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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 February 2017 (23.02.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/030865 A1

(51) International Patent Classification:
H04N 21/435 (2011.01) *H04N 21/6373* (2011.01)
H04N 21/442 (2011.01) *H04N 21/845* (2011.01)
H04N 21/4728 (2011.01)

(21) International Application Number:
PCT/US2016/046317

(22) International Filing Date:
10 August 2016 (10.08.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/205,492 14 August 2015 (14.08.2015) US

(71) Applicant: **VID SCALE, INC.** [US/US]; 200 Bellevue Parkway, Suite 300, Wilmington, Delaware 19809 (US).

(72) Inventors: **RAMASWAMY, Kumar**; 71 Sayre Drive, Princeton, New Jersey 08540 (US). **COOPER, Jeffrey Allen**; 11 Toth Lane, Rocky Hill, New Jersey 08553 (US).

(74) Agent: **IRVINE, Robert J.**; Invention Mine LLC, 216 South Jefferson Street, Suite 102, Chicago, Illinois 60661 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: METHOD AND SYSTEMS FOR DISPLAYING A PORTION OF A VIDEO STREAM

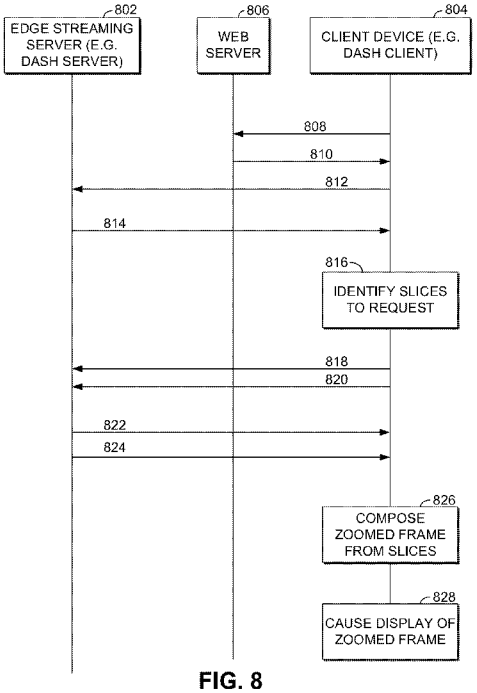


FIG. 8

(57) **Abstract**: Systems and methods are described to enable video clients to zoom in to a region or object of interest without substantial loss of resolution. In an exemplary method, a server transmits a manifest, such as a DASH MPD, to a client device. The manifest identifies a plurality of sub-streams, where each sub-stream represents a respective spatial portion of a source video. The manifest also includes information associating an object of interest with a plurality of the spatial portions. To view high-quality zoomed video, the client requests the sub-streams that are associated with the object of interest and renders the requested sub-streams. In some embodiments, a render point is transmitted to the client to allow for motion of the object of interest.

WO 2017/030865 A1

METHOD AND SYSTEMS FOR DISPLAYING A PORTION OF A VIDEO STREAM

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application is a non-provisional filing of, and claims the benefit under 35 U.S.C. §119(e) from, U.S. Provisional Patent Application Serial No. 62/205,492, filed August 14, 2015 and entitled "METHOD AND SYSTEMS FOR DISPLAYING A PORTION OF A VIDEO STREAM", the entirety of which is incorporated herein by reference.

BACKGROUND

[0002] Digital video signals are commonly characterized by its parameters of i) resolution (luma and chroma resolution or horizontal and vertical pixel dimensions), ii) frame rate, and iii) dynamic range or bit depth (bits per pixel). The resolution of the digital video signals has increased from Standard Definition (SD) through 8K-Ultra High Definition (UHD). The other digital video signal parameters have also increased from 30 frames per second (fps) up to 240 fps, and the bit depth has increased from 8 bit to 10 bit. To transmit digital video signals over a network, MPEG/ITU standardized video compression has undergone several generations of successive improvements in compression efficiency, including MPEG2, MPEG4/H.264, and HEVC/H.265. The technology to display the digital video signals on a consumer device, such as a television or mobile phone, has also increased correspondingly.

[0003] Consumers requesting higher quality digital video on network-connected devices face more bandwidth constraints from the video content delivery network. In an effort to mitigate the effects of bandwidth constraints, several solutions have emerged. Video content is initially captured at a higher resolution, frame rate, and dynamic range. For example, a 4:2:2, 10 bit HD video content is often down-resolved to 4:2:0, 8 bit for distribution. The digital video is encoded and stored at multiple resolutions at a server, and these versions at varying resolutions are made available for retrieval, decoding and rendering by clients with possibly varying capabilities. The digital video gets encoded and stored at multiple resolutions at a server. Adaptive bit rate (ABR) further addresses network congestion. In ABR, a digital video is encoded at multiple bit rates (e.g. choosing the same or multiple lower resolutions, lower frame rates, etc.) and is made available at a server. The client device requests a different bit rate for consumption at periodic intervals based on its calculated available network bandwidth or local computing resources.

SUMMARY

[0004] Described herein are systems and methods related to displaying a portion of a digital video stream. The digital video is encoded and stored at multiple resolutions at a server, and these

versions at varying resolutions are made available for retrieval, decoding and rendering by clients with possibly varying capabilities. The server may make available additional metadata so that clients may request and receive data sufficient to decode and render one or more objects or areas of interest at a high resolution and/or a zoomed scale, where the spatial support for the objects or areas of interest may vary in time.

[0005] In an exemplary method, a server transmits a manifest, such as a DASH MPD, to a client device. The manifest identifies at least one unzoomed stream representing an unzoomed version of a source video. The manifest further identifies a plurality of sub-streams, where each sub-stream represents a respective spatial portion of the source video. The server also transmits, to the client device, information associating at least one object of interest with a plurality of the spatial portions. This information may be provided in the manifest. The server receives, from the client device, a request for at least one of the sub-streams. In response, the server transmits the requested sub-streams to the client device. The sub-streams may be encoded at a higher resolution than the unzoomed stream, allowing for higher-quality video when a client device zooms in on an object of interest represented in the sub-streams.

[0006] The information that associates the at least one object of interest with a plurality of the spatial portions may be provided by including, in the manifest, a syntax element for each sub-stream that identifies at least one object of interest associated with the respective sub-stream.

[0007] In some embodiments, the server also transmits to the client a render point for the object of interest. In instances where the object of interest encompasses less than the entirety of the sub-streams, the render point may be used to indicate which portions of the sub-streams are to be displayed. For example, the render point may represent coordinates of one or more corners of a rectangular region of interest, where the rectangular region of interest is smaller than a complete region represented by all of the sub-streams. The rectangular region of interest is displayed, while portions of the sub-streams that are outside of the rectangular region of interest may not be displayed.

[0008] The rendering reference points may be communicated to the client device. For example, rendering reference points may be transmitted in-band as part of the video streams or video segments, or as side information sent along with the video streams or video segments. One or more rendering reference points may be transmitted in-band in a video stream, such as in an unzoomed stream or in one or more sub-streams. Alternately the rendering reference points may be specified in an out-of-band communication (e.g. as metadata in a manifest such as a DASH MPD).

[0009] In some embodiments, the sub-streams are encoded for adaptive bit rate (ABR) streaming, for example with at least two sub-streams with different bitrates being available for at

least some of the spatial portions. The client may select which sub-stream to request based on network conditions.

[0010] In an exemplary client-side method, a video client receives a manifest, where the manifest identifies an unzoomed stream representing an unzoomed version of a source video. The manifest also identifies a plurality of sub-streams, where each sub-stream represents a respective spatial portion of the source video. The client further receives information associating at least one object of interest with a plurality of the spatial portions. The client device receives a selection (e.g. a user selection entered through a user interface device such as a remote control) of one of the objects of interest. The client device identifies the spatial portions associated with the selected object of interest and retrieves a representative sub-stream for each of the spatial portions. Where there is more than one representative sub-stream for a spatial portion (e.g. with different bitrates), the client device may select which of the representative sub-streams to retrieve based on network conditions. The client device then causes display of a zoomed version of the object of interest by rendering the retrieved sub-streams. The display of the zoomed version may be provided by the client device itself (e.g. on a built-in screen), or the client device may transmit uncompressed video to an external display device (such as a television or monitor).

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] A more detailed understanding may be had from the following description, presented by way of example in conjunction with the accompanying drawings.

[0012] FIG. 1A depicts an example communications system in which one or more disclosed embodiments may be implemented.

[0013] FIG. 1B depicts an example client device that may be used within the communications system of FIG. 1A.

[0014] FIG. 1C depicts an example video encoding and distribution system.

[0015] FIG. 1D depicts example screen resolutions.

[0016] FIG. 1E schematically depicts ABR encoding.

[0017] FIG. 1F depicts an example network entity 190, that may be used within the communication system 100 of FIG. 1A.

[0018] FIG. 2 depicts an example video coding system and distribution system, according to an embodiment.

[0019] FIG. 3 depicts example coding resolutions, in accordance with an embodiment.

[0020] FIG. 4A depicts an example of a video zoom operation, in accordance with an embodiment.

[0021] FIG. 4B depicts a second example of video zoom operation, in accordance with an embodiment.

[0022] FIG. 5 depicts an example of a digital video with an object of interest, in accordance with an embodiment.

[0023] FIG. 6 is a message sequence diagram depicting encoding and delivery of content to a client in accordance with an embodiment.

[0024] FIG. 7 is a message sequence diagram depicting a second example of encoding and delivery of content to a client in accordance with an embodiment.

[0025] FIG. 8 is a message sequence diagram depicting an example communications process, in accordance with an embodiment.

[0026] FIG. 9 illustrates a video having a plurality of spatial portions, at least some of the spatial portions having associated sub-streams to enable zoomed display of an object of interest.

DETAILED DESCRIPTION

[0027] A detailed description of illustrative embodiments will now be provided with reference to the various Figures. Although this description provides detailed examples of possible implementations, it should be noted that the provided details are intended to be by way of example and in no way limit the scope of the application. The systems and methods relating to video compression may be used with the wired and wireless communication systems described with respect to FIGS. 1A-1C. As an initial matter, these wired and wireless systems will be described.

[0028] FIG. 1A is a diagram of an example communications system 100 in which one or more disclosed embodiments may be implemented. The communications system 100 may be a multiple access system that provides content, such as voice, data, video, messaging, broadcast, and the like, to multiple wireless users. The communications system 100 may enable multiple wired and wireless users to access such content through the sharing of system resources, including wired and wireless bandwidth. For example, the communications systems 100 may employ one or more channel-access methods, such as code division multiple access (CDMA), time division multiple access (TDMA), frequency division multiple access (FDMA), orthogonal FDMA (OFDMA), single-carrier FDMA (SC-FDMA), and the like. The communications systems 100 may also employ one or more wired communications standards (e.g.: Ethernet, DSL, radio frequency (RF) over coaxial cable, fiber optics, and the like.

[0029] As shown in FIG. 1A, the communications system 100 may include client devices 102a, 102b, 102c, and/or 102d, Radio Access Networks (RAN) 103/104/105, a core network 106/107/109, a public switched telephone network (PSTN) 108, the Internet 110, and other networks 112, and communication links 115/116/117, and 119, though it will be appreciated that

the disclosed embodiments contemplate any number of client devices, base stations, networks, and/or network elements. Each of the client devices 102a, 102b, 102c, 102d may be any type of device configured to operate and/or communicate in a wired or wireless environment. By way of example, the client device 102a is depicted as a tablet computer, the client device 102b is depicted as a smart phone, the client device 102c is depicted as a computer, and the client device 102d is depicted as a television.

[0030] The communications systems 100 may also include a base station 114a and a base station 114b. Each of the base stations 114a, 114b may be any type of device configured to wirelessly interface with at least one of the WTRUs 102a, 102b, 102c, 102d to facilitate access to one or more communication networks, such as the core network 106/107/109, the Internet 110, and/or the networks 112. By way of example, the base stations 114a, 114b may be a base transceiver station (BTS), a Node-B, an eNode B, a Home Node B, a Home eNode B, a site controller, an access point (AP), a wireless router, and the like. While the base stations 114a, 114b are each depicted as a single element, it will be appreciated that the base stations 114a, 114b may include any number of interconnected base stations and/or network elements.

[0031] The base station 114a may be part of the RAN 103/104/105, which may also include other base stations and/or network elements (not shown), such as a base station controller (BSC), a radio network controller (RNC), relay nodes, and the like. The base station 114a and/or the base station 114b may be configured to transmit and/or receive wireless signals within a particular geographic region, which may be referred to as a cell (not shown). The cell may further be divided into sectors. For example, the cell associated with the base station 114a may be divided into three sectors. Thus, in one embodiment, the base station 114a may include three transceivers, i.e., one for each sector of the cell. In another embodiment, the base station 114a may employ multiple-input multiple output (MIMO) technology and, therefore, may utilize multiple transceivers for each sector of the cell.

[0032] The base stations 114a, 114b may communicate with one or more of the client devices 102a, 102b, 102c, and 102d over an air interface 115/116/117, or communication link 119, which may be any suitable wired or wireless communication link (e.g., radio frequency (RF), microwave, infrared (IR), ultraviolet (UV), visible light, and the like). The air interface 115/116/117 may be established using any suitable radio access technology (RAT).

[0033] More specifically, as noted above, the communications system 100 may be a multiple access system and may employ one or more channel-access schemes, such as CDMA, TDMA, FDMA, OFDMA, SC-FDMA, and the like. For example, the base station 114a in the RAN 103/104/105 and the client devices 102a, 102b, 102c may implement a radio technology such as

Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access (UTRA), which may establish the air interface 115/116/117 using wideband CDMA (WCDMA). WCDMA may include communication protocols such as High-Speed Packet Access (HSPA) and/or Evolved HSPA (HSPA+). HSPA may include High-Speed Downlink Packet Access (HSDPA) and/or High-Speed Uplink Packet Access (HSUPA).

[0034] In another embodiment, the base station 114a and the client devices 102a, 102b, 102c may implement a radio technology such as Evolved UMTS Terrestrial Radio Access (E-UTRA), which may establish the air interface 115/116/117 using Long Term Evolution (LTE) and/or LTE-Advanced (LTE-A).

[0035] In other embodiments, the base station 114a and the client devices 102a, 102b, 102c may implement radio technologies such as IEEE 802.16 (i.e., Worldwide Interoperability for Microwave Access (WiMAX)), CDMA2000, CDMA2000 1X, CDMA2000 EV-DO, Interim Standard 2000 (IS-2000), Interim Standard 95 (IS-95), Interim Standard 856 (IS-856), Global System for Mobile communications (GSM), Enhanced Data rates for GSM Evolution (EDGE), GSM EDGE (GERAN), and the like.

