March 2023

Direct View Rendering with Head Tracking Feedback for 3D Video Calling

Anonymous

Recommended Citation

Anonymous, "Direct View Rendering with Head Tracking Feedback for 3D Video Calling", Technical Disclosure Commons, (March 08, 2023)
https://www.tdcommons.org/dpubs_series/5722

Direct View Rendering with Head Tracking Feedback for 3D Video Calling

ABSTRACT

When participants engage in a video call, discrepancies between the viewing direction from which a participant views other participants and the displayed view of the other participants can lead to a less than satisfactory experience. This disclosure describes techniques to select a subset of available cameras based on head and/or eye movements of participants in a video call to render a corresponding direct view. Cameras capture images of a viewing user that is viewing a display on which a 3D video of a second user (e.g., generated from images captured by a subset of a plurality of cameras) is displayed. Per techniques of this disclosure, with user permission, the head and/or eye movements of the viewing user are tracked based on the captured images of the viewing user. Relationship between the tracked movements of the viewing user and a view of the display of the device is determined. The view of the first display is then updated to render a 3D video based on a subset of individual cameras of a second user's device that match the movement (which corresponds to an updated viewing position) of the viewing user. The view on the first display is updated to show the second user from a corresponding perspective. The techniques can provide a more accurate experience during a three-dimensional (3D) video conference. Suitable techniques such as machine learning can be utilized to predict the user's movement and adjust the view accordingly.

KEYWORDS

- Artificial reality
- 3D video
- Stereoscopic display
- Telepresence
- Viewing perspective
- Video calling
- Head tracking
- Eye tracking

BACKGROUND

Artificial reality is a form of reality that has been adjusted in some manner before presentation to a user, which may include, e.g., a virtual reality (VR), an augmented reality (AR), a mixed reality (MR), a hybrid reality, or some combination and/or derivatives thereof. In this regard, AR, VR, MR, and hybrid reality devices often provide content through visual mechanisms, such as through a display (e.g., a three-dimensional autostereoscopic display) or headset (e.g., glasses). For example, a 3D autostereoscopic display may direct different scenes to each of a viewer's eyes to generate a perception of 3D depth in images.

Many artificial reality devices utilize cameras to present information to a user and may execute various AR operations and simulations. For example, three-dimensional (3D) calling may provide an AR experience for one or more parties to a call or video conference. Such operations may be extremely challenging to execute, since the AR devices (e.g., 3D autostereoscopic displays) may receive, incorporate, and accurately convey images in a manner that provides seamless and realistic output for users.

Three dimensional (3D) displays allow users in separate locations to engage in conversations that approximate a natural face-to-face conversation. For example, 3D displays may use binocular disparity (e.g., the difference in image location of an object seen by the left and right eyes, resulting from the eyes' horizontal separation) to convey depth information from two-dimensional images. For example, without wearing any specialized glasses, each of the user's eyes may see a different view.

When capturing an image or executing operations based on user input, an AR device (e.g., a 3D autostereoscopic display) may synchronize a user's view with a camera or another image capturing module. As a user moves from left to right, the user may get a sense of looking around the displayed content. A movement of a user's head in one direction may allow the user to view a right side of the content and a movement of the user's head in the other direction may allow the user to view a left side of the content. Moreover, motion parallax may provide a user with a monocular depth cue arising from relative velocities of objects moving across the retinae of a moving user.

Discrepancies between the views may lead to erroneous image information that may affect any operations utilizing the image, and/or the user's experience using the AR device (e.g., the 3D autostereoscopic display). Accordingly, there is a need to operate image capturing modules accurately and efficiently on artificial reality devices such as 3D autostereoscopic displays.

DESCRIPTION

The present disclosure generally relates to techniques for operating video conferencing devices based on one or more head and/or eye movements. The described techniques relate to the operation of artificial reality devices, e.g., three-dimensional (3D displays). Per the techniques, head and/or eye movements of a viewer (e.g., a party to a 3D call) are detected and/or predicted. The head/eye. Movements are used to update one or more images (e.g., of another party to the 3D call) presented by an autostereoscopic display.

In various examples, a system (e.g., artificial reality device) may include a camera configured to track at least one of head movements or eye movements of a user, a Red Green Blue (RGB) camera array configured to capture one or more images, an autostereoscopic display (e.g., 3D) for providing images captured by the RGB cameras, a processor and a non-transitory memory including computer-executable instructions to be executed by the processor. The computer- executable instructions may cause the artificial reality device to track the head and eye movements of the viewer and display images from different RGB cameras according to the viewer's head and eye movements.

In some examples, motion prediction, a feedback loop, and/or the viewer's current head or eye position may be utilized to predict future head movements of the viewer. The viewer's current head position may be sent to the RGB camera array. Based on the predicted head position, a subset of the images captured by the cameras (e.g., 10 out of 100) are streamed. The subset of images captured by the cameras (e.g., streamed images) are changed based on changes in the viewer's head position. Compression techniques can be utilized to reduce the bandwidth usage of the subset of cameras.

The RGB cameras may include the camera configured to track the head movements and/or the eye movements of the user. Cameras may be display-mounted cameras or may be stand-alone cameras. In some examples, cameras may be head-mounted cameras. AR devices in accordance with some examples include one or more outward facing cameras, and one or more inward facing cameras. An artificial reality device may further comprise glasses, a headset, a display, a microphone, a speaker, and any of a combination of peripherals, and computing systems. **A** trained machine learning model may be utilized to track or predict head and/or eye movements of a user. A viewer's head and eye movements may be used to identify and display images from different RGB cameras.

In some examples, a plurality of RGB cameras may include an array of cameras (e.g., 5 to 100 cameras). Each camera of the array of cameras may be identical to the other cameras and may be synchronized with the other cameras. For example, an array of 100 cameras may be arranged (e.g., in close proximity) linearly or in a two-dimensional (2D**)** grid.

An increased number of cameras may be associated with additional processing. In order to reduce bandwidth needed to transmit images associated with the plurality of RGB cameras, a subset of the cameras may be selected based on actual or predicted head movements of a viewer (e.g., a party to the videoconference).

For example, as a viewer moves their head from left to right (e.g., as they view a 3D display), the system may track their head movement in order to determine the content for display. Latency and/or required bandwidth may be reduced by tracking the head movements of a viewer. The head movements of the viewer may be transmitted to/from the viewer's side of the videoconference to the caller's side of the video conference. Based on the head movements of the viewer, the system may select a subset of the caller's array of cameras. For example, the subset of the array of cameras may comprise 5 of the cameras, 10 of the cameras, 5% of the cameras, 10% of the cameras, etc. For example, if the viewer is looking at a right side of the content being displayed on the viewer's screen, then it may not be necessary to capture the content from the caller's side that may be displayed on the left side of the viewer's display. Therefore, the system may only capture and/or transmit the images associated with the subset of cameras on the caller's side that relate to the content being viewed by the viewer.

