



(51) International Patent Classification:

G06F 17/30 (2006.01) G06F 3/01 (2006.01)
G06F 17/40 (2006.01) A61B 3/113 (2006.01)
G06F 19/28 (2011.01) H04N 21/278 (2011.01)

(21) International Application Number:

PCT/US2017/035057

(22) International Filing Date:

30 May 2017 (30.05.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/357,259	30 June 2016 (30.06.2016)	US
62/374,687	12 August 2016 (12.08.2016)	US
15/280,962	29 September 2016 (29.09.2016)	US
15/280,933	29 September 2016 (29.09.2016)	US
15/280,947	29 September 2016 (29.09.2016)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:

US 15/280,962 (CON)
Filed on 29 September 2016 (29.09.2016)

(71) Applicant: SONY INTERACTIVE ENTERTAINMENT INC. [JP/JP]; 1-7-1 Konan, Minato-ku, Tokyo, 108-0075 (JP).

(72) Inventor: CASTLEMAN, Dennis D.; c/o Sony Interactive Entertainment America LLC, 2207 Bridgepointe Pkwy., San Mateo, California 94404 (US).

(74) Agent: WAWRZY尼亚K, Richard E. et al.; Fitch, Even, Tabin & Flannery, LLP, 120 S. LaSalle Street, Suite 1600, Chicago, Illinois 60603 (US).

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

(54) Title: APPARATUS AND METHOD FOR GAZE TRACKING

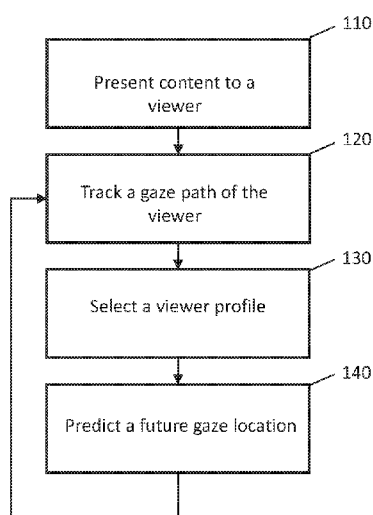


FIG. 1

(57) Abstract: A method for gaze tracking is provided. One embodiment of the method for gaze tracking, comprising: presenting a content to a viewer via a playback device, tracking a gaze path of the viewer with a gaze tracker coupled to the playback device, selecting a viewer profile from a plurality of viewer profiles stored in a viewer profile database based on the gaze path of the viewer, and predicting a future gaze location of the viewer based on the viewer profile. Systems perform similar steps and non-transitory computer readable storage mediums each store one or more computer programs are also provided.

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))*

APPARATUS AND METHOD FOR GAZE TRACKING

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of and claims benefit from United States
5 Patent Application No. 15/280,962, filed on September 29, 2016, entitled
“APPARATUS AND METHOD FOR GAZE TRACKING,” which claims the benefit
of United States Provisional Patent Application No. 62/357,259, filed on June 30,
2016, entitled “APPARATUS AND METHOD FOR CAPTURING AND
10 DISPLAYING SEGMENTED CONTENT”, and which also claims the benefit of
United States Provisional Patent Application No. 62/374,687, filed on August 12,
2016, entitled “APPARATUS AND METHOD FOR PROVIDING AND
DISPLAYING CONTENT”, the entire content and disclosures of which are all
hereby fully incorporated by reference herein in their entireties.

This application is related to United States Patent Application No. 15/280,933,
15 filed on September 29, 2016, entitled “APPARATUS AND METHOD FOR
CAPTURING AND DISPLAYING SEGMENTED CONTENT”, by inventor Dennis
D. Castleman, and identified by Attorney Docket No. 139592 [SCEA16001US01], the
entire content and disclosure of which is hereby fully incorporated by reference herein
in its entirety.

20 This application is also related to United States Patent Application No.
15/280,947, filed on September 29, 2016, entitled “APPARATUS AND METHOD
FOR PROVIDING AND DISPLAYING CONTENT”, by inventor Dennis D.
Castleman, and identified by Attorney Docket No. 139593 [SCEA16003US01], the
entire content and disclosure of which is hereby fully incorporated by reference herein
25 in its entirety.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to motion tracking.
30

2. Discussion of the Related Art

Gaze tracking is the process of tracking a viewer's gaze location using
sensors. The gaze location may then be used as an input to a computer system.

SUMMARY OF THE INVENTION

One embodiment provides a method for gaze tracking comprising: presenting a content to a viewer via a playback device, tracking a gaze path of the viewer with a gaze tracker coupled to the playback device, selecting a viewer profile from a plurality of viewer profiles stored in a viewer profile database based on the gaze path of the viewer, and predicting a future gaze location of the viewer based on the viewer profile.

Another embodiment provides a system for gaze tracking, comprising: a viewer profile database, a communication device, and a processor coupled to the viewer profile database and the communication device. The processor being configured to: present a content to a viewer via a playback device coupled to the communication device, track a gaze path of the viewer with a gaze tracker coupled to the playback device, select a viewer profile from a plurality of viewer profiles stored in the viewer profile database based on the gaze path of the viewer, and predict a future gaze location of the viewer based on the viewer profile.

Another embodiment provides a non-transitory computer readable storage medium storing one or more computer programs configured to cause a processor based system to execute steps comprising: presenting a content to a viewer via a playback device, tracking a gaze path of the viewer with a gaze tracker coupled to the playback device, selecting a viewer profile from a plurality of viewer profiles stored in a viewer profile database based on the gaze path of the viewer, and predicting a future gaze location of the viewer based on the viewer profile.

Another embodiment provides a system for gaze tracking, comprising: a display device, a gaze tracker, a communication device, and a processor coupled to the display device, the gaze tracker, and the communication device. The processor being configured to: present a content to a viewer via a display device, track a gaze path of the viewer with the gaze tracker, select a viewer profile from a plurality of viewer profiles stored in a viewer profile database based on the gaze path of the viewer, predict a future gaze location of the viewer based on the viewer profile, and retrieve, with the communication device, a portion of the content from a content server based on the future gaze location.

A better understanding of the features and advantages of various embodiments

of the present invention will be obtained by reference to the following detailed description and accompanying drawings which set forth an illustrative embodiment in which principles of embodiments of the invention are utilized.

5

BRIEF DESCRIPTION OF THE DRAWINGS

The above and other aspects, features and advantages of embodiments of the present invention will be more apparent from the following more particular description thereof, presented in conjunction with the following drawings wherein:

FIG. 1 is a flow diagram illustrating a method for tracking gaze in accordance with some embodiments of the present invention;

10

FIG. 2 is a block diagram illustrating a system for tracking gaze in accordance with some embodiments of the present invention;

FIG. 3 is an illustration of gaze paths in accordance with some embodiments of the present invention;

15

FIG. 4 is a process diagram illustrating a process for tracking gaze in accordance with some embodiments of the present invention;

FIG. 5 is a flow diagram illustrating a method for providing content in accordance with some embodiments of the present invention; and

FIG. 6 is a flow diagram illustrating a method for providing content in accordance with some embodiments of the present invention.