[0036] The base station 114b in FIG. 1A may be a wired router, a wireless router, Home Node B, Home eNode B, or access point, as examples, and may utilize any suitable wired transmission standard or RAT for facilitating wireless connectivity in a localized area, such as a place of business, a home, a vehicle, a campus, and the like. In one embodiment, the base station 114b and the client devices 102c, 102d may implement a radio technology such as IEEE 802.11 to establish a wireless local area network (WLAN). In another embodiment, the base station 114b and the client devices 102c, 102d may implement a radio technology such as IEEE 802.15 to establish a wireless personal area network (WPAN). In yet another embodiment, the base station 114b and the client devices 102c, 102d may utilize a cellular-based RAT (e.g., WCDMA, CDMA2000, GSM, LTE, LTE-A, and the like) to establish a picocell or femtocell. In yet another embodiment, the base station 114b communicates with client devices 102a, 102b, 102c, and 102d through communication links 119. As shown in FIG. 1A, the base station 114b may have a direct connection to the Internet 110. Thus, the base station 114b may not be required to access the Internet 110 via the core network 106/107/109.

[0037] The RAN 103/104/105 may be in communication with the core network 106/107/109, which may be any type of network configured to provide voice, data, applications, and/or voice over internet protocol (VoIP) services to one or more of the client devices 102a, 102b, 102c, 102d. As examples, the core network 106/107/109 may provide call control, billing services, mobile location-based services, pre-paid calling, Internet connectivity, video distribution, and the like,

and/or perform high-level security functions, such as user authentication. Although not shown in FIG. 1A, it will be appreciated that the RAN 103/104/105 and/or the core network 106/107/109 may be in direct or indirect communication with other RANs that employ the same RAT as the RAN 103/104/105 or a different RAT. For example, in addition to being connected to the RAN 103/104/105, which may be utilizing an E-UTRA radio technology, the core network 106/107/109 may also be in communication with another RAN (not shown) employing a GSM radio technology.

[0038] The core network 106/107/109 may also serve as a gateway for the client devices 102a, 102b, 102c, 102d to access the PSTN 108, the Internet 110, and/or other networks 112. The PSTN 108 may include circuit-switched telephone networks that provide plain old telephone service (POTS). The Internet 110 may include a global system of interconnected computer networks and devices that use common communication protocols, such as the transmission control protocol (TCP), user datagram protocol (UDP) and IP in the TCP/IP Internet protocol suite. The networks 112 may include wired and/or wireless communications networks owned and/or operated by other service providers. For example, the networks 112 may include another core network connected to one or more RANs, which may employ the same RAT as the RAN 103/104/105 or a different RAT.

[0039] Some or all of the client devices 102a, 102b, 102c, 102d in the communications system 100 may include multi-mode capabilities, i.e., the client devices 102a, 102b, 102c, 102d may include multiple transceivers for communicating with different wired or wireless networks over different communication links. For example, the WTRU 102c shown in FIG. 1A may be configured to communicate with the base station 114a, which may employ a cellular-based radio technology, and with the base station 114b, which may employ an IEEE 802 radio technology.

[0040] FIG. 1B depicts an example client device that may be used within the communications system of FIG. 1A. In particular, FIG. 1B is a system diagram of an example client device 102. As shown in FIG. 1B, the client device 102 may include a processor 118, a transceiver 120, a transmit/receive element 122, a speaker/microphone 124, a keypad 126, a display/touchpad 128, a non-removable memory 130, a removable memory 132, a power source 134, a global positioning system (GPS) chipset 136, and other peripherals 138. It will be appreciated that the client device 102 may represent any of the client devices 102a, 102b, 102c, and 102d, and include any sub-combination of the foregoing elements while remaining consistent with an embodiment. Also, embodiments contemplate that the base stations 114a and 114b, and/or the nodes that base stations 114a and 114b may represent, such as but not limited to transceiver station (BTS), a Node-B, a site controller, an access point (AP), a home node-B, an evolved home node-B (eNodeB), a home

evolved node-B (HeNB), a home evolved node-B gateway, and proxy nodes, among others, may include some or all of the elements depicted in FIG. 1B and described herein.

[0041] The processor 118 may be a general purpose processor, a special purpose processor, a conventional 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. The processor 118 may perform signal coding, data processing, power control, input/output processing, and/or any other functionality that enables the client device 102 to operate in a wired or wireless environment. The processor 118 may be coupled to the transceiver 120, which may be coupled to the transmit/receive element 122. While FIG. 1B depicts the processor 118 and the transceiver 120 as separate components, it will be appreciated that the processor 118 and the transceiver 120 may be integrated together in an electronic package or chip.

[0042] The transmit/receive element 122 may be configured to transmit signals to, or receive signals from, a base station (e.g., the base station 114a) over the air interface 115/116/117 or communication link 119. For example, in one embodiment, the transmit/receive element 122 may be an antenna configured to transmit and/or receive RF signals. In another embodiment, the transmit/receive element 122 may be an emitter/detector configured to transmit and/or receive IR, UV, or visible light signals, as examples. In yet another embodiment, the transmit/receive element 122 may be configured to transmit and receive both RF and light signals. In yet another embodiment, the transmit/receive element may be a wired communication port, such as an Ethernet port. It will be appreciated that the transmit/receive element 122 may be configured to transmit and/or receive any combination of wired or wireless signals.

[0043] In addition, although the transmit/receive element 122 is depicted in FIG. 1B as a single element, the client device 102 may include any number of transmit/receive elements 122. More specifically, the client device 102 may employ MIMO technology. Thus, in one embodiment, the WTRU 102 may include two or more transmit/receive elements 122 (e.g., multiple antennas) for transmitting and receiving wireless signals over the air interface 115/116/117.

[0044] The transceiver 120 may be configured to modulate the signals that are to be transmitted by the transmit/receive element 122 and to demodulate the signals that are received by the transmit/receive element 122. As noted above, the client device 102 may have multi-mode capabilities. Thus, the transceiver 120 may include multiple transceivers for enabling the client device 102 to communicate via multiple RATs, such as UTRA and IEEE 802.11, as examples.

[0045] The processor 118 of the client device 102 may be coupled to, and may receive user input data from, the speaker/microphone 124, the keypad 126, and/or the display/touchpad 128 (e.g., a liquid crystal display (LCD) display unit or organic light-emitting diode (OLED) display unit). The processor 118 may also output user data to the speaker/microphone 124, the keypad 126, and/or the display/touchpad 128. In addition, the processor 118 may access information from, and store data in, any type of suitable memory, such as the non-removable memory 130 and/or the removable memory 132. The non-removable memory 130 may include random-access memory (RAM), read-only memory (ROM), a hard disk, or any other type of memory storage device. The removable memory 132 may include a subscriber identity module (SIM) card, a memory stick, a secure digital (SD) memory card, and the like. In other embodiments, the processor 118 may access information from, and store data in, memory that is not physically located on the client device 102, such as on a server or a home computer (not shown).

[0046] The processor 118 may receive power from the power source 134, and may be configured to distribute and/or control the power to the other components in the client device 102. The power source 134 may be any suitable device for powering the WTRU 102. As examples, the power source 134 may include one or more dry cell batteries (e.g., nickel-cadmium (NiCd), nickel-zinc (NiZn), nickel metal hydride (NiMH), lithium-ion (Li-ion), and the like), solar cells, fuel cells, a wall outlet and the like.

[0047] The processor 118 may also be coupled to the GPS chipset 136, which may be configured to provide location information (e.g., longitude and latitude) regarding the current location of the client device 102. In addition to, or in lieu of, the information from the GPS chipset 136, the WTRU 102 may receive location information over the air interface 115/116/117 from a base station (e.g., base stations 114a, 114b) and/or determine its location based on the timing of the signals being received from two or more nearby base stations. It will be appreciated that the client device 102 may acquire location information by way of any suitable location-determination method while remaining consistent with an embodiment. In accordance with an embodiment, the client device 102 does not comprise a GPS chipset and does not acquire location information.

[0048] The processor 118 may further be coupled to other peripherals 138, which may include one or more software and/or hardware modules that provide additional features, functionality and/or wired or wireless connectivity. For example, the peripherals 138 may include an accelerometer, an e-compass, a satellite transceiver, a digital camera (for photographs or video), a universal serial bus (USB) port, a vibration device, a television transceiver, a hands free headset, a Bluetooth® module, a frequency modulated (FM) radio unit, a digital music player, a media player, a video game player module, an Internet browser, and the like.

[0049] FIG. 1C depicts an example video encoding system, in accordance with an embodiment. In particular, FIG. 1C depicts the example system 160. The example system 160 includes a full resolution input video source 162, an adaptive bitrate encoder 164, a streaming server 166, a network 168, and client devices 169. The example system 160 may be implemented in the context of the example communication system 100 depicted in FIG. 1A. For example, both the adaptive bitrate encoder 164 and the streaming server 168 may be entities in any of the networks depicted in the communication system 100. The client devices 169 may be the client devices 102a-d depicted in the communication system 100.

[0050] In accordance with an embodiment, the adaptive bitrate encoder or transcoder 164 receives an uncompressed or compressed input video stream from source 162 and encodes or transcodes the video stream into a plurality of representations 165. Each of the representations may differ from the others in a property such as resolution, frame rate, bit rate, and the like. The adaptive bitrate encoder 164 communicates the encoded video streams 165 to the streaming server 166. The streaming server 166 transmits an encoded video stream via the network to the client devices. The transmission may take place over any of the communication interfaces, such as the communication link 115/116/117 or 119.

[0051] FIG. 1D provides an illustration 170 of different image resolutions. The example image resolutions, listed from lowest resolution to highest resolution, include standard definition (SD), full high definition (FHD), 4K Ultra High Definition (UHD), and 8K UHD, although other resolutions may also be available. As an aid to understanding the present disclosure, higher-resolution video is generally illustrated herein using a larger rectangle (as if each pixel were of equal size), though it should be understood that display size is not necessarily correlated with image resolution.

[0052] FIG. 1E provides a schematic illustration of ABR encoding. As illustrated in FIG. 1E, a 4K UHD source video is converted to three other encodings with three different resolutions. For example, the source video may be downconverted to a stream ABR-1 (182), which may be, for example, a 1080p HD video; a stream ABR-2 (184), which may be, for example, a standard definition (SD) stream; and a stream ABR-3 (186), which may be a still lower-resolution stream (e.g. for use under conditions of network congestion). Each of the ABR encoded versions of the source video are transmitted to the streaming server for further transmission to client devices based in part on client device capability and network congestion. Thus, the highest spatial resolution that is available is not always delivered to the client devices.

[0053] FIG. 1F depicts an example network entity 190 that may be used within the communication system 100 of FIG. 1A. As depicted in FIG. 1F, network entity 190 includes a

communication interface 192, a processor 194, and non-transitory data storage 196, all of which are communicatively linked by a bus, network, or another communication path 198.

[0054] Communication interface 192 may include one or more wired communication interfaces and/or one or more wireless-communication interfaces. With respect to wired communication, communication interface 192 may include one or more interfaces such as Ethernet interfaces, as an example. With respect to wireless communication, communication interface 192 may include components such as one or more antennae, one or more transceivers/chipsets designed and configured for one or more types of wireless (e.g., LTE) communication, and/or any other components deemed suitable by those of skill in the relevant art. And further with respect to wireless communication, communication interface 192 may be equipped at a scale and with a configuration appropriate for acting on the network side—as opposed to the client side—of wireless communications (e.g., LTE communications, Wi-Fi communications, and the like). Thus, communication interface 192 may include the appropriate equipment and circuitry (perhaps including multiple transceivers) for serving multiple mobile stations, UEs, or other access terminals in a coverage area.

[0055] Processor 194 may include one or more processors of any type deemed suitable by those of skill in the relevant art, some examples including a general-purpose microprocessor and a dedicated DSP.

[0056] Data storage 196 may take the form of any non-transitory computer-readable medium or combination of such media, some examples including flash memory, read-only memory (ROM), and random-access memory (RAM) to name but a few, as any one or more types of non-transitory data storage deemed suitable by those of skill in the relevant art could be used. As depicted in FIG. 1F, data storage 196 contains program instructions 197 executable by processor 194 for carrying out various combinations of the various network-entity functions described herein.

[0057] In some embodiments, the network-entity functions described herein are carried out by a network entity having a structure similar to that of network entity 190 of FIG. 1F. In some embodiments, one or more of such functions are carried out by a set of multiple network entities in combination, where each network entity has a structure similar to that of network entity 190 of FIG. 1F. In various different embodiments, network entity 190 is—or at least includes—one or more of the encoders, one or more of (one or more entities in) RAN 103, (one or more entities in) RAN 104, (one or more entities in) RAN 105, (one or more entities in) core network 106, (one or more entities in) core network 107, (one or more entities in) core network 109, base station 114a, base station 114b, Node-B 140a, Node-B 140b, Node-B 140c, RNC 142a, RNC 142b, MGW 144,

MSC 146, SGSN 148, GGSN 150, eNode-B 160a, eNode-B 160b, eNode-B 160c, MME 162, serving gateway 164, PDN gateway 166, base station 180a, base station 180b, base station 180c, ASN gateway 182, MIP-HA 184, AAA 186, and gateway 188. And certainly other network entities and/or combinations of network entities could be used in various embodiments for carrying out the network-entity functions described herein, as the foregoing list is provided by way of example and not by way of limitation.

[0058] FIG. 2 depicts an example video coding and distribution system, according to an embodiment. In particular, FIG. 2 depicts the example system 200. The example system 200 includes components analogous to those depicted in the example ABR system 160 of FIG. 1C, such as a full resolution input video source 262, an adaptive bitrate encoder 264 generating traditional ABR streams 265, a streaming server 266, a network 268, and client devices 269. In addition, system 200 further includes a zoom coding encoder 204. The zoom coding encoder 204 receives a source video stream from the full resolution video source 262, either in uncompressed or a previously compressed format. Zoom coding encoder 204 encodes or transcodes the source video stream into a plurality of zoom coded sub-streams, wherein each of the zoom coded sub-streams encodes a spatial portion (e.g. a segment, a slice, a quadrant, or other division) representing an area smaller than the complete area of the overall source video. In embodiments using transcoding that convert a video stream from one compressed format to another, a decoding process is performed that brings the video back to the uncompressed domain at its full resolution followed by the re-encoding process of creating new compressed video streams representing different resolutions, bit rate or frame rates.

[0059] The zoom coded sub-streams 206 may be i) encoded at a resolution and quality of the source video stream, or similar to the ABR encoding, and/or ii) encoded into a plurality of resolutions. The zoom coded sub-streams 206 are transmitted to the streaming server 266 for further transmission to the client devices 269. In some embodiments, the ABR encoder and the zoom coding encoder are the same encoder, configured to encode the source video into the ABR streams and the zoom coded sub-streams.

