



US 20230342388A1

(19) **United States**

(12) **Patent Application Publication**  
**RAHMAN et al.**

(10) **Pub. No.: US 2023/0342388 A1**

(43) **Pub. Date: Oct. 26, 2023**

(54) **VISUAL SEARCH IN REAL WORLD USING OPTICAL SEE-THROUGH HEAD MOUNTED DISPLAY WITH AUGMENTED REALITY AND USER INTERACTION TRACKING**

*G06F 16/332* (2006.01)  
*G06F 16/583* (2006.01)  
*G06F 16/951* (2006.01)  
*G06F 16/955* (2006.01)  
*G02B 27/01* (2006.01)  
*G06F 3/01* (2006.01)

(71) Applicant: **QUALCOMM Incorporated**, San Diego, CA (US)

(52) **U.S. Cl.**  
CPC ..... *G06F 16/532* (2019.01); *G06T 19/006* (2013.01); *G06F 16/332* (2019.01); *G06F 16/583* (2019.01); *G06F 16/951* (2019.01); *G06F 16/9554* (2019.01); *G02B 27/017* (2013.01); *G06F 3/013* (2013.01); *G06F 3/017* (2013.01); *G02B 2027/0138* (2013.01); *G02B 2027/014* (2013.01); *G02B 2027/0187* (2013.01)

(72) Inventors: **Md Sazzadur RAHMAN**, San Jose, CA (US); **Kexi LIU**, San Diego, CA (US); **Martin RENSCHLER**, San Diego, CA (US)

(21) Appl. No.: **18/344,512**

(22) Filed: **Jun. 29, 2023**

**Related U.S. Application Data**

(63) Continuation of application No. 17/304,739, filed on Jun. 24, 2021, now Pat. No. 11,734,336, which is a continuation of application No. 16/414,667, filed on May 16, 2019, now Pat. No. 11,068,531, which is a continuation of application No. 16/181,175, filed on Nov. 5, 2018, now Pat. No. 10,372,751, which is a continuation of application No. 14/151,664, filed on Jan. 9, 2014, now Pat. No. 10,152,495.

(60) Provisional application No. 61/867,579, filed on Aug. 19, 2013.

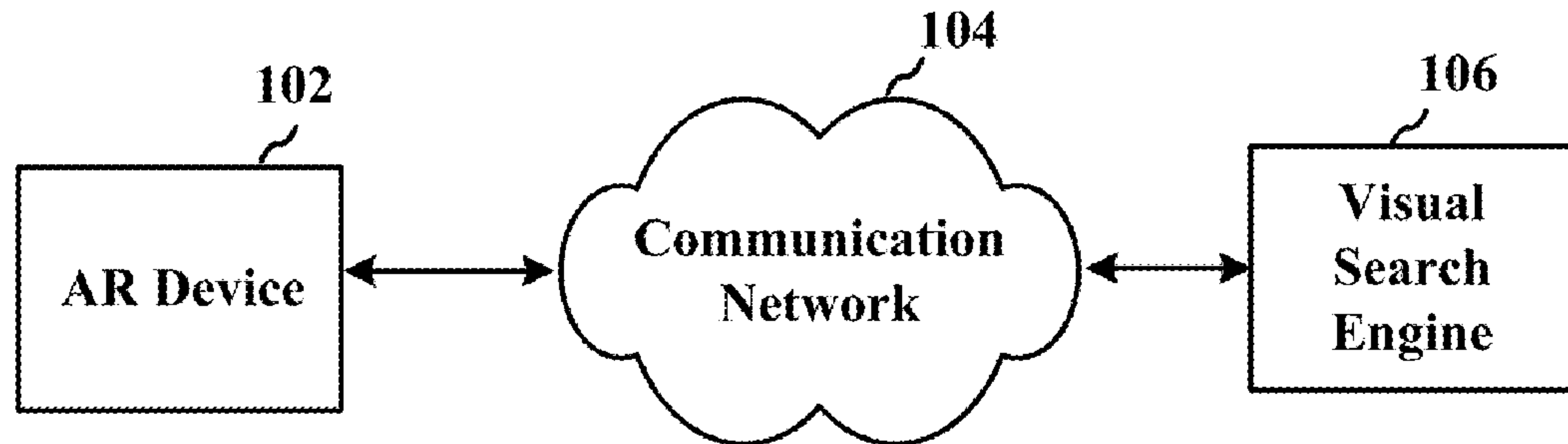
**Publication Classification**

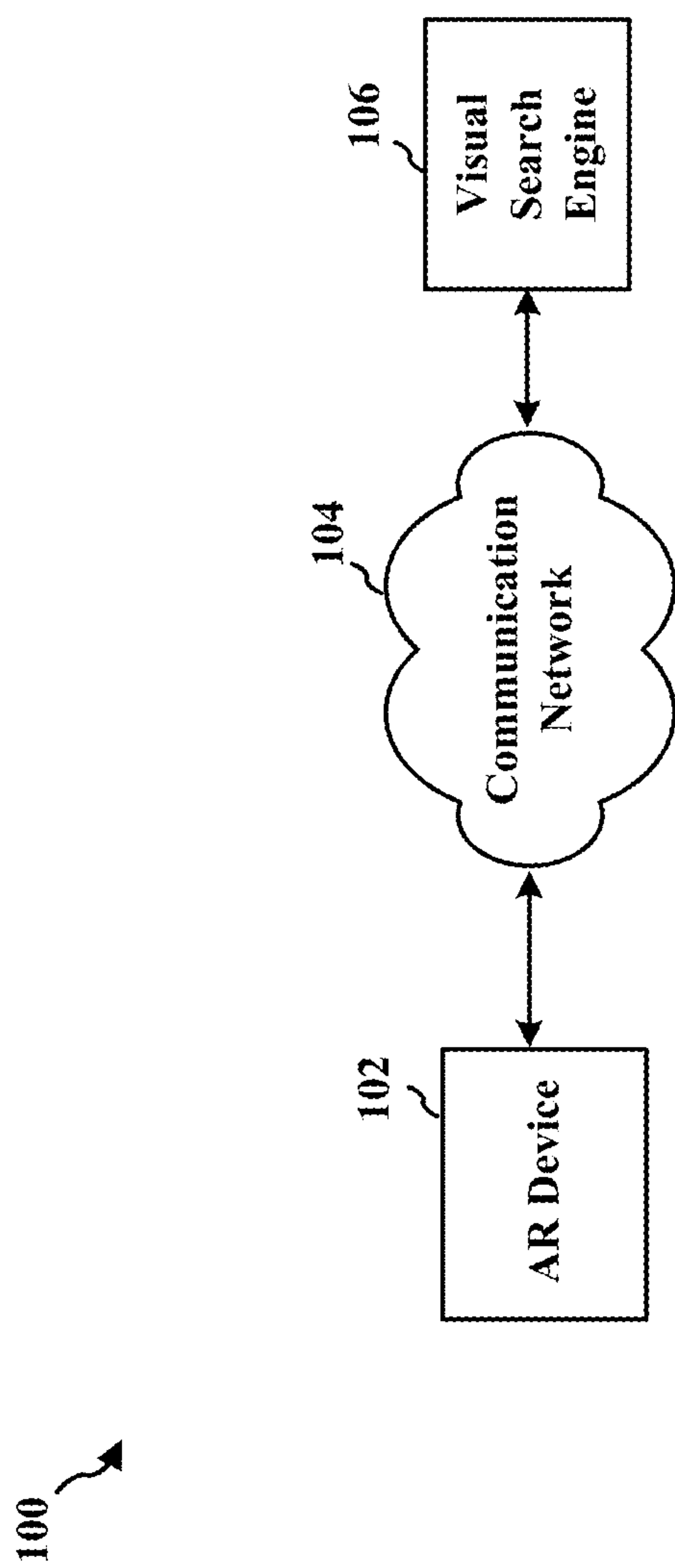
(51) **Int. Cl.**  
*G06F 16/532* (2006.01)  
*G06T 19/00* (2006.01)