20

DETAILED DESCRIPTION

Gaze tracking may comprise eye tracking and/or head tracking. In some embodiments, gaze tracking may be used to determine a viewer's point of fixation in a displayed image and/or in a scene of the real world. In virtual reality (VR) and augmented reality (AR) application, the system may use the viewer's gaze location to determine the content to display to the user.

25

Referring first to FIG. 1, a method for tracking gaze is shown. The steps in FIG. 1 may generally be performed by a processor-based device such as one or more of a computer system, a server, a cloud-based server, a content host, a streaming service host, a media server, a playback device, a display device, and the like. In some embodiments, the steps in FIG. 1 may be performed by one or more of the server 210 and the playback device 220 described with reference to FIG. 2, the server described

30

with reference to FIG. 4, and/or other similar devices.

In step 110, the system presents a content to a viewer via a playback device. In some embodiments, the content may comprise one or more of a movie, a TV show, a video clip, prerecorded video content, streaming video content, live-streamed video content, and the like. In some embodiments, the content may comprise a single video stream or a plurality of video streams captured by one or more of a stereoscopic camera system, a panoramic camera system, a surround view camera system, a 360-degree camera system, an omnidirectional camera system, and the like. In some embodiments, the content may be streamed from a content server to the playback device. In some embodiments, the playback device may comprise one or more of a game console, a personal computer, a tablet computer, a television, a head mounted display (“HMD”), an augmented reality device, a virtual reality device, a wearable device, a portable user device, a smartphone, etc. In some embodiments, the playback device may comprise a device and/or system of the type described in United States Patent Application No. 15/085,887, filed on March 30, 2016, entitled “Head-Mounted Display Tracking,” the entire disclosure of which is hereby fully incorporated by reference herein in its entirety.

In step 120, the system tracks a gaze path of the viewer. In some embodiments, the gaze path may be tracked by a gaze tracker coupled to the playback device. In some embodiments, the gaze tracker may be configured track one or more of the viewer’s eye movement, head movement, and body movement. In some embodiments, the gaze tracker may comprise one or more of an image sensor, an optical reflector sensor, a range sensor, an electromyography (EMG) sensor, and an optical flow sensor for detecting eye and/or head movement. In some embodiments, the gaze tracker may comprise an Inertial Measurement Unit (IMU) that measures and reports a body's specific force, angular rate, and/or magnetic field surrounding the body, using a combination of accelerometers and gyroscopes, sometimes also magnetometers. In some embodiments, the gaze tracker may be coupled to an HMD and/or a wearable device that allows the sensor to detect the motion of the user’s head or eyes via the motion of the HMD and/or wearable device. In some embodiments, the gaze tracker may comprise an optical sensor for detecting one or more of a head motion and eye-motion of the user. In some embodiments, the sensor may be coupled to an HMD and/or a wearable device or comprise a relatively stationary device that

captures data from the viewer from a distance. In some embodiments, the gaze path of the viewer may be tracked by any conventional eye tracking, head tracking, and/or motion tracking sensors.

5 In embodiments, the gaze path of the viewer may comprise a record of a plurality of gaze locations and/or directions of the viewer detected over time. In some embodiments, the gaze path of the viewer may comprise gaze locations and/or directions timestamped according to the timestamp of the content viewed by the viewer. In some embodiments, the gaze locations in a gaze path may each be associated with one or more frames of the content. In some embodiments, the gaze
10 path may comprise a two-dimensional gaze path. In some embodiments, the gaze path may comprise gaze locations comprising a coordinate in the display area of the content. In some embodiments, the gaze path may comprise a three-dimensional gaze paths. For example, the content may comprise immersive content in which viewers may navigation through a three-dimension space with body movement and/or input
15 devices. In some embodiments, the gaze path comprises x, y, and z coordinates corresponding to the head and/or eyes of the viewer and a viewing direction.

In step 130, the system selects a viewer profile based on the gaze path of the viewer tracked in step 120. In some embodiments, the viewer profile may be selected from a plurality of viewer profiles associated with the content stored in a viewer
20 profile database. In some embodiments, each viewer profile may comprise an expected gaze path associated with the content. In some embodiments, a viewer profile may be associated with a segment of the content and/or the entire duration of the content. In some embodiments, the expected gaze path may comprise expected fixation points in one or more frames of the content. In some embodiments, the
25 viewer profile may be selected based on matching the user's gaze path tracked in step 120 with the gaze paths in the plurality of viewer profiles during the same time period in the content. In some embodiments, the tracked gaze path may be compared to the viewer profiles by determining the amount of deviation between the tracked gaze path and the gaze paths in the viewer profiles and selecting the viewer profile with the least
30 amount of deviation.

In some embodiments, the viewer profiles may comprise one or more viewer profiles generated based on gaze paths of a plurality of users who viewed the content. For example, the system may record gaze paths of a plurality of viewers that viewed

the same content. The gaze paths may then be grouped by their similarity to each other and a viewer profile may be created for each cluster of similar gaze paths. In some embodiments, the viewer profiles may comprise one or more default viewer profiles created during a post-production process of the content. For example, a
5 producer of the content may manually input a default gaze path based on where he/she believes the viewer's fixation locations will likely be. In some embodiments, prior to a content being release, one or more viewers may view the content during the post production processor to generate a default viewer profile. In some embodiments, the viewer profiles may comprise one or more viewer profiles generated based on
10 tracking a movement of one or more objects in the content. For example, a viewer profile associated with a soccer game may comprise a gaze path that is associated with the movement of the soccer ball during gameplay. In another example, a viewer profile may comprise a gaze path that follows the fastest moving object in the frame. In yet another example, a viewer profile may comprise a gaze path that follows a
15 particular actor or actors in the content. In some embodiments, the content may be processed with an object recognition engine to identify one or more objects in the content. The locations of the identified objects may then be used to generate a gaze path of a viewer profile.

In some embodiments, when a content is initially shown to a viewer, the
20 system may use a default viewer profile to predict the viewer's future gaze location and the viewer profile associated with the viewer may be subsequently adjusted in step 130 when gaze tracking data becomes available. In some embodiments, when the content is initially shown to a viewer, a viewer profile may be selected base on the viewer's demographic information such as age, gender, location, etc. In some
25 embodiments, one or more the default viewer profiles may be generated during the post-production process of the content.

In step 140, the system predicts a future gaze location of the viewer based on the viewer profile selected in step 130. In some embodiments, the predicted future gaze location for a frame of the content may correspond to the gaze location and/or
30 direction of the expected gaze path in the viewer profile selected in step 130. In some embodiments, the expected gaze path in the viewer profile may comprise gaze direction information and the system first identify a portion of the content display area that corresponds to the expected gaze direction to predict the gaze location.

In some embodiments, the future gaze location predicted in step 140 may be used by the system to affect the display of the content. In some embodiments, the system may be configured to select a portion of the content based on the future gaze location and buffer the portion of the content at the playback device. In some
5 embodiments, the portion of the content buffered at the playback device may correspond to a focal area associated with the future gaze location. In some embodiments, a system may be configured to provide a low bitrate version of a content item and a portion of a high bitrate version of the content item corresponding to the focal area to a playback device as described in United States Provisional Patent
10 Application No. 62/374,687, filed on August 12, 2016, entitled “APPARATUS AND METHOD FOR PROVIDING AND DISPLAYING CONTENT” the entire disclosure of which is hereby fully incorporated by reference herein in its entirety. The portion of the high bitrate version may be selected and/or buffered at the playback device based on the future gaze location predicted in step 140.

