulation on the display device. For instance, where the user extends his hand cursor 514 and closes his fingers into a fist, this may cause the hand cursor 514 to extend into the scene, and the fingers of the cursor 514 to clasp around the virtual baton 516. Where the user continues to extend his controller 506 forward, and this causes his cursor 514 to continue to extend into the scene, it may be determined that the virtual baton 516 is being manipulated by being moved into the scene, and this result may be displayed on the display 512.

[0088] FIG. 6 depicts example operational procedures for manipulating a virtual object in a virtual space.

[0089] Operation 602 depicts determining at least one controller that the a user utilizes to manipulate the virtual object.

[0090] Operation 604 depicts mapping the controller to a cursor in the virtual space. The mapping may be an association such that a movement of the controller in a scene that is observed by the depth camera and parsed for controller movement corresponds to movement of the cursor.

[0091] Operation 606 depicts determining controller input indicative of the user manipulating the virtual object with the cursor.

[0092] In an embodiment, there is a scale of movement, and determining that controller input indicative of the user manipulating the virtual object with the cursor includes determining that controller input indicative of the user manipulating the virtual object with the cursor in proportion to the scale of movement. Where the scale of movement is 1:2, this may mean that for every inch the user moves the controller, the cursor is to be moved two inches in the same direction. This may be two inches as defined by the virtual world (e.g. the virtual world may be presented smaller than the corresponding physical world to begin with, such that a six-foot tall person appears as a

three-foot tall avatar when displayed on the display device) or a literal two inches, such that this corresponds to a movement of two inches on the display screen.

[0093] In an embodiment, the cursor has at least one dimension, and determining that controller input indicative of the user manipulating the virtual object with the cursor
5 in proportion to the scale of movement, includes determining that controller input indicative of the user manipulating the virtual object with the cursor in proportion to the dimension. Take the example where the controller is the user's hand, and the cursor is a virtual hand attached to a forearm. The cursor may stretch or extend in response to controller manipulation. For instance, if the user extends his hand toward the depth
10 camera, along the dimension of the z-axis, rather than solely moving forward along the z-axis dimension, the hand and forearm may stretch forward along that dimension.

[0094] In an embodiment, a physics model is used to determine how the cursor interacts with the virtual object. For instance, in a given physics model, a virtual object may have a mass and the virtual environment may have a gravity, so rolling a stationary
15 virtual ball will require imparting sufficient force on it with the cursor to overcome the ball's inertia. The physics model may accurately model the physical world on Earth, or some other model may be used. For instance, the gravity in the physics model may be weaker than Earth gravity so that an object bounces higher under otherwise equal circumstances after being dropped to the ground by use of the cursor. This physics model
20 may extend beyond how a cursor interacts with a virtual object to how a virtual object interacts with a second virtual object. For instance, where the user uses the controller to cause the cursor to push a first virtual ball into a second virtual ball, the second virtual ball may begin to roll as a result of that collision between the two virtual balls.

[0095] There may be a virtual collision in virtual space without a corresponding
25 physical collision in physical space. The user can move the controller through physical space in such a manner that it would require the corresponding cursor in virtual space to move through a virtual object or part of the virtual environment that is unmovable, such as a wall, or an extremely heavy boulder, as according to the physics model above. In an embodiment, the cursor is being animated to show movement in the virtual environment,
30 and when such a collision occurs, the cursor stops animating and remains still at the point of collision, until the user uses the controller to manipulate the cursor in a direction that does not involve a collision. In an embodiment, the animation of the cursor continues through the collision so that there will be no disconnect between the user's position of the

controller in the physical environment and the position of the cursor in the virtual environment.

[0096] Operation 608 depicts displaying the result of said manipulation.

[0097] In an embodiment, the result of said manipulation comprises associating
5 the virtual object with a physical object such that a change in the physical object is reflected as a corresponding change of the virtual object. For example, two people in separate physical environments may be interacting in a virtual environment that is a concatenation of a representation of their respective physical environments. The first user may select a virtual shirt and associate it with the second user's physical shirt, and as a
10 result of this association, the second user is shown in the virtual environment to be wearing this virtual shirt rather than the physical shirt. This wearing of the virtual shirt may be shown only to the first user, only to the second user, or to both users.

[0098] Optional operation 610 depicts determining at least one second controller that the a second user utilizes to manipulate the virtual object; mapping the second
15 controller to a second cursor in the virtual space, determining second controller input indicative of the second user manipulating the virtual object with the second cursor, and displaying the result of said manipulation.

[0099] Where multiple users are interacting in the same virtual environment, the manipulations of a virtual object by a second user may be displayed to the first user on his
20 display device. In an embodiment, two users may manipulate a single virtual object simultaneously. In an embodiment, the virtual object has a plurality of dimensions and the first user manipulates a first dimension of the plurality of dimensions and the second user manipulates a second dimension of the plurality of dimensions. For instance, the first user may lift an object, and the second user may rotate the object while it is lifted. Further, the
25 two users may operate on the same dimension of a virtual object. One user may impart lateral force on a virtual ball to roll it. The second user may impart a second lateral force in the same direction as the first user, and this additional force applied to the virtual ball will cause it to roll with an increased velocity.