(57) **ABSTRACT**

Techniques can conduct online visual searches through an augmented reality (AR) device having an optical see-through head mounted display (HMD). An apparatus identifies a portion of an object in a field of view of the HMD based on user interaction with the HMD. The portion includes searchable content, such as a barcode. The user interaction may be an eye gaze or a gesture. A user interaction point in relation to the HMD screen is tracked to locate a region of the object that includes the portion and the portion is detected within the region. The apparatus captures an image of the portion. The identified portion of the object does not encompass the entirety of the object. Accordingly, the size of the image is less than the size of the object in the field of view. The apparatus transmits the image to a visual search engine.

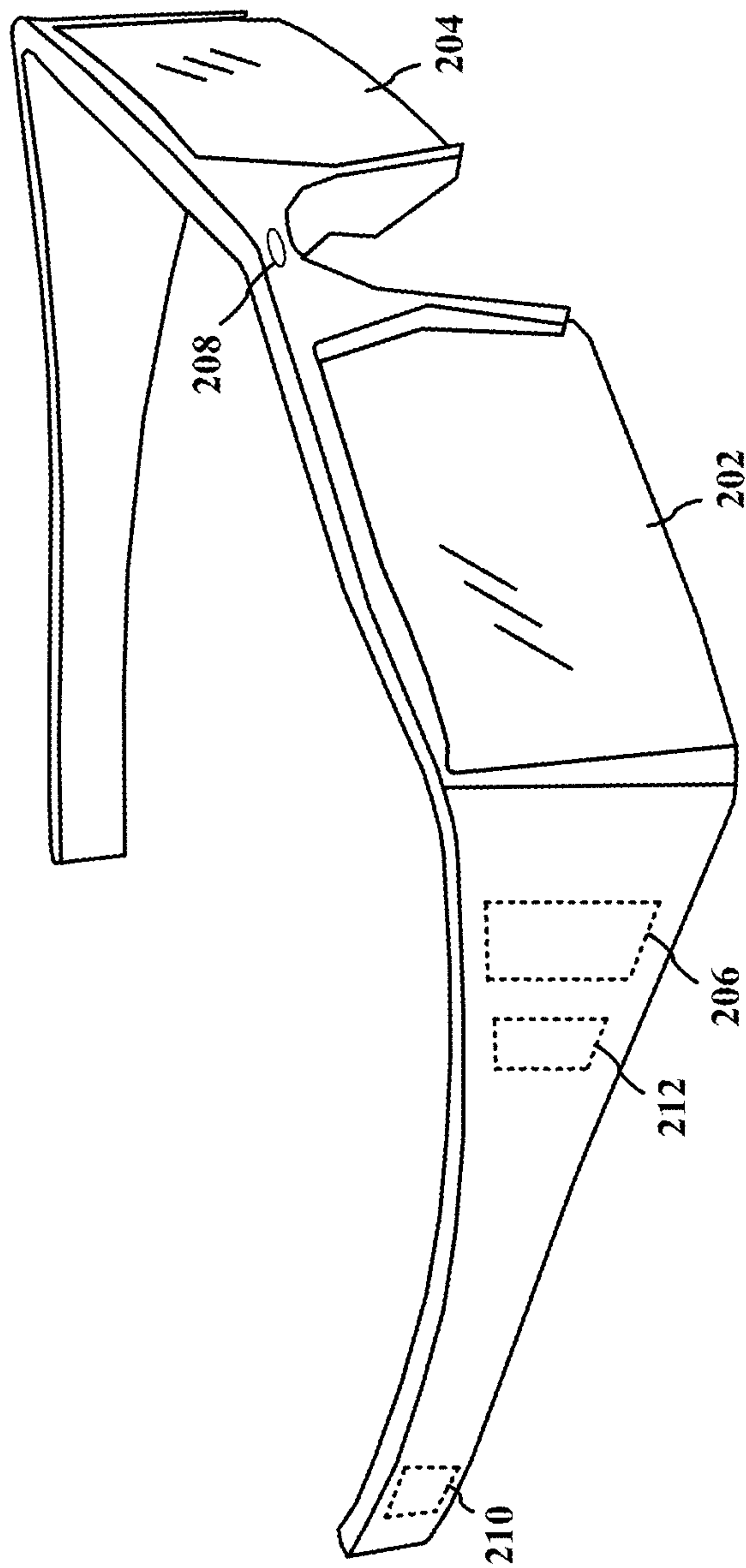
100





**FIG. 1**

200 ↗



**FIG. 2**



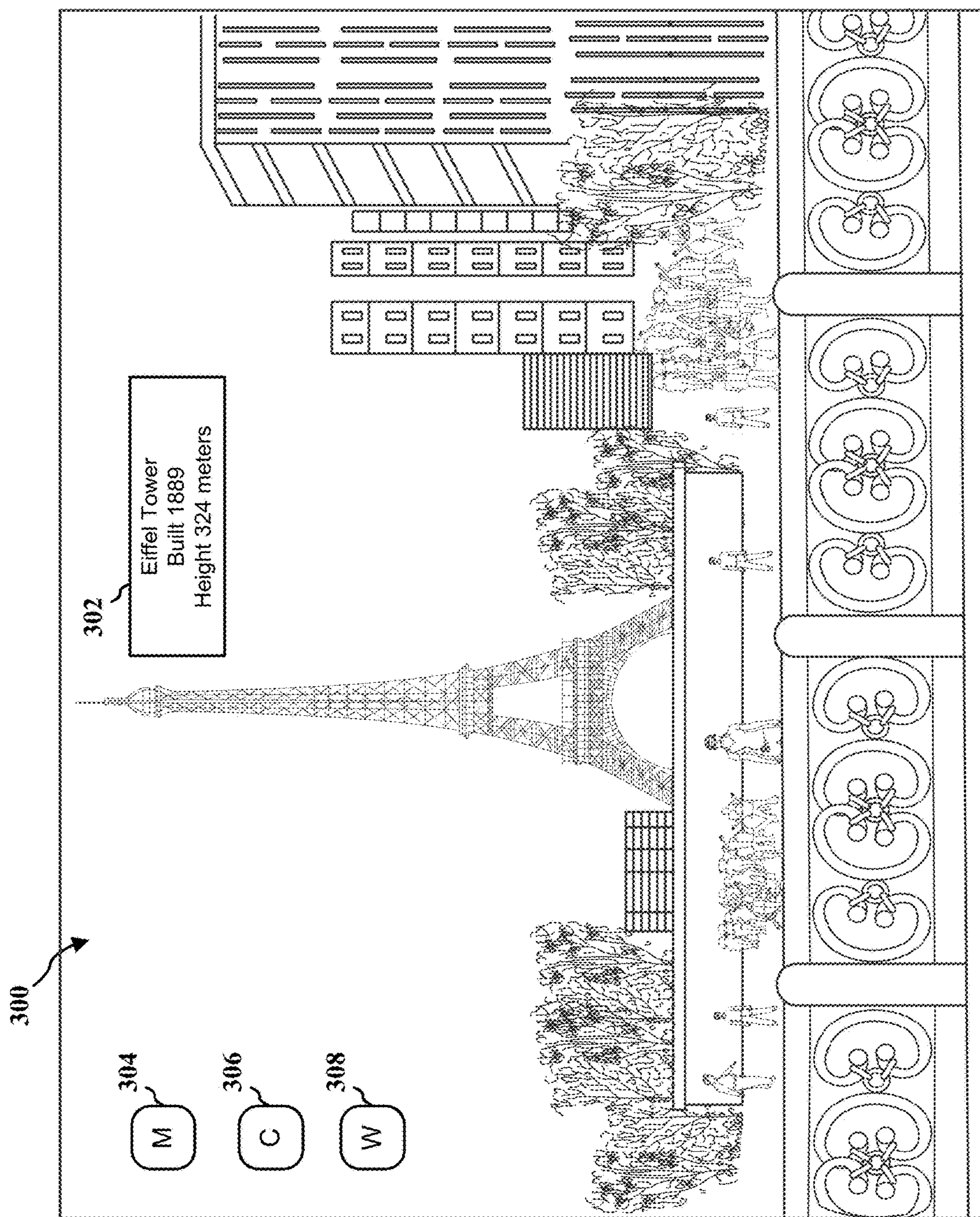


FIG. 3

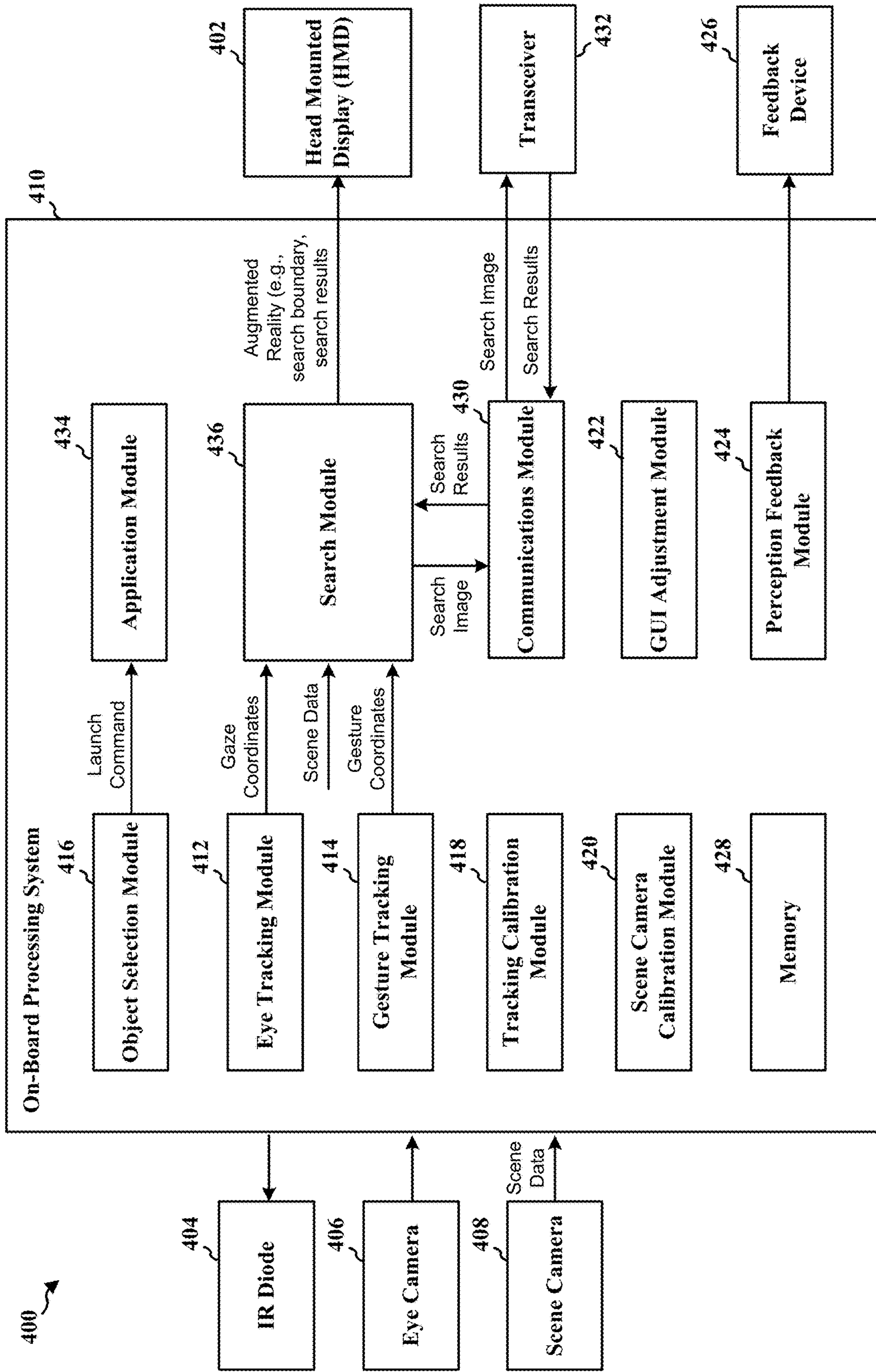


FIG. 4



500 ↗

502 ↗

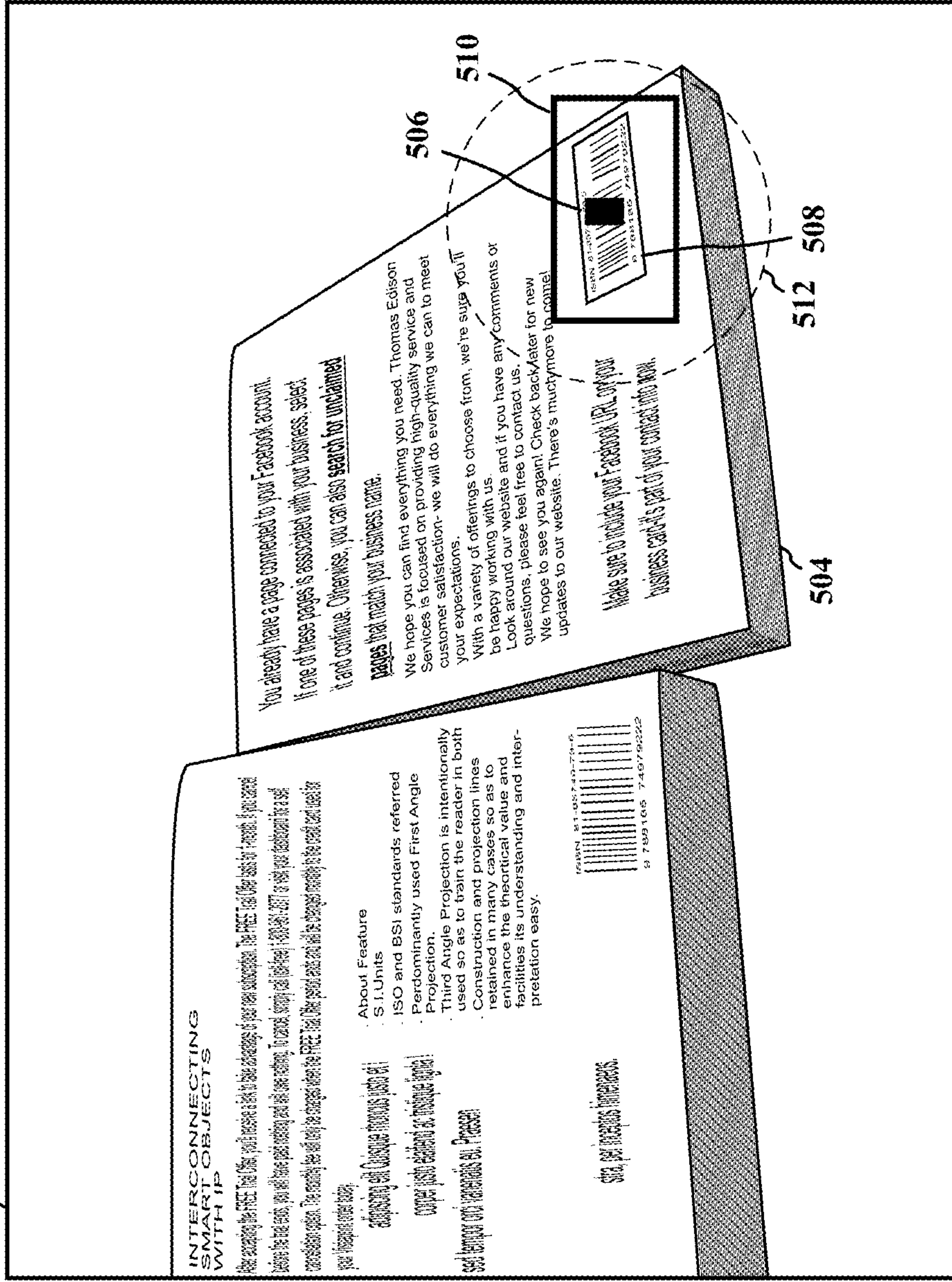


FIG. 5

600

602

**INTERCONNECTING SMART OBJECTS WITH IP**

After receiving the FREE Trial Offer, you'll receive a link to the package of your new subscription. The FREE Trial Offer ends for good. If you cancel within the trial period, you will have your subscription and all your content, including all purchased content, removed from your account for a full cancellation period. The monthly fee will apply for the period when the FREE Trial Offer period ends and will be charged monthly to the credit card used for your subscription order.

**Additional Features:**

- Supporting all unique business units of ISO and BSI standards referred to predominantly used First Angle Projection.
- Third Angle Projection is intended to be used so as to train the reader in construction and projection lines retained in many cases so as to enhance the theoretical value and facilitates its understanding and interpretation easy.

ISBN 978-1-85745-474-4  
9 781857 745744

Stop for frequent interactions.

**604**

**Bluetooth Low Energy: The Developer's Handbook**

Robin Heydon  (Author)

☆☆☆☆☆  (7 customer reviews)

List Price: \$49.99

Price: **\$35.91 & FREE Shipping. Details**

You Save: \$14.08 (28%)

**In Stock.**

We hope to see you again. Check back for updates to our website. There's much more to come!

Makes sure to include your Facebook URL on your business card so it's part of your contact info now.