As another example, as a first viewer moves their viewing angle (e.g., by turning their head about a vertical or horizontal axis, moving their head in a side-to-side motion, etc.), the subset of the cameras capturing images of a second viewer may shift based on the changing viewing angle. Moreover, motion prediction may be used to predict changes in the viewing angle of the first viewer (e.g., movement of the first viewer's head). For example, the motion prediction may be based on the head or eye movements (e.g., eye tracking or pupil tracking) of the first viewer's head. The motion prediction may be based on position, velocity, acceleration, etc. and may comprise predicted movements in a horizonal or vertical direction or about a horizontal or vertical axis.

Artificial intelligence, pattern matching, pattern recognition, or recurring neural networks may be used to predict the viewer's head movements. One or more forms of artificial intelligence may be trained based on one or more sets of collected data. Moreover, forms of artificial intelligence may be continually trained or refined as a user interacts with the system.

In some examples, a first head tracking signal output may be associated with a position of a first viewer's left and right eyes in 3D space. Moreover, a predicted first head tracking signal output may be associated with the position of the first viewer's left and right eyes in 3D space. A second head tracking signal output may be associated with a position of a second viewer's left and right eyes in 3D space, as well as a predicted second head tracking signal. Each of the first and second head tracking signals may be converted, respectively, to a first selected subset of cameras and a second selected subset of cameras, where each selected subset of cameras is based on the corresponding head tracking signal (e.g., actual or predicted). In some examples, conversion of the head tracking signal to the selected subset of cameras may be based on projected geometry (e.g., based on a gaze of the viewer, distance of a viewer to a display, etc.).

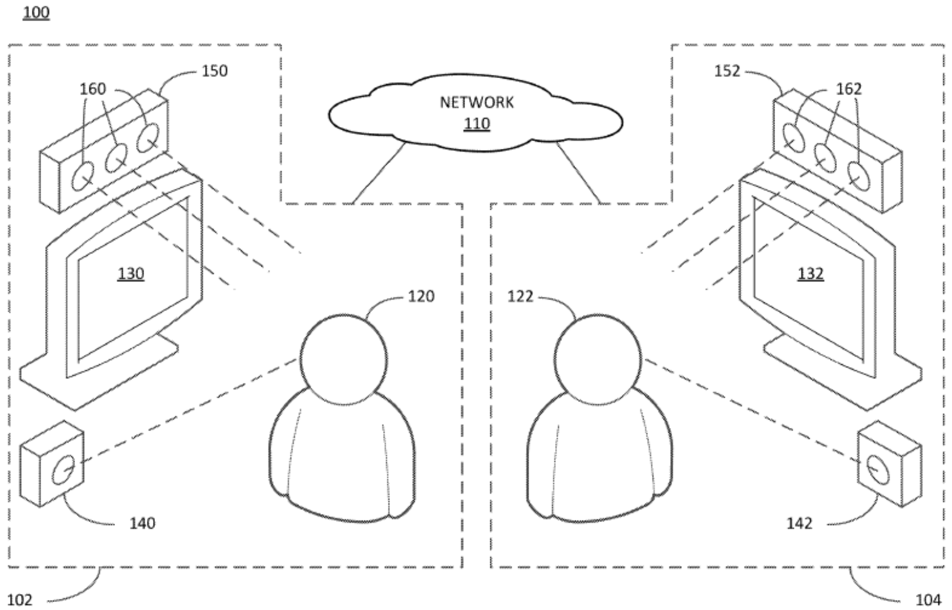

FIG. 1

FIG. 1 illustrates an example videoconferencing system 100. A first viewer and a first portion of a videoconferencing system and a second viewer and a second portion of the videoconferencing system are shown.

The video conferencing system 100 includes a first portion 102, a second portion 104, and a network 110. The videoconferencing system 100 may comprise any number of portions. It is illustrated in FIG. 1 with a first portion 102 and a second portion 104 as an illustrative example. The first portion 102 and the second portion 104 and any combination of components included in the videoconferencing system 100 may be directly or indirectly connected via network 110, e.g., a wired or wireless network.

The first portion 102 of the video conferencing system 100 is viewed by a first user 120 and includes a first display 130, a first tracking camera 140, a first camera array 150, and a plurality of first individual cameras 160. The second portion 104 of the video conferencing system 100 is viewed by a second user 122 and includes a second display 132, a second tracking camera 142, a second camera array 152, and a plurality of second individual cameras 162.

The first user 120 views the first display 130, e.g., a 3D display such as a 3D stereoscopic display. The first tracking camera 140 tracks the head or eye movements or position of the first user 120. First tracking camera 140 may comprise one or more cameras, including an inward-facing camera. An inward- facing camera may track the head or eye movements of the first user 120, and may therefore provide important information into a point of view and/or area of focus of the first user 120. In some examples, the video conferencing system 100 may utilize a camera module to track the gaze of the first user 120. The direction of the gaze of the first user 120 may correspond to one or more focusing operations, including selecting one or more of the plurality of second individual cameras 162 from the second camera array 152.

In some examples, a relationship between the movement of the gaze of the first user 120 and a view captured by one or more of the plurality of first individual cameras 160 may be identified. For example, one or more machine learning, optimization and/or triangulation techniques may be utilized to determine a relationship between the camera position, the user's eye position, eye direction, and the observed scene to determine a viewpoint which the first user 120 is looking. One or more machine learning techniques may be applied to tailor the relationships to the user's particular gazing habits and/or adapt to adjustments from positions of the cameras.

A first head tracking signal may be associated with a current position of the first user 120, e.g., head movements and/or left and right eye positions of the first user 120 in 3D space. A first predicted head tracking signal may be associated with a future position of the first user 120 (e.g., predicted head movements and/or predicted left and right eye positions of the first user 120 in 3D space). The first head tracking signal and/or first predicted head tracking signal may be converted to a selected subset of the plurality of second individual cameras 162, where the selected subset of the plurality of second individual cameras 162 is based on the corresponding first head tracking signal (e.g., actual or predicted). Conversion of the first head tracking signal to the selected subset of the plurality of second individual cameras 162 may be based on projected geometry (e.g., based on a gaze of the first user 120, distance of the first user 120 to the first display 130, etc.).

The second user 122 views the second display 132, e.g., a 3D display such as a 3D stereoscopic display. The second tracking camera 142 may track the head or eye movements or position of the second user 122. The second tracking camera 142 may comprise one or more cameras, including an inward-facing camera. An inward-facing camera may track the head or eye movements of the second user 122, and may therefore provide important information into a point of view and/or area of focus of the second user 122. In some examples, the video conferencing system 100 may utilize a camera module to track the gaze of the second user 122. The direction of the gaze of the second user 122 may correspond to one or more focusing operations, including selecting one or more of the plurality of first individual cameras 160 from the first camera array 150.

A relationship between the movement of the gaze of the second user 122 and a view captured by one or more of the plurality of second individual cameras 162 may be identified. For example, one or more machine learning, optimization and/or triangulation techniques may be utilized to determine a relationship between the camera position, the user's eye position, eye direction, and the observed scene to determine a viewpoint which the second user 122 is looking. One or more machine learning techniques may be applied, as described herein, to tailor the relationships to the user's particular gazing habits and/or adapt to adjustments from positions of the cameras.

