(54) Title: METHODS, SYSTEMS, AND MEDIA FOR GENERATING AND RENDERING IMMERSIVE VIDEO CONTENT

400
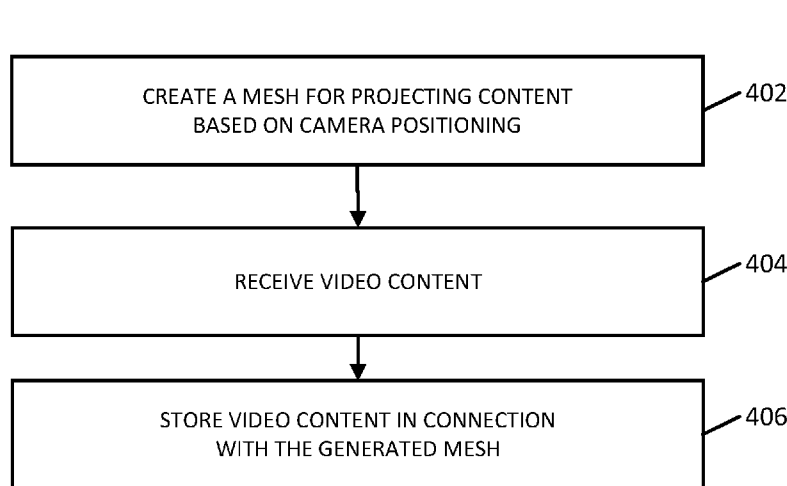
FIG. 4

(57) **Abstract:** Methods, systems, and media for generating and rendering immersive video content are provided. In some embodiments, the method comprises: receiving information indicating positions of cameras in a plurality of cameras; generating a mesh on which video content is to be projected based on the positions of the cameras in the plurality of cameras, wherein the mesh is comprised of a portion of a faceted cylinder, and wherein the faceted cylinder has a plurality of facets each corresponding to a projection from a camera in the plurality of cameras; receiving video content corresponding to the plurality of cameras; and transmitting the video content and the generated mesh to a user device in response to receiving a request for the video content from the user device.

METHODS, SYSTEMS, AND MEDIA FOR
GENERATING AND RENDERING IMMERSIVE VIDEO CONTENT

Cross-Reference to Related Applications

[0001]   This application claims the benefit of United States Patent Application
No. 15/840,361, filed December 13, 2017, which is hereby incorporated by reference herein in
its entirety.

Technical Field

[0002]   The disclosed subject matter relates to methods, systems, and media for
generating and rendering immersive video content.

Background

[0003]   Many users enjoy viewing immersive video content, such as virtual reality
content, three-dimensional content, 180-degree content, or 360-degree content, that can provide a
viewer with an immersive experience. Additionally, these users may enjoy viewing computer-
generated content in an immersive format, such as gaming videos or animations.

[0004]   Immersive video content is frequently rendered by applying video content to a
mesh, such as a fish-eye mesh or a three-dimensional equirectangular mesh, which causes the
video content to be rendered as if it were filling a portion of a sphere rather than as appearing
flat. However, rendering computer-generated content using a fish-eye projection or a three-
dimensional equirectangular projection can be resource intensive, particularly when rendering
graphics in real-time.

[0005]   Accordingly, it is desirable to provide new methods, systems, and media for
generating and rendering immersive video content.

Summary

[0006]   Methods, systems, and media for generating and rendering immersive video
content are provided.

[0007]   In accordance with some embodiments of the disclosed subject matter, a method
for generating immersive video content is provided, the method comprising: receiving

information indicating positions of cameras in a plurality of cameras; generating a mesh on which video content is to be projected based on the positions of the cameras in the plurality of cameras, wherein the mesh is comprised of a portion of a faceted cylinder, and wherein the faceted cylinder has a plurality of facets each corresponding to a projection from a camera in the plurality of cameras; receiving video content corresponding to the plurality of cameras; and transmitting the video content and the generated mesh to a user device in response to receiving a request for the video content from the user device.

[0008]   In some embodiments, generating the mesh comprises stitching the plurality of facets in a side by side orientation.

[0009]   In some embodiments, the method further comprises inserting the generated mesh into a video file that includes the video content prior to transmitting the video file to the user device.

[0010]   In some embodiments, the video content includes a depth map.

[0011]   In some embodiments, the method further comprises converting the generated mesh to a second type of mesh for projecting video content. In some embodiments, the second type of mesh is a fish-eye projection.

[0012]   In some embodiments, the method further comprises: receiving resulting image information associated with a rendered video content item; and causing an indication to be presented to displace one or more of the plurality of cameras based on the information indicating the positions of cameras in the plurality of cameras.

[0013]   In accordance with some embodiments of the disclosed subject matter, a system for generating immersive video content is provided, the system comprising a memory and a hardware processor that, when executing computer executable instructions stored in the memory, is configured to: receive information indicating positions of cameras in a plurality of cameras; generate a mesh on which video content is to be projected based on the positions of the cameras in the plurality of cameras, wherein the mesh is comprised of a portion of a faceted cylinder, and wherein the faceted cylinder has a plurality of facets each corresponding to a projection from a camera in the plurality of cameras; receive video content corresponding to the plurality of cameras; and transmit the video content and the generated mesh to a user device in response to receiving a request for the video content from the user device.

[0014] In some embodiments, the hardware processor is configured to stitch the plurality of facets in a side by side orientation to generating the mesh.

[0015] In some embodiments, the hardware processor is further configured to insert the generated mesh into a video file that includes the video content prior to transmitting the video file to the user device.

[0016] In some embodiments, the video content includes a depth map.

[0017] In some embodiments, the hardware processor is further configured to convert the generated mesh to a second type of mesh for projecting video content.

[0018] In some embodiments, the second type of mesh is a fish-eye projection.

[0019] In some embodiments, the hardware processor is further configured to: receive resulting image information associated with a rendered video content item; and cause an indication to be presented to displace one or more of the plurality of cameras based on the information indicating the positions of cameras in the plurality of cameras.

[0020] In accordance with some embodiments of the disclosed subject matter, a non-transitory computer-readable medium containing computer executable instructions that, when executed by a processor, cause the processor to perform a method for generating immersive video content is provided, the method comprising: receiving information indicating positions of cameras in a plurality of cameras; generating a mesh on which video content is to be projected based on the positions of the cameras in the plurality of cameras, wherein the mesh is comprised of a portion of a faceted cylinder, and wherein the faceted cylinder has a plurality of facets each corresponding to a projection from a camera in the plurality of cameras; receiving video content corresponding to the plurality of cameras; and transmitting the video content and the generated mesh to a user device in response to receiving a request for the video content from the user device.

[0021] In some embodiments, generating the mesh comprises stitching the plurality of facets in a side by side orientation.

[0022] In some embodiments, the method further comprises inserting the generated mesh into a video file that includes the video content prior to transmitting the video file to the user device.

[0023] In some embodiments, the video content includes a depth map.

[0024]   In some embodiments, the method further comprises converting the generated mesh to a second type of mesh for projecting video content.