15 In some embodiments, the portion of the content corresponds to a video stream selected from a plurality of video streams comprising the content. In some embodiments, the content may comprise a plurality of separately encoded video streams and the system may be configured to select one or more streams to provide to the playback device as described in United States Provisional Patent Application No.
20 62/357,259, filed on June 30, 2016, entitled “APPARATUS AND METHOD FOR CAPTURING AND DISPLAYING SEGMENTED CONTENT” the entire disclosure of which is hereby fully incorporated by reference herein in its entirety. The one or more video streams provided to the playback device may be selected based on the future gaze location predicted in step 140.

25 In some embodiments, the system may further determine an insertion location for added content based on the future gaze location. In some embodiments, the added content may comprise one or more of alert information, supplemental information, selection menu, and advertisement content. For example, the system may be configured to insert native advertisement into legacy content based on the predicted
30 gaze location of the viewer. Examples of methods and systems that displays content based on predicted future gaze location according to some embodiments are described with reference to FIGS. 5 and 6 herein.

In some embodiments, after step 140, the process may return to step 120

during the playback of the content. In some embodiments, the system may track the subsequent gaze path of the viewer after selecting the viewer profile and select a different one of the plurality of viewer profiles based on the subsequent gaze path. In some embodiments, the viewer profile selected for a viewer in step 130 may change
5 any number of time as the viewer views the content. In some embodiments, step 130 may be repeated periodically (e.g. 1 second, 2 seconds, 240 frames, etc.). In some embodiments, step 130 may be repeated when a substantial deviation from the expected gaze path is detected.

Referring next to FIG. 2, there is shown a system for gaze tracking that may
10 be used to run, implement, and/or execute any of the methods and techniques shown and described herein in accordance with some embodiments of the present invention. The system includes a server 210 and a playback device 220 communicating over a data connection such as a network.

The server 210 includes a processor 212, a memory 213, and a communication
15 device 214. The server 210 may generally comprise one or more processor-based devices accessible by the playback device via a network such as the Internet. In some embodiments, the content server may comprise one or more of a cloud-based server, a content host, a streaming service host, a media server, a streaming video server, a broadcast content server, a social networking server, and the like. The processor 212
20 may comprise one or more of a control circuit, a central processor unit, a graphical processor unit (GPU), a microprocessor, a video decoder, a video encoder and the like. The memory 213 may include one or more of a volatile and/or non-volatile computer readable memory devices. In some embodiments, the memory 213 stores computer executable code that causes the processor 212 to provide content to the
25 playback device 220. In some embodiments, the communication device 214 may comprise one or more of a network adapter, a data port, a router, a modem, and the like. Generally, the communication device 214 may be configured to allow the processor 212 to communicate with the playback device 220. In some embodiments, the processor 212 may be configured to match a user's gaze path detected by the
30 playback device 220 with a viewer profile stored in the viewer profile database 219. In some embodiments, the server 210 may further comprise a content server configured to provide content to the playback device 220 based on the selected viewer profile. In some embodiments, the server 210 may be configured to stream the content

recorded by a capture device to the playback device 220 in substantially real-time. In some embodiments, the server 210 may be configured to host a plurality of prerecorded content items for streaming and/or downloading to the playback devices 220 on-demand.

5 The viewer profile database 219 may comprise computer readable memory storage storing a plurality of viewer profiles associated with one or more content items. In some embodiments, the viewer profiles may each be associated with a particular content item and comprise an expected gaze path associated with the content. In some embodiments, the expected gaze path may comprise expected
10 fixation points and/or viewing directions in one or more frames of the content. In some embodiments, each viewer profile may be associated with a segment of the content and/or the entire duration of the content.

 In some embodiments, the viewer profiles in the viewer profile database 219 may comprise one or more viewer profiles generated based on gaze paths of a
15 plurality of users who viewed the content. For example, the system may record gaze paths of a plurality of viewers that viewed the same content. The gaze paths may be grouped by their similarity to each other and a viewer profile may be created for each cluster of similar gaze paths. In some embodiments, for broadcasted live-streamed content such as sports events, viewer profiles may be established based on the gaze
20 paths of model viewers who views the content with no or less time delay as compared to the viewers of the broadcasted content. In some embodiments, the viewer profiles in the viewer profile database 219 may comprise one or more default viewer profile created during a post-production process of the content. In some embodiments, prior to a content being released for broadcast, one or more viewers may view the content
25 during the post production processor to generate a default viewer profile. In some embodiments, the viewer profiles in the viewer profile database 219 may comprise one or more viewer profiles generated based on tracking a movement of one or more objects in the content. For example, a viewer profile associated with a soccer game may comprise a gaze path that is associated with the movement of the soccer during
30 gameplay. In another example, a viewer profile in the viewer profile database 219 may comprise a gaze path that follows the fastest moving object in the frame. In yet another example, a viewer profile in the viewer profile database 219 may comprise a gaze path that follows a particular actor or actors in the content. In some

embodiments, the content may be processed with an object recognition engine to identify one or more objects in the content. The locations of the identified objects in the frame of the content may then be used to generate a gaze path for a viewer profile. In some embodiments, one or more viewer profiles in the viewer profile database 219 may be associated with viewer demographic information (e.g. gender, age, location, etc.). If a tracked gaze path is not available, a viewer profile may be selected based on viewer demographic information.

While only one playback device 220 is shown in FIG. 2, in some embodiments, the server 210 may be configured to simultaneously select viewer profiles and/or provide content to a plurality of playback devices 220 via the communication device 214. In some embodiments, the server 210 may be configured to facilitate peer-to-peer transfer of video streams between capture devices and playback devices 220. In some embodiments, the server 210 may be configured add to and/or modify viewer profiles in the viewer profile database 219 based on records of viewer gaze paths received from a plurality of playback devices 220. For example, if a group of viewers shares a similar gaze path for a portion of the content that does not match well with an existing viewer profile in the viewer profile database 219, the server 210 may create a new viewer profile based on the group of similar gaze paths. In some embodiments, the system may use the demographic of viewers to update the demographic associated with one or more viewer profiles. While one server 210 is shown, in some embodiments, functionalities of the server 210 may be implemented on one or more processor-based devices.

The playback device 220 comprises a processor 221, a memory 223, a display device 225, and a sensor device 227. In some embodiments, the playback device 220 may generally comprise a processor-based devices such as one or more of a game console, a media console, a set-top box, a personal computer, a tablet computer, a television, a head mounted display (“HMD”), an augmented reality device, a virtual reality device, a wearable device, a portable user device, a smartphone, etc. The processor 221 may comprise one or more of a control circuit, a central processor unit (CPU), a graphical processor unit (GPU), a microprocessor, a video decoder and the like. The memory 223 may include one or more of a volatile and/or non-volatile computer readable memory devices. In some embodiments, the memory 223 stores computer executable code that causes the processor 221 to track a viewer’s gaze path

with the sensor device 227. In some embodiments, the playback device 220 may be configured predict a further gaze location based on viewer profiles retrieved from the server 210 and/or stored locally. In some embodiments, the playback device 220 may be configured to retrieve content from the server 210 and/or render content locally
5 based on the predicted gaze location of the viewer. In some embodiments, the memory 223 may comprise a buffer for buffering, from the server 210, a portion of the content being view on the display device 225. In some embodiments, the computer executable code stored in the memory 223 may comprise one or more of a computer program, a software program, a playback device firmware, a mobile
10 application, a game and/or media console application, etc.