A second head tracking signal may be associated with a current position of the second viewer 122 (e.g., head movements and/or left and right eye positions of the second viewer 122 in 3D space). A predicted second head tracking signal may be associated with a future position of the second viewer 122 (e.g., predicted head movements and/or predicted left and right eye positions of the second viewer 122 in 3D space). The second head tracking signal and/or second predicted head tracking signal may be converted to a selected subset of the plurality of second individual cameras 162, where the selected subset of the plurality of second individual cameras 1602is based on a corresponding second head tracking signal (e.g., actual or predicted). Conversion of the second head tracking signal to the selected subset of the plurality of second individual cameras 162 may be based on projected geometry (e.g., based on a gaze of the second viewer 122, distance of the second viewer 122 to the second display 132, etc.).

In some examples, the first head tracking signal or the first predicted head tracking signal may be based on identifying a desired region of interest of the first user 120. Based on the first head tracking signal or the first predicted head tracking signal, the videoconferencing system 100 may adjust its focus on the second user 122, e.g., to accommodate for a change in position of the first user 120.

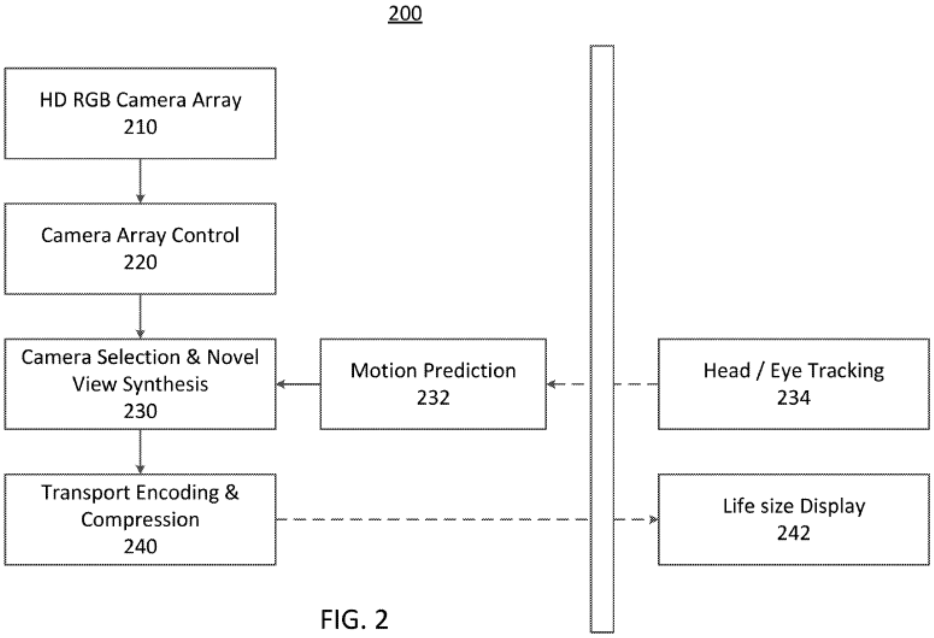

FIG. 2

FIG. 2 illustrates an example videoconferencing system. The videoconferencing system 200 may comprise, for example, one or more of the following functional blocks: High Definition (HD) RGB Camera array 210, camera array control 220, camera selection and novel view synthesis 230, motion prediction 232, head/eye tracking 234, transport encoding and compression 240, and life size display 242.

HD RGB Camera Array block 210 may include a plurality of cameras and may capture one or more images indicative of **a** scene. The scene may comprise a user's view (e.g., viewpoint, head, and/or eye position), for example, and may be captured by and come from an imaging module, such as for example a camera (e.g., first/second tracking camera 140/142, one or more of the plurality of first/second individual cameras 160/162 of first/second camera array 150/152, etc.) associated with the videoconferencing system 200. Moreover, the scene may comprise a view of another user (e.g., a view of another user from a particular viewpoint of the user).

Camera Array Control block 220 may synchronize and compile one or more images captured by the HD RGB Camera Array 210. For example, images from a plurality of cameras of the HD RGB Camera Array 210 may be combined by the Camera Array Control block 220.

Camera selection and novel view synthesis block 230 may select one or more cameras (e.g., one or more of the plurality of first/second individual cameras 160/162 of first/second camera array 150/152) to capture one or more images of a user (e.g., user 120 or user 122) based on a view of another user (e.g., user 120 or user 122). For example, as other user shifts their head or point of view, the cameras used to capture images of the user may shift or change accordingly.

Moreover, camera selection and novel view synthesis block 230 may receive one or more inputs from a motion prediction block 232 and/or a head/eye tracking block 234. The motion prediction block 232 may predict one or more movements of the user (e.g., user 120 or user 122) and the head/eye tracking block 234 may track one or more head or eye movements of the user (e.g., user 120 or user 122). According to some aspects, the motion prediction block 232 may receive one or more tracked head or eye movements from the head/eye tracking block 234 and the predicted one or more movements of the user may be based on the one or more tracked movements.

In order to reduce the size of the image data associated with a user (e.g., in a 3D call), the transport encoding and compression block 240 may use any number of compression techniques. In one example, the transport encoding and compression block 240 may only transmit data associated with one or more cameras as determined by the camera selection and novel view synthesis block 230. Moreover, the transport encoding and compression block 240 may eliminate or reduce any overlapping or redundant data (e.g., overlapping image portions captured by different cameras).

At life size display block 242, the image data from the transport encoding and compression block 240 may be received and formatted for display. For example, images associated with a user may be formatted for display to another user.

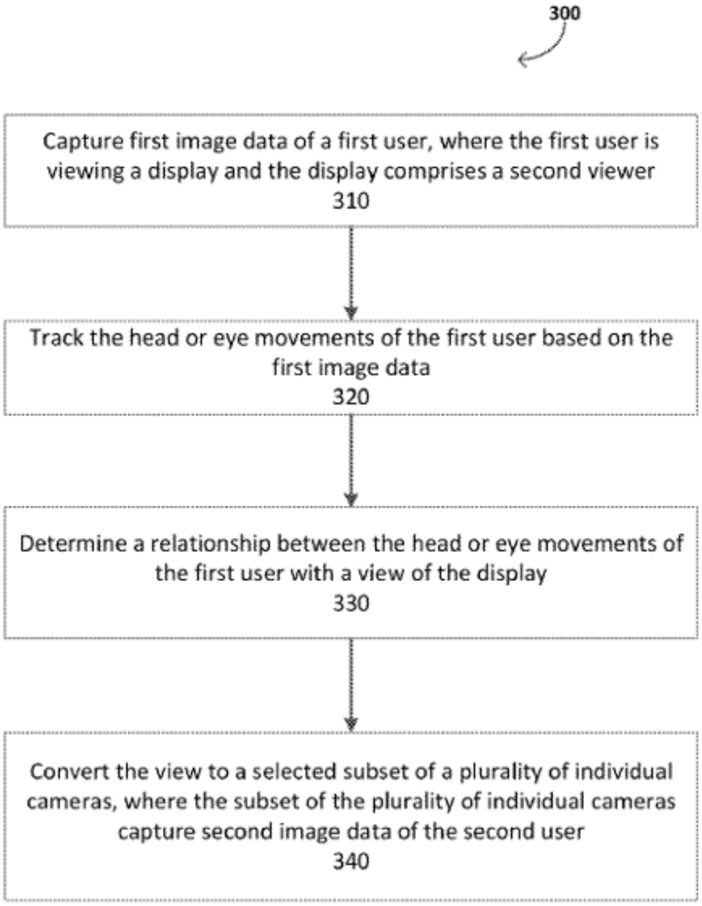

FIGURE 3

FIG. 3 illustrates an example method for video conferencing. According to some aspects, a video conferencing system (e.g., videoconferencing system 100) may facilitate a communication (e.g., a 3D call) between a plurality of users (e.g., first user 120 and second user 122). Each of the users may be located in different physical locations and may be connected to one or more of the other users via a wired or wireless network (e.g., network 100).