[0025]   In some embodiments, the second type of mesh is a fish-eye projection.

[0026]   In some embodiments, the method further comprises: receiving resulting image information associated with a rendered video content item; and causing an indication to be presented to displace one or more of the plurality of cameras based on the information indicating the positions of cameras in the plurality of cameras.

[0027]   In accordance with some embodiments of the disclosed subject matter, a system for generating immersive video content is provided, the system comprising: means for receiving information indicating positions of cameras in a plurality of cameras; means for generating a mesh on which video content is to be projected based on the positions of the cameras in the plurality of cameras, wherein the mesh is comprised of a portion of a faceted cylinder, and wherein the faceted cylinder has a plurality of facets each corresponding to a projection from a camera in the plurality of cameras; means for receiving video content corresponding to the plurality of cameras; and means for transmitting the video content and the generated mesh to a user device in response to receiving a request for the video content from the user device.

Brief Description of the Drawings

[0028]   Various objects, features, and advantages of the disclosed subject matter can be more fully appreciated with reference to the following detailed description of the disclosed subject matter when considered in connection with the following drawings, in which like reference numerals identify like elements.

[0029]   FIGS. 1A and 1B show examples of schematic diagrams for rendering immersive video content in accordance with some embodiments of the disclosed subject matter.

[0030]   FIG. 2 shows a schematic diagram of an illustrative system suitable for implementation of mechanisms described herein for rendering immersive video content in accordance with some embodiments of the disclosed subject matter.

[0031]   FIG. 3 shows a detailed example of hardware that can be used in a server and/or a user device of FIG. 2 in accordance with some embodiments of the disclosed subject matter.

[0032]   FIG. 4 shows an example of a process for processing video streams to be rendered as immersive content in accordance with some embodiments of the disclosed subject matter.

4

[0033]   FIG. 5 shows an example of a process for rendering immersive video content in accordance with some embodiments of the disclosed subject matter.

Detailed Description

[0034]   In accordance with various embodiments, mechanisms (which can include methods, systems, and media) for generating and rendering immersive video content are provided.

[0035]   In some embodiments, the mechanisms described herein can receive or generate video content, and can generate a mesh that can be used to render the video content as immersive content on a user device. For example, the mechanisms described herein can generate a mesh that can be used to render the video content as 180-degree video content, which can allow a viewer of the video content to feel immersed in the video content when viewing the video content. As another example, in some embodiments, the mechanisms described herein can generate a mesh that can be used to render the video content as three-dimensional video content with any suitable spatial extent.

[0036]   In some embodiments, the mechanisms described herein can generate the mesh based on any suitable information associated with one or more cameras generating or recording the video content. For example, in some embodiments, the mechanisms can generate a mesh that includes a group of panels, where each panel corresponds to a field of view from a corresponding group of cameras. In some such embodiments, the panels can be placed (e.g. stitched together) in a side-by-side positioning to generate a portion of a faceted cylinder, representing a cumulative field of view of the cameras in the group of cameras, as described below in connection with FIGS. 1A, 1B, and 4. Note that, in some embodiments, the cameras described herein can be cameras used for generating computer-generated (CG) content, such as gaming videos, animations, graphics, and/or any other suitable type of CG content. In some embodiments, the video content can be CG content that is rendered in real-time or in near real-time.

[0037]   In some embodiments, the video content and the generated mesh can be transmitted to a user device for rendering by the user device. For example, in some embodiments, the generated mesh can be inserted into a video file that includes the video content, and the video file can be transmitted to the user device, as described below in

connection with FIG. 5. Additionally or alternatively, in some embodiments, the generated mesh can be used as an intermediary step prior to creating a fish-eye projection of the video content and/or a three-dimensional equirectangular projection of the video content.

[0038]   In some embodiments, the mechanisms described herein can have the effect of reducing the time needed to render immersive content. In particular, the mechanisms can be used to render CG content, such as animated video and/or visual effects, quickly and efficiently on a user device. Thus, the embodiments described herein can enable efficient use of processor resources on the user device. Additionally, the mechanisms described herein can be used to render immersive content in instances where a rendering application does not support particular types of three-dimensional projection, such as a fish-eye projection or a three-dimension equirectangular projection.

[0039]   Turning to FIG. 1A, an illustrative example 100 of a schematic diagram for recording and/or generating video content that is to be rendered as immersive video content is shown in accordance with some embodiments of the disclosed subject matter. In some embodiments, the immersive video content can be three-dimensional video content, 180-degree video content, or 360-degree video content that allows a viewer to feel immersed within the video content while viewing the video content by allowing the viewer to manipulate a viewpoint of the video content, and/or any other suitable type of video content. In some embodiments, the video content can be generated by a computer-generated (CG) camera that generates CG video content using any suitable technique or combination of techniques.

[0040]   As shown in FIG. 1A, a group 104 of any suitable number of cameras (e.g., two, three, and/or any other suitable number) can be used to generate video content 102 (e.g., that include content represented by the five horizontal squares or cubes). In some embodiments, each camera can be associated with a frustum, such as frustums 106, 108 and/or 110, which can represent a field of view of a corresponding camera. In some such embodiments, each of frustums 106-110 can be represented as a pyramid originating from a center of a corresponding camera. Note that, although not illustrated in FIG. 1A, in some embodiments, the cameras can record or generate video content that is to be presented stereoscopically. For example, in some embodiments, a first camera can be located at a first location (e.g., -30 millimeters, and/or at any other suitable location) and a second camera can be located at a second location (e.g., +30 millimeters, and/or at any other suitable location) offset from the first location. In continuing

this example, the cameras can be positioned (e.g., side by side) such that the combined frustums can capture video content, thereby producing a single output.

[0041]   It should be noted that, in some embodiments, location information associated with each camera can be used to calibrate the location and/or orientation associated with each camera. This can, for example, coordinate the placement of a first camera at a first location of -30 millimeters and a second camera at a second location of +30 millimeters to create resulting image information for each eye of a viewer. In a more particular example, two camera rigs, where each camera rig has at least two cameras, can be configured to generate a single output as follows:

LeftCameraRig: pos.x = -0.032m

LeftCamera: rotation.y = -30°, fov = 60 °, fill 0% to 25% of the viewport

RightCamera: rotation.y = 30°, fov = 60 °, fill 25% to 50% of the viewport

RightCameraRig: pos.x = +0.032m

LeftCamera: rotation.y = -30°, fov = 60 °, fill 50% to 75% of the viewport

RightCamera: rotation.y = 30°, fov = 60 °, fill 75% to 100% of the viewport

[0042]   In some embodiments, an instruction to re-position or otherwise displace one or more cameras by a particular distance can be presented in response to receiving and analyzing the location information associated with each camera. For example, a game engine can render with multiple cameras. In response to receiving the output from the multiple cameras placed at particular positions, the game engine can receive resulting image information associated with a rendered video content item and can cause an indication to be presented to displace one or more of the cameras based on the information indicating the positions of the cameras. In a more particular example, based on the resulting image information, the game engine can present a notification that the position of LeftCamera in LeftCameraRig should be moved from -0.032m to -0.030m.