The display device 225 may comprise a device for displaying content to a viewer. In some embodiments, the display device 225 may comprise one or more of a monitor, a television, a head mounted display (HMD), a virtual reality display device, a wearable device, a display screen, a mobile device, and the like. In some
15 embodiments, the display device 225 may comprise a stereoscopic display having one or more screens.

The sensor device 227 may comprise one or more sensors configured to track the gaze of a viewer of the display device 225. In some embodiments, the sensor device 227 may comprise one or more of an image sensor, an optical reflector sensor,
20 a range sensor, an electromyography (EMG) sensor, and an optical flow sensor for detecting eye and/or head movement. In some embodiments, the sensor device 227 may comprise an IMU that measures and reports a body's specific force, angular rate, and/or magnetic field surrounding the body, using a combination of accelerometers and gyroscopes, sometimes also magnetometers. In some embodiments, the sensor
25 device 227 may be coupled to an HMD and/or a wearable device that allows the sensor to detect the motion of the user's head or eyes via the motion of the HMD and/or wearable device. In some embodiments, the sensor device 227 may comprise a camera for detecting one or more of a head motion and eye-motion of the user. In some embodiments, the sensor device 227 may be coupled to an HMD and/or a
30 wearable device or be a relatively stationary device that captures data from the viewer from a distance.

While the display device 225 is shown as part of the playback device 220, in some embodiments, the display device 225 may comprise a separate device with or

without a separate processor. In some embodiments, the display device 225 may be coupled to the playback device 220 via a wired or wireless communication channel. For example, the playback device 220 may comprise a PC or a game console and the display device 225 may comprise an HMD configured to display content from the playback device 220. In some embodiments, the sensor device 227 may be part of the playback device 220, the display device 225, and/or may be a physically separated device communicating with one or more of the playback device 220 and the display device 225. In some embodiments, one or more of the display device 225 and the sensor device 227 may be integrated with the playback device 220. In some embodiments, the display device 225 may further comprise a processor and/or a memory for at least partially storing the retrieved content and/or the viewer's eye or head movement detected by the sensor device 227.

In some embodiments, the playback device 220 may further include a communication device such as a network adapter, a Wi-Fi transceiver, a mobile data network transceiver, etc. for communicating with the server 210 and/or a capture device. In some embodiments, the playback device 220 may further include one or more user input/output devices such as a motion tracker, buttons, a controller, a keyboard, a display screen, a touch screen and the like for the user to control the selection and playback of content items. In some embodiments, the input devices may further be used by the viewer to navigate in an immersive reality content.

Referring next to FIG. 3, there is shown an illustration of gaze paths in accordance with some embodiments of the present invention. In FIG. 3, the content area 300 represents the display area of a content and gaze paths 301, 303, and 302 represent gaze paths covering the same time period of the content playback and associated with different viewer profiles. In FIG. 3, if a viewer's tracked gaze path corresponds to the gaze path 310, the system may match the viewer to the viewer profile associated with the gaze path 302. The viewer's future gaze locations may then be predicted based on continuing to follow the gaze path 302 with time. In some embodiments, the viewer may be associated with the gaze path 302 until a substantial deviation from the gaze path 302 is detected.

The gaze paths shown in FIG. 3 are provided for illustration only and may not necessarily represent actual gaze paths. While two-dimensional gaze paths are shown, in some embodiments, the gaze paths associated with viewer profiles may comprise

gaze paths in the three-dimension space. For example, the content may comprise immersive content in which viewers may navigation through a three-dimension space. In some embodiments, the gaze locations comprising the gaze path may then each include x, y, and z coordinates from a point of reference in the virtual space and a viewing direction. While the gaze paths in FIG. 3 are shown as thin lines, in some
5 embodiments, a gaze path may comprise a wide band of area and/or viewing directions.

Referring next to FIG. 4, there is shown a process diagram in accordance with some embodiments of the present invention. In some embodiments, the server in FIG.
10 4 may comprise the server 210 described with reference to FIG. 2 or a similar device. The playback device in FIG. 4 may comprise the playback device 220 described with reference to FIG. 2 or a similar device. In FIG. 4, the gaze tracker 410 of the playback device first detects the user's gaze locations and determine a gaze path 412. The gaze path 412 is then provided to the server. In some embodiments, each gaze path data
15 point may be transmitted to the server as they are collected and the server may aggregate the data points to form a gaze path. In step 416, the server selects a viewer profile by matching the gaze path 412 with a plurality of viewer profiles 414. In step 420, the server predicts a further gaze location of the viewer based on the selected viewer profile. In step 430, the system selects a content portion to provide to the
20 playback device based on the predicted gaze location. In some embodiments, the content portion may be retrieved and/or extracted from the content file 432. In some embodiments, the content portion may correspond to one or more video streams comprising the content. In some embodiments, the content portion may comprise a portion of a high bitrate version of the content. In step 434, the content portion is
25 buffered at the playback device. In step 436, the content, including with the content portion, is displayed to the user.

In some embodiments, one or more of the steps 416, 420, and 430 may be performed by the playback device. For example, a plurality of viewer profiles may be preloaded on the playback device. The playback device may be configured to predict
30 gaze locations based on the detected gaze path 412 and the plurality of locally stored viewer profiles. In some embodiments, the playback device may be configured send a request to the server that specifies the content portion based on the predicted gaze location. In some embodiments, instead of buffering retrieved content, the playback

device may be configured to begin rendering the content based on the predicted gaze location using locally stored and/or streamed content files.

Referring next to FIG. 5, a method for providing content is shown. The steps in FIG. 5 may generally be performed by a processor-based device such as a one or more of a head mounted display (“HMD”), an augmented reality device, a virtual reality device, a wearable device, a portable user device, a smartphone, a personal computer, a server device, a television, a tablet computer, a game console, etc. In some embodiments, the steps in FIG. 5 may be performed by one or more of the server 210 and the playback device 220 described with reference to FIG. 2, the server described with reference to FIG. 4, and/or other similar devices.

In step 510, the system predicts a viewing direction of a user viewing a content item. In some embodiments, the viewing direction of the user may comprise one or more of a head direction and an eye direction. In some embodiments, the viewing direction may comprise a predicted direction based the future gaze location predicted in the method described with reference to FIG. 1 or a similar method. In some embodiments, a viewer gaze path comprising a history of the viewer’s viewing directions may be compared to a plurality of gaze paths in viewer profiles. If the viewer’s gaze path matches a gaze path in a viewer profile, future viewing directions of the viewer may be predicted based on the matching gaze path.

The content item may comprise a plurality of video streams captured by one or more of a panoramic camera system, a surround view camera system, a 360-degree camera system, an omnidirectional camera system, and the like. In some embodiments, the content item may comprise a segmented content corresponding one or more of a panorama video, an immersive video, and/or an omnidirectional video. In some embodiments, each of the plurality of video streams may be captured by a separate image sensor of a camera assembly. In some embodiments, the content item may comprise pre-recorded or real-time streaming content. The content item may comprise a plurality of individually encoded video streams hosted on a server accessible to a playback device.