**504**

ISBN 978-1-85745-474-4  
9 781857 745744

FIG. 6

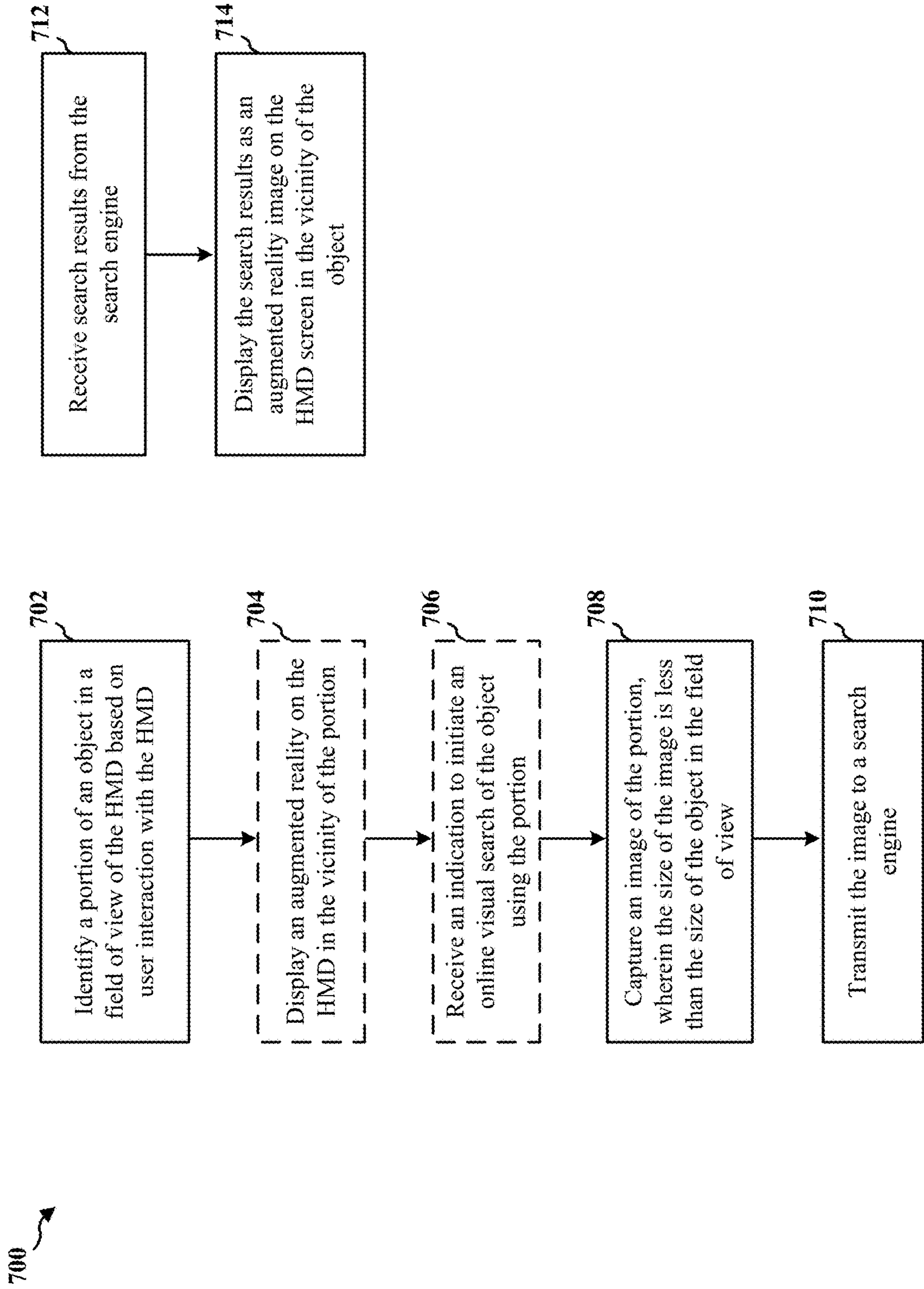


FIG. 7



800 ↗

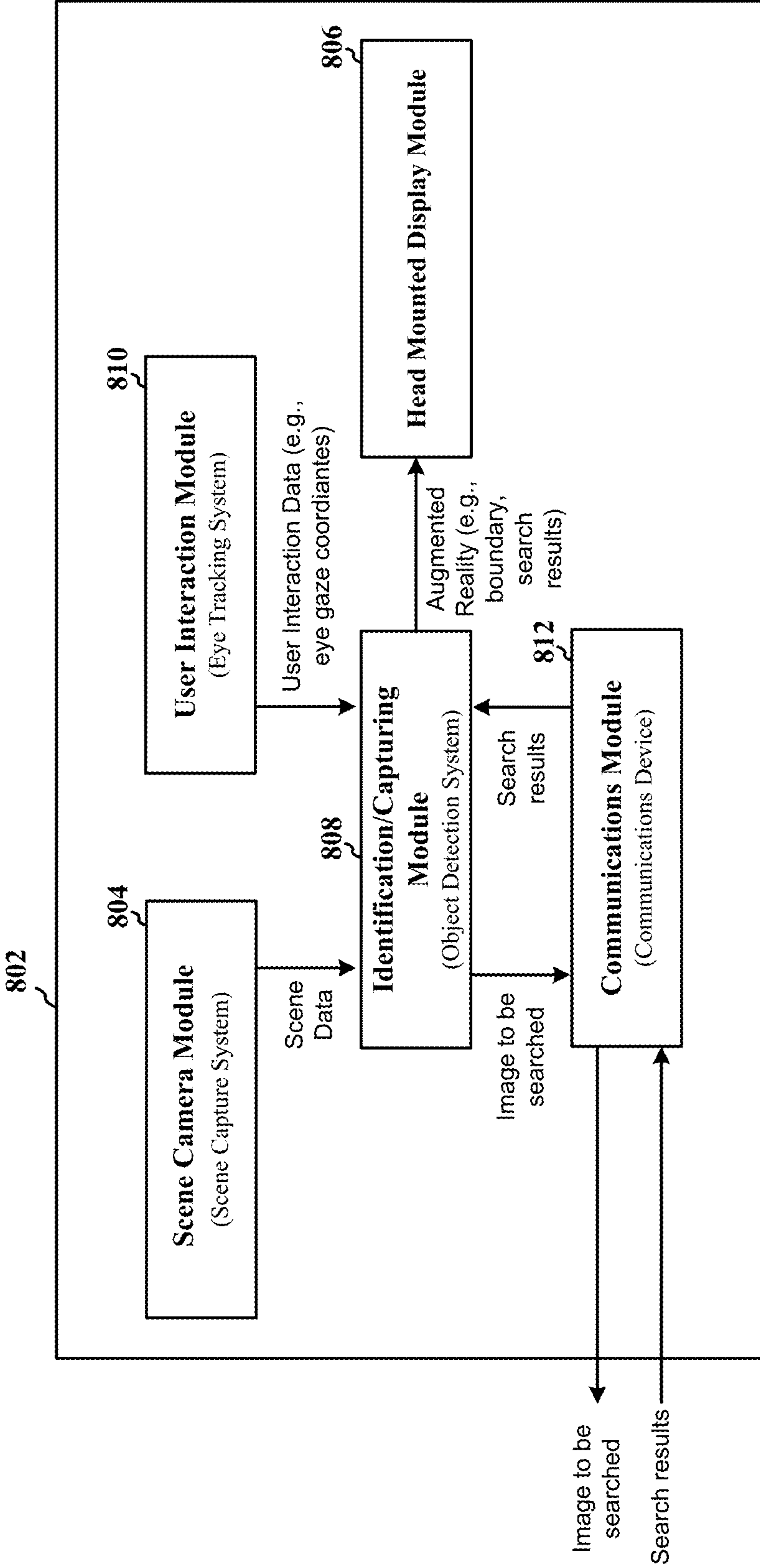


FIG. 8

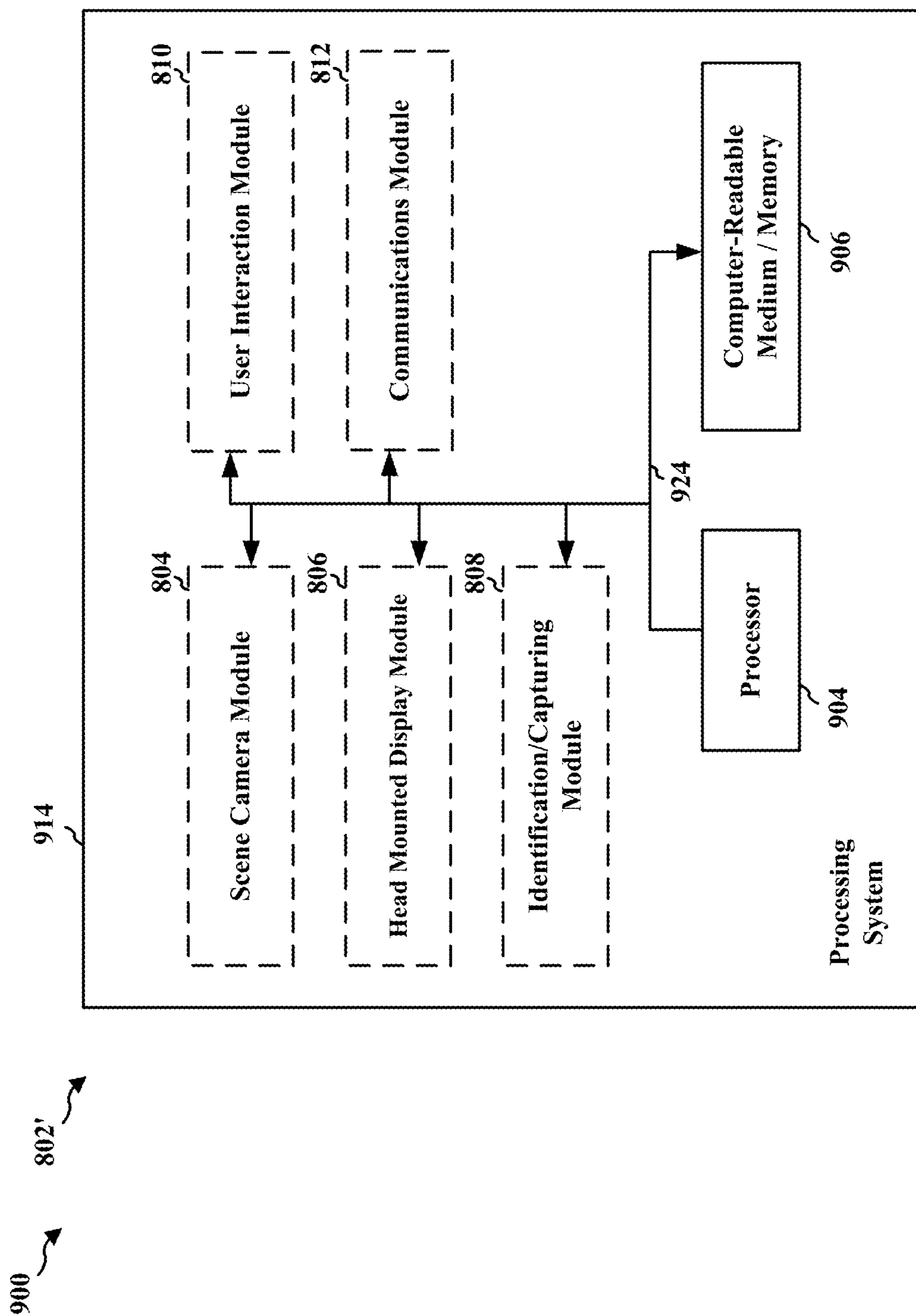


FIG. 9



**VISUAL SEARCH IN REAL WORLD USING  
OPTICAL SEE-THROUGH HEAD MOUNTED  
DISPLAY WITH AUGMENTED REALITY  
AND USER INTERACTION TRACKING**

CROSS-REFERENCE TO RELATED  
APPLICATION(S)

**[0001]** This application claims the benefit of U.S. Provisional Application Ser. No. 61/867,579, entitled “Improving Visual Search in Real World Using Optical See-Through HMD with Augmented Reality and Eye Tracking” and filed on Aug. 19, 2013, which is expressly incorporated by reference herein in its entirety and for all purposes.

**[0002]** This application is a continuation of U.S. patent application Ser. No. 17/304,739, filed Jun. 24, 2021, which is a continuation of U.S. patent application Ser. No. 16/414,667, filed May 16, 2019, which is a continuation of U.S. patent application Ser. No. 16/181,175, filed Nov. 5, 2018, which is a continuation of U.S. patent application Ser. No. 14/151,664, filed Jan. 9, 2014, all of which are expressly incorporated by reference in their entirety and for all purposes.

FIELD

**[0003]** The present disclosure relates generally to augmented reality (AR) devices, e.g., AR eyeglasses, having optical see-through head mounted displays (HMD) and user interaction capability, and more particularly, to conducting online visual searches in real world using such AR devices. AR is a technology in which a user’s view of the real world is enhanced with additional information generated from a computer model. The enhancements may include labels, 3D rendered models, or shading and illumination changes. AR allows a user to work with and examine the physical real world, while receiving additional information about the objects in it.

BACKGROUND

**[0004]** AR devices typically include an optical see-through HMD and one or more user input mechanisms that allow users to simultaneously see and interact with their surroundings while interacting with applications, such as e-mail and media players. User input mechanisms may include one or more of gesture recognition technology, eye tracking technology, and other similar mechanisms.

**[0005]** In AR devices, also referred to herein as HMD systems, virtual objects displayed in the HMD screen augment real world scenes viewed by a user through the HMD. For example, a person in the field of view of a user may be augmented with her name, an artwork may be augmented with descriptive information, and a book may be augmented with its price and so on.