[0060] FIG. 3 depicts example coding resolutions, in accordance with an embodiment. In particular FIG. 3 depicts an overview of coding 300. The overview includes a digital source video 302, an ABR encoder 304, a zoom coding encoder 306, ABR streams 308-312, and zoom coded sub-streams 314-320. The digital source video 302 is depicted as having four quadrants, the top left has diagonal cross hatching, the top right has vertical and horizontal lines, the bottom left has diagonal lines, and the bottom right is dotted. The full resolution of the source digital video 302 is 3840 horizontally by 2160 vertically (4K × 2K). The four quadrants are shown by way of example

as the digital video source may be divided into any number of areas in any arrangement, including segments of different sizes and shapes. The digital source video 302 is received by the ABR encoder 304 and the zoom coding encoder 306. The ABR encoder 304 processes the digital source video into three different ABR streams 308, 310 and 312. Each ABR stream is of a different resolution. For example, the ABR stream 308 is encoded in 2K × 1K (specifically 1920 × 1080), has the highest resolution, and is depicted as the largest area. The ABR stream 312 is encoded in 500 × 250 (specifically 480 × 270), has the lowest resolution, and is depicted as the smallest area. While the ABR streams in example 300 vary the resolution, other attributes, including but not limited to bit rate and frames per second, can also varied by themselves on in conjunction with other digital video attributes. The zoom coded sub-streams 314, 316, 318, and 320 are each encoded in a 2K × 1K resolution (specifically 1920 × 1080), matching the resolution of the corresponding regions in the digital source video 302.

[0061] In an embodiment, a client device is streaming digital video via the system 200, and a source video is being encoded (or has previously been encoded and stored at a streaming server) as depicted in the FIG. 3. The client device can receive, can decode, and can display any of the ABR streams 308, 310, 312 that depict the source video encoded at varying digital video parameters. The client device can zoom in on a portion of the (e.g. decoded) traditional ABR streams 308, 310, or 312. However, the whole ABR stream is transmitted over the network, including portions which the client device may not display (e.g. portions that are outside the boundary of the display when the video is zoomed in). Also, the resulting zoomed image portion displayed on the client device is likely to appear pixelated or otherwise to display a lower resolution. Using embodiments disclosed herein, however, the client device can zoom in on a portion of the video stream by requesting a one or more of the zoom coded sub-streams 314, 316, 318, and 320 corresponding to the portion of the video stream requested by the client device. The client device can, for example, request to see the top left portion of the digital video corresponding to the diagonally cross-hatched area. In response, the streaming server transmits the zoom coded sub-stream 314 over the network to the client device. Thus, the portion of the video display requested by the client device is transmitted over the network and the resulting display is of a higher quality than a zoomed-in version of an ABR stream. A separate video camera or source video is not required to provide a high quality video stream to the client device.

[0062] To facilitate a client device receiving a zoom coded video stream via the network, the streaming server may be configured to notify the client device via a profile communication file of available streams. For example, the profile communication file may be a manifest file, a session

description file, a media presentation description (MPD) file, a DASH MPD, or another suitable representation for describing the available streams.

[0063] In some embodiments the source video is a sports event, a replay of a sports event, an action sequence, a surveillance security video, a movie, or a television broadcast.

[0064] FIG. 4A depicts examples of zooming video streams, in accordance with an embodiment. In particular, FIG. 4A depicts the process to display an area of interest 402 of the source video 302. The area of interest 402 may include an object of interest in the video scene, which may be a stationary object or a moving object. (Henceforth, the terms object of interest and area of interest are used interchangeably in the present disclosure.) As described in conjunction with FIG. 3, the source video 302 is encoded with the ABR encoder 304 producing video streams 308, 310, 312 and is further encoded with the zoom coding encoder 306 producing video streams 314, 316, 318, 320. The regions 404 and 406 represent the portion of the encoded streams associated with the area of interest of the source video 302. The displays 406 and 410 represent the video displayed on the client device by zooming in on a traditional ABR stream (406) as compared to the use of zoom coding (410). The area of interest 402 overlaps areas encoded in four different zoom coded sub-streams and has resolution dimensions of 2K x 1K of the original 4K x 2K source video. Thus, the highest resolution available to be displayed on a client device that represents the area of interest is 2K x 1K.

[0065] The ABR encoder 304 is able to provide a zoomed-in view of the area of interest. In this embodiment, the ABR encoder 304 produces three ABR encoded streams 308, 310, and 312. The ABR stream 308 has a resolution of 2K x 1K and includes the region 404 that corresponds to the area of interest 402. However, the portion of the video stream corresponding to the area of interest has resolution dimensions of approximately 1K x 500 (specifically 960 × 540). Whether the ABR stream 308 or a separate stream representative of the region 404 is transmitted to the client device, the final displayed video 406, with resolution dimensions of approximately 1K x 500 (specifically 960 × 540), is at a lower resolution than the region of interest 402 in the source video 302, and displayed video 406 must be scaled for display on the client device.

[0066] The zoom coding encoder 306 is also able to provide for a zoomed-in view of the area of interest. In this embodiment, the zoom coding encoder 306 produces four zoom coded sub-streams 314, 316, 318, and 320. Each zoom coded sub-stream has the resolution of 2K x 1K. The region 408 overlaps all four zoom coded sub-streams and has a maximum resolution available of 2K x 1K, the same resolution dimensions available for region of interest 402 in the source video 302.

[0067] The source video 302 may be further divided into smaller portions, or slices, than the quadrants depicted. For example, the source video 302 may be divided using a grid of 8 portions horizontally and 8 portions vertically, or using a different grid of 32 portions horizontally and 16 portions vertically, or some other partitioning into portions. Slice encoding is supported by video encoding standards, such as H.264, H.265, HEVC, and the like. In embodiments where the area of interest does not overlap a region covered by an available zoom coded video sub-stream, not all available zoom coded video sub-streams are transmitted via the network to the client device. Either the client device or a network entity, such as the streaming server, can determine the appropriate subset of the available zoom coded video sub-streams to transmit to the client device to cover the area of interest. For example, if the area of interest 402 of FIG. 4A were of a smaller size and/or were shifted to the left and so did not include video from the top-right or bottom-right portions of the source video, then only zoom coded video sub-streams 314 and 316 are provided to represent the area of interest. In this case, only the streams 314 and 316 are transmitted to the client device, and streams 318 and 320 are not transmitted to the client device, in order to allow the client to decode and display the region of interest 402.

[0068] FIG. 4B depicts a second example of video zoom, in accordance with an embodiment. In particular, FIG. 4B depicts zooming in circumstances where segments of video are divided into a plurality of slices, tiles, or other divisions of area. Each slice, tile, or other division of area (collectively referred to herein as a slice) is independently decodable. Each slice may be individually requested, retrieved, and decoded by a client, in the same manner as described for the alternative zoom coded sub-streams in the preceding examples. In the example of FIG. 4B, a source video 412 with resolution dimensions of 4K x 2K has an area of interest 414 that has resolution dimensions of 2K x 1K. In this embodiment, the source video is encoded into twelve slices, six on the left side and six on the right side. The division into slices may be performed using any video codec (e.g. any video codec supporting independently decodable portions or slices) as known to those of skill in the art. Here, the area of interest overlaps eight slices, and does not include the top two and bottom two video slices. Thus, the client device can request to only receive eight of the total twelve video segments from (e.g. encoded by) the zoom coding encoder. Thus, the client device can display the area of interest in the full resolution of 2K x 1K without scaling the video and without the need to receive all of the available zoom coded segments.

[0069] The zoom coded video sub-streams 314, 316, 318, and 320 or the zoom coded slices depicted in FIG. 4B can be further encoded into other bit rates that have different resolutions, frames per second, and bit depths, thus making multiple representations of each zoom coded sub-stream or segment available to be transmitted to the client device.

[0070] FIG. 5 depicts an example of a digital video with an object of interest, in accordance with an embodiment. In particular, FIG. 5 depicts an example digital video 500. The digital video 500 includes a plurality of video slices 502a, 502b, etc. As illustrated in FIG. 5, the use of zoom coded sub-streams allows a user to view a zoomed version of an object or area of interest that moves such that it overlaps different slices at different times. In an exemplary embodiment, the source video 502 has resolution dimensions of 3840 horizontally by 2160 vertically. Each of video segments 502a, 502b, etc., have approximate resolutions dimensions of 800 horizontally and 333 vertically. The source video 500 may be encoded by various ABR encoders and zoom coding encoders and provide the encoded video streams to a streaming server for further transmitting over a network to a client device.

[0071] An object of interest, depicted by a soccer ball, is located at position 504a (inside slice 502c) at a first time T1. The position of the ball may be represented by a data structure (P1,T1), where P1 represents the position 504a. At time T2, the object of interest is located further up and to the right (in slice 502d) at position 504b, which may be represented by (P2,T2). At time T3, the object of interest is located further up and to the right (in slice 502e) at position 504c, which may be represented by (P3,T3). In response to a user input indicating a desire to zoom in on the object of interest, a client device may initially (for viewing of time period T1) request slice 502c (and, in some embodiments, immediately neighboring slices). The client device, on receiving the requested slices, causes display of a zoomed-in stream that includes the object of interest. To continue providing a zoomed-in view of the object of interest, the client device may subsequently (for viewing of time period T2) request and display slice 502d (and, in some embodiments, immediately neighboring slices). Subsequently, to continue tracking the object of interest, the client device may subsequently (for viewing of time period T3) request and display slice 502e (and, in some embodiments, immediately neighboring slices).

[0072] Selection of the appropriate slices to show the object of interest in context (e.g. with some surrounding background) may be performed on the client device or at the streaming server. Further, the concepts in this disclosure may apply to larger objects, objects which span multiple neighboring slices, objects traversing slices at different speeds, multiple objects, source video streams segmented into smaller segments, and the like.

[0073] A rendering reference point or "render point" may be used to indicate a rendering position associated with one or more positions of the object/area of interest. The rendering reference point may, for example, indicate a position (e.g. a corner or an origin point) of a renderable region which contains the object of interest at some point in time. The rendering reference point may indicate a size or extent of the renderable region. The rendering reference

point may define a bounding box which defines the location and extent of the object/area of interest or of the renderable region containing the object/area of interest. The client may use the rendering reference point information to extract the renderable region from one or multiple zoom coded sub-streams or segments, and may render the region as a zoomed region of interest on the client display. In the first set of video segments, the rendering reference point, (0, 0) is depicted in the bottom left corner of the source video 502. However, the second set of video segments has a rendering reference point of (a, b), and is depicted in the bottom left corner of slice 502f. The rendering reference points may be communicated to the client device. For example, rendering reference points may be transmitted in-band as part of the video streams or video segments, or as side information sent along with the video streams or video segments. Alternately the rendering reference points may be specified in an out-of-band communication (e.g. as metadata in a manifest such as a DASH MPD). A discrete jump in the rendering reference point from (0, 0) to (a, b) as the object transitions from (P1, T1) to (P3, T3) will cause an abrupt change in the location of the object of interest as displayed on the client device. The rendering reference point as communicated to the client may be updated on a frame-by-frame basis, which may allow the client to continuously vary the location of the extracted renderable region, and so the object of interest may be smoothly tracked on the client display. Alternately the rendering reference point may be updated more coarsely in time, in which case the client may interpolate the rendering position between updates in order to smoothly track the object of interest when displaying the renderable region on the client display. The rendering reference point may include two parameters, a vertical distance and a horizontal distance represented by (x, y). The rendering reference points may be communicated, for example, as supplemental enhancement information (SEI) messages to the client device.

[0074] At each subsequent frame, the render reference point may be updated to reflect the global object motion between each frame. When the render reference adjustment is equal to the global motion of the object of interest, the object will appear motionless (e.g. having a relatively constant position relative to the displayed region), as if the camera were panning to keep the object at the same point on the screen. When the motion of the object of interest is underestimated, the object skips backwards on the screen. Conversely, when the motion of the object of interest is overestimated, the object skips forwards between frames. Minimizing the error of the object motion results in smooth rendering.

[0075] In the above scenario, it is assumed that the video display transitions from the first set of video segments (and, in some embodiments, video segments which contain slices in the spatial neighborhood of the first set of video segments) to the second set of video segments (and, in some embodiments, video segments which contain slices in the spatial neighborhood of the second set

of video segments) when the object of interest is at (P2, T2). Therefore, in this embodiment, the render reference point for each frame transmitted is adjusted (e.g. interpolated) to smoothly transition from (0, 0) to (a, b) over the time from T1 to T2. The smooth transition may be linear (e.g. moving the rendering reference point a set distance equally each frame), non-linear (e.g. moving the rendering reference point small amounts closer to time T1, large amounts between times T1-T2 and small amounts closer to time T2), or in any other similar method. In some embodiments, the rendering reference point is transmitted as two coordinates, such as (x, y), and in other embodiments, the rendering reference point is transmitted as a differential from the previous frame.

[0076] FIG. 6 depicts an example process for encoding and delivery of content to a client using adaptive bitrate coding. In step 602, source content is communicated from a content source 604 to an encoder 606. The source content is a compressed or uncompressed stream of digital video. The encoder 606 encodes the video into several representations 608 with different bitrates, different resolutions, and or other different characteristics and transmits those representations 608 to a transport packager 610. The transport packager 610 uses the representations 608 to generate segments of, e.g., a few seconds in duration. The transport packager 610 further generates a manifest (e.g. a DASH MPD) describing the available segments. The generated manifest and the segmented files (collectively 616) are distributed to one or more edge streaming servers 614 through an origin server 612. Subsequent segments (collectively 617) are also distributed to the origin server 612 and/or the edge streaming server 614.

[0077] To view the video, a client 620 visits a web server 618, e.g. by sending an HTTP GET request 622. The web server 618 may send a response 624 that directs or redirects the client 620 to a streaming server such as the edge streaming server 614. The client thus sends a request 626 to the edge streaming server. In response, the edge streaming server sends a manifest (e.g. a DASH MPD) 628 to the client. Based on the client capabilities and network conditions, the client selects an appropriate representation of the content and issues a request 630 for an appropriate segment (e.g. the first segment of recorded content, or the most recent segment of live content). The edge streaming server responds by providing the requested segment 632 to the client. As illustrated at 634, the client may request a subsequent segment of the content (which may be at the same bitrate or a different bitrate from the segment 632), and the subsequent segment is sent to the client at 636.

[0078] FIG. 7 depicts an example process of encoding and delivery of content to a client using zoom coding. In step 702, source content is communicated from a content source 704 to an encoder 706. The source content is a compressed or uncompressed stream of digital video. The encoder

706 encodes the video into several representations 708 of the complete screen area with different bitrates, different resolutions, and or other different characteristics and transmits those representations 708 to a transport packager 710. In addition, the zoom coding encoder encodes the video into several different slice streams (e.g. streams 712, 714) representing different areas of the complete video image. Each of the streams 712 may represent a first encoded slice area of the content, with each of the streams being encoded at a different bitrate, and each of the streams 714 may represent a second encoded slice area of the content, again with each of the streams being encoded at a different bitrate. Depending on the chosen partitioning of the content into slices, additional slice streams representing various encoded bit rates for other content slices may be included, though not shown in the figure.