Method 300 may begin at step 310, where image data may be captured of a first user (e.g., first user 120). The first user (e.g., first user 120) may view a first display (e.g., first display 130). The (e.g., first display 130) may provide image data associated with another user (e.g., second user 122). As the first user (e.g., first user 120) views the (e.g., first display 130), the image data of the first user (e.g., first user 120) may be captured by a first camera (e.g., first tracking camera 140) facing the first user (e.g., first user 120). Moreover, the camera facing the user may be part of a camera array (e.g., first camera array 150) including a plurality of individual cameras (e.g., first individual cameras 160).

A second user (e.g., second user 122) may view a second display (e.g., second display 132). The second display (e.g., second display 132) may provide image data associated with another user (e.g., first user 120). As the second user (e.g., second user 122) views the second display (e.g., second display 132), the image data of the second user (e.g., second user 122) may be captured by a second camera (e.g., second tracking camera 142) facing the second user (e.g., second user 122). Moreover, the second camera (e.g., second tracking camera 142) facing the second user (e.g., second user 122) may be part of a second camera array (e.g., second camera array 152) including a plurality of individual cameras (e.g., second individual cameras 162).

At step 320, method 300 may track the head or eye movements of the first user based on the first image data. For example, the head or eye movements may be traced based on a coordinate mapping in 3D space. Moreover, movements of the head or eye may by tracked by trending recorded data, including comparing current movements to past movements (e.g., artificial intelligence or other learning algorithms). A first predicted head tracking signal may be associated with a future position of the first user (e.g., predicted head movements and/or predicted left and right eye positions of the first user 120 in 3D space).

At step 330, method 300 may identify a relationship between the head or eye movements of the first user with a view of the display. The head or eye movements may be correlated with a point of view and/or an area of focus of the first user (e.g., first user 120). For example, as the first user shifts eye or head position from one direction to another, the user's view or perspective of the first display may change. In order to present a realistic 3D image of the second user, it may be desirable to change the image data presented to the first user based on the change of perspective. One or more machine learning, optimization and/or triangulation techniques may be utilized to determine a relationship between the camera position, the user's eye position, eye direction, and the observed scene to determine a viewpoint which the first user is looking. One or more machine learning techniques may be applied, as described herein, to tailor the relationships to the user's particular gazing habits and/or adapt to adjustments from positions of the cameras.

At step 340, the method 300 may convert the view to a selected subset of a plurality of individual cameras, where the subset of the plurality of individual cameras capture second image data of the second user. For example, based on the view, the head tracking signal, and/or a predicted head tracking signal may be converted to a selected subset of the plurality of second individual cameras, where the selected subset of the plurality of second individual cameras is based on the corresponding head tracking signal (e.g., actual or predicted). Conversion of the head tracking signal to a selected subset of the plurality of second individual cameras may be based on projected geometry (e.g., based on a gaze of the first user 120, distance of the first user 120 to the first display 130, etc.).

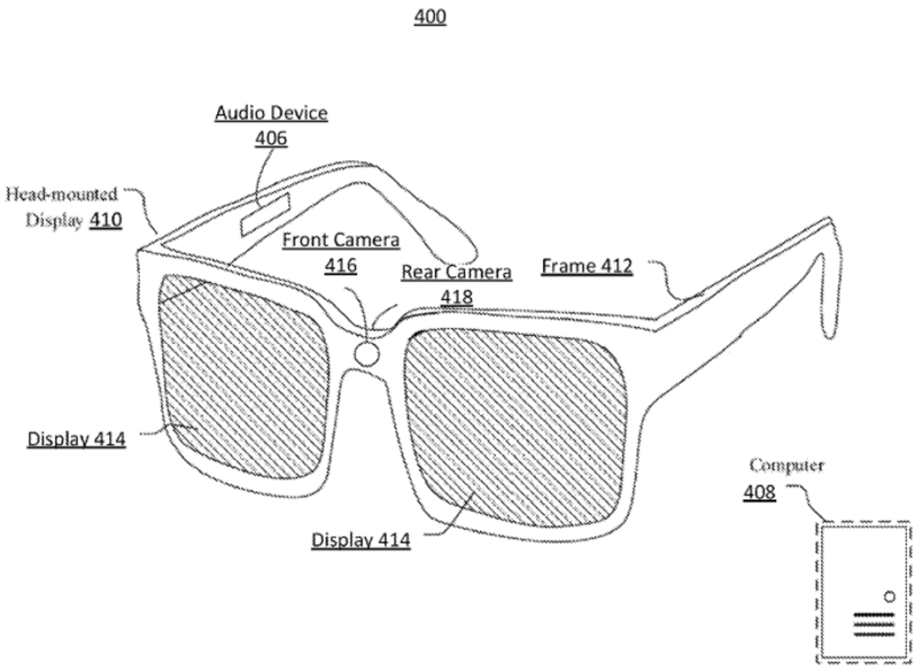

FIGURE 4

FIG. 4 illustrates an artificial reality (AR) system 400. The artificial reality system 400 may include a head-mounted display (HMD) 410 (e.g., glasses) comprising a frame 412, one or more displays 414, and a computing device 408 (also referred to herein as computer 408). The displays 414 may be transparent or translucent allowing a user wearing the HMD 410 to look through the displays 414 to see the real world and displaying visual artificial reality content to the user at the same time. The HMD 410 may include an audio device 406 (e.g., speaker/microphone 38 of FIG. 5) that may provide audio artificial reality content to users.

The HMD 410 may include one or more cameras 416, 418 (also referred to herein as imaging modules 416, 418) which can capture images and videos of environments. In an example, the HMD 410 may include a camera 418 which may be a rear-facing camera tracking movement and/or gaze of a user's eyes. One of the cameras 416 may be a forward- facing camera capturing images and/or videos of the environment that a user wearing the HMD 410 may view.

The HMD 410 may include an eye tracking system to track the vergence movement of the user wearing the HMD 410. In an example, the camera 416 may be the eye tracking system. The HMD 410 may include a microphone of the audio device 406 to capture voice input from the user.

