



(51) International Patent Classification:

H04N 5/225 (2006.01) *H04N 5/232* (2006.01)

(21) International Application Number:

PCT/US2018/020418

(22) International Filing Date:

01 March 2018 (01.03.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/468,773 08 March 2017 (08.03.2017) 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). **KEROFSKY, Louis**; 11347 Carmel Creek, San Diego, California 92130 (US).(74) Agent: **STECK, Jeffrey Alan**; 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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, 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, 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: ZOOM CODING USING SIMULTANEOUS AND SYNCHRONOUS MULTIPLE-CAMERA CAPTURES

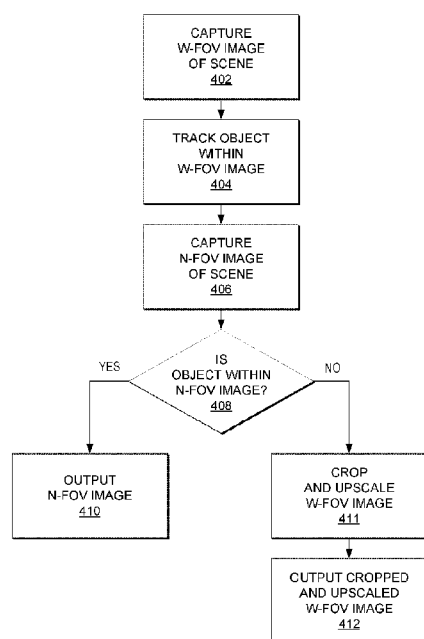


FIG. 4

(57) Abstract: Systems and methods described herein disclose use of simultaneous and synchronous multiple-camera captures for a zoom region. An exemplary method using two field of view (FOV) video streams of a scene, where the second FOV is narrower than the first, comprises: tracking an object captured within the first FOV; responsive to determining that the object is entirely within the second FOV, outputting video corresponding to the second FOV; and responsive to determining that the object is outside the second FOV, outputting a cropped and up-scaled representation of video corresponding to the first FOV. Systems and methods disclosed herein, prior to tracking the object, display video captured for the first FOV and receive user input indicating selection of an object to be tracked in the displayed video for the first FOV.

ZOOM CODING USING SIMULTANEOUS AND SYNCHRONOUS MULTIPLE-CAMERA CAPTURES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application is a non-provisional filing of, and claims benefit under 35 U.S.C. §119(e) from, U.S. Provisional Patent Application Serial No. 62/468,773, entitled “Zoom Coding Using Simultaneous and Synchronous Multiple Camera Captures,” filed March 8, 2017, the entirety of which is incorporated herein by reference.

BACKGROUND

[0002] In many video distribution systems, images are captured by lenses with different optical parameters, such as different visual zoom factors. Different camera optics affect the quality and appearance of the images captured by the cameras. These images may undergo further digital enhancement, such as digitally changing the zoom and field of view (FOV) parameters.

[0003] At times, multiple different video cameras can be used to capture the same event. The multiple different video cameras may have different optical capabilities, such as different optical zoom settings. The multiple video cameras may also be on the same device, such as a smart phone having two cameras oriented in the same direction.

SUMMARY

[0004] An exemplary method comprises capturing a first video of a scene with a first camera on a user device, where the first camera has a first field of view (e.g. a wide-angle field of view) with a first extent (e.g. a field of view between around 64° and 84° in the vertical and/or horizontal direction). A second camera (e.g. a zoom camera) on the user device simultaneously captures a video of a portion of the same scene. The second camera has a second field of view having a second extent that is narrower than the first extent (e.g. less than around 64°). The second extent may be represented in terms of pixels within the first video. The second field of view is contained within the first field of view, such objects within the field of view of the second camera are also within the field of view of the first camera. The method further comprises tracking a position of an object of interest with respect to the first and second fields of view. This may be done using, for example, RFID or other wireless signals, or performing optical tracking using object recognition, among other possibilities. In the exemplary method, a determination is made of whether the object of interest is positioned

entirely within the second field of view. In the method, a video output is provided. If the object of interest is positioned entirely within the second field of view, the video output comprises the second video. Otherwise, a cropped-and-scaled version of the first video is generated such that the video output includes the entirety of the object of interest and such that the extent of the output video corresponds to the second extent, and the cropped-and-scaled version of the first video is provided as the output.

[0005] An exemplary method for providing an output video stream in a device having a first camera and a second camera, the second camera having a field of view that is narrower than and contained within a field of view of the first camera, comprises tracking an object in video captured by the first camera. A determination, based on the tracking, is made on whether the object is entirely within the field of view of the second camera. If the object is within the field of view of the second camera, video captured by the second camera is outputted as the output video stream. If the object is outside the field of view of the second camera, video captured by the first camera is cropped and upscaled to generate a cropped-and-upscaled video of the object, and the generated video is outputted as the output video stream.

[0006] For one embodiment, the method is performed in real time. The method may also include, before tracking the object, displaying, on a screen of the device, video captured by the first camera and receiving user input selecting the object to be tracked in the displayed video. For one embodiment, the output video stream may be displayed on the device. For one embodiment, an indication of the position of the object may be displayed if the object is outside the field of view of the second camera. For one embodiment, tracking the object may be done using RFID or other wireless protocol signals.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIG. 1A is a functional block diagram of a client device that may be used in some embodiments.

[0008] FIG. 1B is a functional block diagram of a network entity that may be used in some embodiments.

[0009] FIG. 2 depicts a client device with multiple video cameras, in accordance with an embodiment.

[0010] FIG. 3A depicts a first system architecture for use of multiple video cameras, in accordance with an embodiment.

[0011] FIG. 3B depicts a second system architecture for use of multiple video cameras, in accordance with an embodiment.

[0012] FIG. 3C depicts a third system architecture for use of multiple video cameras, in accordance with an embodiment.

[0013] FIG. 4 depicts a method, in accordance with an embodiment.

[0014] FIG. 5A depicts selection between W-FOV and a N-FOV views at a first time, in accordance with an embodiment.

[0015] FIG. 5B depicts selection between W-FOV and a N-FOV views at a second time, in accordance with an embodiment.

[0016] FIGs. 6A to 6C depict views of a user interface display on a video capture device, in accordance with embodiments.

[0017] FIGs. 7A and 7B depict embodiments of views of a user interface display on a video capture device.

[0018] FIGs. 8A and 8B depict embodiments of views of a user interface display on a video capture device.

[0019] FIGs. 9A and 9B depict embodiments of views of a user interface display on a video capture device.

DETAILED DESCRIPTION

[0020] 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.

[0021] Note that various hardware elements of one or more of the described embodiments are referred to as “modules” that carry out (i.e., perform, execute, and the like) various functions that are described herein in connection with the respective modules. As used herein, a module includes hardware (e.g., one or more processors, one or more microprocessors, one or more microcontrollers, one or more microchips, one or more application-specific integrated circuits (ASICs), one or more field programmable gate arrays (FPGAs), one or more memory devices) deemed suitable by those of skill in the relevant art for a given implementation. Each described module may also include instructions executable for carrying out the one or more functions described as being carried out by the respective module, and it is noted that those instructions could take the form of or include hardware (i.e., hardwired) instructions, firmware instructions, software instructions, and/or the like, and may be stored in any suitable non-transitory computer-readable medium or media, such as commonly referred to as RAM, ROM, etc.

Exemplary Client and Server Hardware

[0022] Note that various hardware elements of one or more of the described embodiments are referred to as “modules” that carry out (i.e., perform, execute, and the like) various functions that are described herein

in connection with the respective modules. As used herein, a module includes hardware (e.g., one or more processors, one or more microprocessors, one or more microcontrollers, one or more microchips, one or more application-specific integrated circuits (ASICs), one or more field programmable gate arrays (FPGAs), one or more memory devices) deemed suitable by those of skill in the relevant art for a given implementation. Each described module may also include instructions executable for carrying out the one or more functions described as being carried out by the respective module, and it is noted that those instructions may take the form of or include hardware (i.e., hardwired) instructions, firmware instructions, software instructions, and/or the like, and may be stored in any suitable non-transitory computer-readable medium or media, such as commonly referred to as RAM, ROM, etc.

[0023] Exemplary embodiments disclosed herein are implemented using one or more wired and/or wireless network nodes, such as a wireless transmit/receive unit (WTRU) or other network entity.

[0024] FIG. 1A is a system diagram of an exemplary WTRU 102, which may be employed as a client device, video capture device, or other components in embodiments described herein. As shown in FIG. 1A, the WTRU 102 may include a processor 118, a communication interface 119 including 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 sensors 138. It will be appreciated that the WTRU 102 may include any sub-combination of the foregoing elements while remaining consistent with an embodiment.

[0025] 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 WTRU 102 to operate in a wireless environment. The processor 118 may be coupled to the transceiver 120, which may be coupled to the transmit/receive element 122. While FIG. 1A 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.

[0026] The transmit/receive element 122 may be configured to transmit signals to, or receive signals from, a base station over the air interface 116. 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. It will be appreciated that the transmit/receive element 122 may be configured to transmit and/or receive any combination of wireless signals.

[0027] In addition, although the transmit/receive element 122 is depicted in FIG. 1A as a single element, the WTRU 102 may include any number of transmit/receive elements 122. More specifically, the WTRU 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 116.

[0028] 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 WTRU 102 may have multi-mode capabilities. Thus, the transceiver 120 may include multiple transceivers for enabling the WTRU 102 to communicate via multiple RATs, such as UTRA and IEEE 802.11, as examples.

[0029] The processor 118 of the WTRU 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 WTRU 102, such as on a server or a home computer (not shown).

[0030] 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 WTRU 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, and the like.

[0031] 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 WTRU 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 116 from a base station 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 WTRU

102 may acquire location information by way of any suitable location-determination method while remaining consistent with an embodiment.

[0032] 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 sensors such as 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.

[0033] FIG. 1B depicts an exemplary network entity 190 that may be used in embodiments of the present disclosure, for example as an encoder, transport packager, origin server, edge streaming server, web server, or client device as described herein. As depicted in FIG. 1B, 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 other communication path 198.

[0034] 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.

[0035] 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.

[0036] 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. 1B, data storage 196 contains program instructions 197 executable by processor 194 for carrying out various combinations of the various network-entity functions described herein.

Exemplary Video Systems and Methods

[0037] FIG. 2 depicts a client device with two video cameras, in accordance with an embodiment. In alternative embodiments, a client device may have a greater number of video cameras. In some embodiments, client device 202 may be a WTRU such as WTRU 102 of FIG. 1A. The client device 202 includes a front side opposing the back side, the front side having a display and a user interface. The client device 202 includes two cameras as peripherals on the back side, a first camera 206 having a relatively wider field of view (FOV) and a second camera 204 having a relatively narrower FOV that is contained entirely within the FOV of the first camera. The second camera 204 may be a narrow field of view (FOV) camera, and the first camera may be a wide FOV camera. The client device 202 may be a smart phone, a tablet, a dedicated video recording device, or the like. In an exemplary embodiment, the wider FOV (W-FOV) camera 206 includes a 28-mm equivalent lens with optical settings of f/1.8 and the narrower FOV (N-FOV) camera 204 includes a 56-mm equivalent lens with optical settings of f/2.8, thus having a doubling (2x) zoom function over the W-FOV camera 206. In some embodiments, either one or both of the N-FOV and W-FOV camera optics are variable, with each lens able to adjust the zoom within an optical zoom range. The different cameras may also have different zoom capabilities, spatial orientations, or different spatial resolutions.