[0043]   Turning to FIG. 1B, an illustrative example 150 of a schematic diagram for projecting the video content using a mesh based on camera position is shown in accordance with some embodiments of the disclosed subject matter. As illustrated, images created by each camera can be used to create a panel (e.g., panels 156, 158, and/or 160), where each panel spans a field of view corresponding to a field of a view of an associated camera in group 104 of FIG. 1A. For example, in an instance where each camera in group 104 has a horizontal field of view

of 50 degrees and is oriented 50 degrees apart, panels 156-160 can each have a field of view of 50 degrees. In some embodiments, panels 156-160 can be located side-by-side, for example, as the sides of a faceted cylinder. Thus, each of panels 156-160 may correspond to a facet of the faceted cylinder. In some such embodiments, panels 156-160, which can be located side-by-side, can create a portion of a faceted cylinder having a total field of view that is generally equivalent to the sum of the field of view of each camera in the group of cameras 104. For example, in an instance where each camera in group 104 has a horizontal field of view of 50 degrees and where there are three cameras in group 104, panels 156-160 can represent a portion of a faceted cylinder that spans 150 degrees. In continuing this example, FIG. 1B illustrates that a three panel 150 degree mesh can be used to project the flat-projected content from three cameras that each have a horizontal field of view of 50 degrees. That is, a mesh is created that includes a panel per camera that corresponds to the projection panel of the camera. In the case of three cameras as shown in FIGS. 1A and 1B, the created mesh can be a partial cylinder having three facets. In some embodiments, panels 156-160 can each represent a plane in three-dimensional space, with any suitable rotation.

[0044]   Although the embodiments described herein generally relate to a total field of view of a particular number of degrees, such as 180 degrees, this is merely illustrative. It should be noted that, in some embodiments, the total field of view can be an arbitrary value (e.g., 140 degrees, 124 degrees, 250 degrees, etc.). It should also be noted that the vertical field of view and the horizontal field of view can be different values.

[0045]   It should further be noted that, although the embodiments described herein generally relate to a mesh that can be arranged to represent a faceted cylinder, this is merely illustrative. In some embodiments, a panel in a three-dimensional space can be created for each camera, where the panel corresponds to the frustum of the camera, and these panels can be used for any suitable shape. For example, the shape may be created from panels that are not touching or combined such that the panels do not form a continuous shape. In a more particular example, four cameras can be configured to face forward and be angled towards a top-left direction, a top-right direction, a bottom-left direction, and a bottom-right direction, where the frustums of the cameras can overlap at a center position. In continuing this example, the corresponding panels that are created for each of these cameras may have edges that overlap in a three-dimensional space while not having edges that touch.

[0046]   Turning to FIG. 2, an illustrative example 200 of hardware for rendering immersive video content that can be used in accordance with some embodiments of the disclosed subject matter is shown. As illustrated, hardware 200 can include a content server 202, a communication network 204, and/or one or more user devices 206, such as user devices 208 and 210.

[0047]   Content server 202 can be any suitable server(s) for storing and/or providing media content to user devices 206. For example, in some embodiments, content server 202 can store media content, such as videos, television programs, movies, live-streamed media content, audio content, animations, video game content, graphics, and/or any other suitable media content. In some embodiments, content server 202 can transmit media content to user devices 206, for example, via communication network 204. In some embodiments, content server 202 can store video content (e.g., live video content, computer-generated video content, and/or any other suitable type of video content) in association with any suitable information to be used by a client device (e.g., user device 206) to render the video content as immersive content, as shown in and described below in connection with FIG. 4.

[0048]   Communication network 204 can be any suitable combination of one or more wired and/or wireless networks in some embodiments. For example, communication network 204 can include any one or more of the Internet, an intranet, a wide-area network (WAN), a local-area network (LAN), a wireless network, a digital subscriber line (DSL) network, a frame relay network, an asynchronous transfer mode (ATM) network, a virtual private network (VPN), and/or any other suitable communication network. User devices 206 can be connected by one or more communications links (e.g., communications links 212) to communication network 204 that can be linked via one or more communications links (e.g., communications links 214) to content server 202. The communications links can be any communications links suitable for communicating data among user devices 206 and content server 202 such as network links, dial-up links, wireless links, hard-wired links, any other suitable communications links, or any suitable combination of such links.

[0049]   User devices 206 can include any one or more user devices suitable for requesting video content, rendering the requested video content as immersive video content (e.g., as virtual reality content, as three-dimensional content, as 360-degree video content, as 180-degree video content, and/or in any other suitable manner) and/or for performing any other suitable functions.

For example, in some embodiments, user devices 206 can include a mobile device, such as a mobile phone, a tablet computer, a wearable computer, a laptop computer, a virtual reality headset, a vehicle (e.g., a car, a boat, an airplane, or any other suitable vehicle) information or entertainment system, and/or any other suitable mobile device and/or any suitable non-mobile device (e.g., a desktop computer, a game console, and/or any other suitable non-mobile device). As another example, in some embodiments, user devices 206 can include a media playback device, such as a television, a projector device, a game console, desktop computer, and/or any other suitable non-mobile device.

[0050]   Although content server 202 is illustrated as one device, the functions performed by content server 202 can be performed using any suitable number of devices in some embodiments. For example, in some embodiments, multiple devices can be used to implement the functions performed by content server 202.

[0051]   Although two user devices 208 and 210 are shown in FIG. 2 to avoid over-complicating the figure, any suitable number of user devices, and/or any suitable types of user devices, can be used in some embodiments.

[0052]   Content server 202 and user devices 206 can be implemented using any suitable hardware in some embodiments. For example, in some embodiments, devices 202 and 206 can be implemented using any suitable general purpose computer or special purpose computer. For example, a mobile phone may be implemented using a special purpose computer. Any such general purpose computer or special purpose computer can include any suitable hardware. For example, as illustrated in example hardware 300 of FIG. 3, such hardware can include hardware processor 302, memory and/or storage 304, an input device controller 306, an input device 308, display/audio drivers 310, display and audio output circuitry 312, communication interface(s) 314, an antenna 316, and a bus 318.

[0053]   Hardware processor 302 can include any suitable hardware processor, such as a microprocessor, a micro-controller, digital signal processor(s), dedicated logic, and/or any other suitable circuitry for controlling the functioning of a general purpose computer or a special purpose computer in some embodiments. In some embodiments, hardware processor 302 can be controlled by a server program stored in memory and/or storage 304 of a server (e.g., such as content server 202). For example, in some embodiments, the server program can cause hardware processor 302 to transmit a media content item to user device 206, transmit instructions for

presenting rendering a video stream as immersive video content, and/or performing any other suitable actions. In some embodiments, hardware processor 302 can be controlled by a computer program stored in memory and/or storage 304 of user device 206. For example, the computer program can cause hardware processor 302 to render a video stream as immersive video content, and/or perform any other suitable actions.