The augmented reality system 400 may further include a controller (e.g., processor 32 of FIG. 5) comprising a trackpad and one or more buttons. The controller may receive inputs from users and relay the inputs to the computing device 408. The controller may also provide haptic feedback to users. The computing device 408 may be connected to the HMD 410 and the controller through cables or wireless connections.

The computing device 408 may control the HMD 410 and the controller to provide the augmented reality content to and receive inputs from one or more users. In some examples, the controller 418 may be a standalone controller or integrated within the HMD 410. The computing device 408 may be a standalone host computer device, an on-board computer device integrated with the HMD 410, a mobile device, or any other hardware platform capable of providing artificial reality content to and receiving inputs from users. In some examples, HMD 410 may include an artificial reality system/virtual reality system.

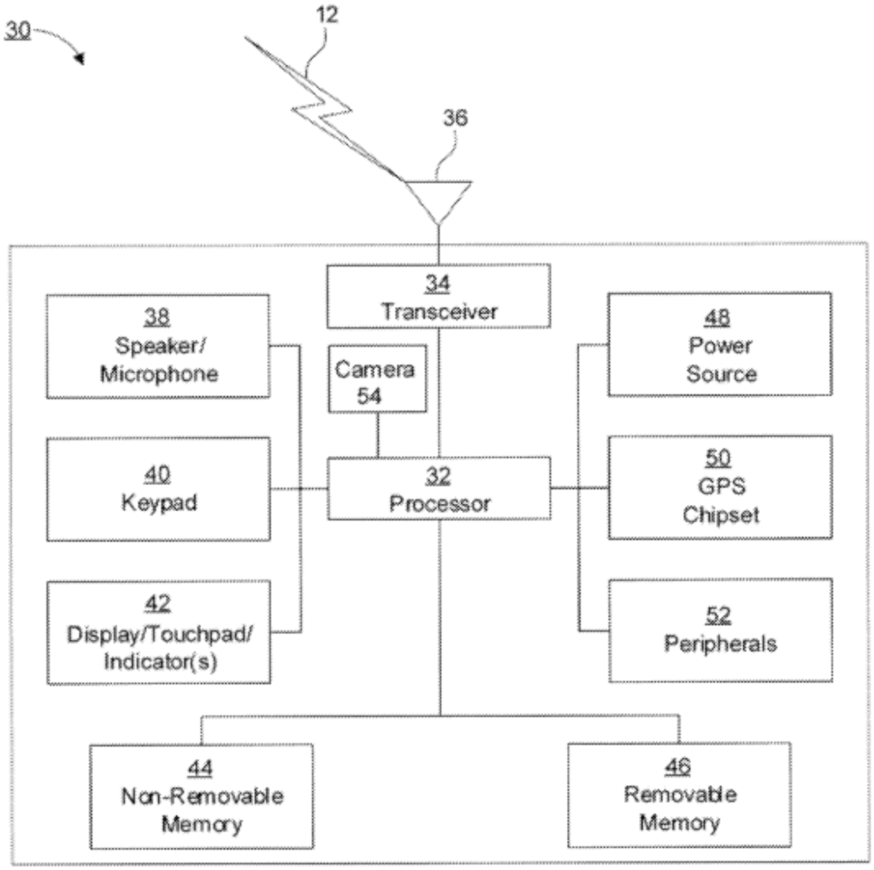

FIGURE 5

FIG. 5 illustrates a block diagram of example hardware/software architecture of a user device or user equipment (UE) 30. As shown in FIG. 5, the UE 30 (also referred to herein as node 30) may include a processor 32, non-removable memory 44, removable memory 46, a speaker/microphone 38, a keypad 40, a display, touchpad, and/or indicators 42, a power source 48, a global positioning system (GPS) chipset 50, and other peripherals 52. The UE 30 may also include a camera 54. In an example, the camera 54 is a smart camera configured to sense images appearing within one or more bounding boxes. The UE 30 may also include communication circuitry, such as a transceiver 34 and a transmit/receive element 36. UE 30 may include any sub-combination of the foregoing elements.

The processor 32 may be a special purpose processor, a digital signal processor (DSP), a plurality of microprocessors, one or more microprocessors in association with a DSP core, a controller, a microcontroller, Application Specific Integrated Circuits (ASICs), Field Programmable Gate Array (FPGAs) circuits, any other type of integrated circuit (IC), a state machine, and the like. In general, the processor 32 may execute computer-executable instructions stored in the memory (e.g., memory 44 and/or memory 46) of the node 30 in order to perform the various required functions of the node. For example, the processor 32 may perform signal coding, data processing, power control, input/output processing, and/or any other functionality that enables the node 30 to operate in a wireless or wired environment. The processor 32 may run application-layer programs (e.g., browsers) and/or radio access-layer (RAN) programs and/or other communications programs. The processor 32 may also perform security operations such as authentication, security key agreement, and/or cryptographic operations, such as at the access-layer and/or application layer for example.

The processor 32 is coupled to its communication circuitry (e.g., transceiver 34 and transmit/receive element 36). The processor 32, through the execution of computer executable instructions, may control the communication circuitry in order to cause the node 30 to communicate with other nodes via the network to which it is connected.

The transmit/receive element 36 may be configured to transmit signals to, or receive signals from, other nodes or networking equipment. For example, the transmit/receive element 36 may be an antenna configured to transmit and/or receive radio frequency (RF) signals. The transmit/receive element 36 may support various networks and air interfaces, such as wireless local area network (WLAN), wireless personal area network (WPAN), cellular, and the like. In another example, the transmit/receive element 36 may be configured to transmit and receive both

RF and light signals. It will be appreciated that the transmit/receive element 36 may be configured to transmit and/or receive any combination of wireless or wired signals.

The transceiver 34 may be configured to modulate the signals that are to be transmitted by the transmit/receive element 36 and to demodulate the signals that are received by the transmit/receive element 36. As noted above, the node 30 may have multi-mode capabilities. Thus, the transceiver 34 may include multiple transceivers for enabling the node 30 to communicate via multiple radio access technologies (RATs), such as universal terrestrial radio access (UTRA) and Institute of Electrical and Electronics Engineers (IEEE 802.11), for example.

The processor 32 may access information from, and store data in, any type of suitable memory, such as the non-removable memory 44 and/or the removable memory 46. For example, the processor 32 may store session context in its memory, as described above. The non-removable memory 44 may include RAM, ROM, a hard disk, or any other type of memory storage device. The removable memory 46 may include a subscriber identity module (SIM) card, a memory stick, a secure digital (SD) memory card, and the like. In some examples, the processor 32 may access information from, and store data in, memory that is not physically located on the node 30, such as on a server or a home computer.

The processor 32 may receive power from the power source 48, and may be configured to distribute and/or control the power to the other components in the node 30. The power source 48 may be any suitable device for powering the node 30. For example, the power source 48 may include one or more dry cell batteries (e.g., nickel-cadmium (NiCd), nickel-zinc (NiZn), nickel metal hydride (NiMH), lithium-ion (Li-ion), etc.), solar cells, fuel cells, and the like.