[0038] In an exemplary embodiment, the device provides a streaming video output of a selected object of interest. In one such embodiment, the first camera captures relatively wider FOV (W-FOV) video, and the position of the selected object is tracked within the W-FOV video. Based on this tracking information, the device determines when the selected object is positioned within the field of view of the second camera, which has a relatively narrower FOV (N-FOV), and when the selected object is positioned only within the field of view of the first, wider-FOV camera. While the selected object is entirely within the field of view of the second camera, the N-FOV video is used to generate the streaming output video, and while the selected object is not entirely within the field of view of the second camera, the W-FOV video is used to generate the streaming output video. While the W-FOV video is used to generate the streaming output video, the W-FOV video is processed by cropping the video down to a zoom region that contains the object of interest and by upscaling the zoom region to match a zoom factor of the output provided from the second field of view camera.

[0039] In some embodiments, the entire spatial extent of the N-FOV video is used as the output video (when the tracked object is within the N-FOV). In other embodiments, the N-FOV video is cropped and/or is upscaled or downscaled for use as the output video (e.g. as a result of image stabilization or to provide a digital zoom effect). In either case, the W-FOV video is cropped and upscaled to match the output generated using the N-FOV video. For example, if the output generated using the N-FOV video is $M \times N$ pixels, then, in an exemplary embodiment, the W-FOV video is cropped and upscaled to $M \times N$ pixels. In one such case, the N-FOV camera has a zoom factor of two as compared to the W-FOV camera. In that case, the output of the

W-FOV camera may be cropped down to a $\frac{1}{2}M \times \frac{1}{2}N$ zoom region and upsampled by a factor of two to give an $M \times N$ output.

[0040] FIG. 3A depicts a first system architecture for use of two video cameras, in accordance with an embodiment. In particular, FIG. 3A depicts the system 300 that includes the N-FOV camera 204, 302 and the W-FOV camera 206, 304 from FIG. 2, and also includes a W-FOV decompression module 306, a tracker module 312, a zoom region selection module 310, a crop and upscale module 314, an output selection module 318, a N-FOV decompression module 308, an alignment and disparity compensation module 316, a filter and smooth module 320, and an output video 322.

[0041] In the embodiment of FIG. 3A, both the W-FOV camera 302 and the N-FOV camera 304 produce compressed video output. The W-FOV camera 302 provides a compressed video output to the W-FOV decompression module 306 and the N-FOV camera 304 provides a compressed video output to the N-FOV decompression module 308. The decompression modules 306, 308 may decompress the compressed video into baseband frames. The decompressed video data obtained from the W-FOV camera 302 is provided to both the tracker module 312 and the crop and upscale module 314. The decompressed video data obtained from the N-FOV camera 304 is provided to the alignment and disparity compensation module 316.

[0042] FIG. 3B depicts a second system architecture for use of two video cameras, in accordance with an embodiment. FIG. 3B depicts the system architecture 330 that is similar to the system 300 of FIG. 3A; however, the W-FOV camera 332 and the N-FOV camera 334 output uncompressed video, and the respective decompression modules 304, 314 are not used to decompress the video into baseband frames. In such an embodiment, the W-FOV camera 332 provides uncompressed video to the tracker module 312 and the crop and upscale module 314 without using a W-FOV decompression module. Additionally, the N-FOV camera 334 may provide uncompressed video to the alignment and disparity compensation module 316 without using an N-FOV decompression module. In some embodiments, only one of the W-FOV camera 332 and the N-FOV camera 334 provides compressed video output, and the other one provides uncompressed video output. In such an embodiment, the system 330 may be modified to include a decompression module 304, 314 for the camera providing the compressed video, as disclosed in FIG. 3A.

[0043] Returning to FIG. 3A, the tracker module 312 determines location coordinates of the selected object within the W-FOV camera 302 view as the object moves within the camera view. The tracking of the object may be based on image recognition or other object tracking techniques. In some embodiments, the position of the object is determined on a frame-by-frame basis. In other embodiments, the position of the object is determined less frequently. In some embodiments, the tracker module 312 is assisted in tracking the selected object based on inputs from a radio frequency identification device (RFID) tag on the object. For example, an athlete may be wearing an RFID tag in communication with an RFID location system. Based on

the determined location of the athlete's RFID tag and the orientation of the cameras, the location of the athlete may be determined.

[0044] The object to be tracked may be selected using a variety of techniques. For example, the tracked object may be determined by a user interacting with a touchscreen display, for example by tapping onto an object or icon displayed on a touchscreen display. Example tracked objects may be a particular person (e.g., a sports player), or a particular region (e.g., a goal, a basketball hoop, a sports ball). Additional user interfaces may be used to select the zoom regions. In some embodiments, a voice recognition module interprets instructions to select a certain person or a certain team's goal. In other embodiments, the tracked object is automatically selected based on preset criteria for tracking objects of interest.

[0045] The zoom region selection module 310 determines a region within the W-FOV video to be used as the zoom region. In an exemplary embodiment, the zoom region is a region that has the same aspect ratio as the N-FOV output video and that is smaller (in linear dimensions) than the N-FOV output video by an amount equal to the relative zoom factor of the N-FOV output video. For example, if the output generated using the N-FOV video is $M \times N$ pixels and the relative zoom factor of the N-FOV camera is z (including any digital zoom performed on the N-FOV prior to output), then, in an exemplary embodiment, the zoom region has a size of $M/z \times N/z$. The zoom region selection module 310 further operates to determine a position of the zoom region within the W-FOV video. In some embodiments, the position of the zoom region is selected such that the tracked object is substantially centered in the zoom region.

[0046] The zoom region selection module 310 then provides data related to the zoom region to the tracker module 312 and to the crop and upscale module 314.

[0047] The crop and upscale module 314 crops the video obtained by the W-FOV camera 302 to include the zoom region. The cropping of the video stream is updated based on inputs received from the tracker module 312. The position of the zoom region may be provided to the crop and upscale module as coordinate points of a bounding box around the zoom region. In some embodiments, each frame from the W-FOV baseband video is cropped. The cropped video is then upscaled by a factor proportional to the ratio of zoom factors, which may be a ratio of the W-FOV focal length and the N-FOV focal length. In the example using the W-FOV camera 302 and the N-FOV camera 304, the upscale factor would be a factor of two, doubling the size of the W-FOV cropped image. The crop and upscale module 314 provides the cropped and upscaled video of the zoom region to the output selection module 318.

[0048] The alignment and disparity compensation module 316 determines the time alignment and location disparity of the N-FOV video data relative to the W-FOV image in order to overlay onto the spatial coordinate system of the W-FOV video. Thus, the N-FOV video may be overlaid on the cropped and upscaled W-FOV video after being aligned. This permits the transition between the selected videos.

[0049] The output selection module 318 makes a determination of whether to output the cropped and upscaled video captured from the W-FOV camera 302 or to output the video captured from the N-FOV camera 304. In response to a determination that the zoom region is captured within the field of view of the N-FOV camera 304, the N-FOV output is selected to be the output and provided to the filter and smooth module 320. If the zoom region is not captured within the N-FOV camera 304, the cropped and upscaled video output of the zoom region captured by the W-FOV camera 302 is selected to be the output video and provided to the filter and smooth module 320.

[0050] The output selection module 318 may operate in different ways depending on the type of information available from the tracker module. For example, in some embodiments, the tracker module may operate to provide a detailed perimeter (e.g. in the form of a polygon) of the tracked object. In such an embodiment, the object may be considered to be outside the N-FOV if any portion of the polygon is outside the N-FOV. In another embodiment, the object may be considered to be outside the N-FOV if at least a threshold percentage of the polygon is outside the N-FOV. In some embodiments, the tracker module may provide only coordinates for the tracked object. In some such embodiments, the object may be considered to be outside the N-FOV if the coordinates are outside the N-FOV. In other such embodiments, a bounding box may be defined and centered on the coordinates, and the object may be considered to be outside the N-FOV if any part of the bounding box is outside the N-FOV. The dimensions of the bounding box may be predefined dimensions, they may be dimensions based on an estimate of the object's apparent size (e.g. generated by the tracker module), or they may be set by a user, among other options.

[0051] The filter and smooth module 320 provides an output video 322 to be displayed. The output video 322 may be generated using either the video captured by the W-FOV camera 302, or the video captured by the N-FOV camera 304, as determined by the output selection module 318. During a transition between the two different video sources, the filter and smooth module 320 processes the video to provide a smooth transition between the video from the W-FOV camera 302 and the N-FOV camera 304.

[0052] In some embodiments, the output video 322 is displayed on a screen of the video recording device. In other embodiments, different video may be displayed on the user-interface display of the video recording device. In one such example, the video recording device may display the video obtained from the W-FOV camera 302. Additional annotations may be displayed over the W-FOV video on the display device. For example, the zoom region may be indicated with a rectangle box around the zoom region, and the location of the N-FOV may be similarly indicated. The output video 322 may be saved on the video capture device for later editing, may be streamed to the Internet or a remote party (or device), or may be output in other ways.

[0053] In some embodiments, the W-FOV video is displayed on a screen of the device, and a visual indicator is also displayed as an overlay on the video to indicate when the tracked object is not within the N-

FOV. The indicator may be in the format of a highlighted box or some other visual cue. The indicator prompts the user to move the camera in the appropriate direction such that the tracked object may be recaptured, or within the field of view, of the N-FOV camera 304. In one such embodiment, the client device capturing the video with both the N-FOV and W-FOV cameras is a single device, such as a smartphone, and the smartphone's display is displaying a representation of video captured from the W-FOV camera, with the indications being displayed on the smartphone's touchscreen display of the zoom region and the N-FOV view.

[0054] In some embodiments, a higher frame rate camera is used with a lower frame rate camera. The cameras may be offset and the output videos of the cameras are combined to create a higher frame rate signal. The combined video is corrected for spatial disparity and lens variability.

[0055] FIG. 3C depicts a third system architecture for use of multiple video cameras, in accordance with an embodiment. In particular, FIG. 3C depicts the system 360 that is similar to the systems 300 and 330 but includes the fine crop and upscale module 362. In some embodiments, video captured by the N-FOV camera 304 is provided to the fine crop and upscale module 362. The fine crop and upscale module 362 receives data from the tracker module 312 regarding the location of the tracked object, similar to the data received by the crop and upscale module 314. The fine crop and upscale module 362 then crops and upscales the video captured by the N-FOV camera 304 to more closely zoom in on the tracked object. The fine crop and upscale module 362 then provides a cropped and zoomed version of the N-FOV video to the output selection module 312 for later output.

[0056] FIG. 4 depicts a method, in accordance with an embodiment. In particular, FIG. 4 depicts the method 400. The steps of method 400 may be accomplished with the systems depicted in FIGs. 3A-C, or other similar system, as known by those with skill in the art. At 402, a W-FOV image of a scene is captured. The W-FOV image may be a frame of video captured by the W-FOV camera 206. The location of the zoom region is tracked within the W-FOV view at 404, and portion of the W-FOV view is determined to be a zoom region (e.g., by the zoom region selection module 308). Additionally, a N-FOV image of the scene is captured at 406. The N-FOV image may be a frame of video captured by the N-FOV camera 204.