Conclusion

30 [0100] While the present disclosure has been described in connection with the preferred aspects, as illustrated in the various figures, it is understood that other similar aspects may be used or modifications and additions may be made to the described aspects for performing the same function of the present disclosure without deviating there from. Therefore, the present disclosure should not be limited to any single aspect, but rather

construed in breadth and scope in accordance with the appended claims. For example, the various procedures described herein may be implemented with hardware or software, or a combination of both. Thus, the methods and apparatus of the disclosed embodiments, or certain aspects or portions thereof, may take the form of program code (i.e., instructions) embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other machine-readable storage medium. When the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus configured for practicing the disclosed embodiments. In addition to the specific implementations explicitly set forth herein, other aspects and implementations will be apparent to those skilled in the art from consideration of the specification disclosed herein. It is intended that the specification and illustrated implementations be considered as examples only.

51050-145

CLAIMS:

1. A method for manipulating a first virtual object in a virtual space, the first virtual object being associated with a physical object in a first physical space, the physical object being separate from a first user, comprising:
 - 5 receiving first depth image data indicative of the first physical space;
generating the first virtual object in the virtual space based on the physical object as represented in the first depth image data;
in response to receiving second depth image data indicative of the first physical space where the physical object is being moved by the first user, displaying a corresponding
10 moving of the first virtual object in the virtual space; and
in response to receiving third depth image data indicative of a second user in a second physical space, the second physical space differing from the first physical space in which the physical object is located, and in response to determining that the third depth image data is indicative of the second user attempting to move a second virtual object in the virtual
15 space and the first virtual object in the virtual space without the second user having moved the physical object in the first physical space, determining to display movement of the second virtual object and not to display movement of the first virtual object.
2. The method of claim 1, wherein the virtual space comprises a representation of the first physical space, a virtual sub-space, a representation of the second physical space, or a
20 combination thereof.
3. The method of claim 1, wherein there is a scale of movement, and wherein displaying a corresponding moving of the first virtual object comprises:
displaying the corresponding moving of the first virtual object based on an amount that the physical object is moved modified by the scale of movement.

51050-145

4. The method of claim 1, wherein the virtual space is presented to the first user in a first-person perspective or a third-person perspective.

5. A system for interacting with a first virtual object in a virtual environment, the first virtual object being associated with a physical object in a first physical space, the

5 physical object being separate from a first user, comprising:

a memory bearing instructions that, upon execution by a processor, cause the system at least to:

generate the first virtual object based in the virtual environment on the physical object as represented in first depth image data;

10 in response to receiving second depth image data indicative of the first physical space where the physical object is being moved by the first user, display a corresponding moving of the first virtual object in the virtual environment; and

in response to receiving third depth image data indicative of a second user in a second physical space, the second physical space differing from the first physical space in
15 which the physical object is located, and in response to determining that the third depth image data is indicative of the second user attempting to move a second virtual object in the virtual environment and the first virtual object in the virtual environment without the second user having moved the physical object in the first physical space, determine to display movement of the second virtual object and not to display movement of the first virtual object.

20 6. A computer-readable storage medium having stored thereon computer executable instructions that upon execution on a computer, cause the computer to perform operations comprising:

receiving first depth image data indicative of a first physical space;

generating a first virtual object in a virtual space based on a physical object as
25 represented in the first depth image data;

51050-145

in response to receiving second depth image data indicative of the first physical space where the physical object is being moved by a first user, displaying a corresponding moving of the first virtual object in the virtual space; and

5 in response to receiving third depth image data indicative of a second user in a second physical space, the second physical space differing from the first physical space in which the physical object is located, and in response to determining that the third depth image data is indicative of the second user attempting to move a second virtual object in the virtual space and the first virtual object in the virtual space without the second user having moved the physical object in the first physical space, determining to display movement of the second
10 virtual object and not to display movement of the first virtual object.

7. The system of claim 5, wherein the virtual environment comprises a representation of the first physical space, a virtual sub-space, a representation of the second physical space, or a combination thereof.

8. The system of claim 5, wherein there is a scale of movement, and wherein the
15 instructions that, upon execution by the processor, cause the system to at least display a corresponding moving of the first virtual object further cause the system at least to:

display the corresponding moving of the first virtual object based on an amount that the physical object is moved modified by the scale of movement.

9. The system of claim 5, wherein the virtual environment is presented to the first
20 user in a first-person perspective or a third-person perspective.

10. The computer-readable storage medium of claim 6, wherein the virtual space comprises a representation of the first physical space, a virtual sub-space, a representation of the second physical space, or a combination thereof.

11. The computer-readable storage medium of claim 6, wherein there is a scale of
25 movement, and wherein displaying a corresponding moving of the first virtual object comprises:

51050-145

displaying the corresponding moving of the first virtual object based on an amount that the physical object is moved modified by the scale of movement.

12. The computer-readable storage medium of claim 6, wherein the virtual space is presented to the first user in a first-person perspective or a third-person perspective.

5 13. The method of claim 1, further comprising:

in response to receiving fourth depth image data indicative of the physical object being moved in the first physical space by the first user, determining not to display a corresponding moving of the first virtual object where the moving of the first virtual object conflicts with a boundary of the virtual space.