The processor 32 may also be coupled to the GPS chipset 50, which may be configured to provide location information (e.g., longitude and latitude) regarding the current location of the node 30. The node 30 may acquire location information by way of any suitable location-determination method.

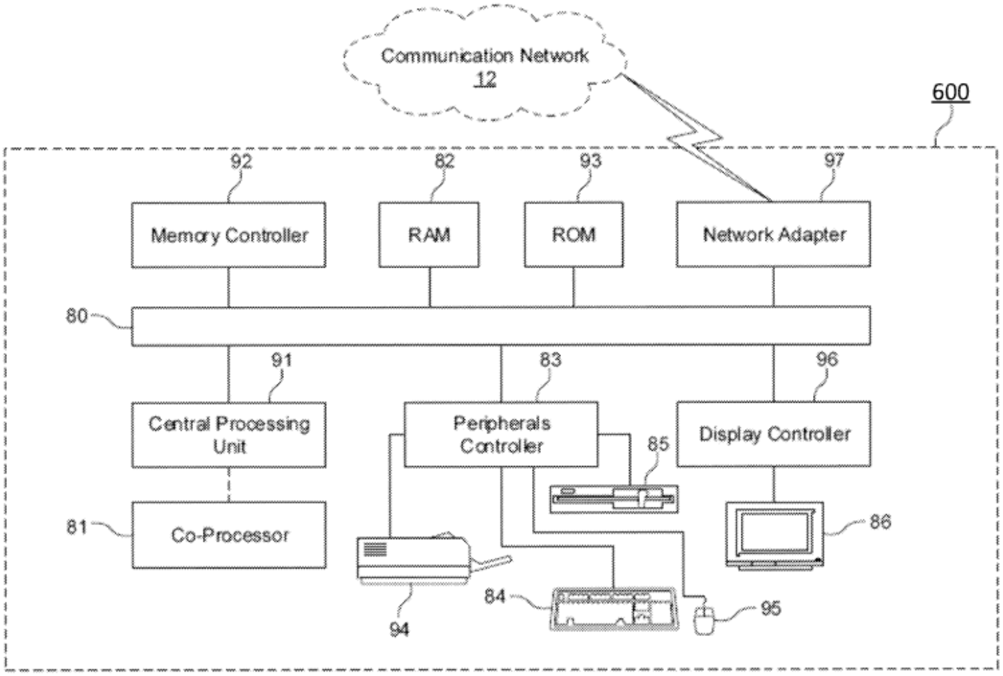

FIGURE 6

FIG. 6 illustrates a block diagram of an example computing system 600 which may be used to implement components of the system or be part of the UE 30. The computing system 600 may comprise a computer or server and may be controlled primarily by computer readable instructions, which may be in the form of software, wherever, or by whatever means such software is stored or accessed. Such computer readable instructions may be executed within a processor, such as central processing unit (CPU) 91, to cause computing system 600 to operate. In many workstations, servers, and personal computers, central processing unit 91 may be implemented by a single-chip CPU called a microprocessor. In other machines, the central processing unit 91 may comprise multiple processors. Coprocessor 81 may be an optional processor, distinct from main CPU 91, that performs additional functions or assists CPU 91.

In operation, CPU 91 fetches, decodes, and executes instructions, and transfers information to and from other resources via the computer's main data-transfer path, system bus 80. Such a system bus connects the components in computing system 600 and defines the medium for data exchange. System bus 80 typically includes data lines for sending data, address lines for sending addresses, and control lines for sending interrupts and for operating the system bus. An example of such a system bus 80 is the Peripheral Component Interconnect (PCI) bus.

Memories coupled to system bus 80 include RAM 82 and ROM 93. Such memories may include circuitry that allows information to be stored and retrieved. ROMs 93 generally contain stored data that cannot easily be modified. Data stored in RAM 82 may be read or changed by CPU 91 or other hardware devices. Access to RAM 82 and/or ROM 93 may be controlled by memory controller 92. Memory controller 92 may provide an address translation function that translates virtual addresses into physical addresses as instructions are executed. Memory controller 92 may also provide a memory protection function that isolates processes within the system and isolates system processes from user processes. Thus, a program running in a first mode may access only memory mapped by its own process virtual address space; it cannot access memory within another process's virtual address space unless memory sharing between the processes has been set up.

In addition, computing system 600 may contain peripherals controller 83 responsible for communicating instructions from CPU 91 to peripherals, such as printer 94, keyboard 84, mouse 95, and disk drive 85.

Display 86, which is controlled by display controller 96, is used to display visual output generated by computing system 200. Such visual output may include text, graphics, animated graphics, and video. Display 86 may be implemented with a cathode-ray tube (CRT)-based video display, a liquid-crystal display (LCD)-based flat-panel display, gas plasma-based flat-panel display, or a touch-panel. Display controller 96 includes electronic components required to generate a video signal that is sent to display 86.

Further, computing system 600 may contain communication circuitry, such as for example a network adaptor 97, that may be used to connect computing system 200 to an external communications network, such as network 12 of FIG. 5, to enable the computing system 200 to communicate with other nodes (e.g., UE 30) of the network.

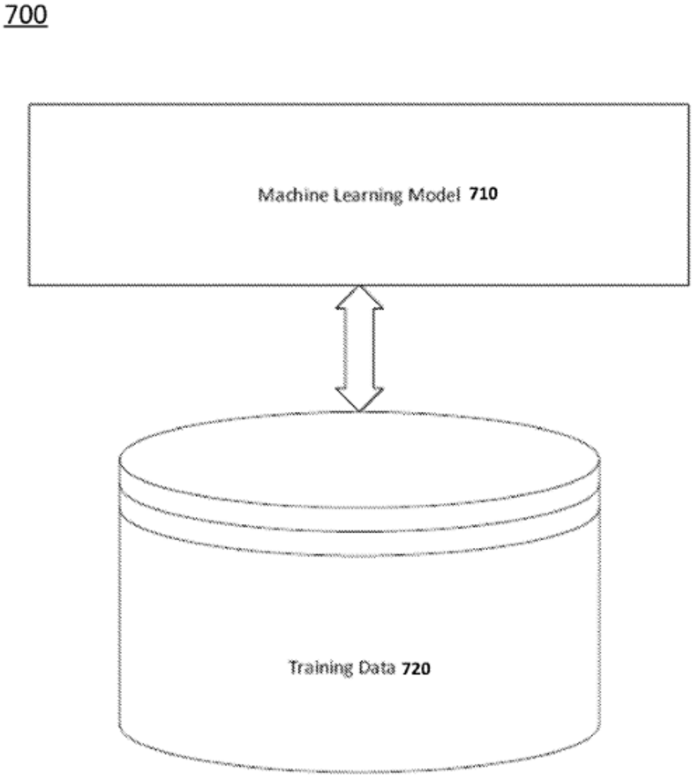

FIGURE 7

FIG. 7 illustrates a framework 700 employed by a software application (e.g., algorithm). The framework 700 may be hosted remotely. Alternatively, the framework 700 may reside within the UE 30 shown in FIG. 5 and/or be processed by the computing system 600 shown in

FIG. 6. Framework 700 includes a machine learning model 710. Machine learning model 710 can training data stored in a database 720.