[0057] For one embodiment, a process 408 determines if the tracked object is within the N-FOV image. If the tracked object is within the N-FOV image, the N-FOV image is output at 410. If the zoom region is not within the N-FOV image, a cropped and upscaled W-FOV image is generated in step 411 and output at 412. While FIG. 4 illustrates an embodiment in which the W-FOV image is cropped and upscaled only if the object is not within the N-FOV, in other embodiments, the W-FOV video is cropped and upscaled regardless of whether the object is within the N-FOV (although that cropped and upscaled video may not be output). In some embodiments, a filtered and smoothed video is output during a transition between outputting the N-FOV video and the W-FOV video.

[0058] In some embodiments, the N-FOV camera capturing the video may include a variable zoom. The zoom settings of the N-FOV camera may be changed to attempt to keep the tracked object within the N-FOV. For example, the N-FOV camera may zoom out if the tracked object has moved outside of the N-FOV. If the N-FOV camera optics are unable to change any further to obtain the tracked object within the N-FOV view, a cropped and upscaled W-FOV video may then be output. The size of the zoom region may be adjusted for different zoom settings of the N-FOV camera.

[0059] The methods 300, 330, 360, 400 of FIGs. 3A, 3B, 3C, and 4 may be performed in real time as the video is being captured by the client device, without intermediate storage of the W-FOV and N-FOV videos. For one embodiment, the methods 300, 330, 360, 400 of FIGs. 3A, 3B, 3C, and 4 may be performed within milli-seconds. For one embodiment, the methods 300, 330, 360, 400 of FIGs. 3A, 3B, 3C, and 4 may capture video with a N-FOV camera and a W-FOV camera and store the video in memory for analysis (which may include cropping, upscaling, filtering, and smoothing) at a later point in time. A non-transitory, computer-readable medium capable of storing video captured by a N-FOV camera and video captured by a W-FOV camera may be used. For one embodiment, outputting the output video stream comprises recording the output video stream.

[0060] FIGs. 5A and 5B illustrate an embodiment in which a basketball player 501 has been selected as a tracked object. In FIG. 5A, a W-FOV camera 582 captures W-FOV video of a scene, and a N-FOV camera 584 captures N-FOV video of the scene at a first time t1. Frame 500 of W-FOV video and frame 504 of N-FOV video are captured substantially simultaneously. Frame 500 of W-FOV video, on the left, includes images of two basketball players, and frame 504 of the N-FOV video, on the right, encompasses only one of those basketball players, tracked player 501. The location of the N-FOV within the W-FOV is depicted at 502 with a dashed rectangle. While the players and the camera are in the relative positions illustrated in FIG. 5A, with player 501 being within the N-FOV, video frames captured by the N-FOV camera are used to generate the output video 506. While a cropped and upscaled image 508 of tracked player 501 may be generated from the W-FOV frame, that image may be expected to have lower quality (represented schematically by pixelation) than the image captured using the N-FOV camera.

[0061] In the example of FIG. 5B, the W-FOV camera 582 and the N-FOV camera 584 respectively capture a frame 550 of W-FOV video and a frame 554 of N-FOV video at a second time t2. Frame 550 has been captured shortly after frame 500, with the player 501 who was within the N-FOV in FIG. 5A having moved towards the left within the W-FOV image, such that the tracked player 501 is no longer within the N-FOV 502 and thus is not visible in frame 554. The player's movement is tracked in the W-FOV video using video object tracking to determine the player's position in frame 550. While the players and the camera are in the relative positions illustrated in FIG. 5B, with player 501 being outside the N-FOV, video frames captured by the W-FOV camera are used to generate the output video. In using frame 550 to generate the W-FOV, a

zoom region is determined based on the position of player 501, and areas of frame 550 outside the zoom region are cropped away. The remaining zoom region is upsampled to match the zoom-characteristics of the N-FOV camera, resulting in cropped-and-upsampled frame 558. The cropping and upscaling processes generate cropped-and-upsampled frame 558, which is included as frame 560 of the output video. Although cropped and upsampled frame 558 may be of a lower quality than frame 554, frame 558 is chosen for the output video because it includes the tracked object of interest.

[0062] FIGs. 6A to 6C depict views of a user interface display on a video capture device 650, which may be a WTRU. The video capture device 650 includes both a N-FOV camera and a W-FOV camera.

[0063] FIG. 6A depicts a first view of the user interface, in accordance with an embodiment. In the embodiment of FIG. 6A, the video capture device 650 has a screen 652 on which the W-FOV video is being displayed in real time (as the video is captured). Also displayed on the screen 652 is a rectangle 662 that indicates the field of view of the N-FOV camera, which is narrower than and contained within the field of view of the W-FOV camera. In this example, the N-FOV camera has a zoom factor of two compared to the W-FOV camera, and the two cameras have the same aspect ratio, so the rectangle 662 subtends half the width and half the height of the W-FOV video. In this example, the N-FOV camera and W-FOV camera have fixed orientations relative to one another, so the rectangle 662 remains in the same position with respect to the W-FOV video.

[0064] Two basketball players are displayed on the screen 652 in the W-FOV video. In this example, as illustrated in FIG. 6B, the user has selected one of those players, namely player 601, to be tracked. The selection of player 601 may be made by the user touching the position of player 601 on the screen 652. While the player 601 is being tracked, in this embodiment, a rectangle 602 is displayed to indicate to the user the position of the zoom region. This indication allows the user to turn the device to better align the zoom region 602 with the N-FOV indicated by rectangle 662. Under the conditions illustrated in FIG. 6B, the tracked object 601 is within the N-FOV, so the N-FOV camera is used to generate the output video (e.g. for storage or for streaming to another device).

[0065] Under the conditions illustrated in FIG. 6C, however, the tracked player 601 has moved to the left (and the user of the device has failed to follow by turning the device). The position of the rectangle 602 indicating the zoom region has moved accordingly. Because the tracked player 601 is now at least partially outside the N-FOV, video from the W-FOV camera is used to generate the output video. In particular, the zoom region may be cropped and upsampled to provide the output video. In some embodiments, the switching of the source of the output video may be transparent to the user. In such embodiments, the user can try to follow the tracked object while the device automatically and transparently selects which camera to use for the output video.

[0066] FIGs. 7A and 7B depict views of a user interface display on a video capture device 750, which may be a WTRU. The video capture device 750 includes both a N-FOV camera and a W-FOV camera. In the embodiment of FIG. 7A, the video capture device 750 has a screen 752 on which the W-FOV video is being displayed in real time (as it is captured). Also displayed on the screen 752 is a rectangle 762 that indicates the field of view of the N-FOV camera. The user has selected player 601 to be tracked. In this embodiment, a tracking module generates a bounding box representing an approximation of the extent of the tracked object (in this case player 601) around the tracked player, and that bounding box is displayed on the screen as rectangle 702. In this embodiment, while the bounding box is included entirely within the N-FOV, the N-FOV camera is used to generate the output video. In some embodiments, the N-FOV camera is used to generate the output video as long as at least a threshold percentage (e.g., 90%) of the bounding box is within the N-FOV.

[0067] Under the conditions illustrated in FIG. 7B, however, the tracked player 601 has moved to the left. The position of the rectangle 702 indicating the bounding box has been updated accordingly. In this example, the extent of the bounding box has changed as the posture of the tracked player 601 has changed, but in some embodiments, the extent of the bounding box may be fixed (e.g., using predetermined or user-defined dimensions). Because the bounding box is now at least partially outside the N-FOV, video from the W-FOV camera is used to generate the output video. In particular, the zoom region (not displayed in FIGs. 7A-7B) may be cropped and upscaled to provide the output video.

[0068] FIGs. 8A and 8B illustrate an embodiment under the same video recording conditions as those in FIGs. 7A and 7B. In the embodiment of FIGs. 8A and 8B, however, the zoomed output video is displayed on the user interface. Initially, as illustrated in FIGs. 7A, tracked player 601 is within the N-FOV, and high-quality zoomed video from the N-FOV camera is displayed on the screen 852 of the video recording device 850. Subsequently, the tracked player moves out of the N-FOV, and, in response, the device automatically switches to output of cropped and upscaled video on the screen 852. Due to the upscaling, this output video may have a lower quality than the video captured using the native resolution of the N-FOV camera (represented schematically by the pixelation in FIG. 8B). However, the tracked player 601 remains within the output video. In some embodiments, an indicator such as the arrow 854 may be displayed on the screen 852 to indicate to the user which way to turn the device such that the tracked player 601 will again be within the N-FOV. The display of the video capture device may return to displaying a representation of the N-FOV camera video after the user reacquires the zoom region within the view of the N-FOV camera.

[0069] FIGs. 9A and 9B illustrate an embodiment under the same video recording conditions as those in FIGs. 7A and 7B. In the embodiment of FIGs. 9A and 9B, however, the device 950 displays the N-FOV video on the display 952 while the tracked object is within the N-FOV, and the device switches to display of the W-FOV video when the tracked object is not within the W-FOV. This allows the user to more easily

reorient the device such that the tracked object is again within the N-FOV. Initially, as illustrated in FIG. 7A, tracked player 601 is within the N-FOV, and high-quality zoomed video from the N-FOV camera is displayed on the screen 952 of the video recording device 950. Subsequently, the tracked player moves out of the N-FOV, and, in response, the device automatically switches to display of the W-FOV video on the screen. In this embodiment, the device further displays a rectangle 962 representing the N-FOV. The device may further display an indication, such as rectangle 702, to identify which object is being tracked. In other embodiments, the device may display an indication of the zoom region (such as the rectangle 602 of FIGs. 6B-6C). The device 602 may switch back to displaying the N-FOV video if the tracked object is reacquired within the N-FOV.

[0070] The foregoing examples have focused on embodiments using a device that has two cameras with different fields of view. In alternative embodiments, a device with more than two cameras may be used. For example, an object may be tracked to identify which cameras out of a plurality of cameras have that object within their respective fields of view. An output video may be generated by selecting which one of those cameras will be used to provide the output any particular time. In some such embodiments, the output is selected from the camera with the narrowest field of view that still includes the tracked object within its field of view. Video from cameras with a wider field of view may be cropped and upscaled as appropriate to match the field of view of cameras with a narrower field of view.