In some embodiments, after step 510, the system may determine whether one or more video streams of the content item falls within the predicted view area of the user viewing the content item. In some embodiments, the user’s predicted view area may correspond to an area around a fixation point corresponding to the viewing

direction predicted based on the selected viewer profile. In some embodiments, the viewing direction may be based on a user's input via one or more input devices. For example, a user may use a joystick, a touchpad, a mouse, direction key, and the like to change the viewing direction. The view area may then correspond to the display area associated with the viewing direction. In some embodiments, if only one video stream falls within the view area, the system may only retrieve and/or decode the select video stream to display to the user. In some embodiments, the system may perform steps 520-540 only if two or more video streams fall within the view area of the user.

In step 520, the system selects two or more video streams of the content item. The selection of the video streams may be based on the viewing direction of the user predicted in step 510 and directional data associated with the plurality of video streams of the content item. The directional data of the plurality of video streams may be recorded by the capture system that records the content item and included as part of the content item file. In some embodiments, the capture system may include one or more IMUs for recording the orientation of the captured system during the recording of the content item. The IMU data associated with the current section of the content item may be compared with the predicted viewing direction of the user to determine the video streams that corresponds to the predicted viewing direction of the user.

In some embodiments, selecting the two or more video streams comprises loading the selected video streams to a buffer or cache to decode. In some embodiments, selecting the two or more video streams comprises retrieving the selected streams from a server that hosts the content item. In some embodiments, steps 510 and 520 may be performed by the content server, the playback device, and/or a third device.

In step 530, the system decodes the two more video streams selected in step 520 to form two or more decoded video streams. In some embodiments, the decoding may be performed by an image or video decoder hardware and/or software module on the playback device.

In step 540, the system stitches the two or more decoded video streams to form a combined image. In some embodiments, the video streams may be stitched according to known stitching methods such as methods comprising one or more of aligning, warping, and blending of the video streams. In some embodiments, the aligning of the images may be performed at the playback device via feature and/or

keypoint detection.

In some embodiments, the system may further be configured to selectively stitch portions of the video streams based on the predicted view area of the user. For example, if only a portion of the overlapping area of the two video streams falls within the predicted view area of the user, the system may only blend the portion of the overlaying area that is estimated to be visible to the user based on the prediction. In some embodiments, the degree of blending may further depend on the area's position in the field of view. For example, a rough (e.g. less computationally demanding) blending may be performed if the overlapping area falls in the peripheral area of the user's field of view and a more extensive blending may be performed if the overlapping area falls near the expected fixation point of the user. In some embodiments, the stitching of the two or more decoded video streams further comprises combining images based on warp and distortions associated with a playback device.

In step 550, the system causes the combined image generated in step 540 to be displayed to the user. In some embodiments, the image may be displayed on a flat screen display device or a head mounted display. In some embodiments, prior to step 550, the system further adjusts the combined image based on the parameters associated with the display device. In some embodiments, the system may generate stereoscope images for a 3D and/or virtual reality display device. In some embodiments, the system may adjust the resolution of the combined image based on the display resolution of the display device. In some embodiments, the system may adjust the images based on warp and distortion associated with the display device.

In some embodiments, steps 510-550 may be continuously repeated as the user changes his/her viewing direction while viewing the content item. In some embodiments, the viewer profiles selected to predict user's viewing direction in step 510 may change any number of times during a viewing of the content item. In some embodiments, depending on the viewing direction predicted in step 510, the video streams selected, decoded, and stitched in steps 520-540 may change over time.

Referring next to FIG. 6, a method for providing content is shown. The steps in FIG. 6 may generally be performed by a processor-based device such as one or more of a computer system, a server, a cloud-based server, a content host, a streaming service host, a media server, and the like. In some embodiments, the steps in FIG. 6

may be performed by one or more of the server 210 and the playback device 220 described with reference to FIG. 2, the server described with reference to FIG. 4, and/or other similar devices.

5 In step 610, the system predicts a future gaze location a viewer of a content item. In some embodiments, the gaze location of the user may comprise one or more of a head direction and an eye direction. In some embodiments, the future gaze location may be predicted based the method described with reference to FIG. 1 or similar methods. In some embodiments, a gaze path comprising a history of the viewer's gaze locations may be compared to a plurality of gaze paths in viewer
10 profiles. If the viewer's gaze path matches a viewer profile gaze a path, future gaze locations the viewer may be predicted based on the matching gaze path. In some embodiments, steps 610 may be performed by the content server, the playback device, and/or a third device.

In step 620, the playback device retrieves a low bit rate version of a content
15 item. In some embodiments, a playback device sends a content request to a server hosting the content item in step 620 to retrieve the content item. The low bit rate version of the content item may comprise a reduced bit rate version of the content item generated by a content provider and/or the hosting service. In some embodiments, step 620 may occur prior to step 610 and the low bit rate version of the
20 content item may begin to be downloaded, buffered, and/or viewed prior to the focal area of the viewer being determined or predicted.

In step 630, the playback device retrieves a portion of a high bit rate version of the content item. In some embodiments, the playback device sends a content request identifying the focal area of the viewer predicted in step 610 to a server to retrieve the
25 portion of the high bit rate version of the content item. In some embodiments, the server may be configured to predict a focal area and send the corresponding portion of the high bit rate version of the content item to the playback device. Generally, the retrieved portion may comprise a spatial portion of the content selected based on the predicted focal area of the viewer. In some embodiments, the retrieved portion may
30 comprise a short temporal segment of an area of the content item (e.g. milliseconds, seconds, frames, etc.). In some embodiments, the portion of the high bit rate version of the content item may be retrieved in a video stream separately encoded from the low bit rate version of the content item retrieved in step 620. In some embodiments,

the low bit rate version of the content item may buffer ahead of the retrieval of the high bit rate version of the content item.

In step 640, the system combines the portion of the high bit rate version of the content item with the low bit rate version of the content item to generate a combined image. In some embodiments, in step 640, the system first decodes the portion of the high bit rate version of the content item retrieved in step 630 and the low bit rate version of the content item retrieved in step 620. In some embodiments, if the resolution and/or framerate of the low and high bit rate versions of the content item are different, the system may first adjust the resolution and/or framerate of at least one of the versions prior to combining the images.

In step 650, the combined image is displayed on a display device. In some embodiments, the display device may comprise one or more of a monitor, a television set, a projector, a head mounted display (HMD), a virtual reality display device, a wearable device, a display screen, a mobile device, and the like. In some embodiments, prior to step 650, the system may further adjust the combined image based on the display device's specifications. For example, for virtual reality display devices, the system may adjust for the warp and distortions associated with the device.

In some embodiments, steps 610 to 650 may be repeated continuously as a viewer views a content item. In some embodiments, the viewer profiles selected to predict user's focal area in step 610 may change any number of times during a viewing of the content item. In some embodiments, based on the focal area predicted in step 610, different portions of the high bit rate version of the content item may be retrieved in step 630 and combined with the low bit rate version in step 640 over time. In some embodiments, step 620 may occur independently of steps 610 and 630.