Training data 720 may include attributes of thousands of objects. For example, the object may be a hand position. Attributes may include but are not limited to the size, shape, orientation, position of a hand, etc. The training data 720 employed by the machine learning model 710 may be fixed or updated periodically. Alternatively, the training data 720 may be updated in real-time based upon the evaluations performed by the machine learning model 710 in a non-training mode. This is illustrated by the double-sided arrow connecting the machine learning model 710 and stored training data 720.

In operation, the machine learning model 710 may evaluate attributes of images/videos obtained by hardware (e.g., of the AR system 400, UE 30, etc.). For example, the camera 416 and/or the camera 418 of AR system 400 and/or camera 54 of the UE 30 shown in FIG. 5 senses and captures an image/video, such as for example hand positions, hand movements (e.g., gestures) and/or other objects, appearing in or around a bounding box of a software application. The attributes of the captured image (e.g., captured image of a gesture(s)) are then compared with respective attributes of stored training data 720 (e.g., prestored training gestures). The likelihood of similarity between each of the obtained attributes (e.g., of the captured image of a gesture(s)) and the stored training data 720 (e.g., prestored training gestures) is given a confidence score. In an example, if the confidence score exceeds a predetermined threshold, the attribute is included in an image description (e.g., thumbs up gesture, thumbs down gesture, etc.) that is ultimately communicated to the user via a user interface of a computing device (e.g., UE 30, computing device). In another example, the description may include a certain number of attributes which exceed a predetermined threshold to share with the user. The sensitivity of sharing more or less attributes can be customized based upon the needs of the particular user.

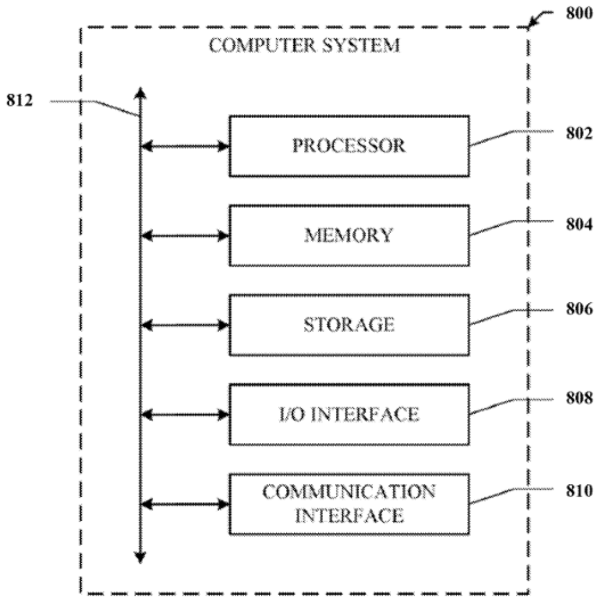

FIGURE 8

FIG. 8 illustrates an example computer system 800. In various examples, one or more computer systems 800 perform one or more steps of one or more methods described or illustrated herein. In particular examples, one or more computer systems 800 provide functionality described or illustrated herein. In various examples, software running on one or more computer systems 800 performs one or more steps of one or more methods described or illustrated herein or provides functionality described or illustrated herein. Various examples may include one or more portions of one or more computer systems 800. Herein, reference to a computer system may encompass a computing device, and vice versa, where appropriate. Moreover, reference to a computer system may encompass one or more computer systems, where appropriate.

This disclosure contemplates any suitable number of computer systems 800. This disclosure contemplates computer system 800 taking any suitable physical form. As example and not by way of limitation, computer system 800 may be an embedded computer system, a system-on-chip (SOC), a single-board computer system (SBC) (such as, for example, a computer-on-module (COM) or system-on-module (SOM)), a desktop computer system, a laptop or notebook computer system, an interactive kiosk, a mainframe, a mesh of computer systems, a mobile telephone, a personal digital assistant (PDA), a server, a tablet computer system, or a combination of two or more of these. Where appropriate, computer system 800 may include one or more computer systems 800; be unitary or distributed; span multiple locations; span multiple machines; span multiple data centers; or reside in a cloud, which may include one or more cloud components in one or more networks. Where appropriate, one or more computer systems 800 may perform without substantial spatial or temporal limitation one or more steps of one or more methods described or illustrated herein. As an example, and not by way of limitation, one or more computer systems 800 may perform in real time or in batch mode one or more steps of one or more methods described or illustrated herein. One or more computer systems 800 may perform at different times or at different locations one or more steps of one or more methods described or illustrated herein, where appropriate.

In various examples, computer system 800 includes a processor 802, memory 804, storage 806, an input/output (I/O) interface 808, a communication interface 810. and a bus 812. Although this disclosure describes and illustrates a particular computer system having a particular number of particular components in a particular arrangement, this disclosure contemplates any suitable computer system having any suitable number of any suitable components in any suitable arrangement.

In various examples, processor 802 includes hardware for executing instructions, such as those making up a computer program. As an example, to execute instructions, processor 802 may retrieve (or fetch) the instructions from an internal register, an internal cache, memory 804, or storage 806; decode and execute them; and then write one or more results to an internal register, an internal cache, memory 804, or storage 806.

In particular examples, processor 802 may include one or more internal caches for data, instructions, or addresses. This disclosure contemplates processor 802 including any suitable number of any suitable internal caches, where appropriate. As an example, processor 802 may include one or more instruction caches, one or more data caches, and one or more translation lookaside buffers (TLBs). Instructions in the instruction caches may be copies of instructions in memory 804 or storage 806, and the instruction caches may speed up retrieval of those instructions by processor 802. Data in the data caches may be copies of data in memory 804 or storage 806 for instructions executing at processor 802 to operate on; the results of previous instructions executed at processor 802 for access by subsequent instructions executing at processor 802 or for writing to memory 804 or storage 806; or other suitable data. The data caches may speed up read or write operations by processor 802. The TLBs may speed up virtual-address translation for processor 802. In particular examples, processor 802 may include one or more internal registers for data, instructions, or addresses. This disclosure contemplates processor 802 including any suitable number of any suitable internal registers, where appropriate. Where appropriate, processor 802 may include one or more arithmetic logic units (ALUs); be a multi-core processor; or include one or more processors 802. Although this disclosure describes and illustrates a particular processor, this disclosure contemplates any suitable processor.

In various examples, memory 804 includes main memory for storing instructions for processor 802 to execute or data for processor 802 to operate on. As an example, computer system 800 may load instructions from storage 806 or another source (such as, for example, another computer system 800) to memory 804. Processor 802 may then load the instructions from memory 804 to an internal register or internal cache. To execute the instructions, processor 802 may retrieve the instructions from the internal register or internal cache and decode them. During or after execution of the instructions, processor 802 may write one or more results (which may be intermediate or final results) to the internal register or internal cache. Processor 802 may then write one or more of those results to memory 804.