[0071] 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 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

What is Claimed:

1. A method of providing an output video stream in a device having a first camera and a second camera, the second camera having a field of view that is narrower than and contained within a field of view of the first camera, the method comprising:
 - tracking an object in video captured by the first camera;
 - determining, based on the tracking, whether the object is entirely within the field of view of the second camera;
 - while the object is within the field of view of the second camera, outputting video captured by the second camera as the output video stream; and
 - while the object is outside the field of view of the second camera, (i) cropping and upscaling the video captured by the first camera to generate a cropped-and-upscaled video of the object and (ii) outputting the cropped-and-upscaled video as the output video stream.
2. The method of claim 1, further comprising:
 - displaying, on a screen of the device, the video captured by the first camera; and
 - in the displayed video, displaying an indication of a position of the tracked object and an indication of the field of view of the second camera.
3. The method of claim 1, wherein the method is performed in real time.
4. The method of claim 1, further comprising, before tracking the object:
 - displaying, on a screen of the device, the video captured by the first camera; and
 - receiving user input selecting the object to be tracked in the displayed video.
5. The method of claim 1, wherein outputting video captured by the second camera as the output video stream occurs while at least a threshold percentage of a bounding box around the object is within the field of view of the second camera.
6. The method of claim 1, wherein the up-scaling is performed to match an optical zoom of the second camera.
7. The method of claim 1, further comprising:
 - while the object is within the field of view of the second camera, displaying the video captured by the second camera on a screen of the device; and

while the object is outside the field of view of the second camera, displaying on the screen (i) the video captured by the first camera and (ii) an indication of the field of view of the second camera.

8. The method of 7 further comprising, while the object is outside the field of view of the second camera, displaying on the screen an indication of the position of the object.
9. The method of claim 1, wherein outputting the output video stream comprises recording the output video stream.
10. The method of claim 1, wherein outputting the output video stream comprises displaying the output video stream on a screen of the device.
11. A video capture device, comprising:
 - a first camera;
 - a second camera having a field of view that is narrower than and contained within a field of view of the first camera;
 - a processor; and
 - a non-transitory, computer-readable medium storing instructions that are operative, when executed on the processor, to perform the functions of:
 - tracking an object in video captured by the first camera;
 - determining, based on the tracking, whether the object is entirely within the field of view of the second camera;
 - while the object is within the field of view of the second camera, outputting video captured by the second camera as an output video stream; and
 - while the object is outside the field of view of the second camera, (i) cropping and upscaling the video captured by the first camera and (ii) outputting the cropped and upscaled video as the output video stream.
12. The device of claim 11, further comprising:
 - a display; and
 - a user interface capable of receiving user selections of objects,wherein the processor is further operative, before tracking the object, to perform the functions of:
 - displaying, on the display, the video captured by the first camera; and
 - receiving user input selecting the object to be tracked in the displayed video.
13. The device of claim 11, further comprising a Radio Frequency Identification (RFID) receiver, wherein the processor is further operative to track the object using the RFID receiver.

14. The device of claim 11, further comprising a transceiver, wherein outputting video includes transmitting the output video stream to a remote device.
15. The device of claim 11, wherein outputting the output video stream comprises recording the output video stream.