[0079] The transport packager 710 uses the representations 708, 712, 714 to generate segments of, e.g., a few seconds in duration. The transport packager 710 further generates a manifest (e.g. a DASH MPD) describing the available segments, including the segments that represent the entire screen and segments that represent only a slice area of the screen. The generated manifest and the segmented files (collectively 716) are distributed to one or more streaming servers such as edge streaming server 720 through an origin server 718.

[0080] To view the video, a client 724 visits a web server 722, e.g. by sending an HTTP GET request 726. The web server 722 may send a response 728 that directs or redirects the client 724 to the edge streaming server 720. The client thus sends a request 730 to the edge streaming server. In response, the edge streaming server sends a manifest (e.g. a DASH MPD) 732 to the client. Based on the client capabilities and network conditions, the client selects an appropriate representation of the normal (unzoomed) content and issues a request 734 for an appropriate segment (e.g. the first segment of recorded content, or the most recent segment of live content). The edge streaming server responds by providing the requested unzoomed segment 736 to the client. The client may request, receive, parse, decode and display additional unzoomed segments in addition to the segment 736 shown in the diagram.

[0081] In response to user input indicating selection of an object or region of interest, the client device 724 may issue a request 738 for one or more sub-streams that are associated with an object or region of interest. In some embodiments, the client device identifies the streams to be requested based on, e.g. information such as render point information which may be provided in the manifest or in-band in the video streams. In other embodiments, the client device identifies the object or region of interest and forms a request based on the identified object or region of interest, and the identification of appropriate streams for that object or region of interest is made at the server side.

Such server-identified streams or segments may then be returned by the server to the client in response to the request.

[0082] The appropriate slice stream or streams 740 are sent to the client device 724, and the client device decodes and combines the streams 740 to provide a zoomed version of the object or region of interest. The client may request and receive the stream or streams 740 at a bitrate appropriate to the capabilities of the client device and the current network conditions using ABR techniques.

[0083] In accordance with an embodiment, more than one object of interest can be tracked and displayed. For example, at a certain point in time, a first object may be associated with a first set of slices such that a client must retrieve the slices of the first set in order to recover and render a view (e.g. a zoomed view) of the first object. At the same time, a second object may be associated with a second set of slices such that the client must retrieve the slices of the second set in order to recover and render a view (e.g. a zoomed view) of the second object. The first set of slices may be completely different than, partially overlapping with, or fully overlapping with the second set of slices. Moreover the amount of overlap between the first and second set of slices may change with time as the underlying objects move. The render point information for each set may be independently encoded for each such set and may be contained in different slices or the same slice. The receiver must retrieve the appropriate rendering point (corresponding to a current zoom coded object) and apply the render point offset accordingly.

[0084] As the object, or objects, of interest move through the screen, there may be changes to the sets of slices that represent the new zoomed view. The manifest may be updated to signal such changes, or a completely new manifest can be created. The client device may use the updated manifest information to appropriately request the set of slices that represent the updated view. Alternately the changes may be signaled in-band in the video stream, or in side information such as a render point metadata file retrievable by the client.

[0085] The request for streams may correspond to a particular area of the video or correspond to an object ID. For example, if the video source is video of a soccer (a.k.a. football) game, examples of different objects may include a goal box, a ball, or a player. The objects may be detected via any means, including image detection (e.g. detecting the rectangular dimensions of a goal, the round shape of a ball, or numbers on a uniform, etc.), spatial information encoded in the source video (e.g. correlation between camera position and a stationary object's position, sensor information being transmitted from a soccer ball, etc.) or any other similar method. In this embodiment, a client device can request to receive zoom coded sub-streams associated with an object of interest such as the ball. The request may also include a magnitude of context to include

with the ball, such that the ball comprises a certain percentage of the display, or the like. The magnitude of context may be specified as a rendering area size specified as a horizontal dimension and a vertical dimension in pixels, for example. A network entity, or the client device, can determine the appropriate zoom coded sub-streams to frame the object of interest, and may communicate to the streaming server which zoom coded sub-streams to send to the client device. The client device receives the zoom coded sub-streams and the appropriate rendering information to display the zoomed-in video stream.

[0086] The spatial regions of the object, or objects, of interest may be determined at the streaming server, at the client device, a separate network entity, or a combination of the above examples. In one embodiment, the server-side creates the arbitrary spatial region, such as mapping the streams to slices for encoding. In another embodiment, the client-device side creates or assembles the arbitrary spatial regions by, for example, decoding more than one spatial content portion (e.g. more than one slice or video segment) from the server and combining parts of the decoded spatial content portions in order to create or assemble the desired spatial region. In yet another embodiment, a hybrid server-side/player-side creates the arbitrary spatial regions.

[0087] Other variations can be applied to the various zoom coding examples, including the following:

- Zoom-coded regions can include variations of frame rate, chroma resolution, and bit depth characteristics. For example, the ABR Streams for Each Segment as shown in FIG. 4B may be encoded using such variations.
- Various techniques for storage and transport layer packaging of zoom coded sequences may be applied. For example, the zoom coded sub-streams or segments may be packaged using MPEG-2 transport stream segments, or using an ISO Base Media file format.
- Zoom-coded sequences or segments may be created with additional bit depth for specific spatial regions. The regions of enhanced bit depth may correspond to the areas or objects of interest, for example.
- Two-way interaction may be used to optimize for client side display capabilities.
- Creation of special effects may be provided, such as slow motion and zoom.

[0088] FIG. 8 depicts an example communications process, in accordance with an embodiment. In particular, FIG. 8 depicts a DASH-Type exchange between a streaming server 802 and a client device 804 to receive a zoom coded sub-stream. The client device 804 sends a request 808 to a web server 806 for streaming services, and the web server at 810 directs or

redirects the client device 804 to the edge streaming server 802. In response to a request 812 from the client device 804, the edge streaming server sends an extended MPD 814, with zoom coded information, to the client device 804. The client device parses the extended MPD in order to determine what objects/areas of interest are available and also to determine in step 816 the slices to request for each object. The client sends requests for the appropriate slices (e.g. requesting a first slice at 818 and requesting a second slice at 820). The requested slices may be a subset of the available slices, and/or may be requested by requesting the video segments which contain the slices. The edge streaming server sends each requested slice of the video stream to the client device (e.g. sending the first slice at 822 and the second slice at 824), and the client device renders the zoom coded frame for the specific object in step 826. In step 828, the client device causes display of the zoom coded frame, e.g. by displaying the frame on a built-in screen or by transmitting information representing the frame to an external display. Composition of the zoom coded frame by the client may comprise receiving, decoding, and/or rendering the requested slices for an object/area of interest. The client may render a subset of the pixels of the requested slices, as determined by a current render point and/or a rendering area size or context magnitude indication for the object. The DASH-type message may include additional extensions to support tracking of multiple objects with overlapping slices.

[0089] Zoom coding may be enabled using MPEG-DASH. MPEG-DASH (ISO/IEC 23009-1:2014) is an ISO standard that defines an adaptive streaming protocol for media delivery over Internet Protocol (IP) Networks. An exemplary process for performing zoom coding using MPEG-DASH may be performed as follows. A determination is made that a zoom coded representation is available and how to access that content. This information is signaled to the DASH client using syntax in the MPD descriptor. Per Amendment 2 of the ISO DASH standard, the MPD may provide a "supplementary stream." This supplementary stream may be utilized for zoom coding. A spatial relationship descriptor (SRD) syntax element can describe a spatial portion of an image (see Annex H of ISO 23009-1 AM2).

[0090] An object render point provided in the video bitstream is used to render the zoomed section for the object being tracked. The zoomed section may, for example, be rendered with uniform motion or with interpolated motion as described herein. The object (or objects) render point may be sent in user data for one or more slices as an SEI message. For example, the SEI message may be as defined in a video coding standard such as AVC/H.264 or HEVC/H.265. Zero or more objects may be signaled per slice.

[0091] Exemplary Slice User Data for object render points includes the following parameters:

Object_ID: Range 0-255. This syntax element provides a unique identifier for each object.

Object_x_position[n]: For each object ID n, the x position of the object bounding box.

Object_y_position[n]: For each object ID n, the y position of the object bounding box.

Object _x_size_in_slice[n] : For each object ID n, the x_dimension of the object bounding box.

Object _y_size_in_slice[n] : For each object ID n, the y_dimension of the object bounding box.

[0092] The object bounding box represents a rectangular region that encloses the object. The object bounding box may also enclose some amount of surrounding context to be rendered with the object. The x,y position may indicate, for example, the upper left corner position of the object bounding box. The object position and size may pertain to the portion of the object contained in the slice that contains the user data.

[0093] The video depicted in FIG. 5 (described in further detail above) may be used in the implementation of zoom coding of a video with resolution of 4K, or 3840×2160. The video of FIG. 5 illustrated in FIG. 9, with each slice (spatial portion) of the video being assigned a number ranging from 1 to 30. The 4K video is encoded with H.264 compression into thirty independent H.264 slices. Each slice is 768×360 pixels. The native full image is scaled down to HD 1920×1080 and provided as the normal unzoomed stream for a client device to display. Additionally, each of the thirty segments is encoded in the native 768×360 resolution. The encoder tracks an object as shown in the figure moving across the scene. The subset of slices is signaled to the client via the MPD SRD descriptor. For each slice, an Adaptation Set with SRD descriptor is provided.

[0094] In order to support tracking and zooming of more than one object, the SRD descriptor syntax may be extended to allow the client device to determine which slices are needed for rendering the given objects. The Object_ID (consistent with the slice SEI information) is included in the SRD, added to the end of the "value" syntax for SRD. If multiple objects are associated with the slice, then multiple Object_ID's may be added to the end of the SRD value syntax. When OBJECT_IDs are used, the Spatial_Set_ID may also be in the stream. After the Spatial_Set_ID parameter in the SRD, up to 256 Object_IDs can be included. An example of two SupplementalProperty SRD elements are shown below.

Example 1: SRD with 1 object (Object_ID 5)

```
<SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
      value="0,x16,y16,768,360,3840,2160,1,5"/>
```

Example 2: SRD with 5 objects (Object_ID 2, 4, 7, 9, and 14)

```
<SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
      value="0,x16,y16,768,360,3840,2160,1,2,4,7,9,14"/>
```

[0095] As illustrated in these examples, the SupplementalProperty syntax element may be used to provide an association between particular spatial portions of a video (e.g. particular slices) and particular objects of interest. Thus, Example 1 above provides an association between the slice numbered 16 and an object numbered 5, and Example 2 above provides an association between the slice numbered 16 and an objects numbered 2, 4, 7, 9, and 14. To provide a zoomed view of a selected object (e.g. a user-selected object), a client device may request all slices that are associated with the selected object.

[0096] In accordance with an embodiment, a full SRD example is described. In the full SRD, xM,yM (e.g. x16,y16), refers to the x,y position of the slice origin. In an actual SRD, these may be pixel values. For example, x16,y16 would be equal to 0, 1080.

[0097] As each frame is encoded (e.g. with H.264, HEVC, or the like) a user data entry may be inserted for each slice and may be used to provide the object position information. The client uses this information provide a smoothly rendered picture with the object being tracked on the client display. As the object moved across and/or into different slices, the MPD may be updated with a new list of slices for the client to access. This MPD change may include a sequence access point (SAP) in the DASH segments.

[0098] Following, is an example of an MPD with zoomed content slices 16, 17, 18, 21, 22, 23, 26, 27, 28 associated with a single tracked object:

```
<?xml version="1.0" encoding="UTF-8"?>
<MPD
  xmlns="urn:mpeg:dash:schema:mpd:2011"
  type="static"
  mediaPresentationDuration="PT10S"
  minBufferTime="PT1S"
  profiles="urn:mpeg:dash:profile:isoff-on-demand:2011">

  <ProgramInformation>
    <Title>Example of a DASH Media Presentation Description using Spatial
Relationship Description to indicate that a video is a zoomed part of
another</Title>
  </ProgramInformation>

  <Period>

    <!-- Panorama Video -->
    <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="main"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,0,0,1920,1080,1920,1080"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="1920"
height="1080" bandwidth="1055223" startWithSAP="1">
        <BaseURL> panorama_video.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="839-990"/>
      </Representation>
    </AdaptationSet>
```

```
    <!-- Zoomed Video -->
   <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="supplementary"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,x16,y16,768,360,3840,2160,1,1"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="768"
height="360" bandwidth="1055223" startWithSAP="1">
        <BaseURL> zoomed_video_slice16.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="838-989"/>
      </Representation>
    </AdaptationSet>

    <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="supplementary"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,x17,y17,768,360,3840,2160,1,1"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="768"
height="360" bandwidth="1055223" startWithSAP="1">
        <BaseURL> zoomed_video_slice17.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="838-989"/>
      </Representation>
    </AdaptationSet>

    <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="supplementary"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,x18,y18,768,360,3840,2160,1,1"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="768"
height="360" bandwidth="1055223" startWithSAP="1">
        <BaseURL> zoomed_video_slice18.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="838-989"/>
      </Representation>
    </AdaptationSet>

    <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="supplementary"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,x21,y21,768,360,3840,2160,1,1"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="768"
height="360" bandwidth="1055223" startWithSAP="1">
        <BaseURL> zoomed_video_slice21.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="838-989"/>
      </Representation>
    </AdaptationSet>

    <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="supplementary"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,x22,y22,768,360,3840,2160,1,1"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="768"
height="360" bandwidth="1055223" startWithSAP="1">
        <BaseURL> zoomed_video_slice22.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="838-989"/>
      </Representation>
    </AdaptationSet>

    <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
```

```
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="supplementary"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,x23,y23,768,360,3840,2160,1,1"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="768"
height="360" bandwidth="1055223" startWithSAP="1">
        <BaseURL> zoomed_video_slice23.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="838-989"/>
      </Representation>
    </AdaptationSet>

    <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="supplementary"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,x26,y26,768,360,3840,2160,1,1"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="768"
height="360" bandwidth="1055223" startWithSAP="1">
        <BaseURL> zoomed_video_slice26.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="838-989"/>
      </Representation>
    </AdaptationSet>

    <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="supplementary"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,x27,y27,768,360,3840,2160,1,1"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="768"
height="360" bandwidth="1055223" startWithSAP="1">
        <BaseURL> zoomed_video_slice27.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="838-989"/>
      </Representation>
    </AdaptationSet>

    <AdaptationSet segmentAlignment="true" subsegmentAlignment="true"
subsegmentStartsWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="supplementary"/>
      <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014"
value="0,x28,y28,768,360,3840,2160,1,1"/>
      <Representation mimeType="video/mp4" codecs="avc1.42c033" width="768"
height="360" bandwidth="1055223" startWithSAP="1">
        <BaseURL> zoomed_video_slice28.mp4</BaseURL>
        <SegmentBase indexRangeExact="true" indexRange="838-989"/>
      </Representation>
    </AdaptationSet>
```

[0099] Although features and elements are described above in particular combinations, one of ordinary skill in the art will appreciate that each feature or element can be used alone or in any combination with the other features and elements. In addition, the methods described herein may be implemented in a computer program, software, or firmware incorporated in a computer-readable medium for execution by a computer or processor. Examples of computer-readable media include electronic signals (transmitted over wired or wireless connections) and computer-readable storage media. Examples of computer-readable storage media include, but are not limited to, a read only memory (ROM), a random access memory (RAM), a register, cache memory,

semiconductor memory devices, magnetic media such as internal hard disks and removable disks, magneto-optical media, and optical media such as CD-ROM disks, and digital versatile disks (DVDs). A processor in association with software may be used to implement a radio frequency transceiver for use in a WTRU, UE, terminal, base station, RNC, or any host computer.