In some embodiments, the system may further be configured to predict a view area of the viewer and retrieve only a portion of the low bit rate content based on a view area of the viewer in step 620. The view area of the viewer may be predicted based on matching the viewer to a viewer profile similar to the prediction of the focal area in step 610. The view area of the viewer may generally refer to the area of the content that is visible to the user but may or may not be in focus in the viewer's field of vision.

In some embodiments, one or more of the embodiments, methods, approaches,

and/or techniques described above may be implemented in one or more computer programs or software applications executable by a processor based apparatus or system. By way of example, such processor based apparatus or systems may comprise a computer, entertainment system, game console, workstation, graphics workstation, server, client, portable device, pad-like device, etc. Such computer program(s) may be used for executing various steps and/or features of the above-described methods and/or techniques. That is, the computer program(s) may be adapted to cause or configure a processor based apparatus or system to execute and achieve the functions described above. For example, such computer program(s) may be used for implementing any embodiment of the above-described methods, steps, techniques, or features. As another example, such computer program(s) may be used for implementing any type of tool or similar utility that uses any one or more of the above described embodiments, methods, approaches, and/or techniques. In some embodiments, program code macros, modules, loops, subroutines, calls, etc., within or without the computer program(s) may be used for executing various steps and/or features of the above-described methods and/or techniques. In some embodiments, the computer program(s) may be stored or embodied on a computer readable storage or recording medium or media, such as any of the computer readable storage or recording medium or media described herein.

Therefore, in some embodiments the present invention provides a computer program product comprising a medium for embodying a computer program for input to a computer and a computer program embodied in the medium for causing the computer to perform or execute steps comprising any one or more of the steps involved in any one or more of the embodiments, methods, approaches, and/or techniques described herein. For example, in some embodiments the present invention provides one or more non-transitory computer readable storage mediums storing one or more computer programs adapted or configured to cause a processor based apparatus or system to execute steps comprising: presenting a content to a viewer via a playback device, tracking a gaze path of the viewer with a gaze tracker coupled to the playback device, selecting a viewer profile from a plurality of viewer profiles stored in a viewer profile database based on the gaze path of the viewer, and predicting a future gaze location of the viewer based on the viewer profile.

While the invention herein disclosed has been described by means of specific

embodiments and applications thereof, numerous modifications and variations could be made thereto by those skilled in the art without departing from the scope of the invention set forth in the claims.

CLAIMS

What is claimed is:

1. A method for gaze tracking, comprising:
 - 5 presenting a content to a viewer via a playback device;
 - tracking a gaze path of the viewer with a gaze tracker coupled to the playback device;
 - selecting a viewer profile from a plurality of viewer profiles stored in a viewer profile database based on the gaze path of the viewer; and
 - 10 predicting a future gaze location of the viewer based on the viewer profile.
2. The method of claim 1, wherein the viewer profile comprises an expected gaze path associated with the content.
- 15 3. The method of claim 1, wherein the plurality of viewer profiles comprises one or more viewer profiles generated based on gaze paths of a plurality of users who viewed the content.
4. The method of claim 1, wherein the plurality of viewer profiles comprises
 - 20 one or more default viewer profile created during a post-production process of the content.
5. The method of claim 1, wherein the plurality of viewer profiles comprises one or more viewer profiles generated based on tracking a movement of one or more
 - 25 objects in the content.
6. The method of any one of claims 1-5, further comprising:
 - selecting a portion of the content based on the future gaze location; and
 - buffering the portion of the content at the playback device.
- 30 7. The method of claim 6, wherein the portion of the content corresponds to a focal area associated with the future gaze location.

8. The method of claim 6, wherein the portion of the content corresponds to a video stream selected from a plurality of video streams comprising the content.

9. The method of any one of claims 1-5, further comprising:
5 determining an insertion location for added content based on the future gaze location.

10. The method of any one of claims 1-5, further comprising:
tracking a subsequent gaze path of the viewer after selecting the viewer
10 profile; and
selecting a different one of the plurality of viewer profiles based on the subsequent gaze path.

11. A system for gaze tracking, comprising:
15 a viewer profile database;
a communication device; and
a processor coupled to the viewer profile database and the communication device, the processor being configured to:
present a content to a viewer via a playback device coupled to the
20 communication device;
track a gaze path of the viewer with a gaze tracker coupled to the playback device;
select a viewer profile from a plurality of viewer profiles stored in the viewer profile database based on the gaze path of the viewer; and
25 predict a future gaze location of the viewer based on the viewer profile.

12. The system of claim 11, wherein the viewer profile comprises an expected gaze path associated with the content.

30 13. The system of claim 11, wherein the plurality of viewer profiles comprises one or more viewer profiles generated based on gaze paths of a plurality of users who viewed the content.

14. The system of claim 11, wherein the plurality of viewer profiles comprises one or more default viewer profile created during a post-production process of the content.

5 15. The system of claim 11, wherein the plurality of viewer profiles comprises one or more viewer profiles generated based on tracking a movement of one or more objects in the content.

10 16. The system of any one of claims 11-15, wherein the processor is further configured to:
 select a portion of the content based on the future gaze location; and
 buffer the portion of the content at the playback device.

15 17. The system of claim 16, wherein the portion of the content corresponds to a focal area associated with the future gaze location.

 18. The system of claim 16, wherein the portion of the content corresponds to a video stream selected from a plurality of video streams comprising the content.

20 19. The system of any one of claims 11-15, wherein the processor is further configured to:
 determine an insertion location for added content based on the future gaze location.

25 20. The system of any one of claims 11-15, wherein the processor is further configured to:
 track a subsequent gaze path of the viewer after selecting the viewer profile;
 and
 select a different one of the plurality of viewer profiles based on the
30 subsequent gaze path.

 21. A non-transitory computer readable storage medium storing one or more computer programs configured to cause a processor based system to execute steps

comprising:

- presenting a content to a viewer via a playback device;
- tracking a gaze path of the viewer with a gaze tracker coupled to the playback device;
- 5 selecting a viewer profile from a plurality of viewer profiles stored in a viewer profile database based on the gaze path of the viewer; and
- predicting a future gaze location of the viewer based on the viewer profile.

22. A system for gaze tracking, comprising:

- 10 a display device;
- a gaze tracker;
- a communication device; and
- a processor coupled to the display device, the gaze tracker, and the communication device, the processor being configured to:
- 15 present a content to a viewer via the display device;
- track a gaze path of the viewer with the gaze tracker;
- select a viewer profile from a plurality of viewer profiles stored in a viewer profile database based on the gaze path of the viewer;
- predict a future gaze location of the viewer based on the viewer profile;
- 20 and
- retrieve, with the communication device, a portion of the content from a content server based on the future gaze location.

23. The system of claim 22, wherein the plurality of viewer profiles comprises
25 one or more viewer profiles generated based on gaze paths of a plurality of users who viewed the content.

24. The system of claim 22, wherein the plurality of viewer profiles comprises
one or more default viewer profile created during a post-production process of the
30 content.

25. The system of claim 22, wherein the plurality of viewer profiles comprises
one or more viewer profiles generated based on tracking a movement of one or more

objects in the content.