**[0006]** With such an HMD system, a user can perform a visual search to obtain more information about a specific object visible in a real-world scene, i.e., within the field of view of the user. During the visual search, the HMD system generally takes a snapshot of the scene in the field of view, conducts a search using an online visual search engine (e.g. Google Goggles), retrieves the results, and displays the results. The results are displayed on the HMD screen to provide an augmented reality scene to the user that includes the real-world scene, with the object just searched visible through the HMD, and the augmented reality search result.

**[0007]** A problem with this visual search approach is that other objects may be present in the real-world scene resulting in inaccurate search results and redundant or extraneous data transfer to the online visual search engine. To work around this issue, a user may have to move closer to the object of interest so that the object of interest approximately covers her entire field of view. Although this may work for visual searching with a Smartphone, wherein a user zooms in on the object to be searched and takes a picture, this approach is not desirable for optical see-through HMD systems in that it requires the user to move and thus may not provide a good user experience.

SUMMARY

**[0008]** In an aspect of the disclosure, a method, an apparatus, and a computer program product for conducting online visual searches through an augmented reality (AR) device having an optical see-through head mounted display (HMD) are disclosed. An apparatus identifies a portion of an object in a field of view of the HMD based on user interaction with the HMD. The portion includes searchable content, such as a barcode. The user interaction may be an eye gaze or a gesture. A user interaction point in relation to the HMD screen is tracked to locate a region of the object that includes the portion and the portion is detected within the region. The apparatus captures an image of the portion. The identified portion of the object does not encompass the entirety of the object. Accordingly, the size of the image is less than the size of the object in the field of view. The apparatus transmits the image to a visual search engine.

BRIEF DESCRIPTION OF THE DRAWINGS

**[0009]** FIG. 1 is a diagram illustrating an online visual search system for conducting online searches of objects visible through an AR device having an optical see-through HMD.

**[0010]** FIG. 2 is an illustration of an AR device in the form of a pair of eyeglasses.

**[0011]** FIG. 3 is an illustration of a real-world scene through an optical see-through HMD with augmented reality.

**[0012]** FIG. 4 is a diagram illustrating elements of an AR device.

**[0013]** FIG. 5 is an illustration of a real-world scene visible to a user through an optical see-through HMD, including an object of interest to be searched on line.

**[0014]** FIG. 6 is an illustration of an augmented reality scene visible to a user that includes the real-world scene of FIG. 5 and the augmented reality search result.