CLAIMS

We claim:

1. A method of streaming digital video over a network, the method comprising:

transmitting a manifest to a client device, where the manifest identifies at least (i) an unzoomed stream representing an unzoomed version of a source video and (ii) a plurality of sub-streams, each sub-stream representing a respective spatial portion of the source video;

transmitting, to the client device, information associating at least one object of interest with a plurality of the spatial portions;

receiving, from the client device, a request for at least one of the sub-streams; and

transmitting the requested sub-streams to the client device.

2. The method of claim 1, wherein the information associating the at least one object of interest with a plurality of the spatial portions is provided in the manifest.

3. The method of claim 2, wherein the information associating the at least one object of interest with a plurality of the spatial portions comprises, in the manifest, a syntax element for each sub-stream identifying at least one object of interest associated with the respective sub-stream.

4. The method of any of claims 1-3, further comprising transmitting a render point for the at least one object of interest to the client device.

5. The method of claim 4, wherein the render point is transmitted in the unzoomed stream.

6. The method of any of claims 1-5, wherein the sub-streams are encoded at a resolution higher than the unzoomed stream.

7. The method of any of claims 1-5, wherein the manifest identifies, for each of the spatial portions of the source video represented in the manifest, at least two sub-streams encoded at different bitrates.

8. The method of any of claims 1-7, wherein the manifest is a DASH media presentation description (MPD).

9. A video client method comprising, at video client device:

receiving a manifest, where the manifest identifies at least (i) an unzoomed stream representing an unzoomed version of a source video and (ii) a plurality of sub-streams, each sub-stream representing a respective spatial portion of the source video;

receiving information associating at least one object of interest with a plurality of the spatial portions;

in response to a selection of one of the objects of interest, (i) identifying the spatial portions associated with the selected object of interest and (ii) retrieving a representative sub-stream for each of the spatial portions; and

causing display of a zoomed version of the object of interest by rendering the retrieved sub-streams.

10. The method of claim 9, wherein the information associating the at least one object of interest with a plurality of the spatial portions is received in the manifest.

11. The method of claim 10, wherein the information associating the at least one object of interest with a plurality of the spatial portions comprises, in the manifest, a syntax element for each sub-stream identifying at least one object of interest associated with the respective sub-stream.

12. The method of any of claims 9-11, further comprising receiving a render point for the selected object of interest, wherein the zoomed version includes an area of the retrieved sub-streams determined by the render point.

13. The method of claim 12, wherein the render point is received in the unzoomed stream.

14. The method of any of claims 9-13, wherein the manifest identifies, for each of the spatial portions of the source video represented in the manifest, at least two sub-streams encoded at different bitrates, and wherein the method comprises selecting one of the at least two sub-streams based on network conditions.

15. The method of any of claims 9-14, wherein the manifest is a DASH media presentation description (MPD).

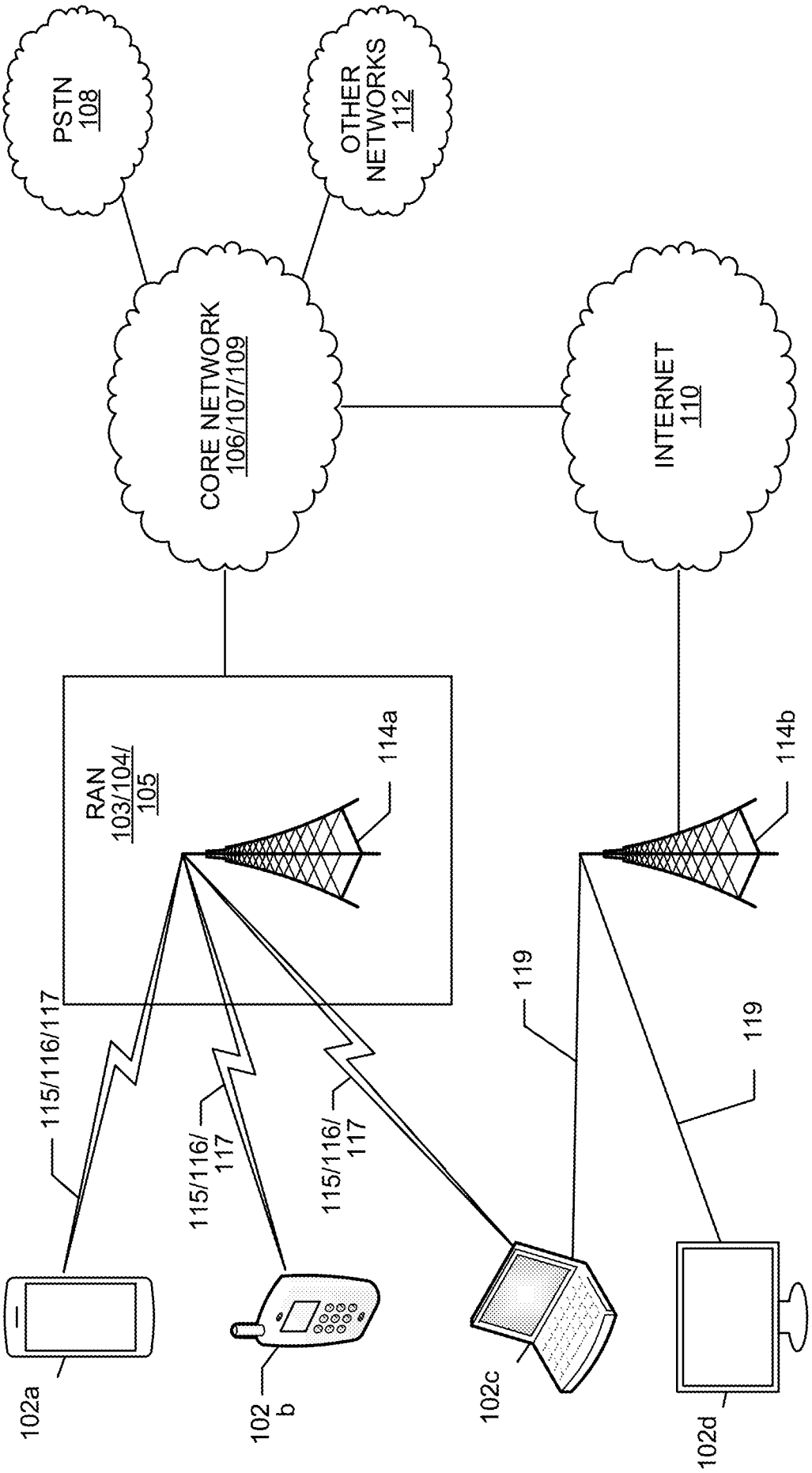

PSTN
108
OTHER
NETWORKS
112
CORE NETWORK
106/107/109
INTERNET
110
RAN
103/104/
105
114a
114b
115/116/117
115/116/
117
115/116/
117
119
119
102a
102
b
102c
102d
100

FIG. 1A

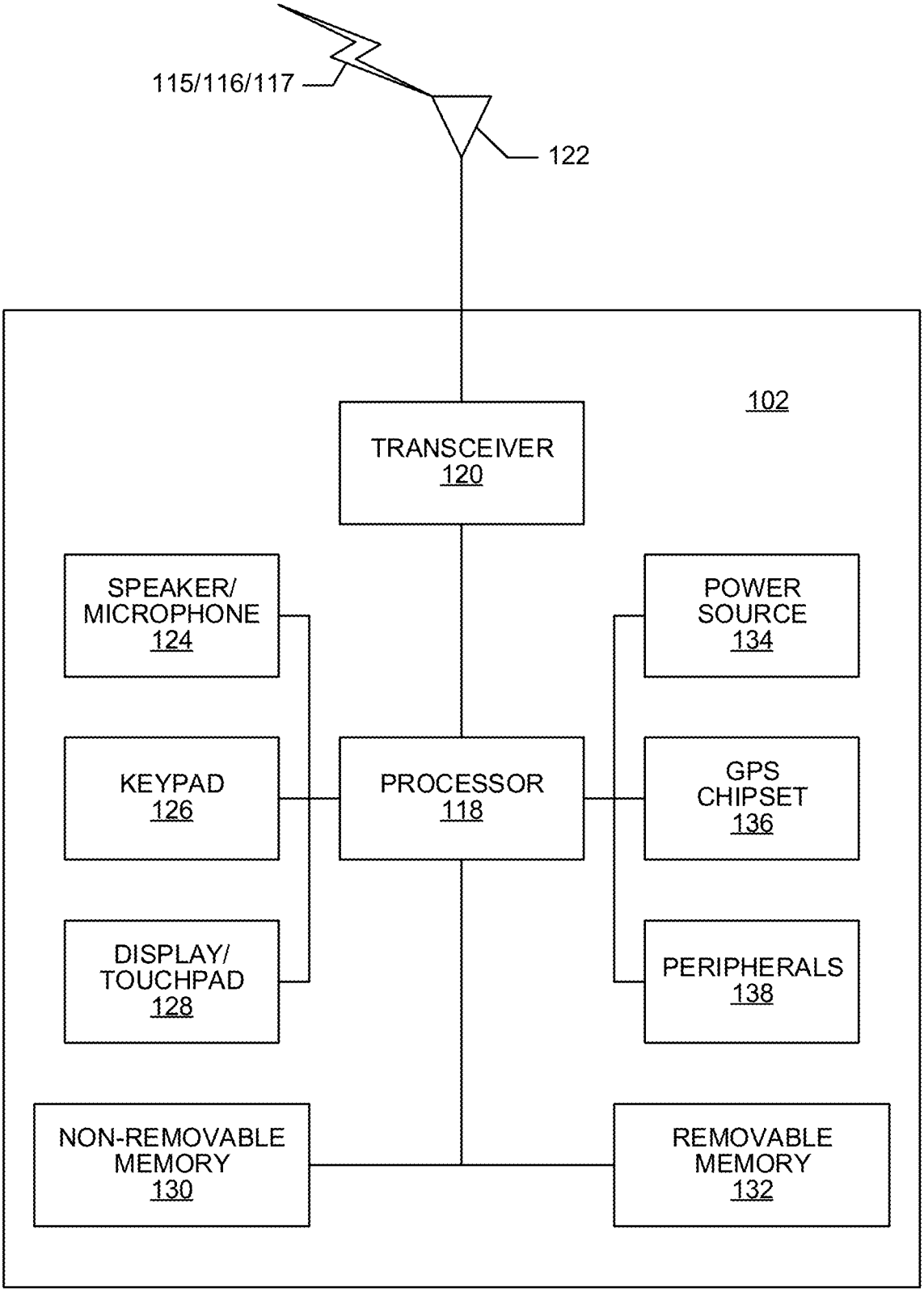

115/116/117
122
102
TRANSCEIVER
120
SPEAKER/
MICROPHONE
124
KEYPAD
126
DISPLAY/
TOUCHPAD
128
PROCESSOR
118
POWER
SOURCE
134
GPS
CHIPSET
136
PERIPHERALS
138
NON-REMOVABLE
MEMORY
130
REMOVABLE
MEMORY
132

FIG. 1B

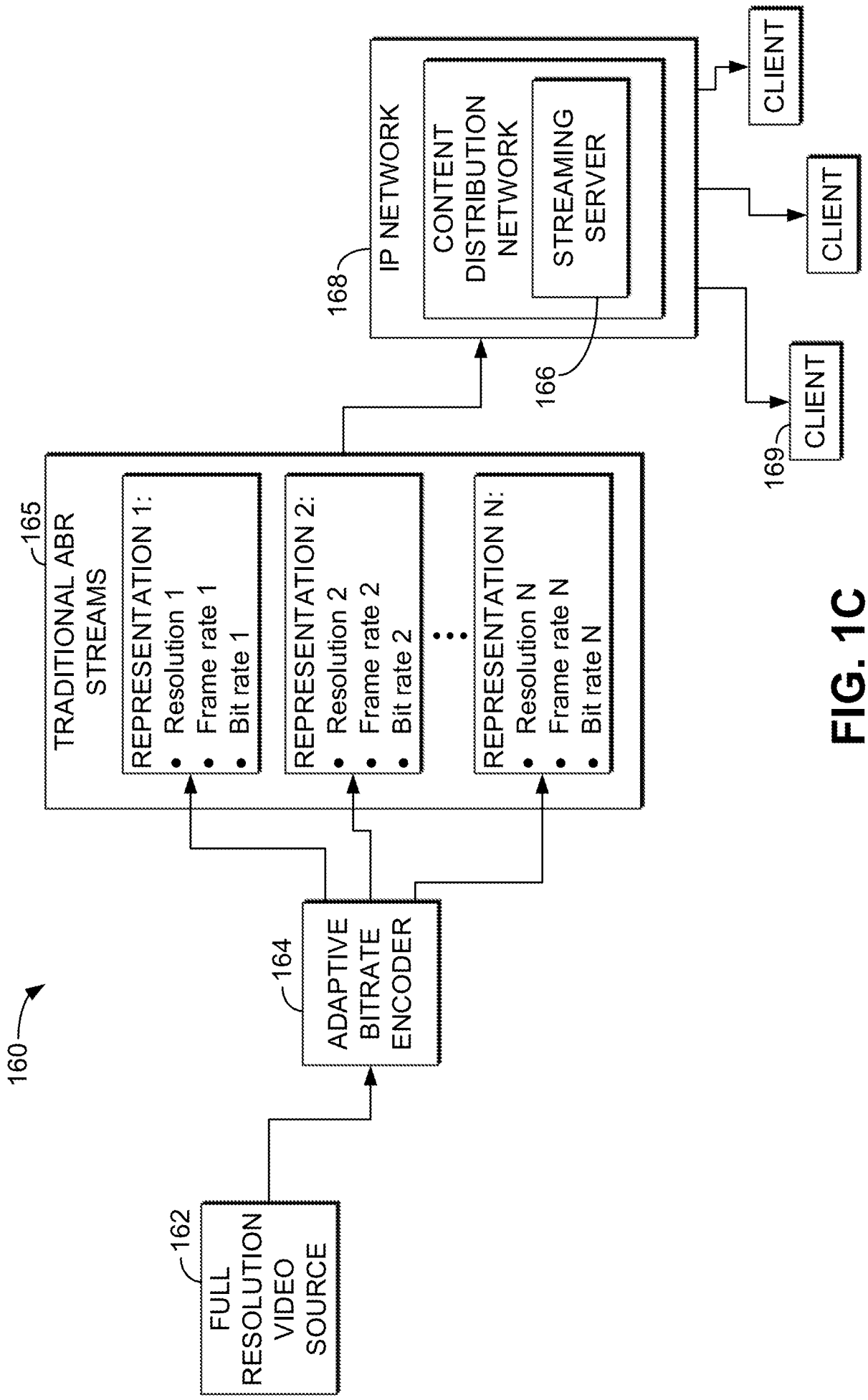

160
162
FULL RESOLUTION VIDEO SOURCE
164
ADAPTIVE BITRATE ENCODER
165
TRADITIONAL ABR STREAMS
REPRESENTATION 1:
• Resolution 1
• Frame rate 1
• Bit rate 1
REPRESENTATION 2:
• Resolution 2
• Frame rate 2
• Bit rate 2
REPRESENTATION N:
• Resolution N
• Frame rate N
• Bit rate N
166
168
IP NETWORK
CONTENT DISTRIBUTION NETWORK
STREAMING SERVER
CLIENT
CLIENT
169
CLIENT