26. The system of any one of claims 22-25, wherein the processor is further configured to:

5 buffer a portion of the content that is selected based on the future gaze location.

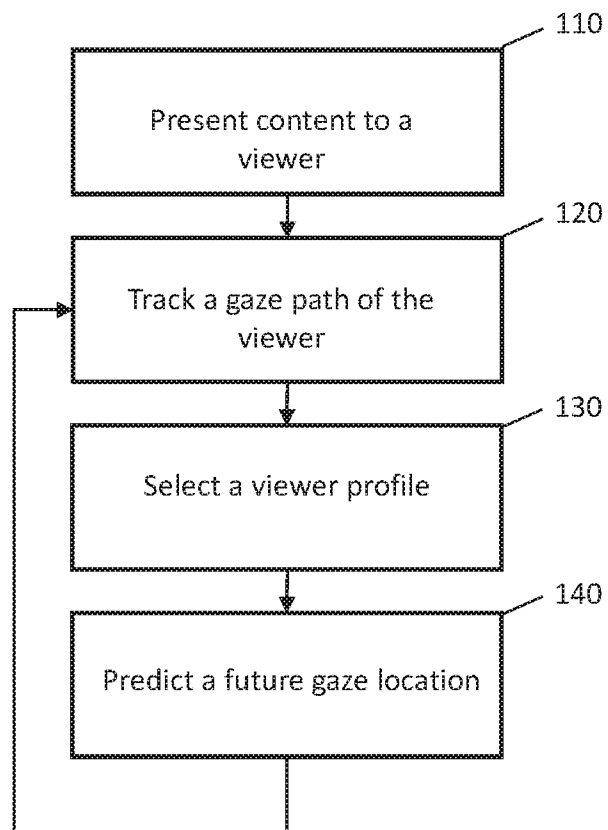
27. The system of any one of claims 22-25, wherein the processor is further configured to:

10 track a subsequent gaze path of the viewer after selecting the viewer profile;
and

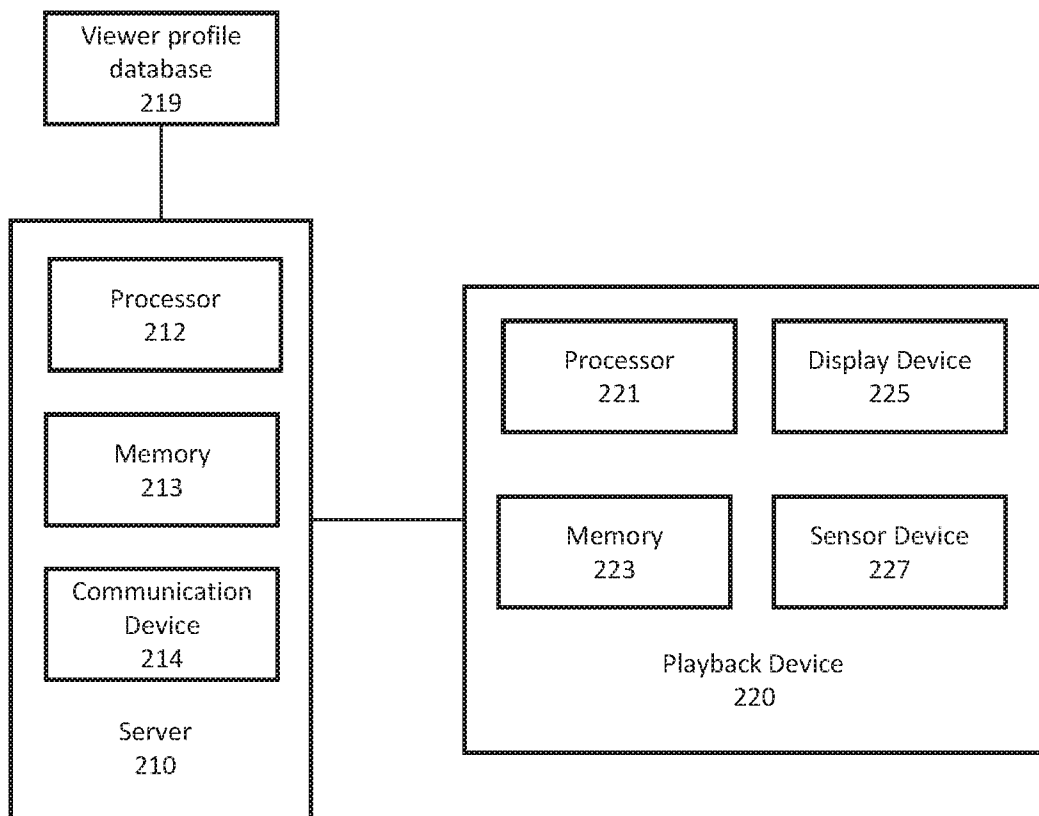
 select a different one of the plurality of viewer profiles based on the subsequent gaze path.

15

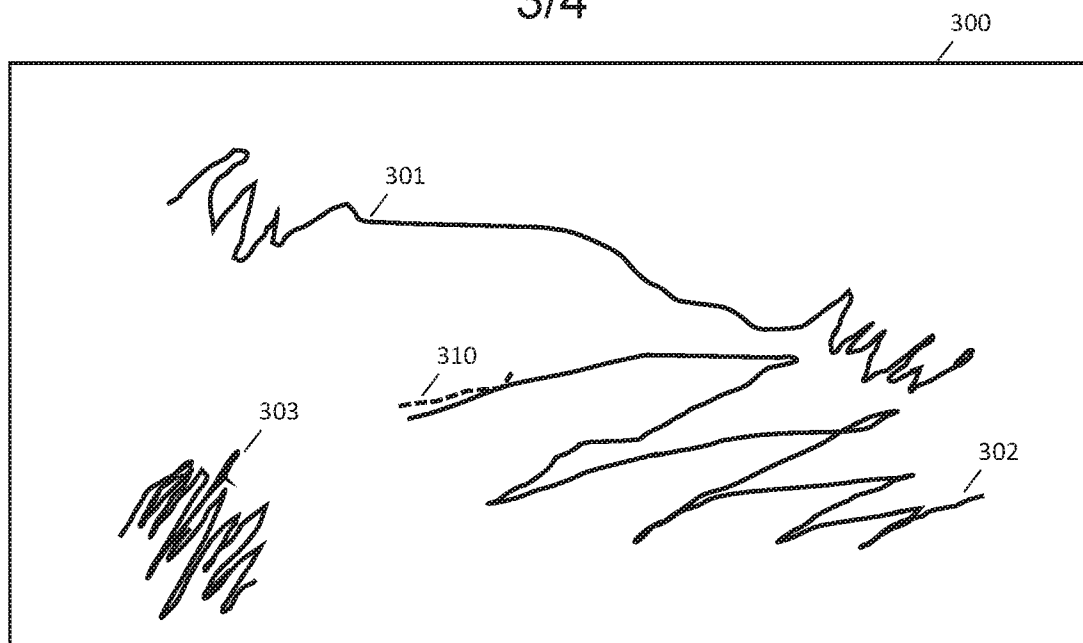
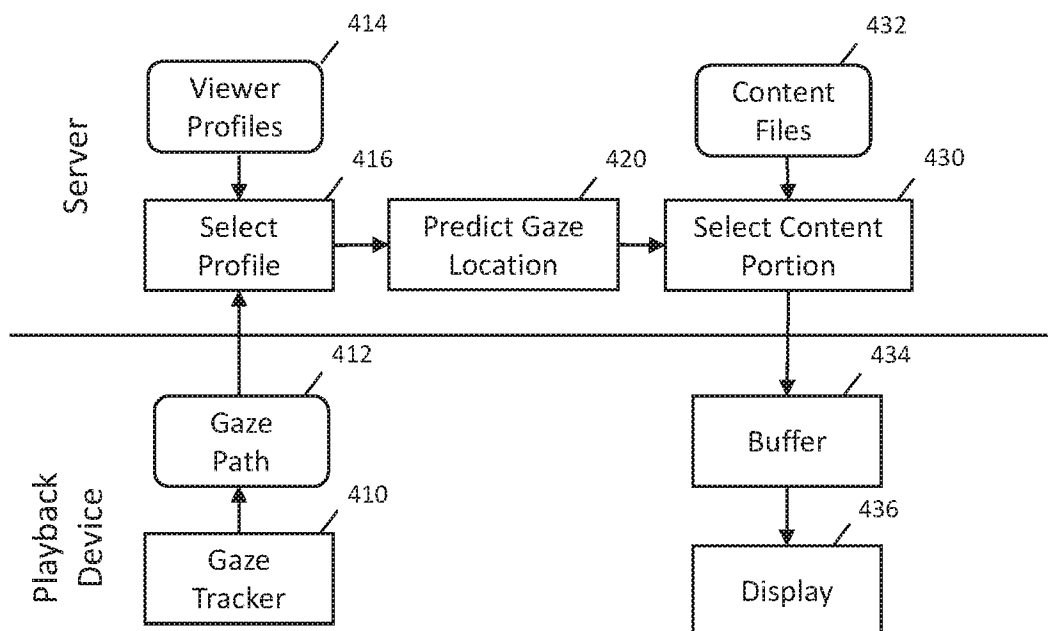
1/4