1/12

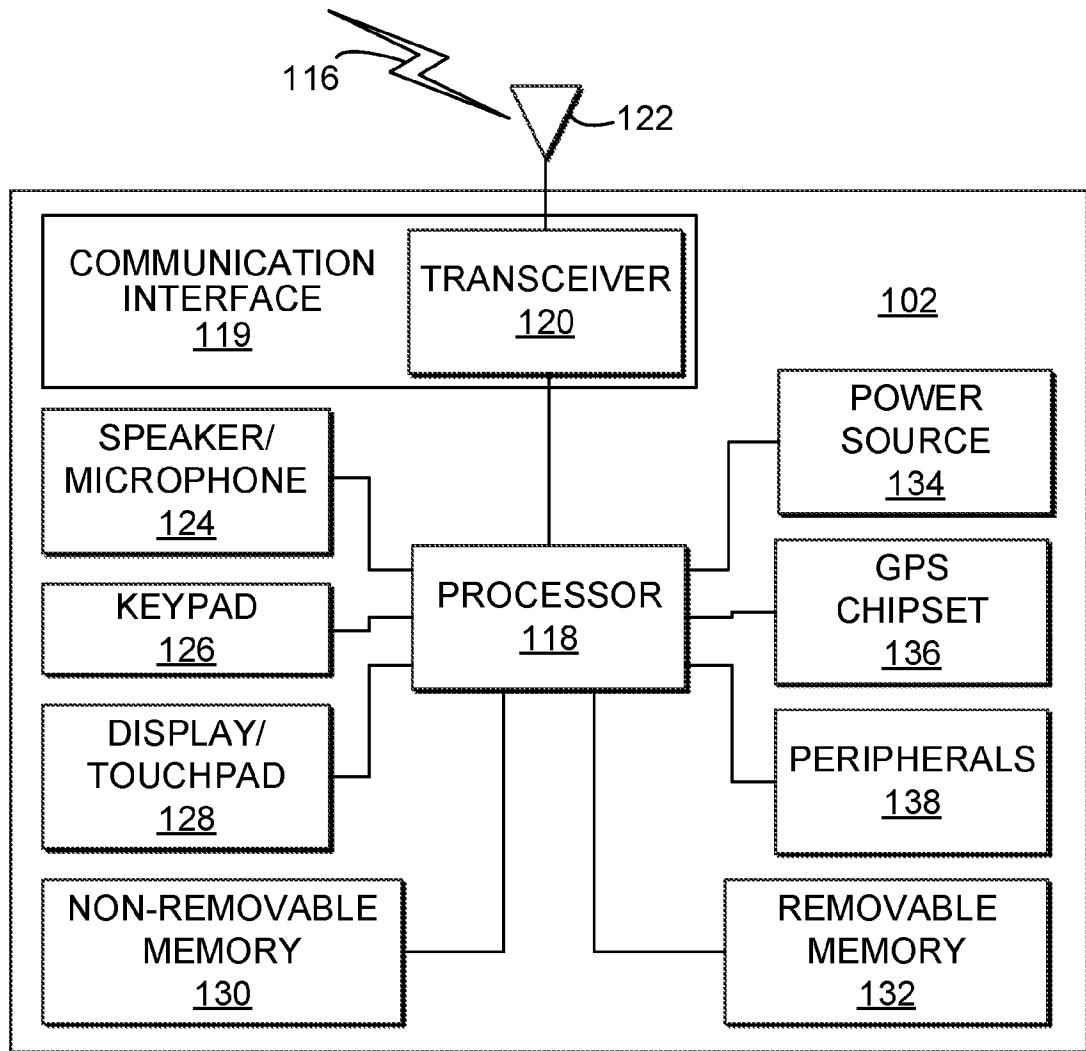


FIG. 1A

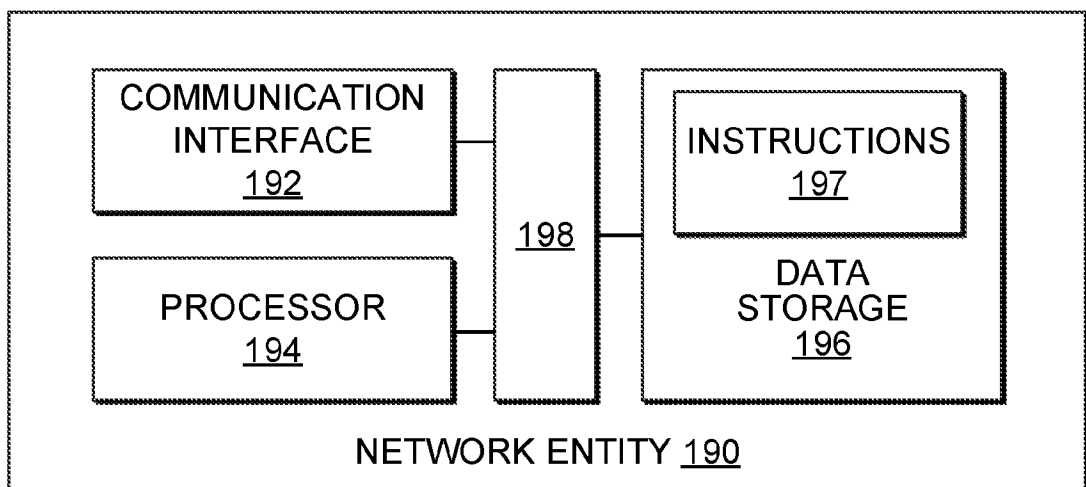


FIG. 1B

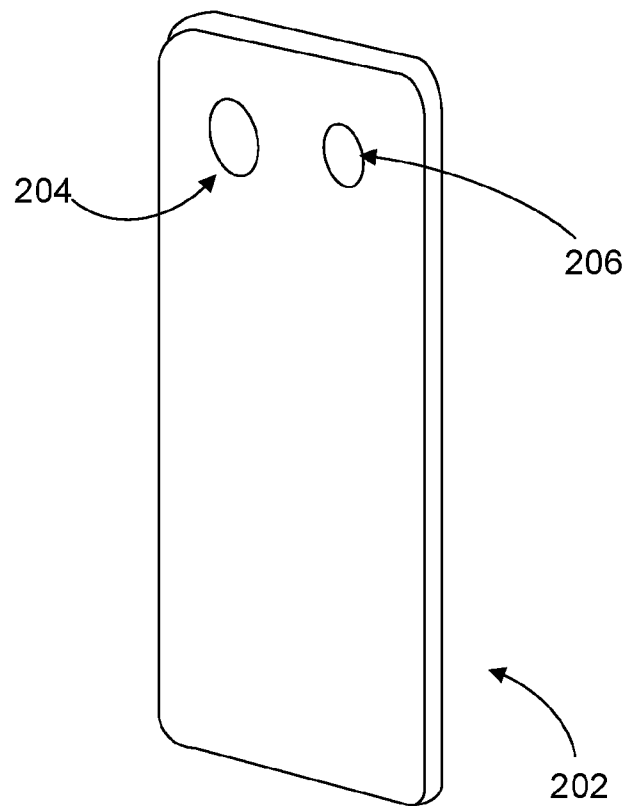


FIG. 2