FIG. 1C

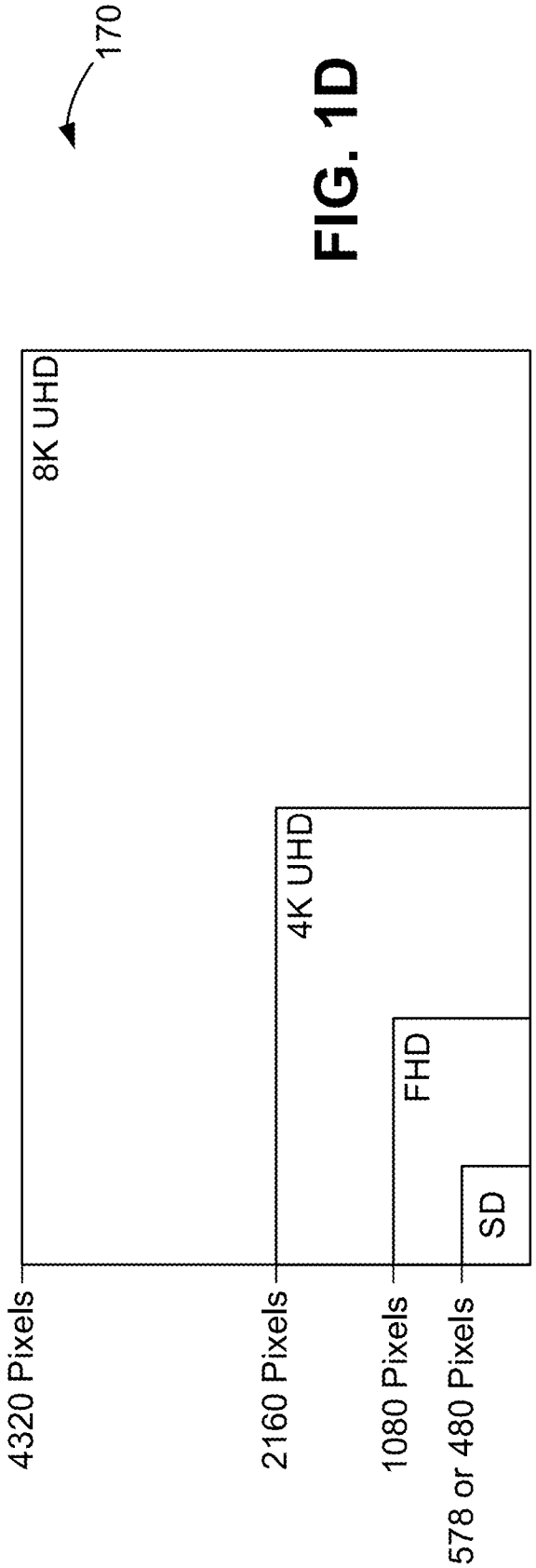

170
8K UHD
4K UHD
FHD
SD
4320 Pixels
2160 Pixels
1080 Pixels
578 or 480 Pixels

FIG. 1D

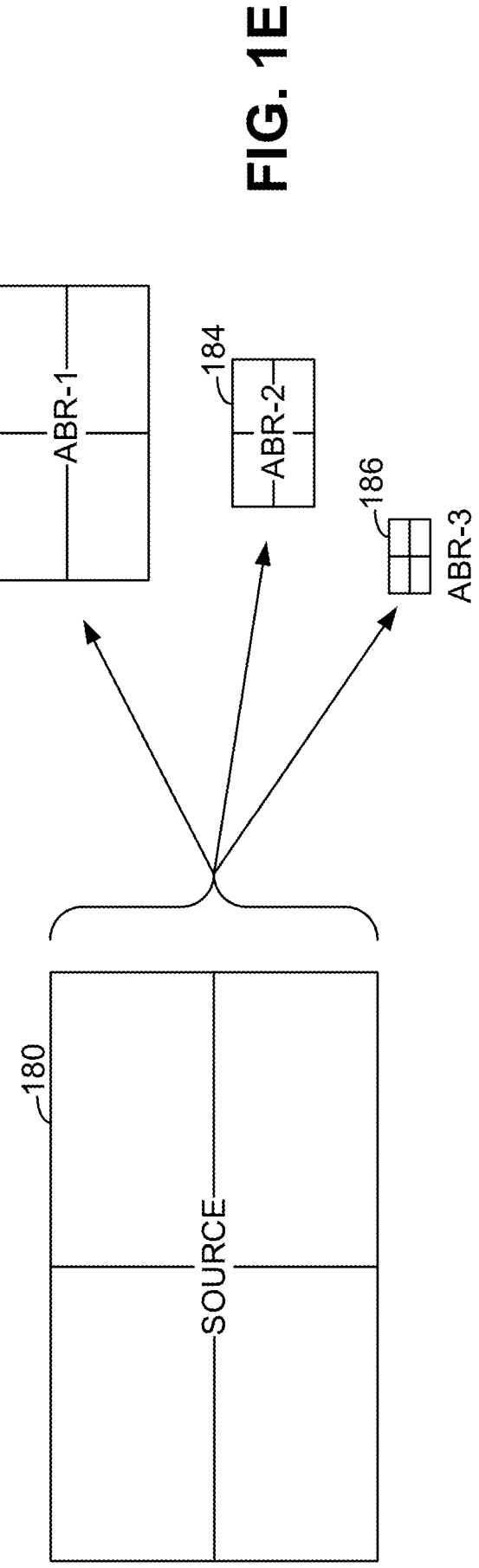

182
ABR-1
184
ABR-2
186
ABR-3
180
SOURCE

FIG. 1E

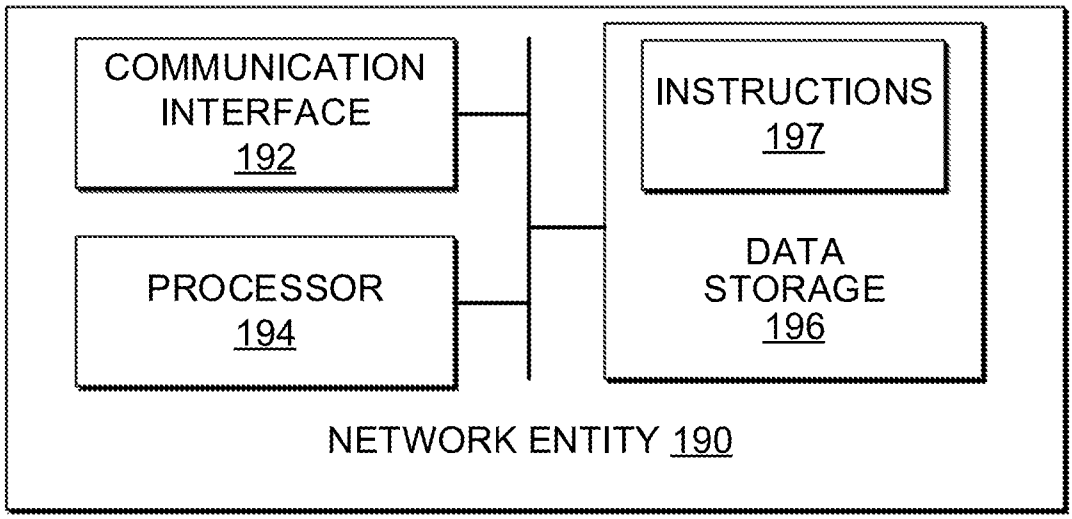

COMMUNICATION INTERFACE 192
INSTRUCTIONS 197
PROCESSOR 194
DATA STORAGE 196
NETWORK ENTITY 190

FIG. 1F

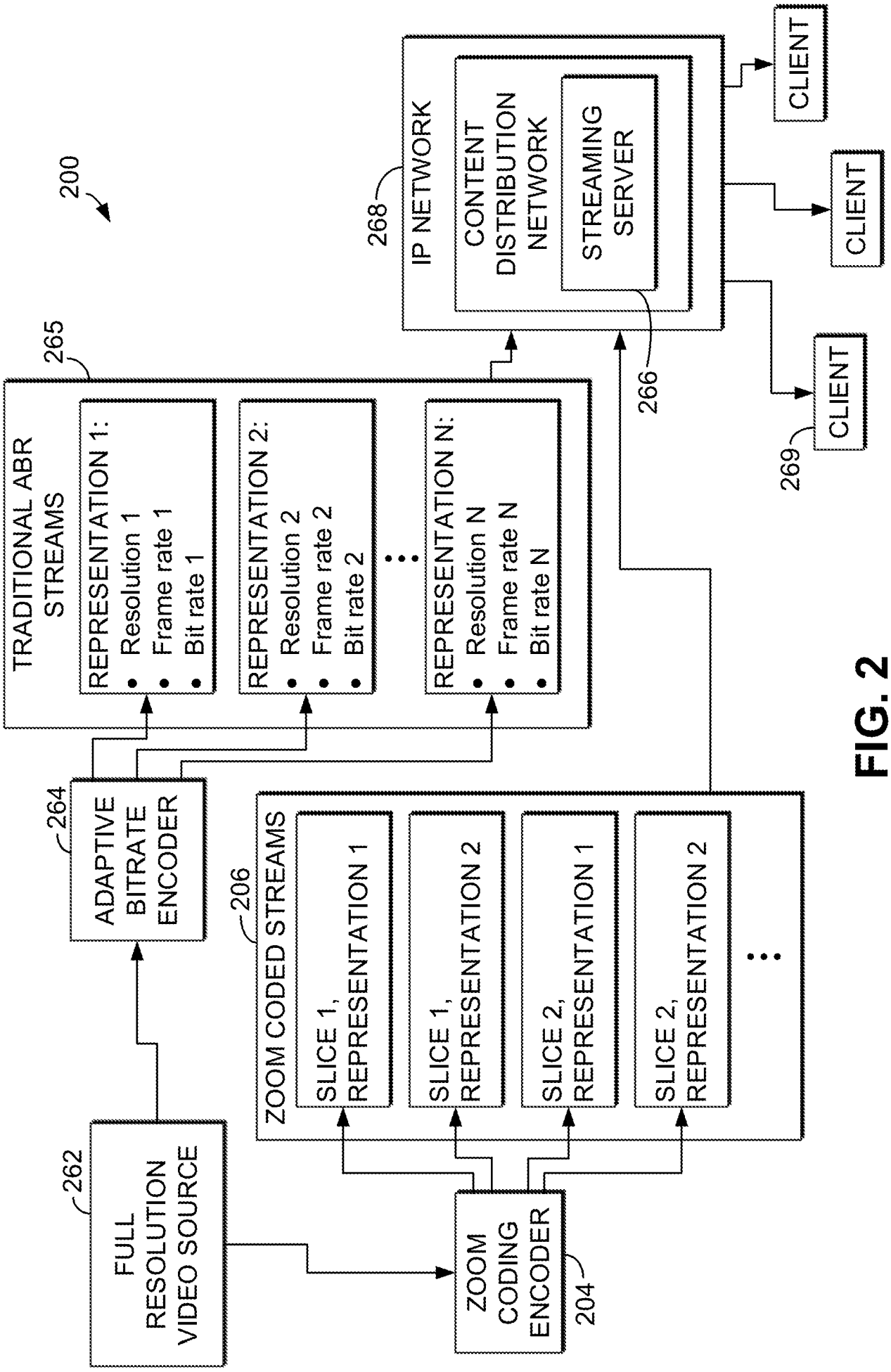

200
262
FULL RESOLUTION VIDEO SOURCE
204
ZOOM CODING ENCODER
206
ZOOM CODED STREAMS
SLICE 1, REPRESENTATION 1
SLICE 1, REPRESENTATION 2
SLICE 2, REPRESENTATION 1
SLICE 2, REPRESENTATION 2
264
ADAPTIVE BITRATE ENCODER
265
TRADITIONAL ABR STREAMS
REPRESENTATION 1:
• Resolution 1
• Frame rate 1
• Bit rate 1
REPRESENTATION 2:
• Resolution 2
• Frame rate 2
• Bit rate 2
REPRESENTATION N:
• Resolution N
• Frame rate N
• Bit rate N
266
268
IP NETWORK
CONTENT DISTRIBUTION NETWORK
STREAMING SERVER
CLIENT
CLIENT
CLIENT
269