[0054]   Memory and/or storage 304 can be any suitable memory and/or storage for storing programs, data, media content, and/or any other suitable information in some embodiments. For example, memory and/or storage 304 can include random access memory, read-only memory, flash memory, hard disk storage, optical media, and/or any other suitable memory.

[0055]   Input device controller 306 can be any suitable circuitry for controlling and receiving input from one or more input devices 308 in some embodiments. For example, input device controller 306 can be circuitry for receiving input from a touchscreen, from a keyboard, from a mouse, from one or more buttons, from a voice recognition circuit, from a microphone, from a camera, from an optical sensor, from an accelerometer, from a temperature sensor, from a near field sensor, and/or any other type of input device.

[0056]   Display/audio drivers 310 can be any suitable circuitry for controlling and driving output to one or more display/audio output devices 312 in some embodiments. For example, display/audio drivers 310 can be circuitry for driving a touchscreen, a flat-panel display, a cathode ray tube display, a projector, a speaker or speakers, and/or any other suitable display and/or presentation devices.

[0057]   Communication interface(s) 314 can be any suitable circuitry for interfacing with one or more communication networks, such as network 204 as shown in FIG. 2. For example, interface(s) 314 can include network interface card circuitry, wireless communication circuitry, and/or any other suitable type of communication network circuitry.

[0058]   Antenna 316 can be any suitable one or more antennas for wirelessly communicating with a communication network (e.g., communication network 204) in some embodiments. In some embodiments, antenna 316 can be omitted.

[0059]   Bus 318 can be any suitable mechanism for communicating between two or more components 302, 304, 306, 310, and 314 in some embodiments.

[0060]   Any other suitable components can be included in hardware 300 in accordance with some embodiments.

[0061]   Turning to FIG. 4, an illustrative example 400 of a process for generating immersive video content is shown in accordance with some embodiments of the disclosed subject matter. In some embodiments, blocks of process 400 can be performed by any suitable device, such as a content server (e.g., content server 202 as shown in and described above in connection with FIG. 2) that stores video content and transmits the stored video content to user devices.

[0062]   Process 400 can begin by generating a mesh on which video content can be projected at 402. Generating the mesh on the content server 202 can avoid incurring computing resources (e.g. processing and/or memory resources) on the user device where computing resources may be limited. In some embodiments, the mesh can be generated based on a number of cameras used to generate the video content and/or based on positions of the cameras used to generate the video content. For example, in some embodiments, a number of panels included in the mesh can be based on a number of cameras used to record or generate the video content, as shown in and described above in connection with FIGS. 1A and 1B. As another example, in some embodiments, a field of view represented by the mesh can be based on fields of view of the cameras used to record or generate the video content. As a more particular example, in an instance where N cameras each have an X degree horizontal field of view, the horizontal field of view represented by the mesh can be $N \times X$. As a specific example, in an instance where the video content is recorded or generated by 3 cameras each having a 50-degree field of view, the mesh can include three panels each representing a 50-degree field of view. Continuing with this example, in some embodiments, the mesh can include three panels arranged to represent 150 degrees of a faceted cylinder. In some embodiments, the panels of the mesh can be arranged or positioned in any suitable manner, such as placed in a side-by-side orientation. In some embodiments, the panels can have any suitable rotational orientation which can be based an orientation of the cameras used to generate the video content.

[0063]   Note that, although the generated mesh is generally described herein as representing a portion of a faceted cylinder, in some embodiments, the generated mesh can have any suitable shape. For example, in some embodiments, one or more panels can be arranged in a top-to-bottom orientation, thereby adjusting a vertical field of view. As another example, in some embodiments, panels can be arranged to cause a particular amount of overlap between each adjacent panel.

[0064] At 404, process 400 can receive video content or any suitable video information. For example, in some embodiments, process 400 can receive video content from a group (e.g., two, three, and/or any other suitable number) of cameras generating CG content, as described above in connection with block 402. In some embodiments, a camera used to generate video content can be any suitable type of camera. For example, in some embodiments, a camera can be a CG camera that generates video content used in video games, animations, and/or any other suitable type of CG video content using any suitable computer graphics technique(s). Note that, in some embodiments, the video content can be a combination of different types of video content. For example, in some embodiments, process 400 can receive a portion of recorded video content and a portion of CG video content, such as an overlay, a graphic, text, and/or any other suitable type of CG content that is to be superimposed on the recorded video content. As a more particular example, in some embodiments, the portion of CG video can include visual effects, titles, credits, and/or any other suitable content that is to be overlaid on other video content.

[0065] In some embodiments, process 400 can receive the video content at any suitable resolution (e.g., 3840x2160, 1920x2160 per eye for stereoscopic content, and/or at any other suitable resolution). In some embodiments, the received video content can include any suitable depth information. For example, in some embodiments, received CG video content can include a depth map representing a distance of each pixel of the video content. As a more particular example, in some embodiments, the depth map can include different values for pixels of the video content, where each value represents a distance of content of the pixel. In particular, for each pixel of the video content the depth map includes a value representing a distance between a scene object captured by the pixel and the camera used to capture video content of the scene object. In some embodiments, the depth map can be used to render the video content stereoscopically.

[0066] Note that, in some embodiments, in instances where the video content is content rendered by a particular application (e.g., an application associated with a game or a gaming console, and/or any other suitable type of application), process 400 can record the rendered content using any suitable technique or combination of techniques. For example, in some embodiments, process 400 can record the video content using an Application Programming Interface (API) associated with the application.

[0067]   In some embodiments, process 400 can process or manipulate the received video content in any suitable manner. For example, in some embodiments, process 400 can replicate the received video content to generate video content to be rendered stereoscopically. As a more particular example, in some embodiments, process 400 can replicate the received video content to generate two versions of the video content offset by any suitable distance (e.g., 30 millimeters, 60 millimeters, and/or any other suitable offset), where a first version is to be delivered to a left eye of a viewer and a second version is to be delivered to a right eye of the viewer.

[0068]   Process 400 can store the video content in connection with the generated mesh at 406. For example, in some embodiments, process 400 can store the video content and the generated mesh on a content server, such as content server 202, to be delivered at any suitable time to a user device (e.g., in response to a request from the user device, as described below in connection with FIG. 5). In some embodiments, process 400 can store the video content in connection with the generated mesh in any suitable manner. For example, in some embodiments, any suitable data representing the mesh (e.g., locations of panels of the mesh, sizes of panels of the mesh, orientations of panels of the mesh, and/or any other suitable data) can be inserted into a video file associated with the video content.

[0069]   Note that, in some embodiments, the mesh generated at block 402 can be used by process 400 as an intermediate step to generating a different mesh to render video content using a spherical projection, such as a fish-eye projection, a cropped equirectangular projection, and/or any other suitable type of projection. For example, the mesh generated at block 402 can be converted into a cropped equirectangular mesh.

[0070]   Turning to FIG. 5, an illustrative example 500 of a process for rendering immersive video content is shown in accordance with some embodiments of the disclosed subject matter. In some embodiments, blocks of process 500 can be executed by any suitable device, such as a virtual reality headset, a game console, a mobile phone, a tablet computer, a television, and/or any other suitable type of user device.