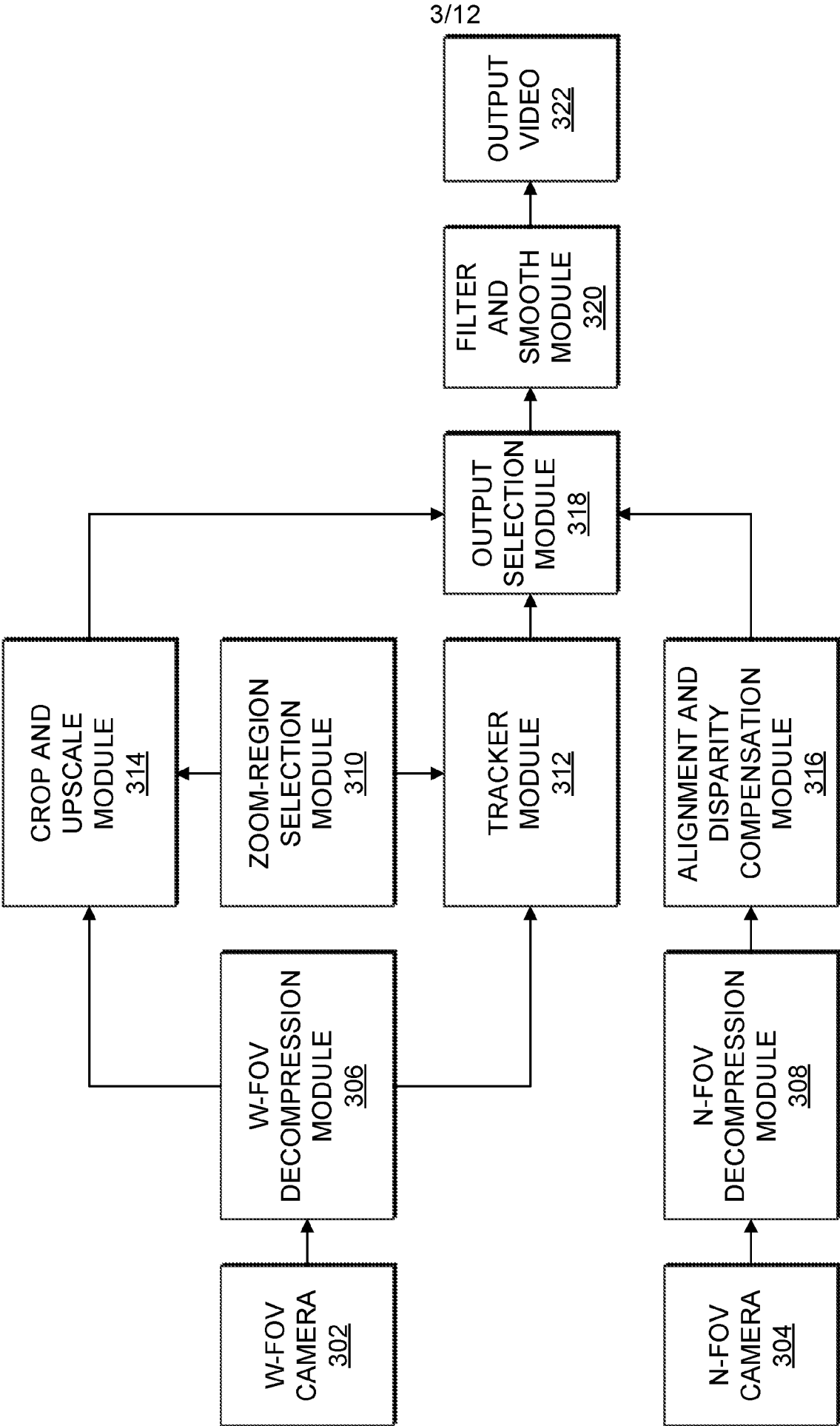


FIG. 3A

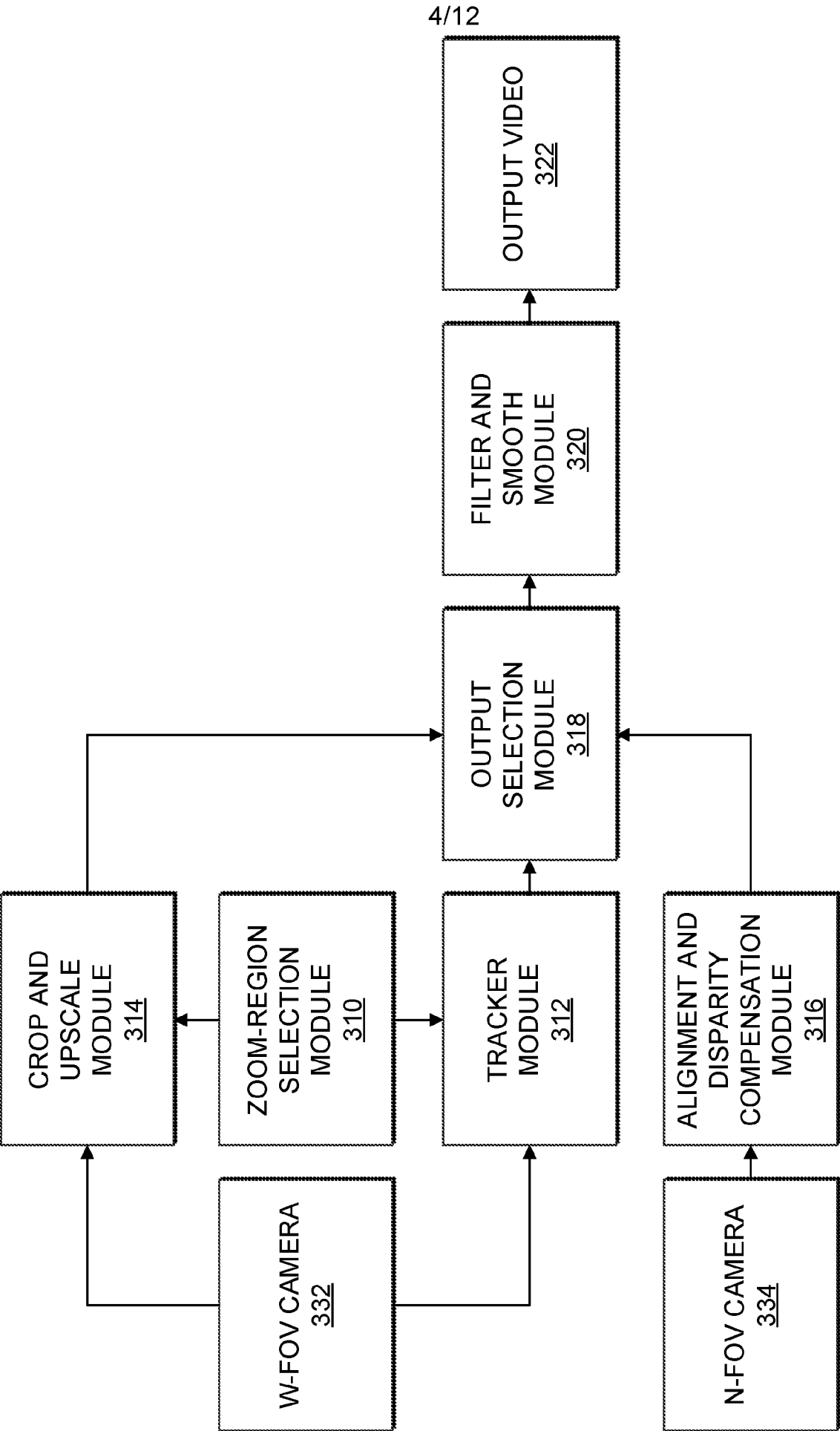


FIG. 3B

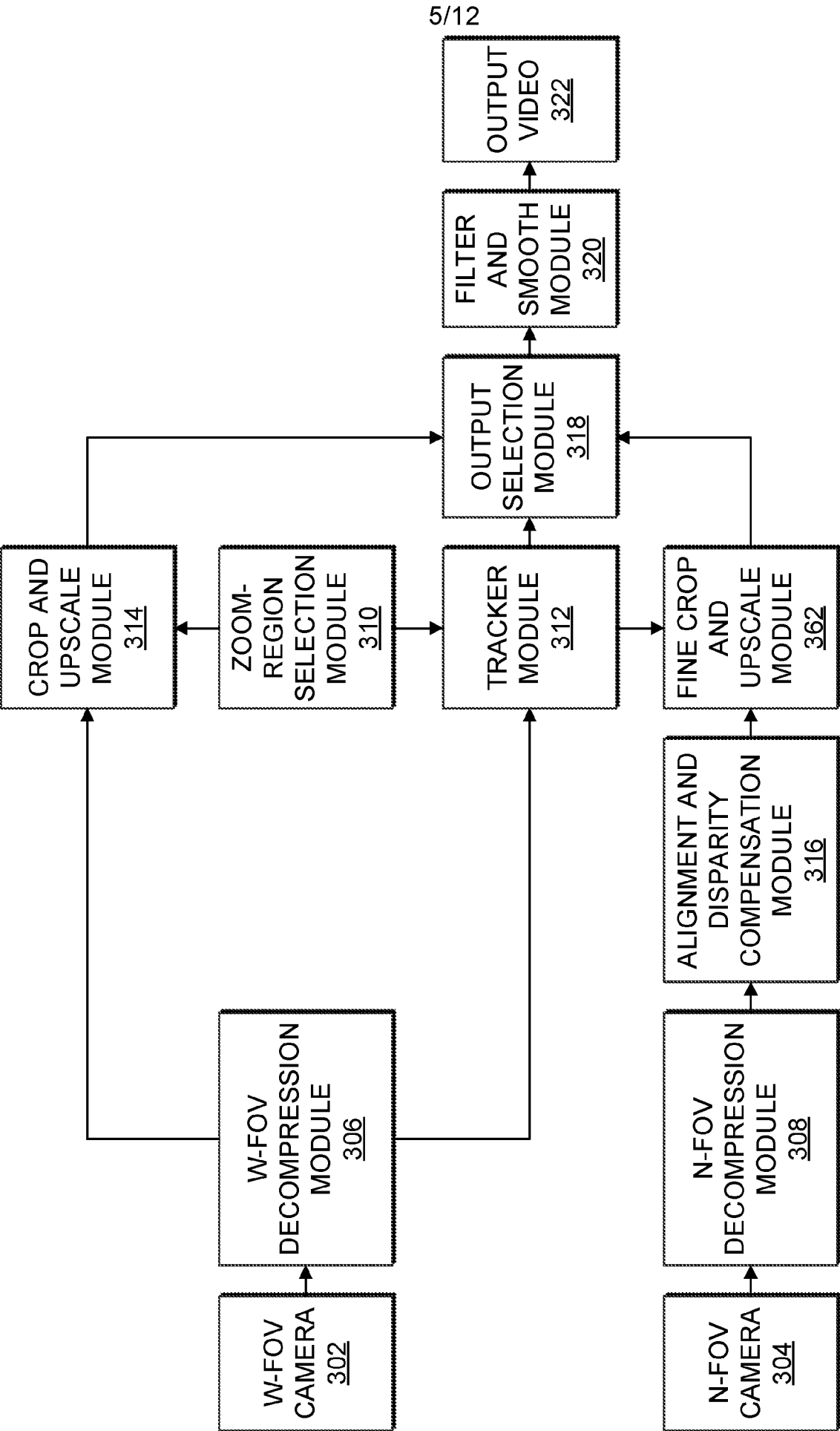
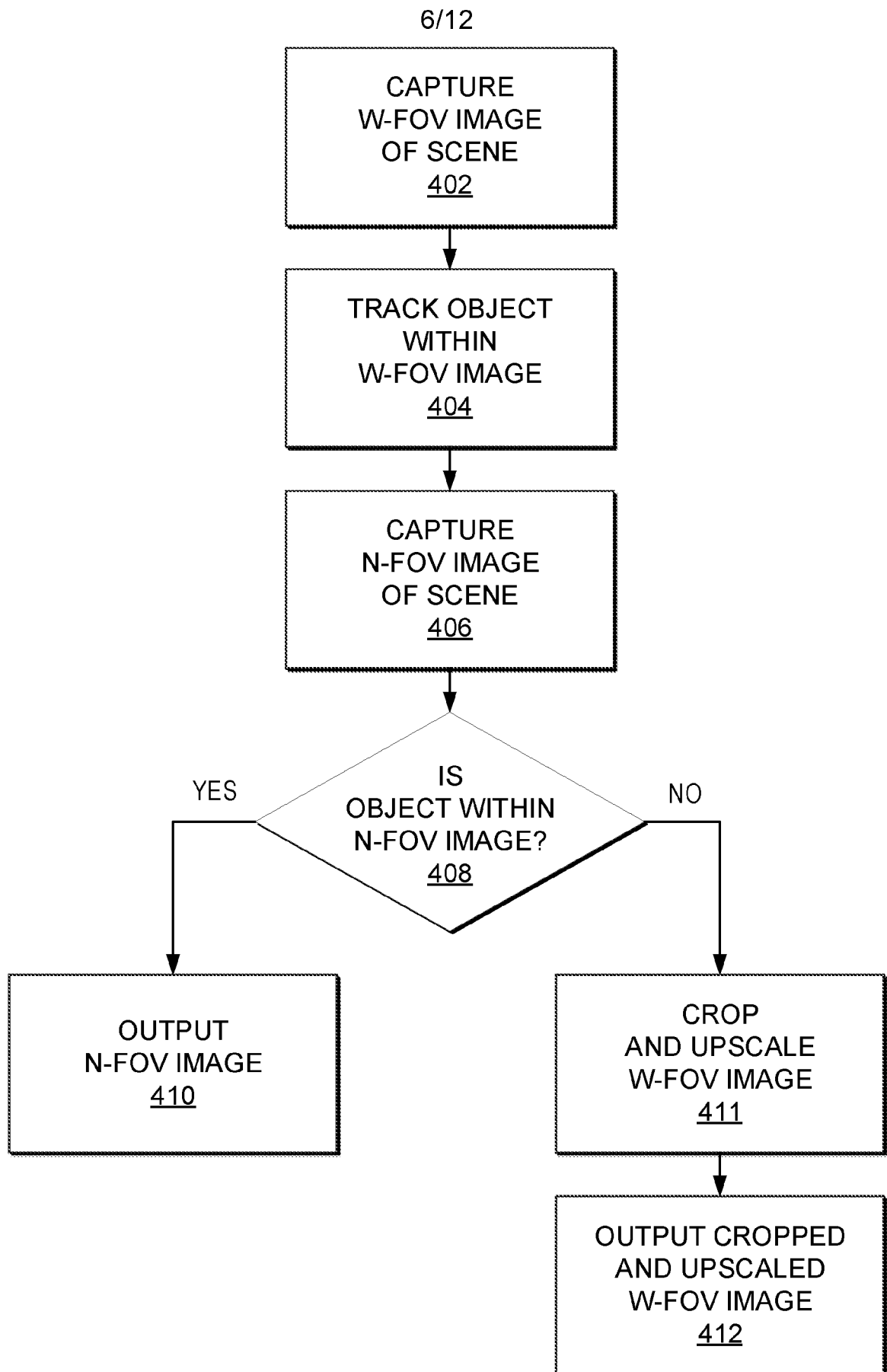