FIG. 2

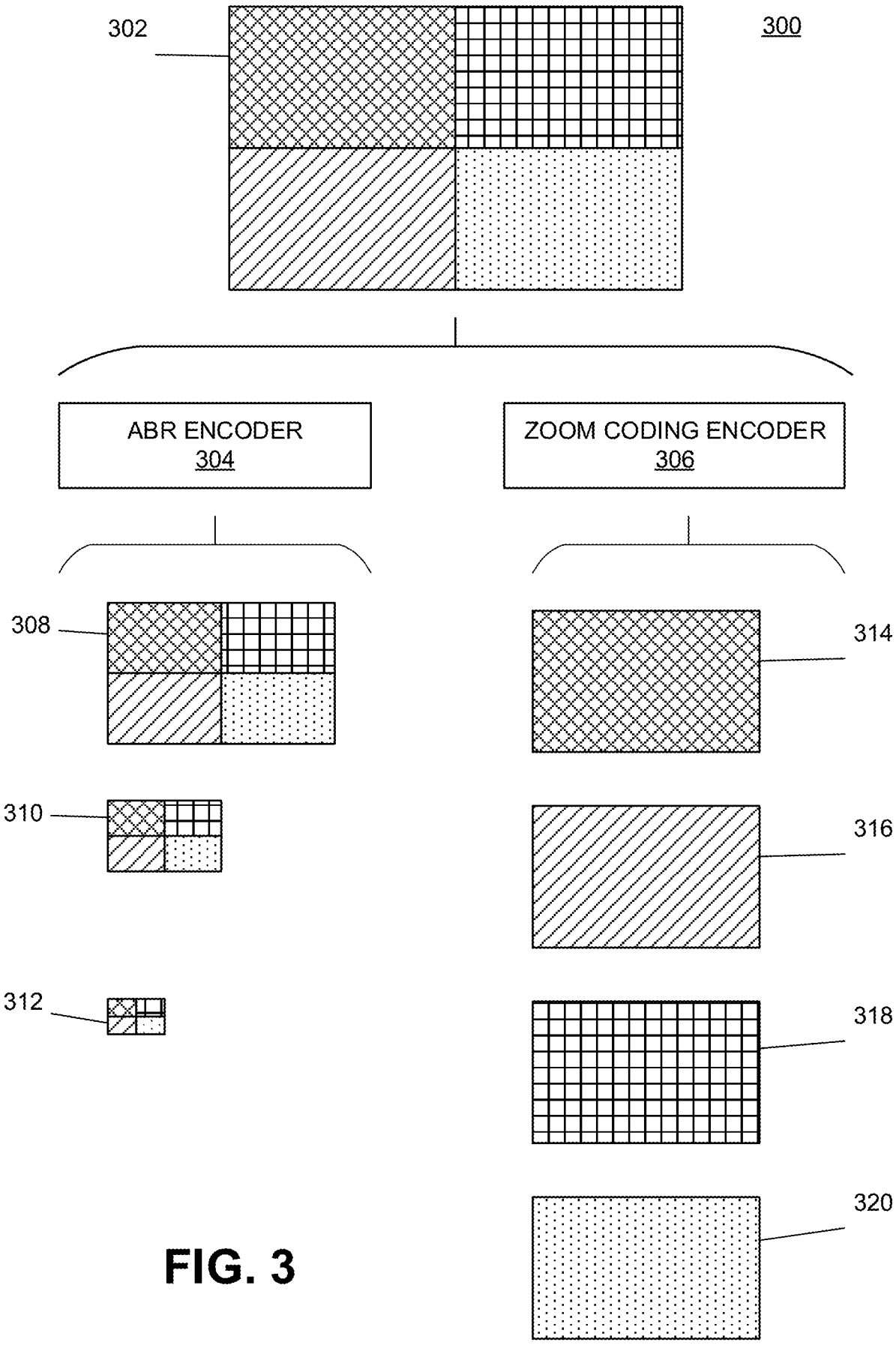

302
300
ABR ENCODER
304
ZOOM CODING ENCODER
306
308
310
312
314
316
318
320

FIG. 3

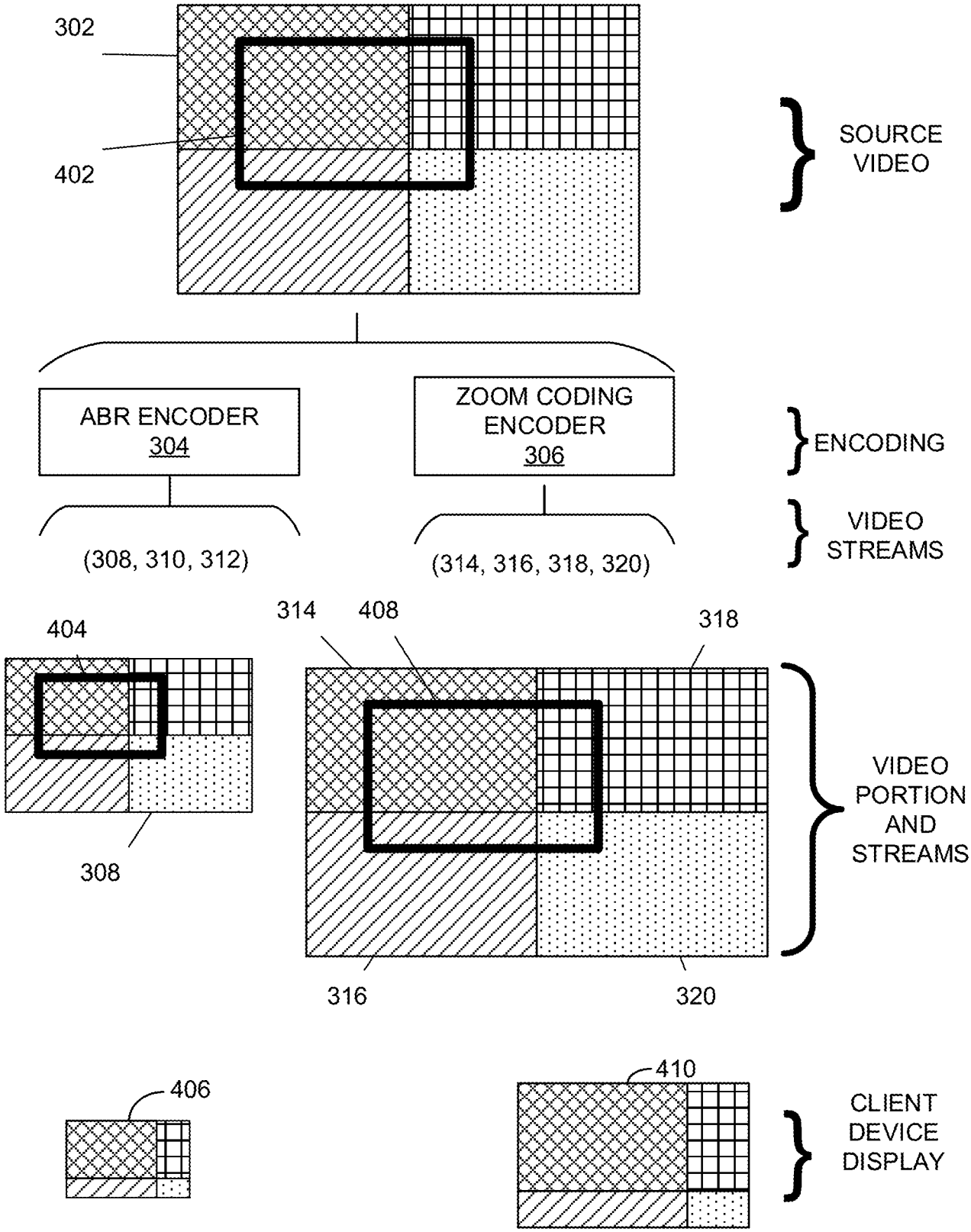

302
402
SOURCE VIDEO
ABR ENCODER 304
ZOOM CODING ENCODER 306
ENCODING
(308, 310, 312)
(314, 316, 318, 320)
VIDEO STREAMS
404
314
408
318
308
316
320
VIDEO PORTION AND STREAMS
406
410
CLIENT DEVICE DISPLAY

FIG. 4A

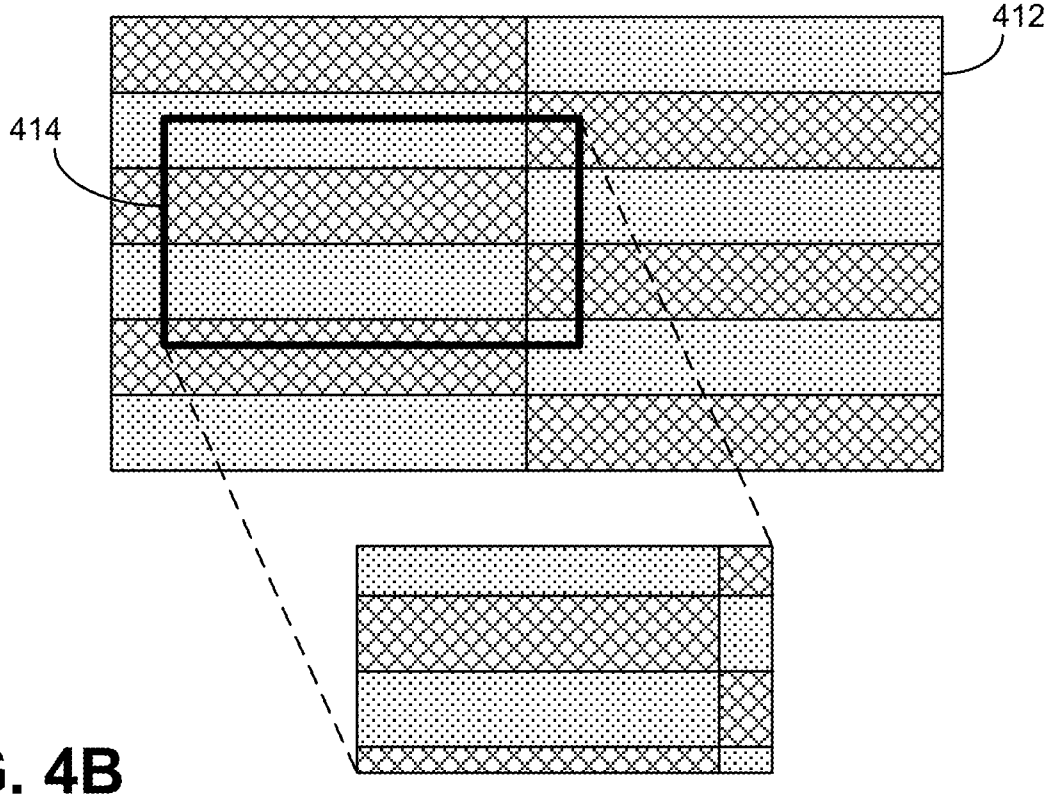


FIG. 4B

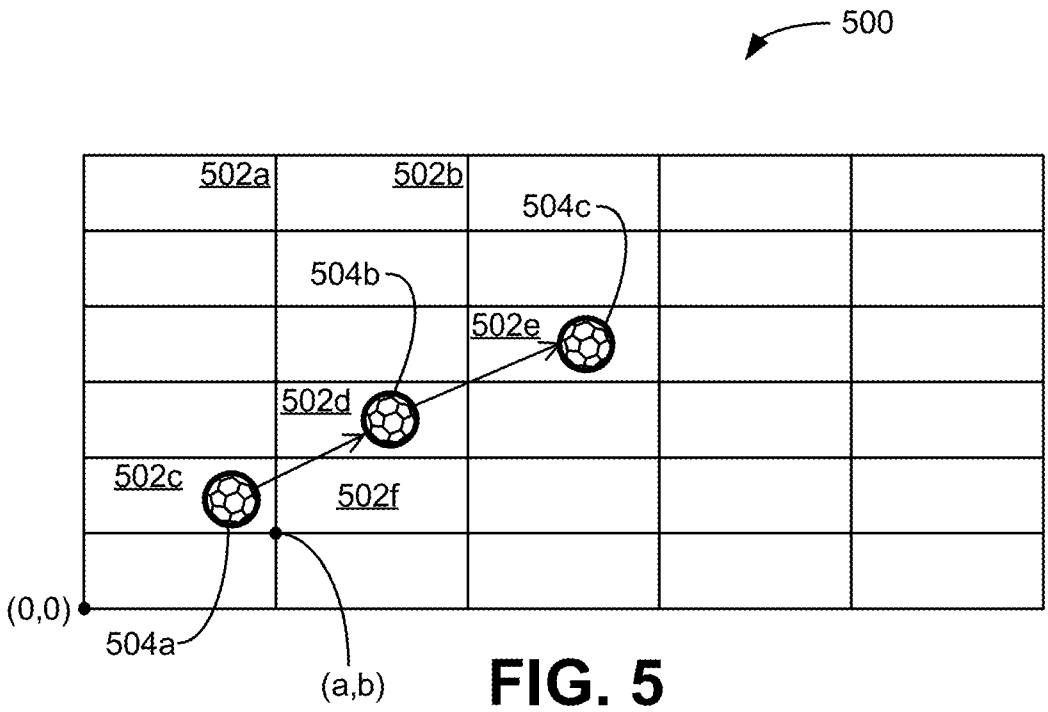


FIG. 5

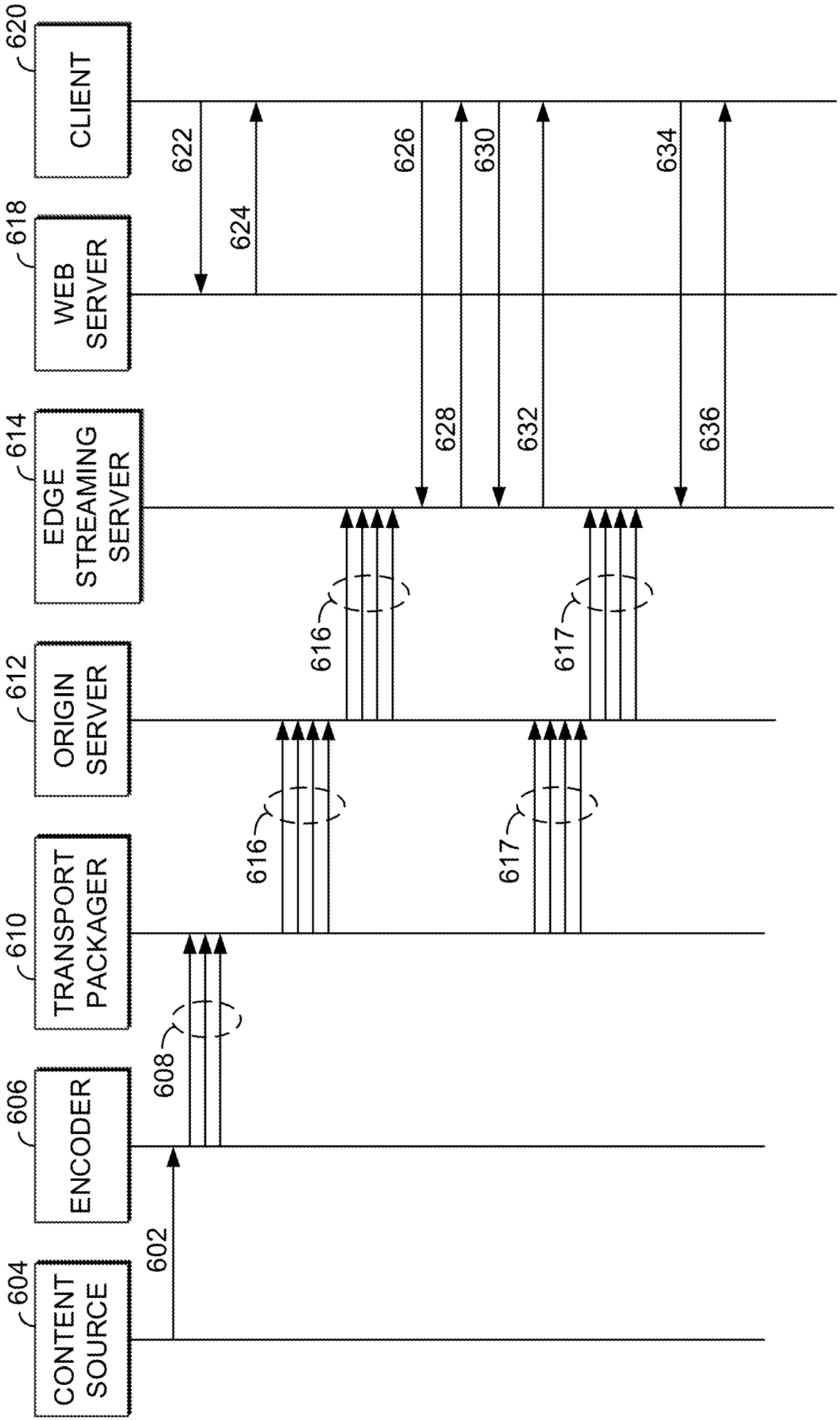

620
CLIENT
618
WEB SERVER
614
EDGE STREAMING SERVER
612
ORIGIN SERVER
610
TRANSPORT PACKAGER
606
ENCODER
604
CONTENT SOURCE
602
608
616
617
622
624
626
628
630
632
634
636