[0071]   Process 500 can begin at 502 by requesting a video stream for presentation on a user device. In some embodiments, the video stream can be any suitable type of video stream, such as a video to be streamed to the user device from a content server, and/or any other suitable type of video stream. In some embodiments, the video stream can be requested on the user device in any suitable manner. For example, in some embodiments, the video stream can be

14

requested in response to a user of the user device selecting the video stream from a page indicating available video content, and/or in any other suitable manner. The request for a video stream may be transmitted from a user device to a content server (e.g., content server 202 as shown in and described above in connection with FIG. 2).

[0072]   Process 500 can receive the video stream and a corresponding mesh to be used to render the video stream at 504 in response to the request transmitted at 502. For example, in some embodiments, the video stream and the corresponding mesh can be received from a content server (e.g., content server 202 as shown in and described above in connection with FIG. 2) that stores the video content and/or that generated the mesh, as described above in connection with FIG. 4. In some embodiments, the video stream and the corresponding mesh can be received in any suitable manner. For example, in some embodiments, the mesh can be inserted into a file that includes the video stream, and process 500 can extract the mesh from the file using any suitable technique or combination of techniques. It should be noted that any suitable approach can be used to inject the corresponding mesh into the video stream.

[0073]   Process 500 can render the video stream on the user device as immersive content at 506 using any suitable technique or combination of techniques. For example, in some embodiments, process 500 can apply or project the video stream onto the mesh to render the video stream as immersive content. As a more particular example, in some embodiments, process 500 can use any suitable technique (e.g., UV mapping, and/or any other suitable technique(s)) for mapping content from the video stream to the mesh as a texture. As another example, in some embodiments, process 500 can render the video stream as stereoscopic content by presenting a first portion of the video stream to a left eye of a viewer of the user device and a second portion of the video stream to a right eye of the viewer of the user device. In some such embodiments, the first portion of the video stream and the second portion of the video stream can be separately applied or projected onto the mesh prior to rendering on the user device.

[0074]   In some embodiments, process 500 can render the video stream as immersive content that can be manipulated or interacted with by a viewer of the video stream. For example, in some embodiments, process 500 can render the video stream in a manner that can allow a viewer to change a viewpoint of the video stream. As a more particular example, in some embodiments, in instances where the viewer is using a virtual reality headset or other wearable computer, process 500 can change a presented viewpoint of the video stream in response to

determining that the viewer has changed an orientation of the viewer's head. As another more particular example, in some embodiments, process 500 can change a presented viewpoint of the video stream in response to determining that the viewer has input a gesture on the user device to manipulate the viewpoint, such as by selecting and dragging a user interface on which the video stream is presented.

[0075]   In some embodiments, at least some of the above described blocks of the processes of FIGS. 4 and 5 can be executed or performed in any order or sequence not limited to the order and sequence shown in and described in connection with the figures. Also, some of the above blocks of FIGS. 4 and 5 can be executed or performed substantially simultaneously where appropriate or in parallel to reduce latency and processing times. Additionally or alternatively, some of the above described blocks of the processes of FIGS. 4 and 5 can be omitted.

[0076]   In some embodiments, any suitable computer readable media can be used for storing instructions for performing the functions and/or processes herein. For example, in some embodiments, computer readable media can be transitory or non-transitory. For example, non-transitory computer readable media can include media such as non-transitory forms of magnetic media (such as hard disks, floppy disks, and/or any other suitable magnetic media), non-transitory forms of optical media (such as compact discs, digital video discs, Blu-ray discs, and/or any other suitable optical media), non-transitory forms of semiconductor media (such as flash memory, electrically programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), and/or any other suitable semiconductor media), any suitable media that is not fleeting or devoid of any semblance of permanence during transmission, and/or any suitable tangible media. As another example, transitory computer readable media can include signals on networks, in wires, conductors, optical fibers, circuits, any suitable media that is fleeting and devoid of any semblance of permanence during transmission, and/or any suitable intangible media.

[0077]   In situations in which the systems described herein collect personal information about users, or make use of personal information, the users may be provided with an opportunity to control whether programs or features collect user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location). In addition, certain data may be treated in one or more ways before it is stored or used, so that personal information is removed. For example, a user's identity may be treated so that no

personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over how information is collected about the user and used by a content server.

[0078]   Accordingly, methods, systems, and media for rendering immersive video content are provided.

[0079]   Although the invention has been described and illustrated in the foregoing illustrative embodiments, it is understood that the present disclosure has been made only by way of example, and that numerous changes in the details of implementation of the invention can be made without departing from the spirit and scope of the invention, which is limited only by the claims that follow. Features of the disclosed embodiments can be combined and rearranged in various ways.

What is claimed is:

1.   A method for generating immersive video content, comprising:

receiving information indicating positions of cameras in a plurality of cameras;

generating a mesh on which video content is to be projected based on the positions of the cameras in the plurality of cameras, wherein the mesh is comprised of a portion of a faceted cylinder, and wherein the faceted cylinder has a plurality of facets each corresponding to a projection from a camera in the plurality of cameras;

receiving video content corresponding to the plurality of cameras; and

transmitting the video content and the generated mesh to a user device in response to receiving a request for the video content from the user device.

2.   The method of claim 1, wherein generating the mesh comprises stitching the plurality of facets in a side by side orientation.

3.   The method of claim 1 or 2, further comprising inserting the generated mesh into a video file that includes the video content prior to transmitting the video file to the user device.

4.   The method of any preceding claim, wherein the video content includes a depth map.

5.   The method of any preceding claim, further comprising converting the generated mesh to a second type of mesh for projecting video content.

6.   The method of claim 5, wherein the second type of mesh is a fish-eye projection.

7.   The method of any preceding claim, further comprising:

receiving resulting image information associated with a rendered video content item; and

causing an indication to be presented to displace one or more of the plurality of cameras based on the information indicating the positions of cameras in the plurality of cameras.

8.   A system for generating immersive video content, the system comprising:

a memory; and

a hardware processor that, when executing computer-executable instructions stored in the memory, is configured to:

receive information indicating positions of cameras in a plurality of cameras;

generate a mesh on which video content is to be projected based on the positions of the cameras in the plurality of cameras, wherein the mesh is comprised of a portion of a faceted cylinder, and wherein the faceted cylinder has a plurality of facets each corresponding to a projection from a camera in the plurality of cameras;

receive video content corresponding to the plurality of cameras; and

transmit the video content and the generated mesh to a user device in response to receiving a request for the video content from the user device.

9.   The system of claim 8, wherein the hardware processor is configured to stitch the plurality of facets in a side by side orientation to generating the mesh.

10.   The system of claim 8 or 9, wherein the hardware processor is further configured to insert the generated mesh into a video file that includes the video content prior to transmitting the video file to the user device.

11.   The system of any of claims 8 to 10, wherein the video content includes a depth map.

12.   The system of any of claims 8 to 11, wherein the hardware processor is further configured to convert the generated mesh to a second type of mesh for projecting video content.