FIG. 3C



400

FIG. 4

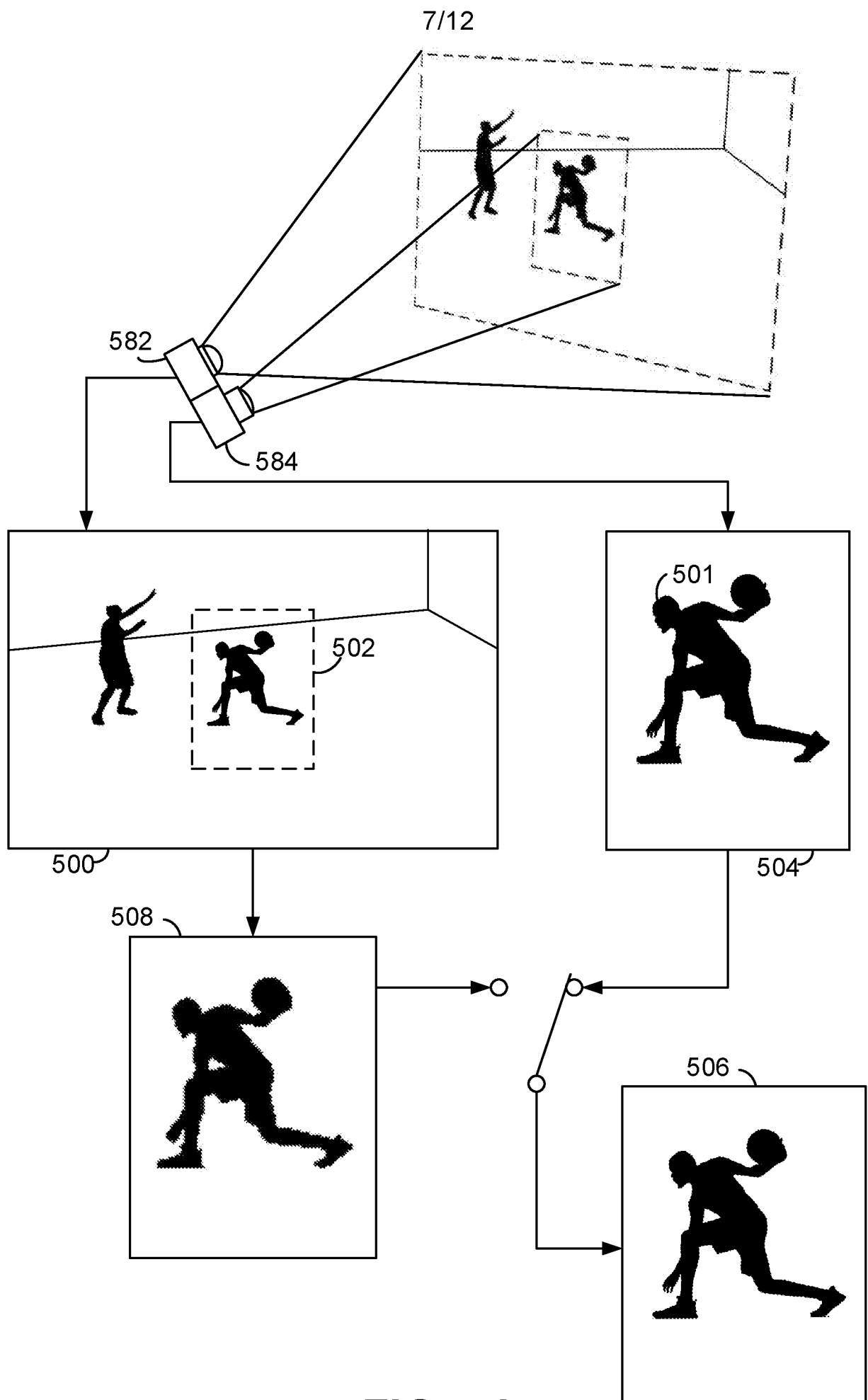


FIG. 5A

8/12

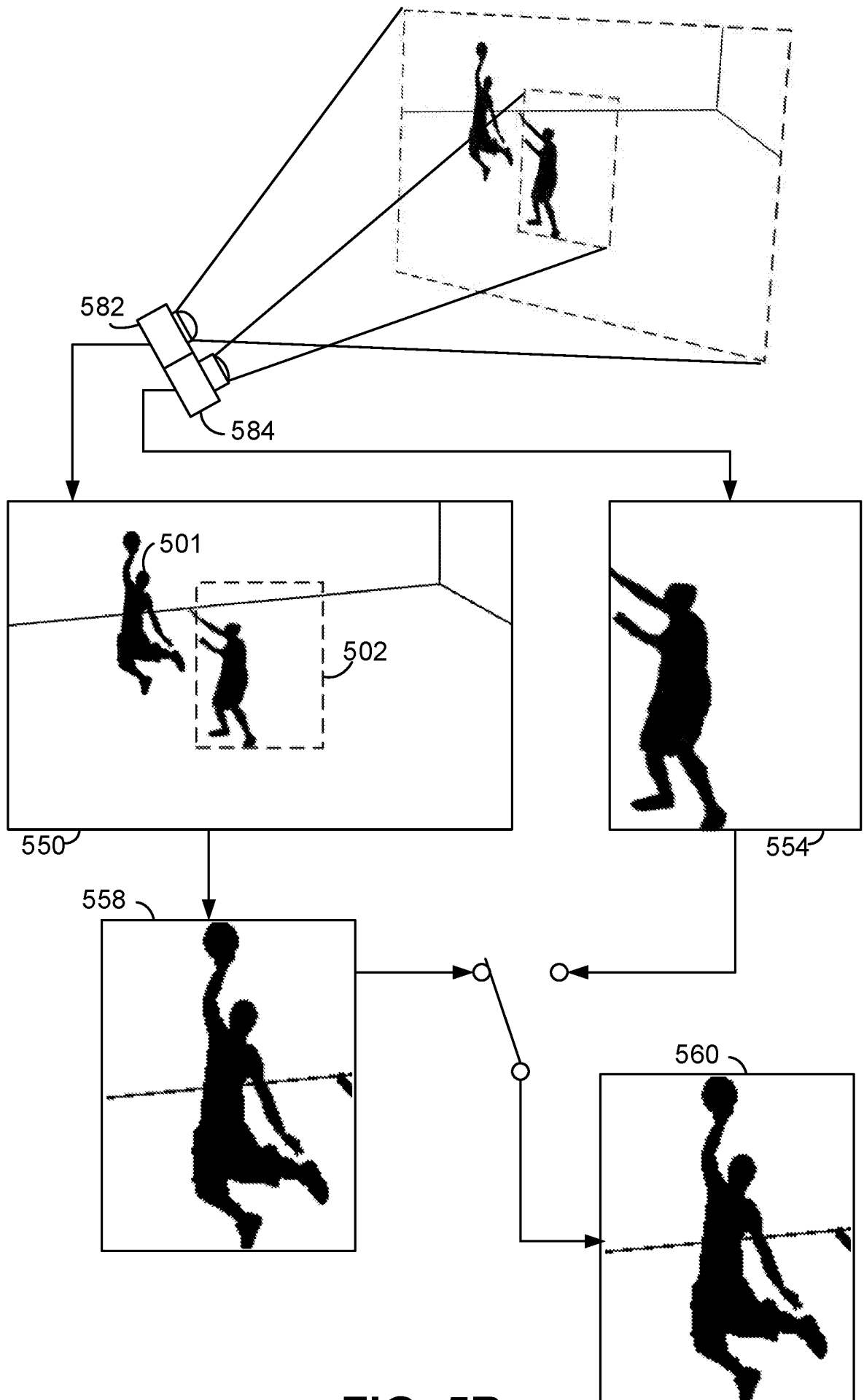
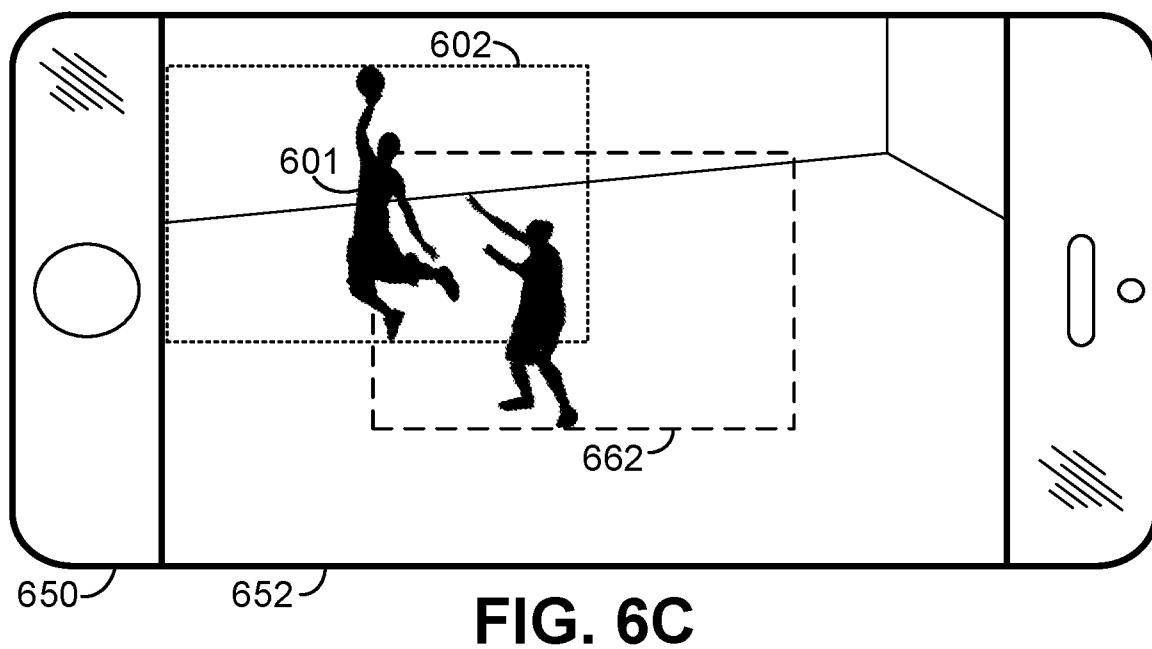
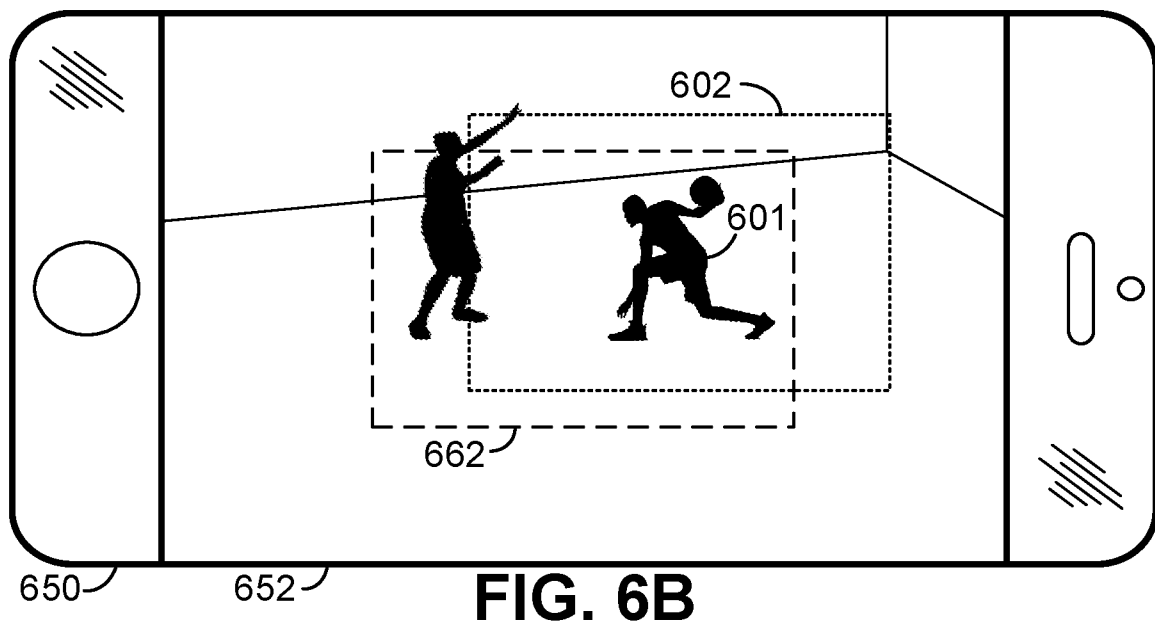
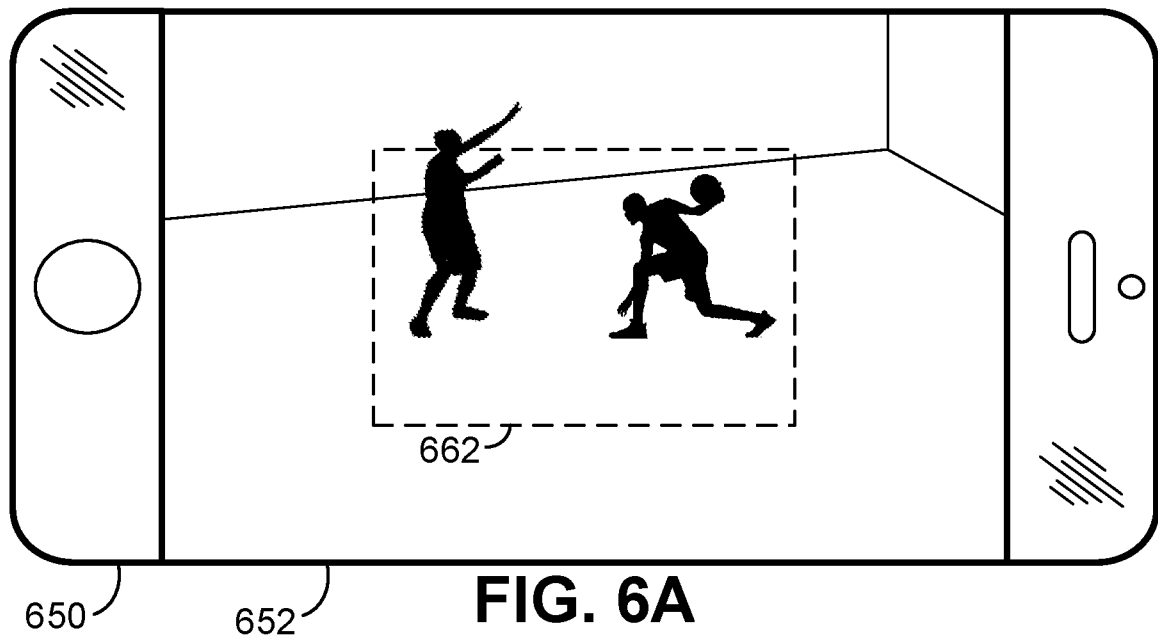