**FIG. 1**

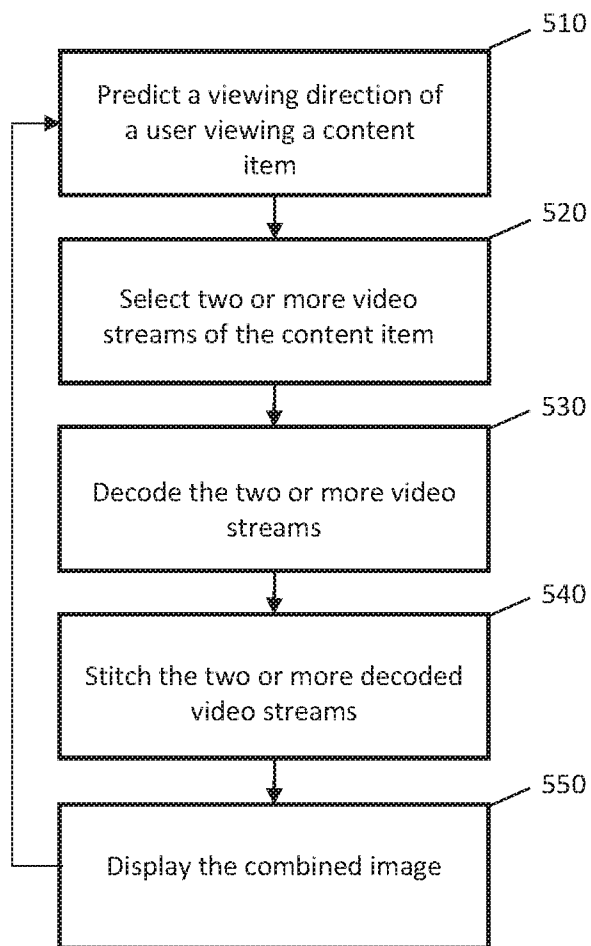
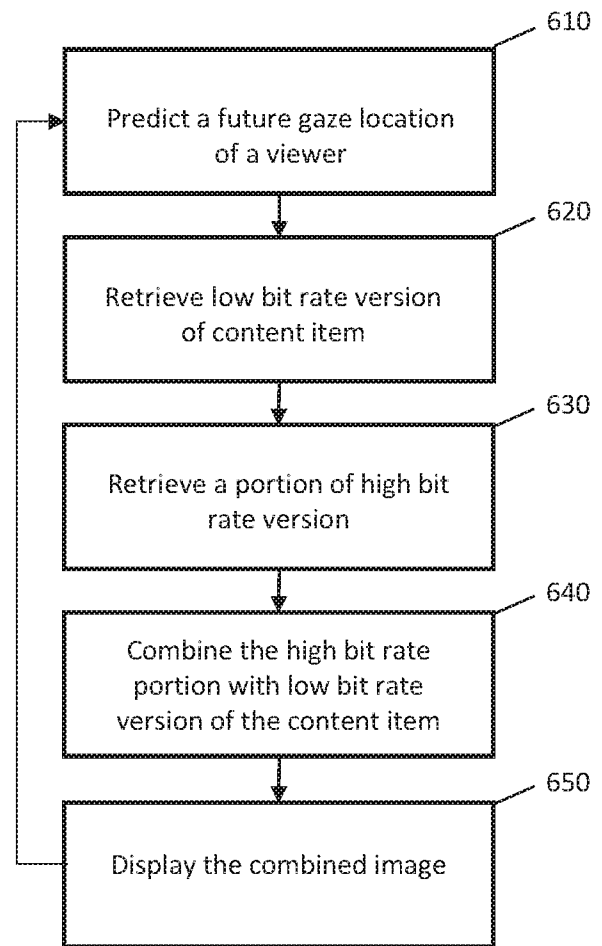
2/4

**FIG. 2**

3/4

**FIG. 3****FIG. 4**

4/4

**FIG. 5****FIG. 6**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US17/35057

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G06F 17/30, 17/40, 19/28, 3/01; A61B3/113, H04N 21/278 (2017.01)
 CPC - G06F 17/30, 17/40, 19/28, 3/01; A61B3/113, H04N 21/278

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)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2005/0273185 A1 (TEIWES, W et al.) 08 December 2005; paragraphs [0036], [0041]	1-27
Y	US 9,462,230 B1 (AMAZON TECHNOLOGIES, INC) 04 October 2016; column 1 lines 42-45, column 2 lines 1-5	1-27
Y	CIG et al. "Gaze-Based Virtual Task Predictor" GAZEIN, ACM; Publication [online]. November 2014 [retrieved 26 July 2017]. Retrieved from the Internet: <URL: https://ui.ku.edu.tr/sezgin_publications/2014/ICMI_GazeIn.pdf >; pp 1-6	3, 5, 6/3, 6/5, 7/6/3, 7/6/5, 8/6/3, 8/6/5, 9/3, 9/5, 10/3, 10/5, 13, 15, 16/13, 16/15, 17/16/13, 17/16/15, 18/16/13, 18/16/15, 19/13, 19/15, 20/13, 20/15, 23, 25, 26/23, 26/25, 27/23, 27/25

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

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

26 July 2017 (26.07.2017)

Date of mailing of the international search report

23 AUG 2017

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774