**[0015]** FIG. 7 is a flow chart of a method of conducting online searches of objects visible through an AR device having an optical see-through HMD.

**[0016]** FIG. 8 is a diagram illustrating elements of an AR device that conducts online searches of objects visible through an optical see-through HMD.

**[0017]** FIG. 9 is a diagram illustrating an example of a hardware implementation for an AR device employing a processing system.

DETAILED DESCRIPTION

**[0018]** The detailed description set forth below in connection with the appended drawings is intended as a description of various configurations and is not intended to represent the



only configurations in which the concepts described herein may be practiced. The detailed description includes specific details for the purpose of providing a thorough understanding of various concepts. However, it will be apparent to those skilled in the art that these concepts may be practiced without these specific details. In some instances, well known structures and components are shown in block diagram form in order to avoid obscuring such concepts.

[0019] Several aspects of visual searching through an AR device having an optical see-through HMD and user interaction tracking will now be presented with reference to various apparatus and methods. These apparatus and methods will be described in the following detailed description and illustrated in the accompanying drawings by various blocks, modules, components, circuits, steps, processes, algorithms, etc. (collectively referred to as “elements”). These elements may be implemented using electronic hardware, computer software, or any combination thereof. Whether such elements are implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system.

[0020] By way of example, an element, or any portion of an element, or any combination of elements may be implemented with a “processing system” that includes one or more processors. Examples of processors include microprocessors, microcontrollers, digital signal processors (DSPs), field programmable gate arrays (FPGAs), programmable logic devices (PLDs), state machines, gated logic, discrete hardware circuits, and other suitable hardware configured to perform the various functionality described throughout this disclosure. One or more processors in the processing system may execute software. Software shall be construed broadly to mean instructions, instruction sets, code, code segments, program code, programs, subprograms, software modules, applications, software applications, software packages, routines, subroutines, objects, executables, threads of execution, procedures, functions, etc., whether referred to as software, firmware, middleware, microcode, hardware description language, or otherwise.