13.   The system of claim 12, wherein the second type of mesh is a fish-eye projection.

14.   The system of any of claims 8 to 13, wherein the hardware processor is further configured to:

receive resulting image information associated with a rendered video content item; and

cause an indication to be presented to displace one or more of the plurality of cameras based on the information indicating the positions of cameras in the plurality of cameras.

15.   A non-transitory computer-readable medium containing computer executable instructions that, when executed by a processor, cause the processor to perform a method for generating immersive video content, the method comprising:

receiving information indicating positions of cameras in a plurality of cameras;

generating a mesh on which video content is to be projected based on the positions of the cameras in the plurality of cameras, wherein the mesh is comprised of a portion of a faceted cylinder, and wherein the faceted cylinder has a plurality of facets each corresponding to a projection from a camera in the plurality of cameras;

receiving video content corresponding to the plurality of cameras; and

transmitting the video content and the generated mesh to a user device in response to receiving a request for the video content from the user device.

16.   The non-transitory computer-readable medium of claim 15, wherein generating the mesh comprises stitching the plurality of facets in a side by side orientation.

17.   The non-transitory computer-readable medium of claim 15 or 16, wherein the method further comprises inserting the generated mesh into a video file that includes the video content prior to transmitting the video file to the user device.

18.   The non-transitory computer-readable medium of any of claims 15 to 17, wherein the video content includes a depth map.

19.   The non-transitory computer-readable medium of any of claims 15 to 18, wherein the method further comprises converting the generated mesh to a second type of mesh for projecting video content.

20.   The non-transitory computer-readable medium of claim 19, wherein the second type of mesh is a fish-eye projection.

21.   The non-transitory computer-readable medium of any of claims 15 to 20, wherein the method further comprises:

receiving resulting image information associated with a rendered video content item; and