FIG. 6

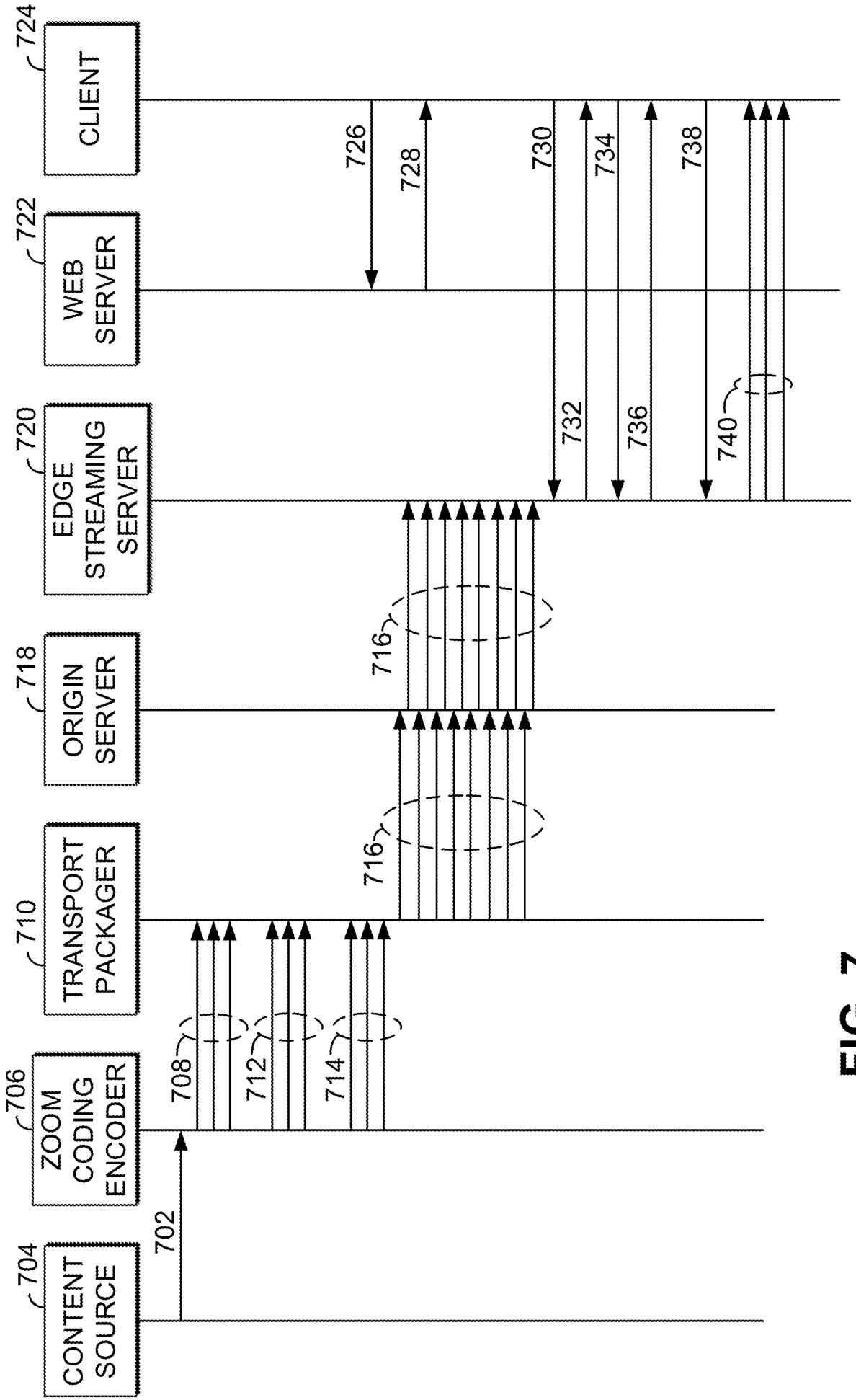

724
CLIENT
722
WEB SERVER
720
EDGE STREAMING SERVER
718
ORIGIN SERVER
710
TRANSPORT PACKAGER
706
ZOOM CODING ENCODER
704
CONTENT SOURCE
702
708
712
714
716
716
726
728
730
732
734
736
738
740

FIG. 7

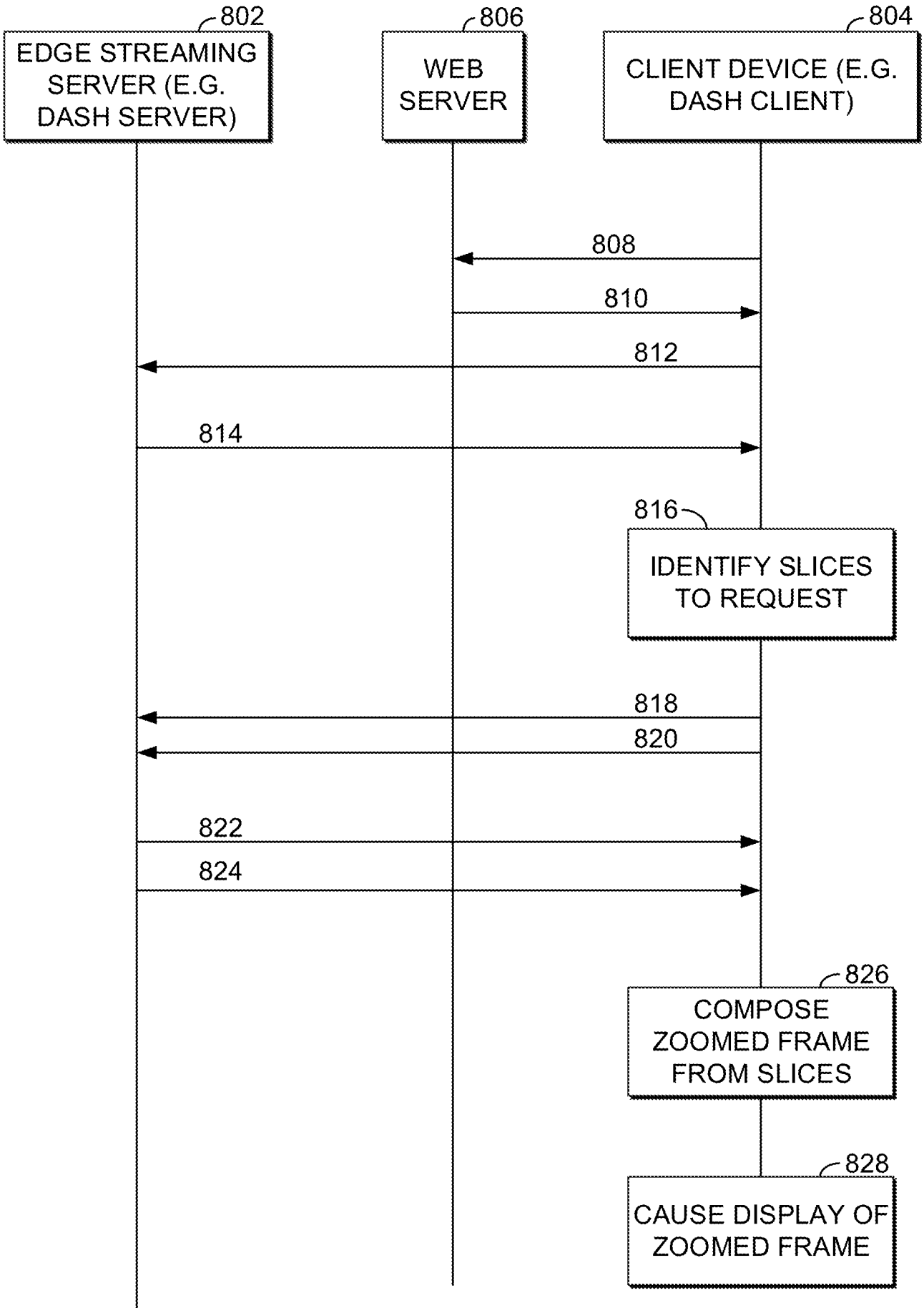

802
EDGE STREAMING SERVER (E.G. DASH SERVER)
806
WEB SERVER
804
CLIENT DEVICE (E.G. DASH CLIENT)
808
810
812
814
816
IDENTIFY SLICES TO REQUEST
818
820
822
824
826
COMPOSE ZOOMED FRAME FROM SLICES
828
CAUSE DISPLAY OF ZOOMED FRAME

FIG. 8

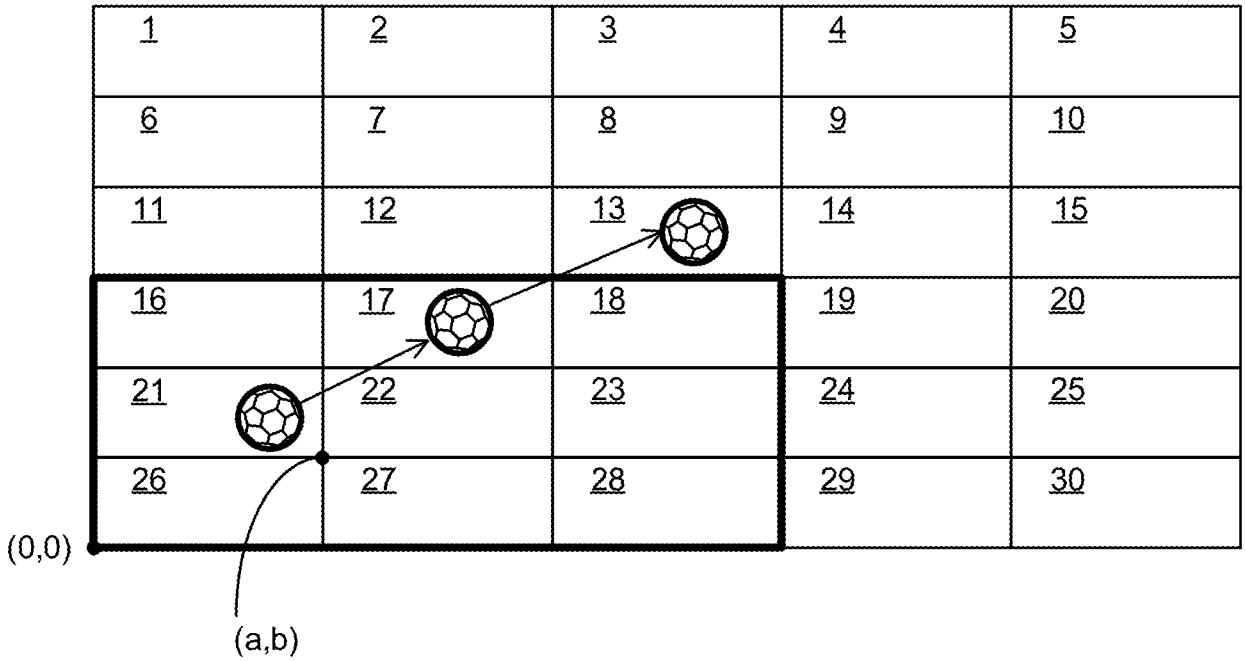

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
(0,0)
(a,b)

FIG. 9

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/046317

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N21/435 H04N21/442 H04N21/4728 H04N21/6373 H04N21/845
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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 2 824 885 A1 (ALCATEL LUCENT [FR]) 14 January 2015 (2015-01-14) figure 1 paragraphs [0001], [0003], [0007], [0008], [0016] - [0019], [0046], [0052] -----	1-15
X	WO 2014/057131 A1 (CANON KK [JP]; DENOUAL FRANCK [FR]; LE FLOCH HERVÉ [FR]; MAZE FRÉDÉRIC) 17 April 2014 (2014-04-17) figures 1b,3,11 page 18, lines 1-15 -----	1-15
X	WO 2015/014773 A1 (KONINKL KPN NV [NL]; TNO [NL]) 5 February 2015 (2015-02-05) page 1, lines 25-30 page 2, line 18 - page 3, line 6 page 13, lines 10-18 ----- -/--	1-15

[X] Further documents are listed in the continuation of Box C.

[X] 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
12 October 2016	19/10/2016

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	Doyle, Walter

1

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

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/046317

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2014/089990 A1 (VAN DEVENTER MATTIJS OSKAR [NL] ET AL) 27 March 2014 (2014-03-27) figures 4,5 -----	1-15
A	ADITYA MAVLANKAR ET AL: "An interactive region-of-interest video streaming system for online lecture viewing", PACKET VIDEO WORKSHOP (PV), 2010 18TH INTERNATIONAL, IEEE, 13 December 2010 (2010-12-13), pages 64-71, XP031899005, DOI: 10.1109/PV.2010.5706821 ISBN: 978-1-4244-9522-1 the whole document -----	1-15

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2016/046317

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 2824885 A1	14-01-2015	NONE	
WO 2014057131 A1	17-04-2014	CN 104904225 A	09-09-2015
		EP 2907313 A1	19-08-2015
		JP 2016500948 A	14-01-2016
		KR 20150070260 A	24-06-2015
		US 2015237166 A1	20-08-2015
		WO 2014057131 A1	17-04-2014
WO 2015014773 A1	05-02-2015	EP 3028472 A1	08-06-2016
		KR 20160032184 A	23-03-2016
		US 2016165309 A1	09-06-2016
		WO 2015014773 A1	05-02-2015
US 2014089990 A1	27-03-2014	CN 103583050 A	12-02-2014
		EP 2719190 A1	16-04-2014
		JP 5905957 B2	20-04-2016
		JP 2014520442 A	21-08-2014
		JP 2016158255 A	01-09-2016
		KR 20140016357 A	07-02-2014
		KR 20150038538 A	08-04-2015
		KR 20150127733 A	17-11-2015
		US 2014089990 A1	27-03-2014
		WO 2012168365 A1	13-12-2012