FIG. 5B



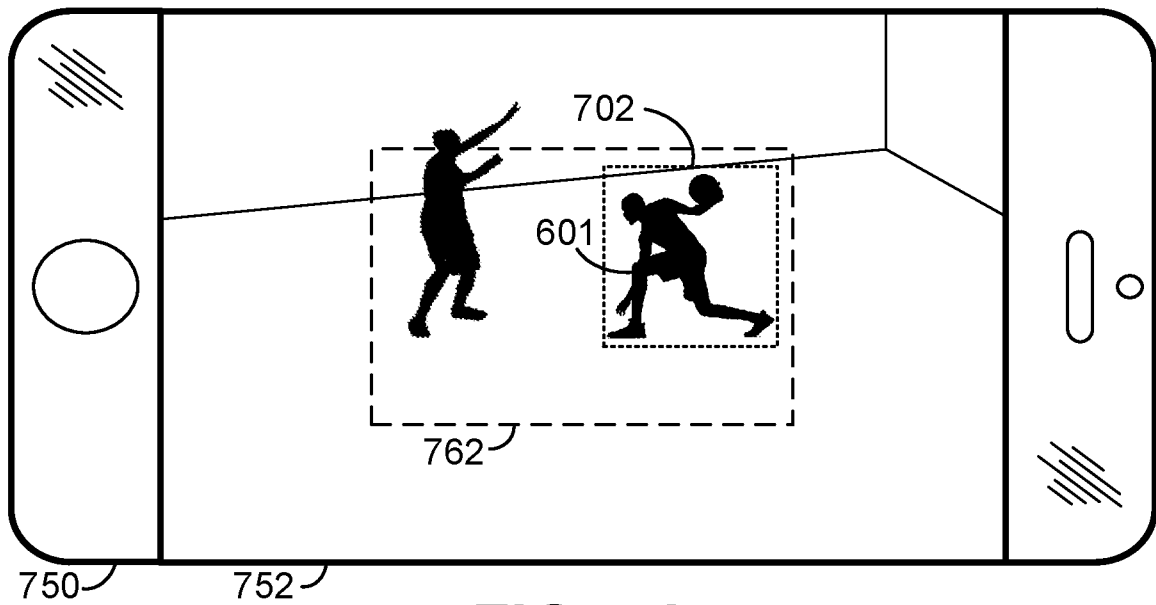


FIG. 7A

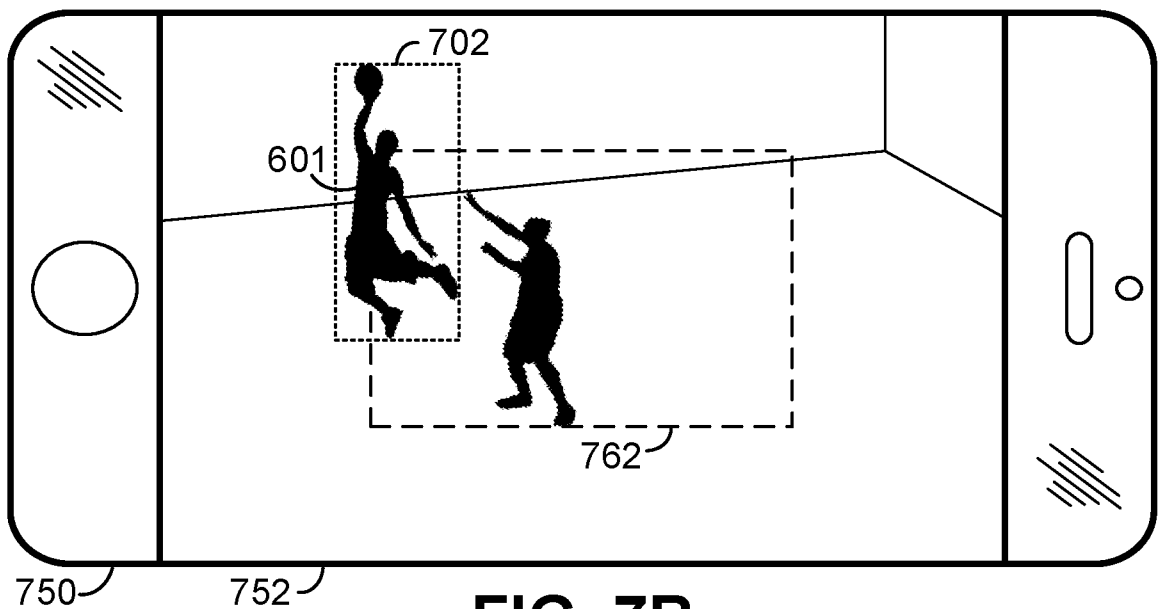


FIG. 7B

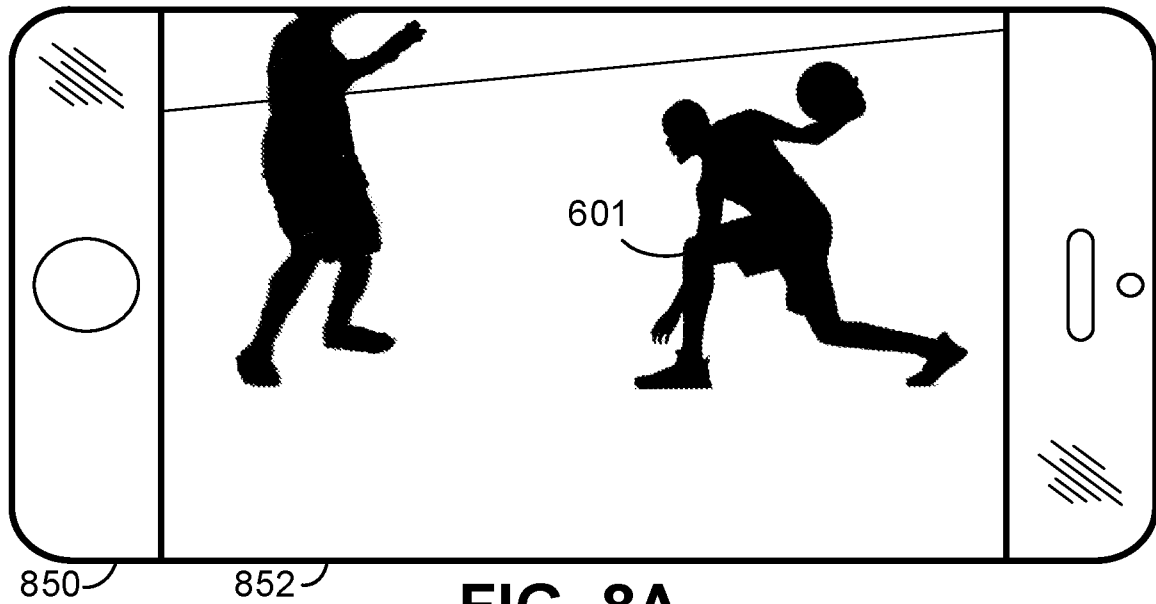


FIG. 8A

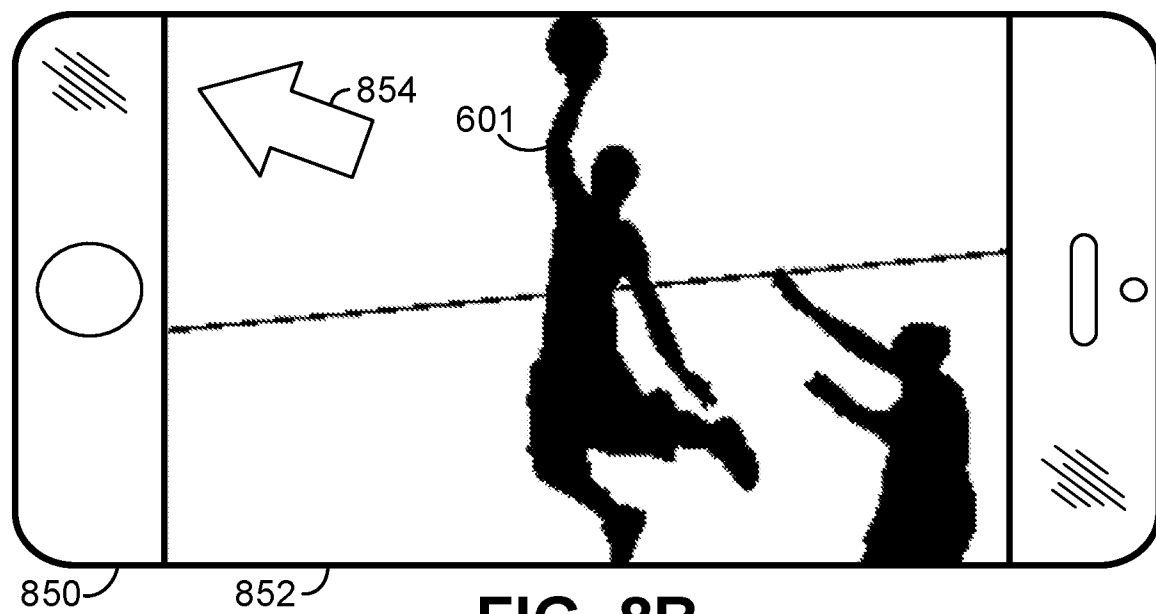


FIG. 8B

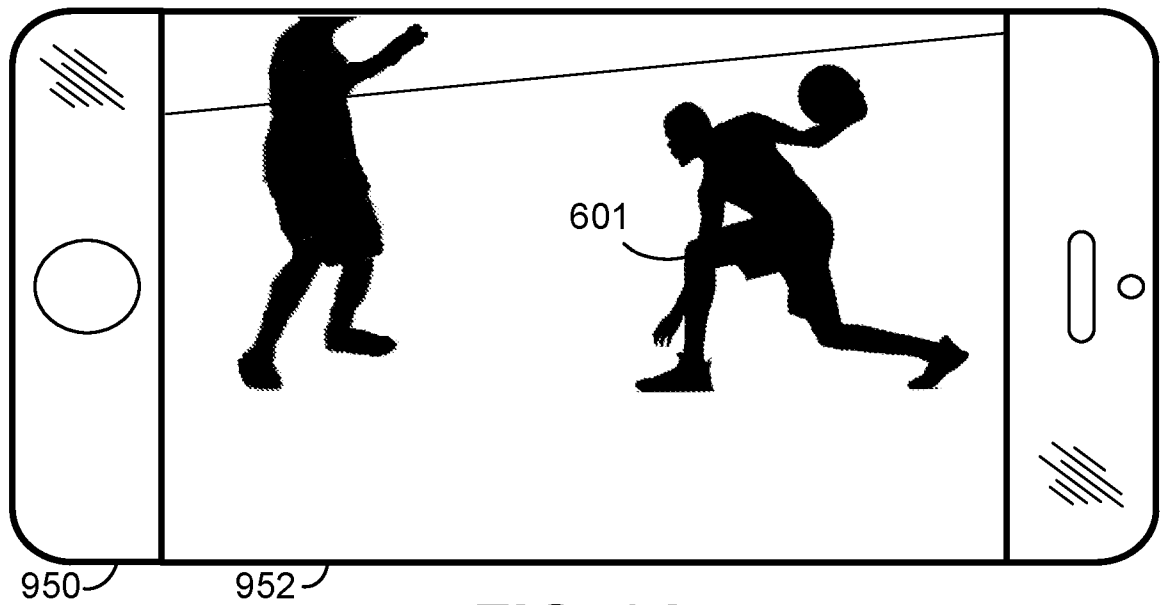


FIG. 9A

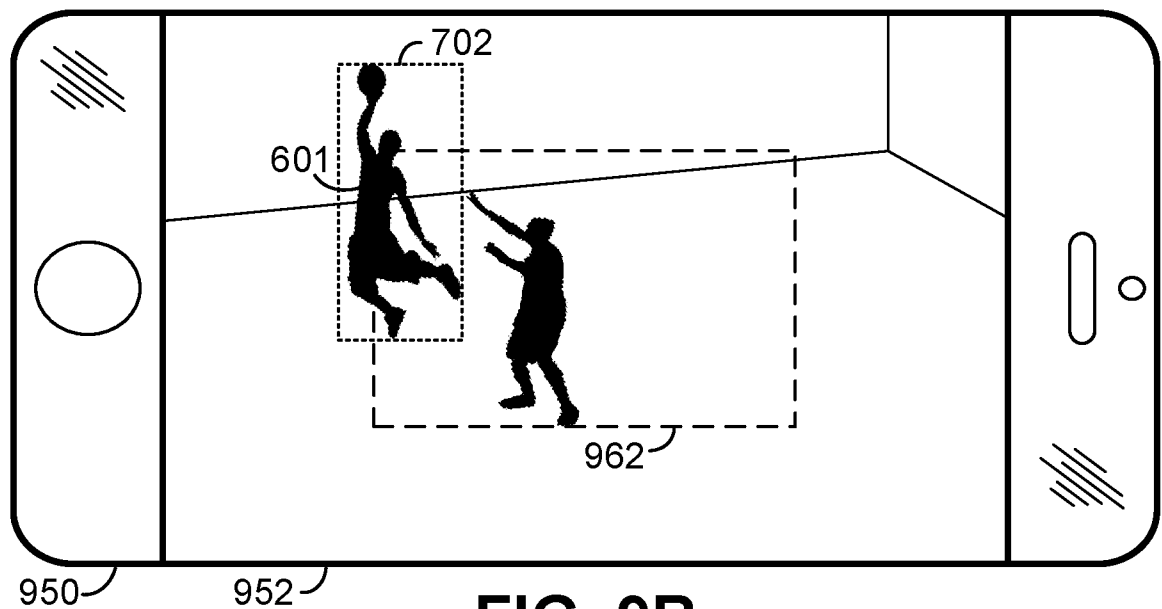


FIG. 9B

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/020418

A. CLASSIFICATION OF SUBJECT MATTER

INV. H04N5/225 H04N5/232
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)

EP0-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2016/381289 A1 (JEON S; JEON S R; JIN Z; KIM J; KIM J G; LIM S; LIM S H; RIM S) 29 December 2016 (2016-12-29) paragraph [0091] - paragraph [0134] figures 1-3B, 9	1-15
Y	US 2010/238262 A1 (KURTZ ANDREW F [US] ET AL) 23 September 2010 (2010-09-23) paragraph [0116] - paragraph [0117] figures 4A-4D	1-15
Y	US 2015/116501 A1 (MCCOY CHARLES [US] ET AL) 30 April 2015 (2015-04-30) paragraph [0084] - paragraph [0086] paragraph [0088] - paragraph [0089] paragraph [0052] figure 3A	1-15
	----- -/--	



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

28 May 2018

Date of mailing of the international search report

05/06/2018

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Rocca, Delphine

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/020418

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2012/300051 A1 (DAIGO KENJI [JP] ET AL) 29 November 2012 (2012-11-29) paragraph [0033] - paragraph [0071] figures 1-6 -----	2,4,7,8, 12
Y	US 2013/230293 A1 (BOYLE CHRISTOPHER T [US] ET AL) 5 September 2013 (2013-09-05) paragraph [0043] paragraph [0131] figures 1,2 -----	13

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2018/020418

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2016381289 A1	29-12-2016	US 2016381289 A1	29-12-2016
		WO 2016208849 A1	29-12-2016

US 2010238262 A1	23-09-2010	NONE	

US 2015116501 A1	30-04-2015	BR 102014026563 A2	11-10-2016
		CN 104601878 A	06-05-2015
		JP 2015089119 A	07-05-2015
		US 2015116501 A1	30-04-2015

US 2012300051 A1	29-11-2012	JP 2013013050 A	17-01-2013
		US 2012300051 A1	29-11-2012

US 2013230293 A1	05-09-2013	AU 2013225635 A1	23-10-2014
		CA 2866131 A1	09-06-2013
		EP 2820840 A1	07-01-2015
		JP 2015513858 A	14-05-2015
		US 2013230293 A1	05-09-2013
		WO 2013131100 A1	06-09-2013