[0021] Accordingly, in one or more exemplary embodiments, the functions described may be implemented in hardware, software, firmware, or any combination thereof. If implemented in software, the functions may be stored on or encoded as one or more instructions or code on a computer-readable medium. Computer-readable media includes computer storage media. Storage media may be any available media that can be accessed by a computer. By way of example, and not limitation, such computer-readable media can comprise a random-access memory (RAM), a read-only memory (ROM), an electrically erasable programmable ROM (EEPROM), compact disk ROM (CD-ROM) or other optical disk storage, magnetic disk storage or other magnetic storage devices, or any other medium that can be used to carry or store desired program code in the form of instructions or data structures and that can be accessed by a computer. Disk and disc, as used herein, includes CD, laser disc, optical disc, digital versatile disc (DVD), and floppy disk where disks usually reproduce data magnetically, while discs reproduce data optically with lasers. Combinations of the above should also be included within the scope of computer-readable media.

[0022] FIG. 1 is a diagram illustrating a visual online search system 100 for conducting online searches of objects visible through an optical see-through HMD with aug-

mented reality, herein referred to as an AR device or an HMD system. The online search system 100 includes an AR device 102, a communications network 104, and a visual search engine 106. The AR device 102 captures an image of an object seen through a HMD and transmits the image to the visual search engine 106 over the communications network 104. The visual search engine 106 conducts an online search of the object captured in the image using known technology and sends the results to the AR device 102 over the communications network 104. The AR device 102 displays the search results as an augmented reality in the HMD.

[0023] FIG. 2 is an illustration of an example AR device 200 in the form of a pair of eyeglasses. The AR device 200 is configured such that the user of the device is able to view real-world scenes through optical see-through HMDs together with content displayed on the HMDs, including both two-dimensional (2D) and three-dimensional (3D) AR content. The AR device 200 may also be configured to allow the user to interact with the content and possibly with remote devices, systems or networks through wireless communication. The AR device may also provide feedback to the user as a result of such interactions, including for example, audio, video or tactile feedback. To these ends, the example AR device 200 includes a pair of optical see-through HMDs 202, 204, an on-board processing system 206, one or more sensors, such as a scene camera 208, one or more eye tracking components (not visible) for each of the right eye and left eye, one or more user-interaction feedback devices 210 and a transceiver 212.

[0024] The processing system 206 and the eye tracking components provide eye tracking capability. Depending on the eye tracking technology being employed, eye tracking components may include one or both of eye cameras and infra-red emitters, e.g. diodes. The processing system 206 and the scene camera 208 provide gesture tracking capability.

[0025] The feedback devices 210 provide perception feedback to the user in response to certain interactions with the AR device. Feedback devices 210 may include a speaker or a vibration device. Perception feedback may also be provided by visual indication through the HMD.

[0026] The transceiver 212 facilitates wireless communication between the processing system 206 and remote devices, systems or networks. For example, the AR device may communicate with remote servers through the transceiver 212 for purposes of remote processing, such as on-line searches through remote search engines.

[0027] As mentioned above, the AR device 200 allows a user to view real-world scenes through optical see-through HMDs together with content displayed on the HMDs. For example, with reference to FIG. 3, as a user is viewing a real-world scene 300 through the optical see-through HMDs 202, 204, the scene camera 208 may capture an image of the scene and send the image to the on-board processing system 206. The processing system 206 may process the image and output AR content 302 for display on the HMDs 202, 204. The content 302 may provide information describing what the user is seeing. In some cases, the processing system 206 may transmit the image through the transceiver 212 to a remote processor (not shown) for processing. The processing system 206 may also display one or more application icons 304, 306, 308 on the HMDs 202, 204 and output application content, such as e-mails, documents, web pages,



or media content such as video games, movies or electronic books, in response to user interaction with the icons.

[0028] User interaction with the AR device 200 is provided by one or more user input mechanisms, such as a gesture tracking module or an eye-gaze tracking module. Gesture tracking is provided by the scene camera 208 in conjunction with a gesture tracking module of the processing system 206. With gesture tracking, a user may attempt to activate an application by placing his finger on an application icon 304, 306, 308 in the field of view of the AR device. The scene camera 208 captures an image of the finger and sends the image to the gesture tracking module. The gesture tracking module processes the image and determines coordinates of a gesture point corresponding to where the user is pointing. The processing system 206 compares the coordinate location of the gesture point to the coordinate location of the icon on the display. If the locations match, or are within a threshold distance of each other, the processing system 206 determines that the user has selected the icon 304, 306, 308 and accordingly, launches the application. Often, such eye-gaze based launching is coupled with another form of input, e.g., gesture, to confirm the user's intention of launching the application.

[0029] Eye-gaze tracking is provided by the eye tracking components (not visible) in conjunction with an eye tracking module of the processing system 206. A user may attempt to activate an application by gazing at an application icon 304, 306, 308 in the field of view of the AR device. The eye tracking components capture images of the eyes, and provide the images to the eye tracking module. The eye tracking module processes the images and determines coordinates of an eye-gaze point corresponding to where the user is looking. The processing system 206 compares the coordinate location of the eye-gaze point to the coordinate location of the icon on the display. If the locations match, or are within a threshold distance of each other, the processing system 206 determines that the user has selected the icon 304, 306, 308 and accordingly, launches the application.

[0030] FIG. 4 is a diagram illustrating elements of an example AR device 400 with optical see-through HMDs 402. The AR device 400 may include one or more sensing devices, such as infrared (IR) diodes 404 facing toward the wearer of the AR device and eye cameras 406 facing toward the wearer. A scene camera 408 facing away from the wearer captures images of the field of view seen by the user through the HMD 402. The cameras 406, 408 may be video cameras. While only one IR diode 404 and one eye camera 406 are illustrated, the AR device 400 typically includes several diodes and cameras for each of the left eye and right eye. A single scene camera 408 is usually sufficient. For ease of illustration only one of each sensor type is shown in FIG. 4.

[0031] The AR device 400 includes an on-board processing system 410, which in turn includes one or more of an eye tracking module 412 and a gesture tracking module 414. An object selection module 416 processes the outputs of the one or more tracking modules to determine user interactions and tracking module accuracy. A tracking calibration module 418 calibrates the one or more tracking modules if the tracking module is determined to be inaccurate.

[0032] The on-board processing system 410 may also include a scene camera calibration module 420, a graphical user interface (GUI) adjustment module 422, a perception feedback module 424, and a search module 436. The scene camera calibration module 420 calibrates the AR device so

that the AR content is aligned with real world objects. The GUI adjustment module 422 may adjust the parameters of GUI objects displayed on the HMD to compensate for eye-tracking or gesture-tracking inaccuracies detected by the object selection module 416. Such adjustments may precede, supplement, or substitute for the actions of the tracking calibration module 418. The feedback module 424 controls one or more feedback devices 426 to provide perception feedback to the user in response to one or more types of user interactions. For example, the feedback module may command a feedback device 426 to output sound when a user selects an icon in the field of view using a gesture or eye gaze. The search module 436 identifies and captures a portion of a real world object seen through the optical see-through HMD for purposes of conducting online searches. The portion to be captured is identified by the search module 436 based on real world scene data provided by the scene camera 408 and one of gaze coordinates provided by the eye tracking module 412, and gesture coordinates provided by the gesture tracking module 414.