causing an indication to be presented to displace one or more of the plurality of cameras based on the information indicating the positions of cameras in the plurality of cameras.
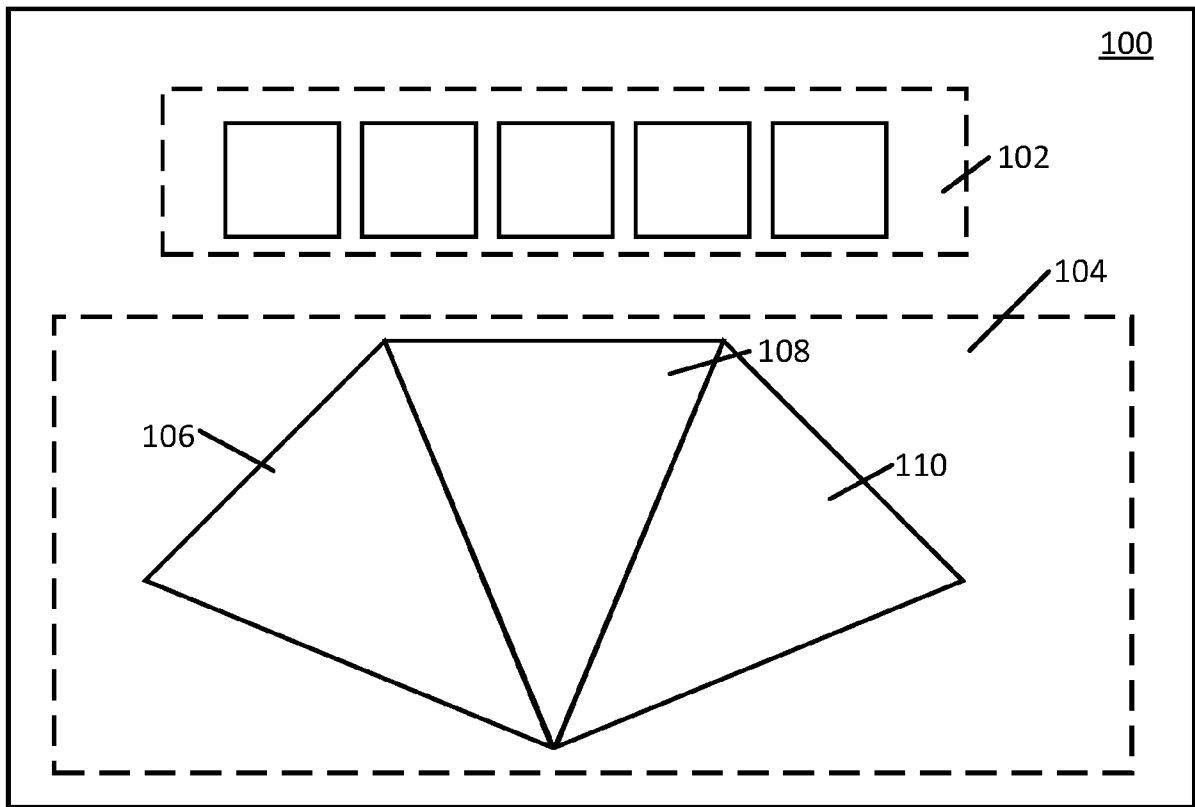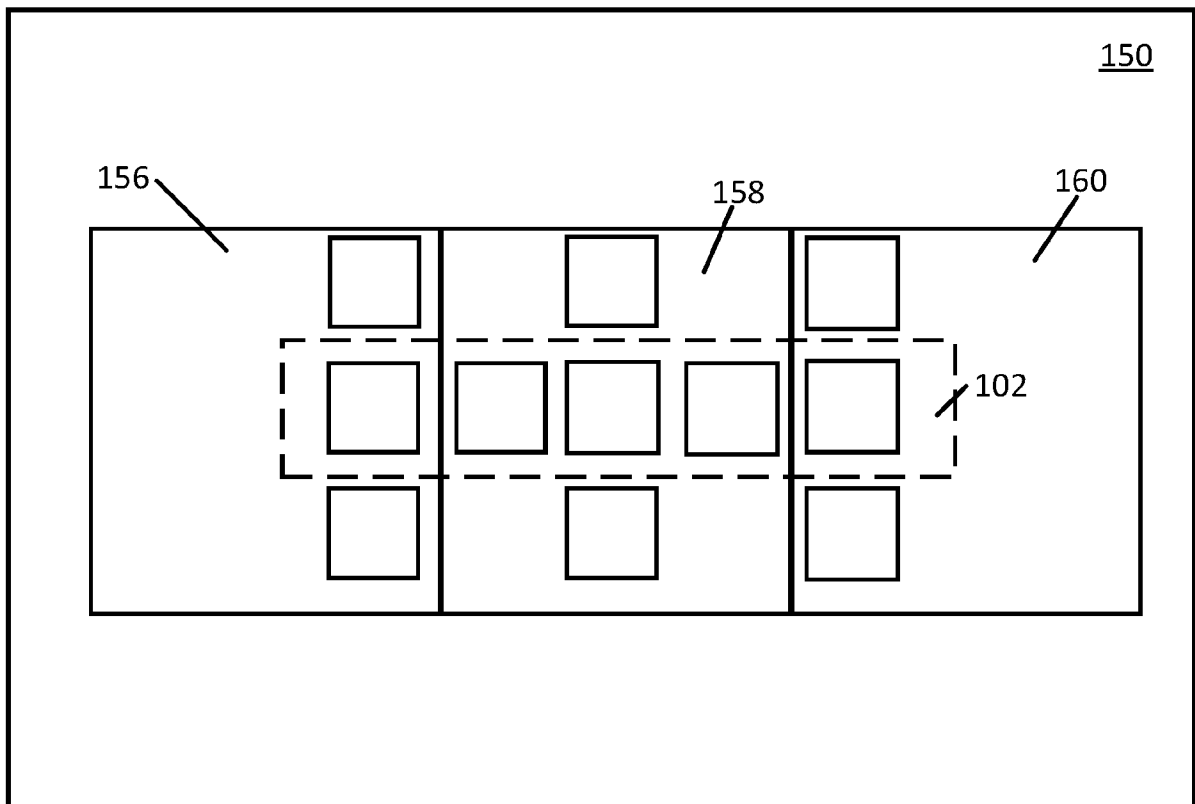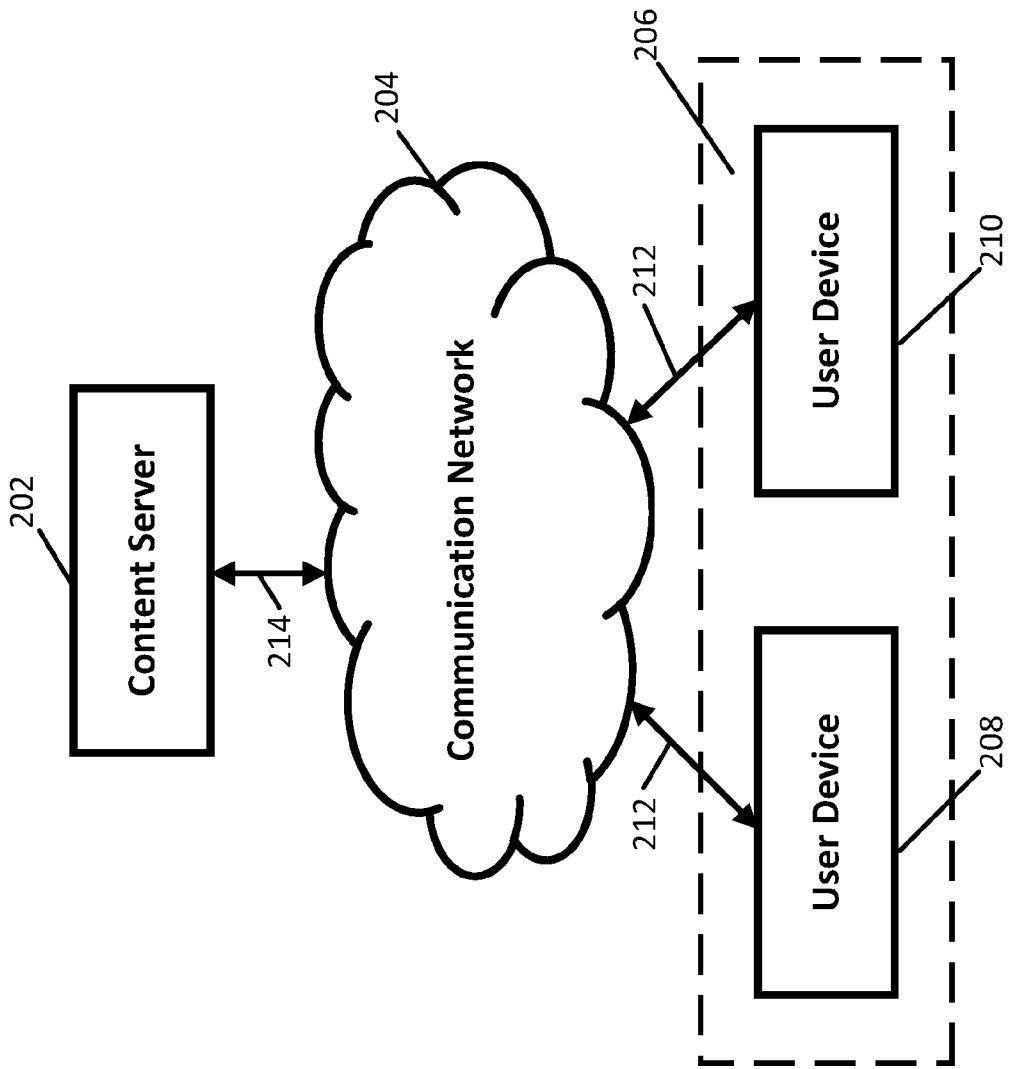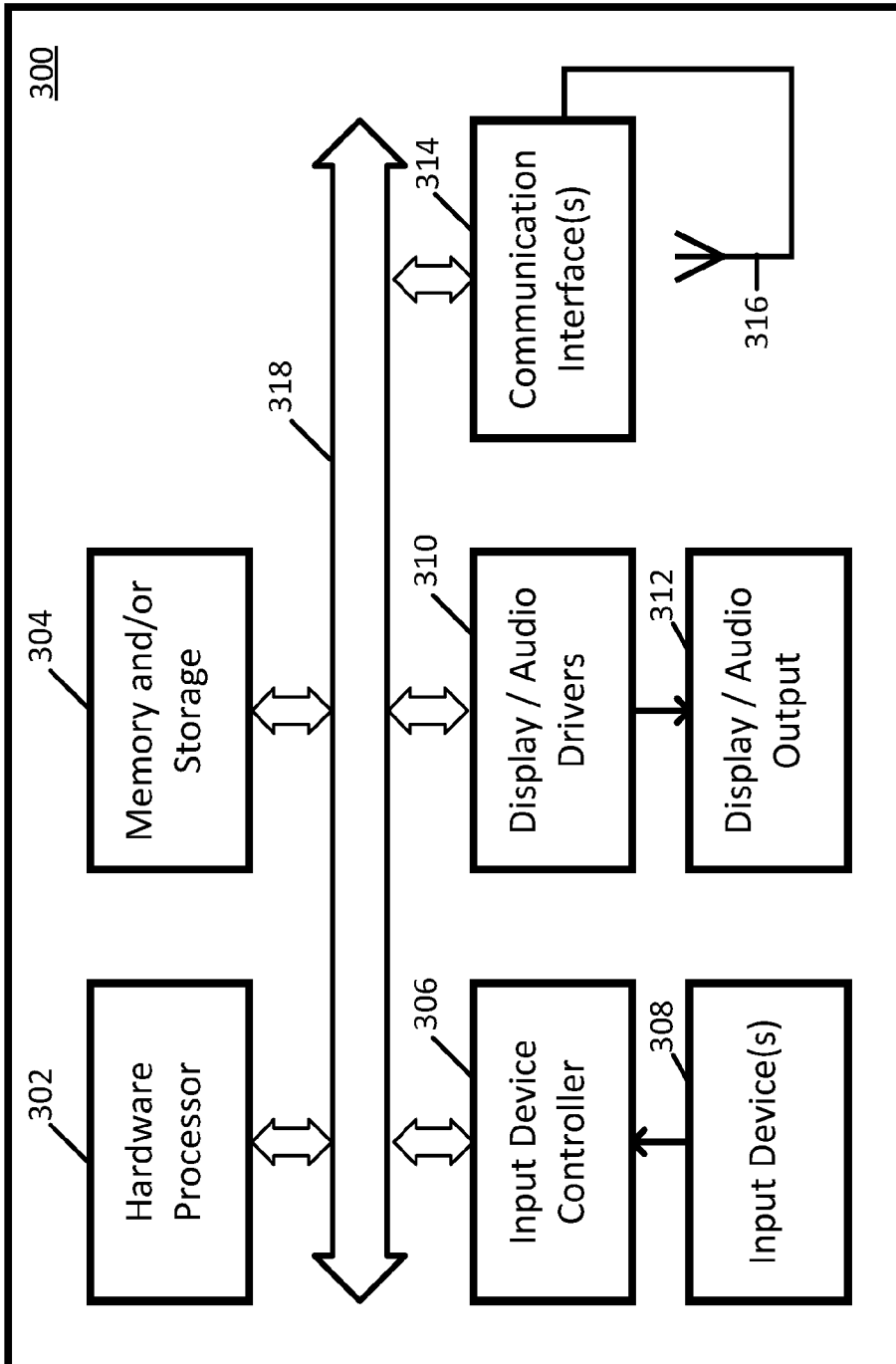
FIG. 1A