In particular examples, processor 802 executes only instructions in one or more internal registers or internal caches or in memory 804 (as opposed to storage 806 or elsewhere) and operates only on data in one or more internal registers or internal caches or in memory 804 (as opposed to storage 806 or elsewhere). One or more memory buses (which may each include an address bus and a data bus) may couple processor 802 to memory 804. Bus 812 may include one or more memory buses, as described below. In some examples, one or more memory management units (MMUs) reside between processor 802 and memory 804 and facilitate accesses to memory 804 requested by processor 802.

In particular examples, memory 804 includes random access memory (RAM). This RAM may be volatile memory, where appropriate. Where appropriate, this RAM may be dynamic RAM (DRAM) or static RAM (SRAM). Moreover, where appropriate, this RAM may be single-ported or multi-ported RAM. This disclosure contemplates any suitable RAM. Memory 804 may include one or more memories 804, where appropriate. Although this disclosure describes and illustrates a particular memory, this disclosure contemplates any suitable memory.

In various examples, storage 806 includes mass storage for data or instructions. As an example, storage 806 may include a hard disk drive (HDD), a floppy disk drive, flash memory, an optical disc, a magneto-optical disc, magnetic tape, or a Universal Serial Bus (USB) drive or a combination of two or more of these. Storage 806 may include removable or non-removable (or fixed) media, where appropriate. Storage 806 may be internal or external to computer system 800, where appropriate. In some examples, storage 806 is non-volatile, solid-state memory. In particular examples, storage 806 includes read-only memory (ROM). Where appropriate, this ROM may be mask-programmed ROM, programmable ROM (PROM), erasable PROM (EPROM), electrically erasable PROM (EEPROM), electrically alterable ROM (EAROM), or flash memory or a combination of two or more of these. This disclosure contemplates mass storage 806 taking any suitable physical form. Storage 806 may include one or more storage control units facilitating communication between processor 802 and storage 806, where appropriate. Where appropriate, storage 806 may include one or more storages 806. Although this disclosure describes and illustrates particular storage, this disclosure contemplates any suitable storage.

In various examples, input/output (I/O) interface 808 includes hardware, software, or both, providing one or more interfaces for communication between computer system 800 and one or more I/O devices. Computer system 800 may include one or more of these I/O devices, where appropriate. One or more of these I/O devices may enable communication between a person and computer system 800. As an example, an I/O device may include a keyboard, keypad, microphone, monitor, mouse, printer, scanner, speaker, still camera, stylus, tablet, touch screen, trackball, video camera, another suitable I/O device or a combination of two or more of these. An I/O device may include one or more sensors. This disclosure contemplates any suitable I/O devices and any suitable I/O interfaces 808 for them. Where appropriate, I/O interface 808 may include one or more device or software drivers enabling processor 802 to drive one or more of these I/O devices. I/O interface 808 may include one or more I/O interfaces 808, where appropriate. Although this disclosure describes and illustrates a particular I/O interface, this disclosure contemplates any suitable I/O interface.

In various examples, communication interface 810 includes hardware, software, or both providing one or more interfaces for communication (such as, for example, packet-based communication) between computer system 800 and one or more other computer systems 800 or one or more networks. As an example, communication interface 810 may include a network interface controller (NIC) or network adapter for communicating with an Ethernet or other wire-based network or a wireless NIC (WNIC) or wireless adapter for communicating with a wireless network, such as a WI-FI network. This disclosure contemplates any suitable network and any suitable communication interface 810 for it.

As an example, computer system 800 may communicate with an ad hoc network, a personal area network (PAN), a local area network (LAN), a wide area network (WAN), a metropolitan area network (MAN), or one or more portions of the Internet or a combination of two or more of these. One or more portions of one or more of these networks may be wired or wireless. As an example, computer system 800 may communicate with a wireless PAN (WPAN) (such as, for example, a BLUETOOTH WPAN), a WI-FI network, a WI-MAX network, a cellular telephone network (such as, for example, a Global System for Mobile Communications (GSM) network), or other suitable wireless network or a combination of two or more of these. Computer system 800 may include any suitable communication interface 810 for any of these networks, where appropriate. Communication interface 810 may include one or more communication interfaces 810, where appropriate. Although this disclosure describes and illustrates a particular communication interface, this disclosure contemplates any suitable communication interface.

In various examples, bus 812 includes hardware, software, or both coupling components of computer system 800 to each other. As an example, bus 812 may include an Accelerated Graphics Port (AGP) or other graphics bus, an Enhanced Industry Standard Architecture (EISA) bus, a front-side bus (FSB), a HYPERTRANSPORT (HT) interconnect, an Industry Standard Architecture (ISA) bus, an INFINIBAND interconnect, a low-pin-count (LPC) bus, a memory bus, a Micro Channel Architecture (MCA) bus, a Peripheral Component Interconnect (PCI) bus, a PCI-Express (PCIe) bus, a serial advanced technology attachment (SATA) bus, a Video Electronics Standards Association local (VLB) bus, or another suitable bus or a combination of two or more of these. Bus 812 may include one or more buses 812, where appropriate. Although this disclosure describes and illustrates a particular bus, this disclosure contemplates any suitable bus or interconnect.

Herein, a computer-readable non-transitory storage medium or media may include one or more semiconductor-based or other integrated circuits (ICs) (such, as for example, field-programmable gate arrays (FPGAs) or application-specific ICs (ASICs)), hard disk drives (HDDs), hybrid hard drives (HHDs), optical discs, optical disc drives (ODDs), magneto-optical discs, magneto-optical drives, floppy diskettes, floppy disk drives (FDDs), magnetic tapes, solid-state drives (SSDs), RAM-drives, SECURE DIGITAL cards or drives, any other suitable computer-readable non-transitory storage media, or any suitable combination of two or more of these, where appropriate. A computer-readable non-transitory storage medium may be volatile, non-volatile, or a combination of volatile and non-volatile, where appropriate.

For the purpose of illustrating the disclosed subject matter, there are shown in the drawings various examples of the disclosed subject matter; however, the disclosed subject matter is not limited to the specific methods, compositions, and devices disclosed. The drawings herein are not necessarily drawn to scale.

CONCLUSION

This disclosure describes techniques to select a subset of available cameras based on head and/or eye movements of participants in a video call to render a corresponding direct view. Cameras capture images of a viewing user that is viewing a display on which a 3D video of a second user (e.g., generated from images captured by a subset of a plurality of cameras) is displayed. Per techniques of this disclosure, with user permission, the head and/or eye movements of the viewing user are tracked based on the captured images of the viewing user. Relationship between the tracked movements of the viewing user and a view of the display of the device is determined. The view of the first display is then updated to render a 3D video based on a subset of individual cameras of a second user's device that match the movement (which corresponds to an updated viewing position) of the viewing user. The view on the first display is updated to show the second user from a corresponding perspective. The techniques can provide a more accurate experience during a three-dimensional (3D) video conference. Suitable techniques such as machine learning can be utilized to predict the user's movement and adjust the view accordingly.