10 14. The system of claim 5, further comprising instructions that upon execution cause the system to perform further operations, comprising:

in response to receiving fourth depth image data indicative of the physical object being moved in the first physical space by the first user, determine not to display a corresponding moving of the first virtual object where the moving of the first virtual object
15 conflicts with a boundary of the virtual environment.

15. The computer-readable storage medium of claim 6, wherein the instructions, when executed, cause the computer to perform further operations, comprising:

in response to receiving fourth depth image data indicative of the physical object being moved in the first physical space by the first user, determining not to display a
20 corresponding moving of the first virtual object where the moving of the first virtual object conflicts with a boundary of the virtual space.

1/8

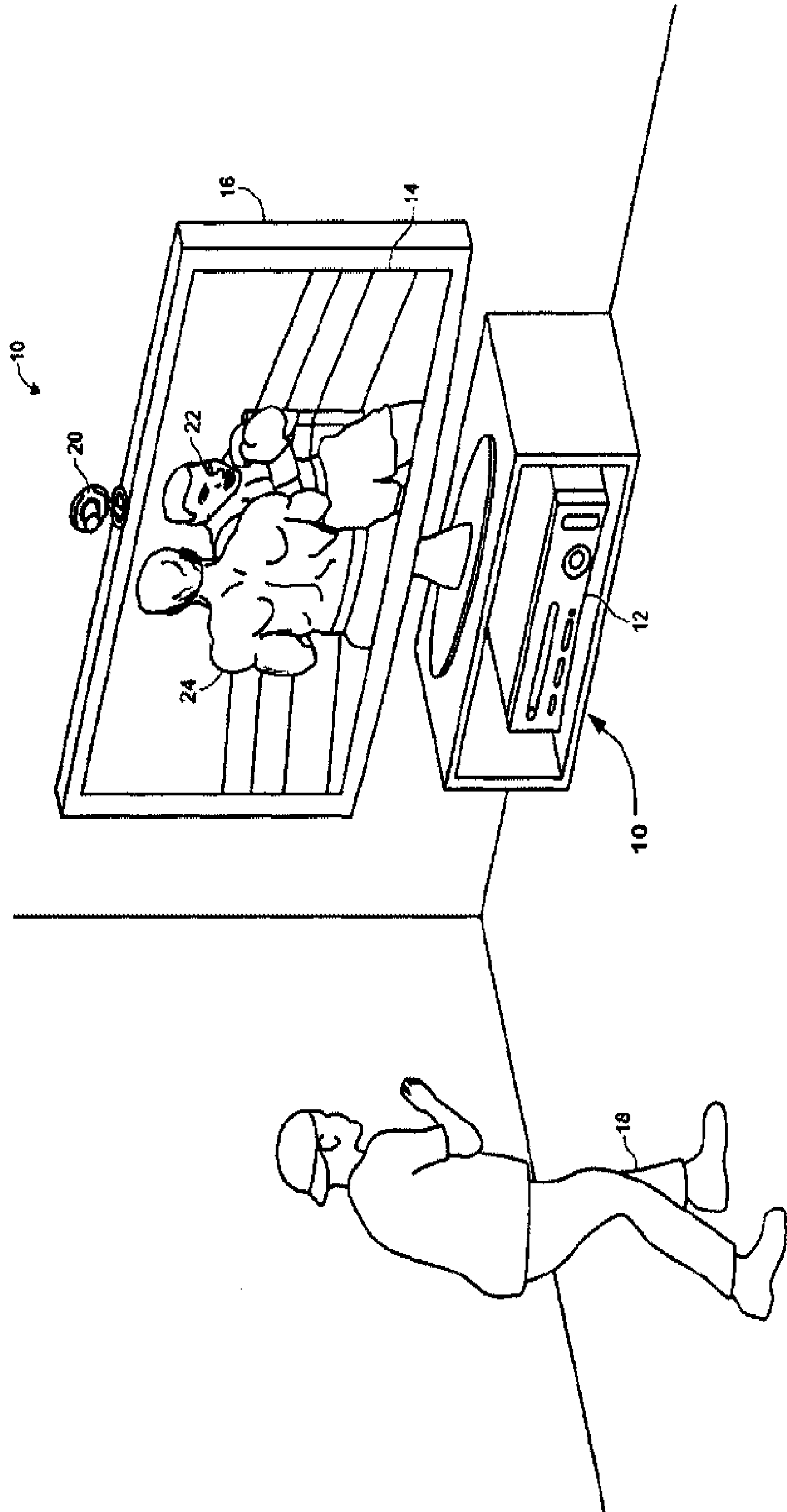


FIG. 1A

2/8

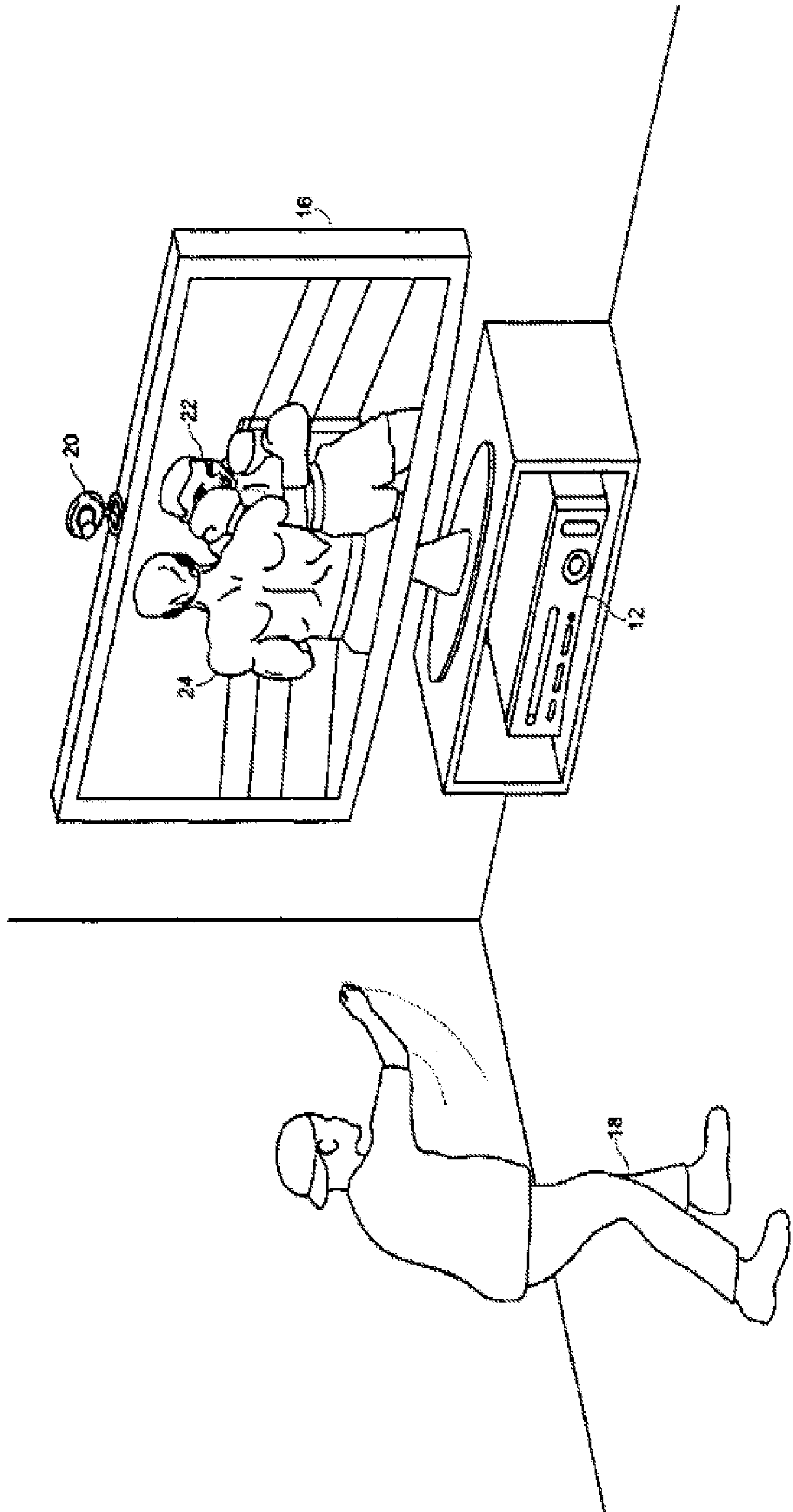


FIG. 1B

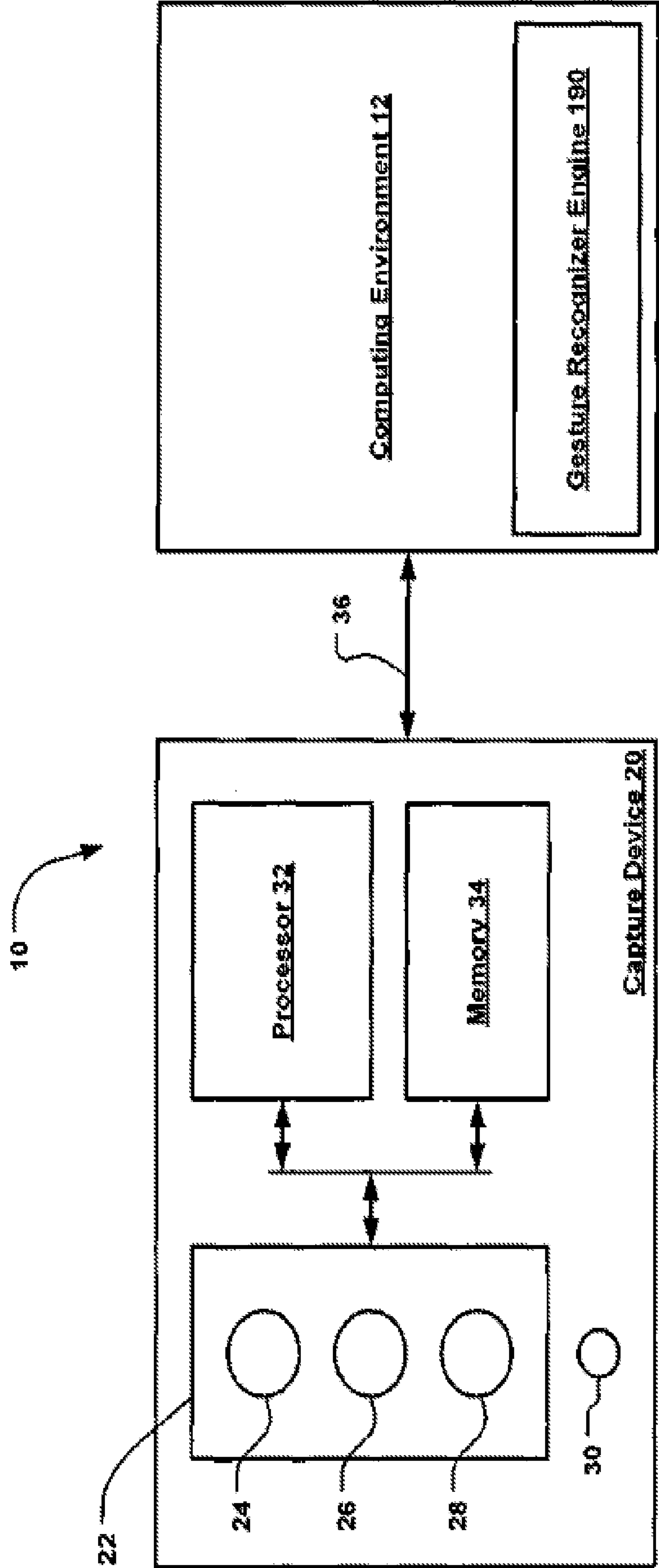


FIG. 2

4/8

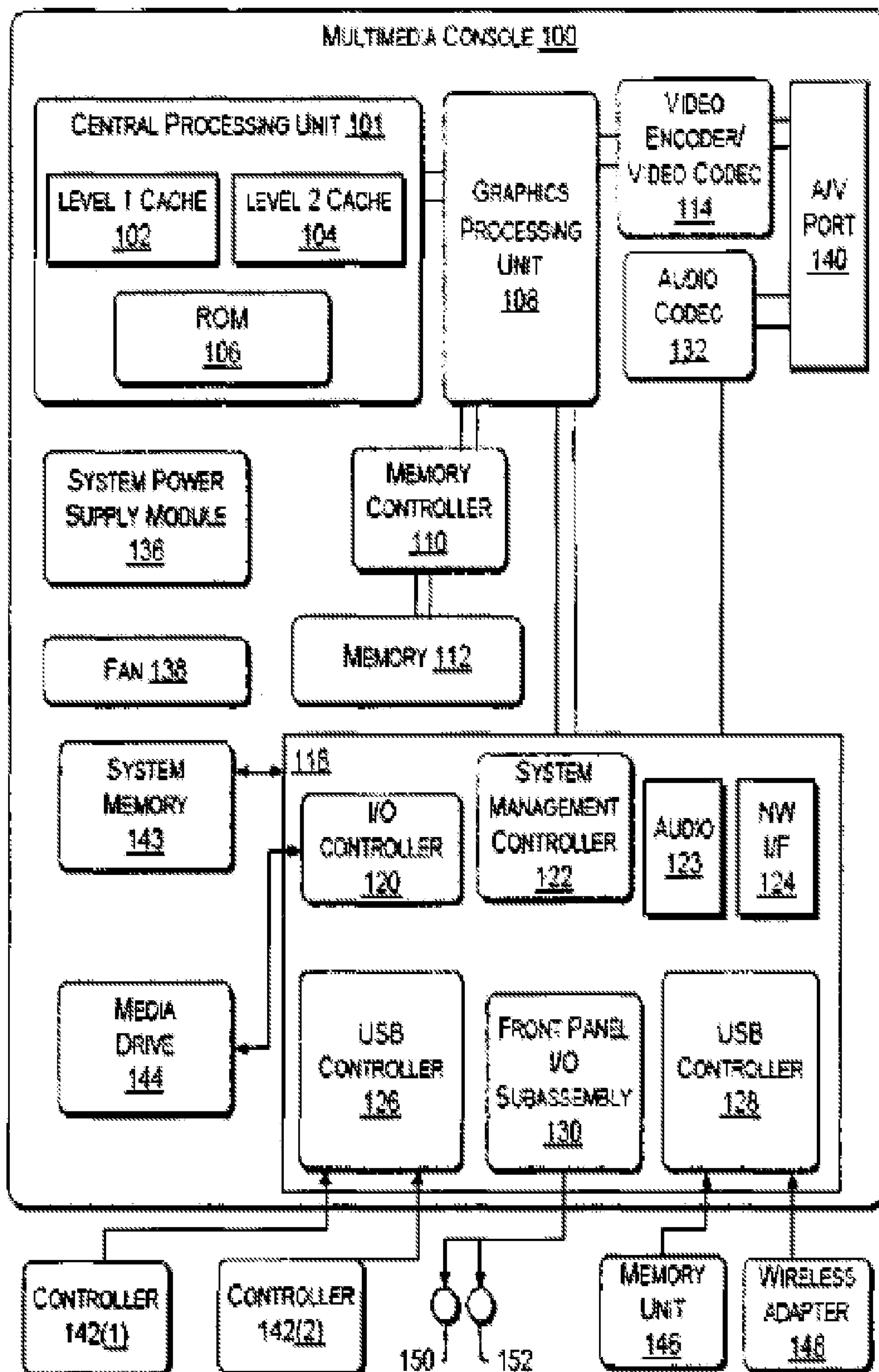
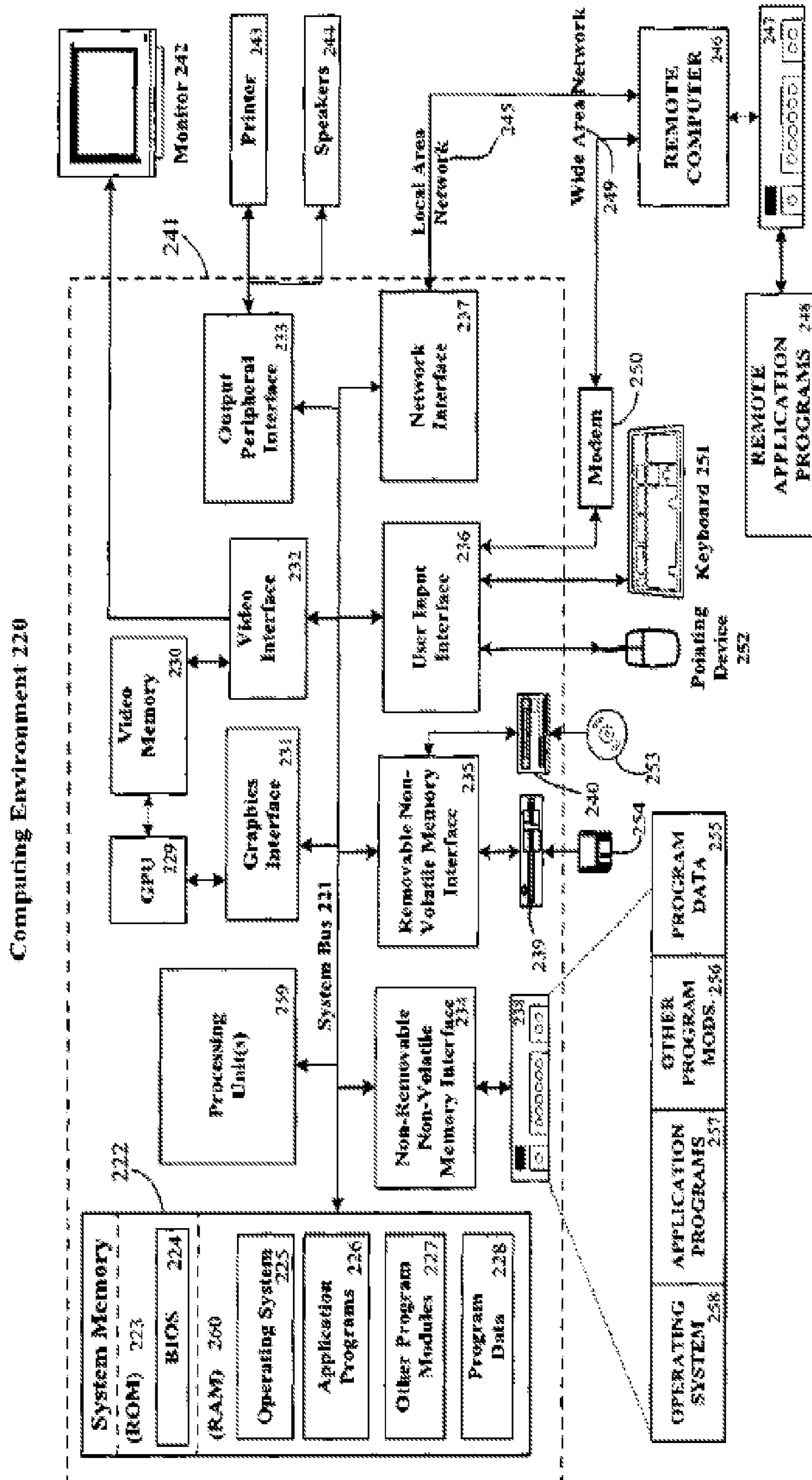