FIG. 1B

200
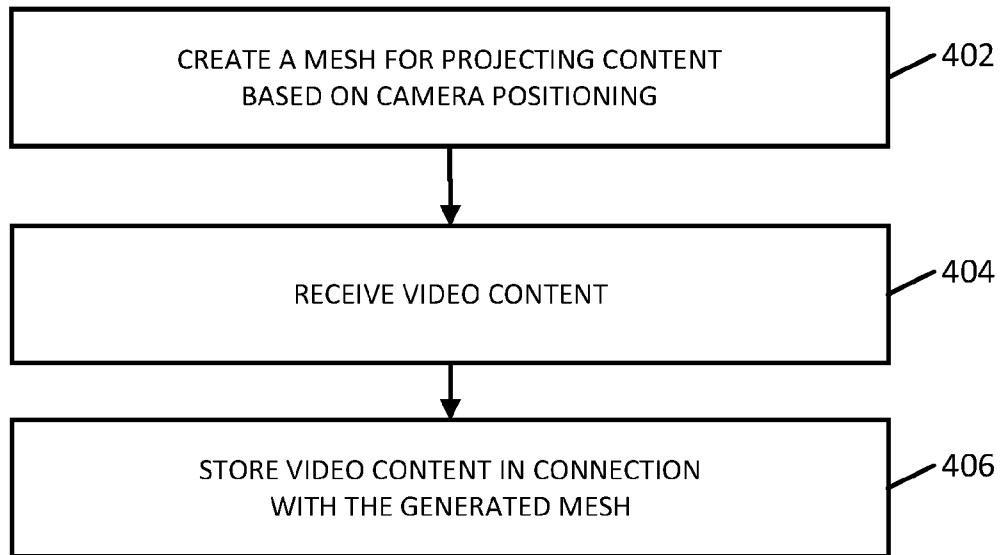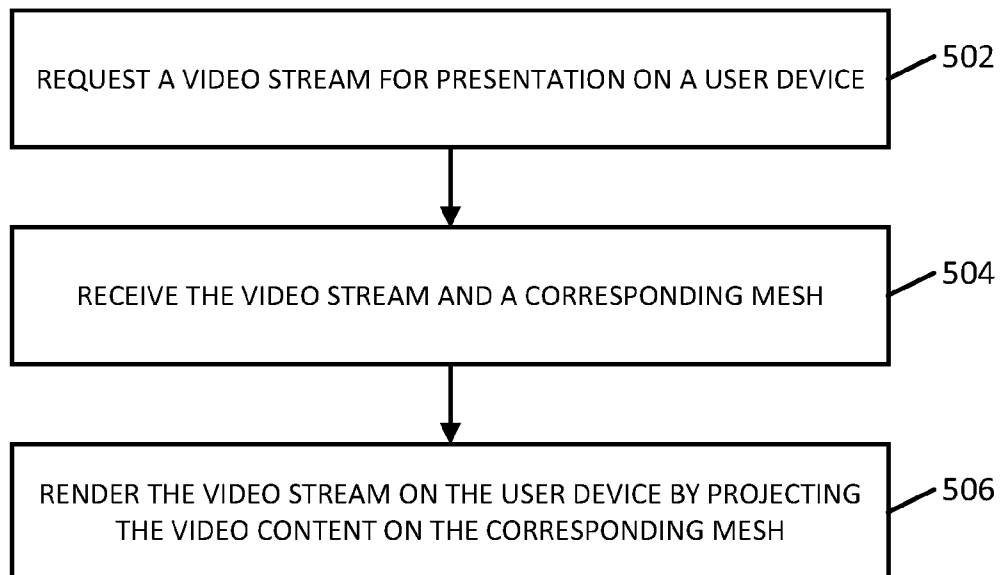
Content Server
202

214

Communication Network
204

212

212

User Device
210
206

User Device
208

FIG. 2

FIG. 3

400

CREATE A MESH FOR PROJECTING CONTENT
BASED ON CAMERA POSITIONING — 402

RECEIVE VIDEO CONTENT — 404

STORE VIDEO CONTENT IN CONNECTION
WITH THE GENERATED MESH — 406

FIG. 4

REQUEST A VIDEO STREAM FOR PRESENTATION ON A USER DEVICE — 502

RECEIVE THE VIDEO STREAM AND A CORRESPONDING MESH — 504

RENDER THE VIDEO STREAM ON THE USER DEVICE BY PROJECTING
THE VIDEO CONTENT ON THE CORRESPONDING MESH — 506

FIG. 5

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

INV. H04N13/194   G06T17/20   H04N13/282   H04N13/296   H04N13/271
     H04N5/232   H04N21/2343   H04N13/279
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N   G06T

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MCCUTCHEN D ET AL: "3DAV Exploration of Video-Based Rendering Technology in MPEG", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, US, vol. 14, no. 3, 1 March 2004 (2004-03-01), pages 348-356, XP011108800, ISSN: 1051-8215, DOI: 10.1109/TCSVT.2004.823395 the whole document -----	1-21
X	US 9 573 062 B1 (LONG JIEYI [US] ET AL) 21 February 2017 (2017-02-21) column 3, lines 36-47 column 8, lines 1-27 column 15, line 16 - column 16, line 25 column 17, lines 8-30 column 18, lines 1-12 column 20, line 58 - column 21, line 8 -----	1-21

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

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 December 2018	18/12/2018

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

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report			Publication date	Patent family member(s)	Publication date
US 9573062	B1		21-02-2017	NONE	