FIG. 3A

5/8

**FIG. 3B**

6/8

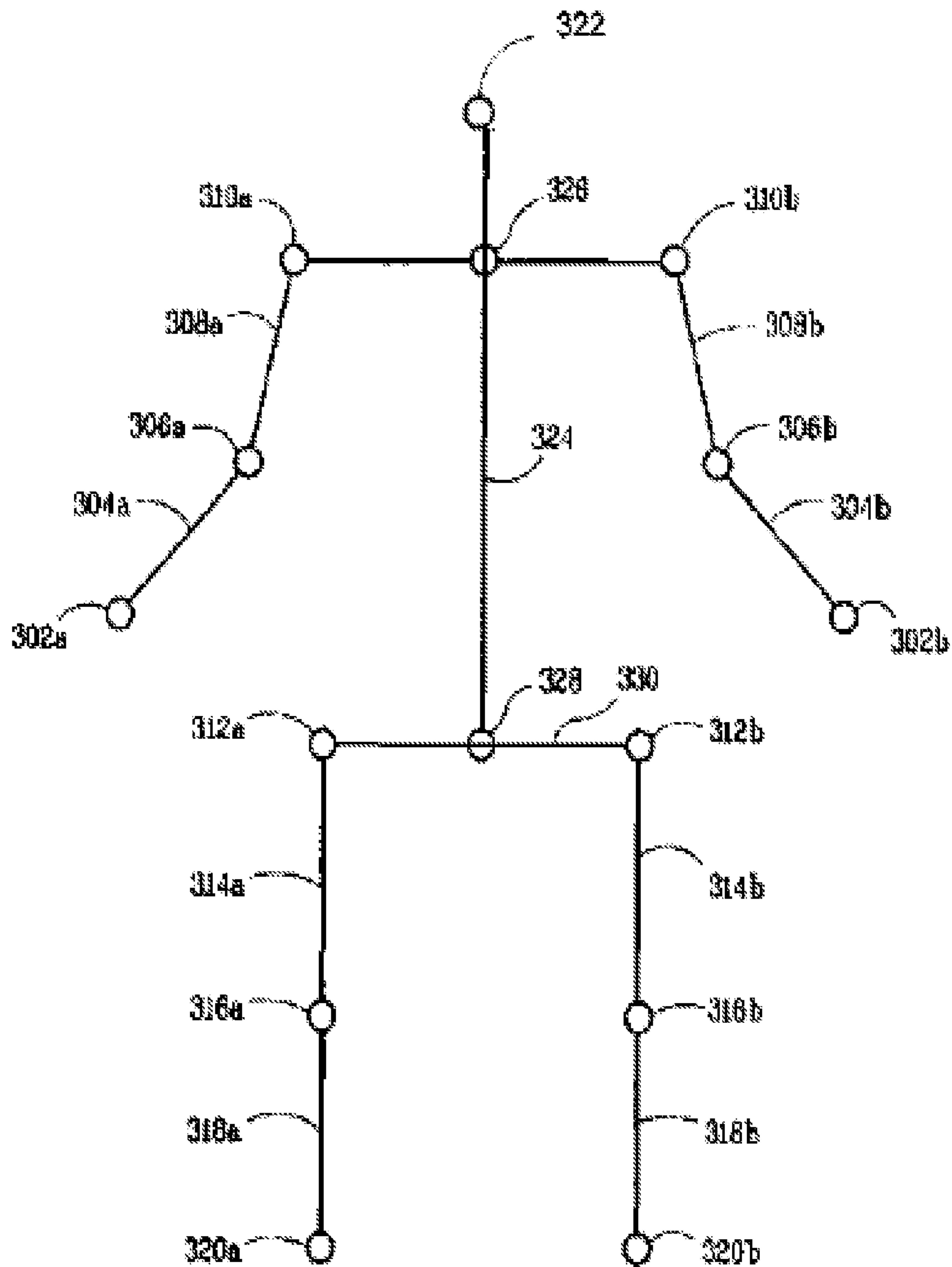


FIG. 4A

7/8

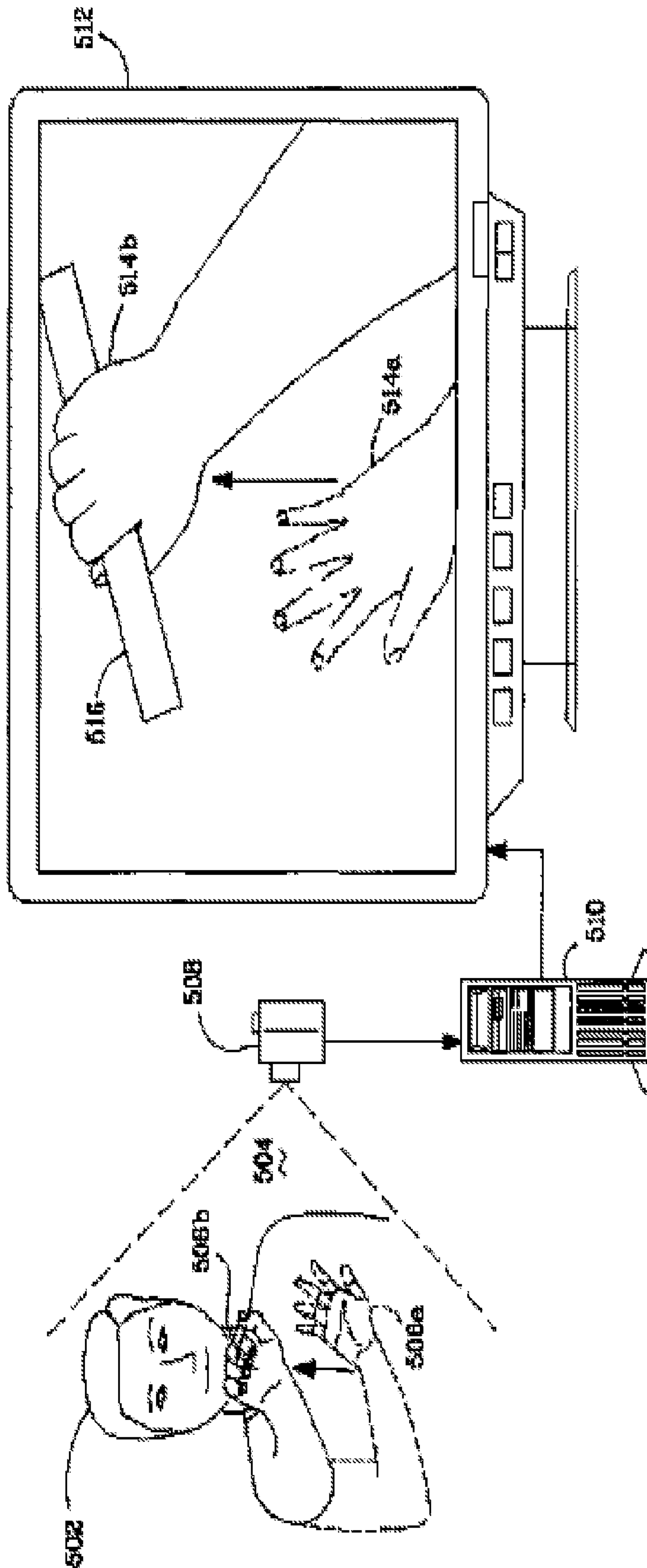


Fig. 5

8/8

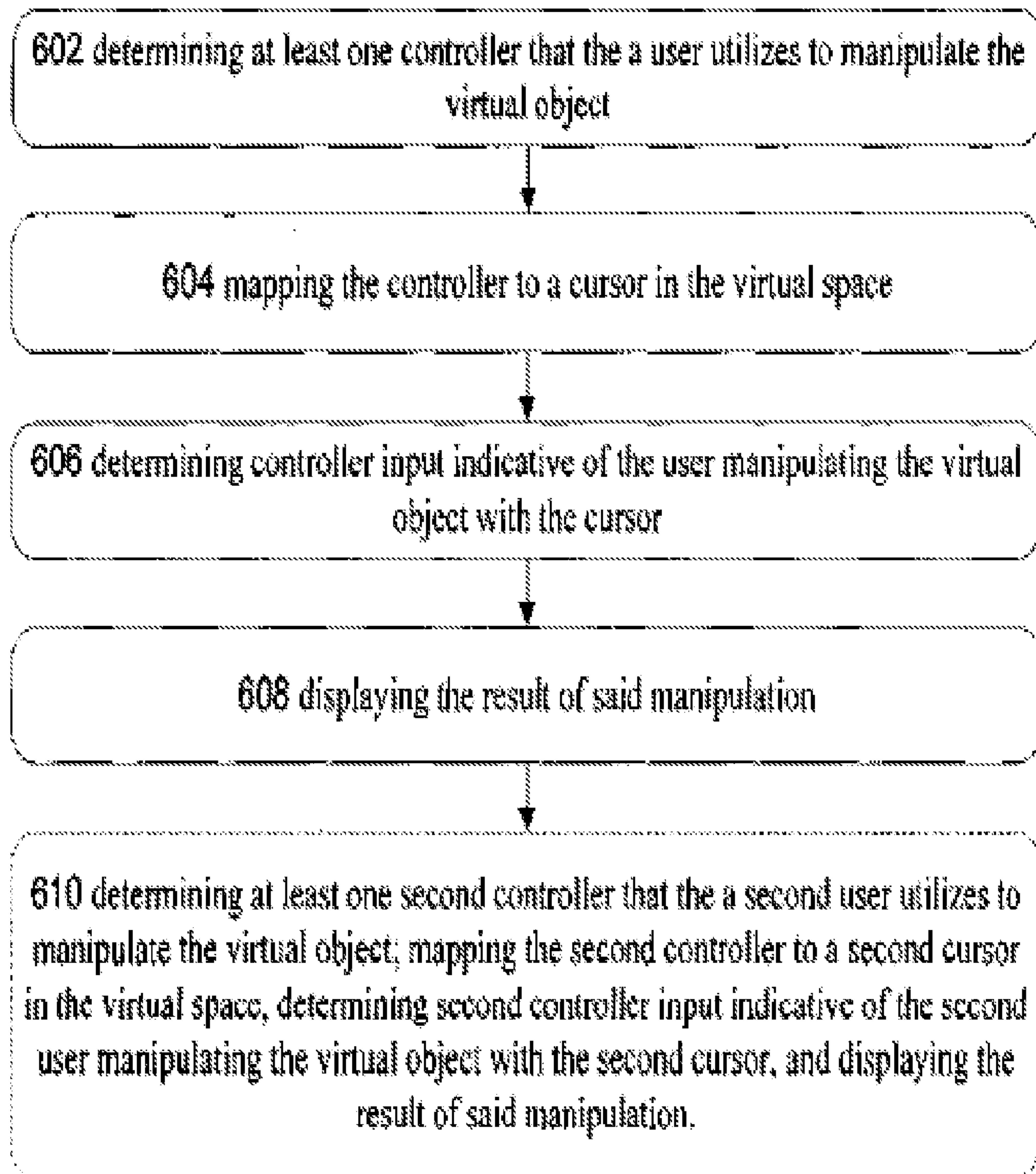


Fig. 6