[0033] The AR device 400 further includes memory 428 for storing program code to implement the foregoing features of the on-board processing system 410. A communications module 430 and transceiver 432 facilitate wireless communications with remote devices, systems and networks. For example, in one implementation, an image of a real-world object may be captured by the search module 436 and transmitted by the communications module 430 and the transceiver 432 to a remote search engine, with subsequent search results being received by the transceiver.

[0034] With further respect to eye tracking capability, the diodes 404 and eye cameras 406, together with the eye tracking module 412, provide eye tracking capability as generally described above. In the example implementation of FIG. 4, the eye tracking capability is based on known infrared technology. One such known technology uses infrared light emitting diodes and infrared sensitive video camera for remotely recording images of the eye. Infrared light output by the diode 404 enters the eye and is absorbed and re-emitted by the retina, thereby causing a "bright eye effect" that makes the pupil brighter than the rest of the eye. The infrared light also gives rise to an even brighter small glint that is formed on the surface of the cornea. The eye tracking module 412 acquires a video image of the eye from the eye camera 406, digitizes it into a matrix of pixels, and then analyzes the matrix to identify the location of the pupil's center relative to the glint's center, as well as a vector between these centers. Based on the determined vector, the eye tracking module 412 outputs eye gaze coordinates defining an eye gaze point (E).

[0035] The scene camera 408, together with the gesture tracking module 414, provide gesture tracking capability using a known technology as generally described above. In the example implementation of FIG. 4, the gesture tracking capability is based on gesture images captured by the scene camera 408. The gesture images are processed by the gesture tracking module 414 by comparing captured images to a catalog of images to determine if there is a match. For example, the user may be pointing at an icon in the field of view. The gesture tracking module 412 may detect a match between the gesture image and a cataloged image of pointing and thereby recognize the gesture as pointing. Upon detection of a recognized gesture, the gesture tracking module 414 processes the captured image further to deter-



mine the coordinates of a relevant part of the gesture image. In the case of finger pointing, the relevant part of the image may correspond to the tip of the finger. The gesture tracking module 414 outputs gesture coordinates defining a gesture point (G).

[0036] The object selection processor 416 functions to determine whether interactions of the user, as characterized by one or more of the eye tracking module 412 and the gesture tracking module 414, correspond to a selection of an object, e.g., application icon, displayed on the HMD 402 and visible in the field of view. If an interaction does correspond to a selection by the user, for example, a selection of an icon to launch an application 434, the object selection processor 416 outputs a command to the application.

[0037] As previously mentioned, the AR device 410 may provide for online searching related to objects seen through the optical see-through HMD. Components of the AR device that facilitate such searching include the scene camera 408, one or more of the user interaction tracking modules, e.g., the eye tracking module 412 and the gesture tracking module 414, the object portion identification module, the search module 436, and/or communication module 430.

[0038] FIG. 5 is an illustration 500 of a real-world scene 502 visible to a user through an optical see-through HMD, including an object 504 of interest to be searched on line. FIG. 6 is an illustration of an augmented reality scene visible to a user that includes the real-world scene of FIG. 5 and the augmented reality search result. In FIG. 5, the real-world scene 502 includes an object of interest 504 to be searched on line. In the example of FIG. 5, the object 504 is a book. The eye tracking module 412 is configured to track movement of the user's eye and to detect when the user is gazing at a portion (e.g., portion 508) of the object 504. In this case, the user is determined to be gazing at a point 506 overlaying the portion 508 of the object (e.g., a book) where the barcode is located.

[0039] The search module 436 receives data from the scene camera 408 corresponding to a captured scene (herein referred to as "scene data") and data from the eye tracking module 412 corresponding to the eye gaze of the user (herein referred to as "gaze coordinates" or "eye gaze data"). The search module 436 processes the scene data and the eye gaze data to identify the portion 508 of the object 504 and to provide an image of the portion. With reference to FIG. 5, the search module 436 determines the portion 508 of the object, and then determines a boundary 510 around the portion. The image of the portion corresponds to the scene that lies within the boundary 510. Accordingly, the size of the image is less than the size of the object 504 as well as the scene data 502 in the field of view.

[0040] Data corresponding to the image of the portion is provided to the communications module 430, which in turn, transmits the image data to the online visual search engine 106 over the communications network 104. The communications module 430 may be, for example, a Bluetooth device. Search results are transmitted back to the AR device, received by the communications module 430 and provided to the search module 436 for display. With reference to FIG. 6, search results 604 are displayed on the HMD screen in the vicinity of the object 504.

[0041] FIG. 7 is a flowchart of a method of conducting an online visual search through an AR device having an optical see-through HMD. At step 702, the AR device identifies a portion 508 of an object 504 in a field of view of the HMD

based on user interaction with the HMD. In one configuration, the user interaction may be an eye gaze 506, in which case the AR device identifies a portion 508 of an object by tracking the eye gaze 506 of the user to determine a location of the eye gaze, locating a region 512 of the object corresponding to the location of the eye gaze, and detecting the portion 508 within the region. In another configuration, the user interaction may be a gesture, in which case the AR device identifies a portion of an object by tracking the gesture of the user to determine a location of the gesture, locating a region of the object corresponding to the location of the gesture, and detecting the portion within the region.

[0042] At step 704, the AR device may display an augmented reality 510 on the HMD in the vicinity of the detected portion 508. The augmented reality 510 may be a boundary around the portion 508.

[0043] At step 706, the AR device may receive an indication to initiate an online visual search of the object 504 using content contained the portion 508. For example, upon seeing the augmented reality 510 that identifies the portion 508 of the object, the user may request that the AR device initiate an online search of the object based on the portion. Such a request may be input by the user through a physical search activation button on the AR device or through user interaction with the HMD screen. In the latter case, the AR device may display a search icon on the HMD screen and the search may be initiated through user interaction, e.g., eye gaze or gesture, with the search icon.

[0044] At step 708, the AR device captures an image of the portion 508. The image may correspond to the area of the object 504 within the boundary 510. Accordingly, the size of the image is less than the size of the object in the field of view. At step 710, the AR device transmits the image to a search engine.

[0045] At step 712, the AR device receives search results from the search engine, and at step 714, displays the search results as an augmented reality image 604 on the HMD screen in the vicinity of the object 504.

[0046] FIG. 8 is a diagram 800 illustrating elements of an AR device 802 that conducts online visual searches. The AR device 802 includes a scene camera module 804, a HMD module 806, an identification/capture module 808, a user interaction module 810 and a communications module 812. These modules may be referred to by different names, as indicated parenthetically, and may correspond to one or more of the modules of FIG. 4. For example, the user interaction module 810 may be the eye tracking module 412 or the gesture tracking module 414, and the identification/capture module 808 may be the search module 436.

[0047] The identification/capture module 808 identifies a portion of an object in a field of view of the HMD based on user interaction with the HMD. To this end, the scene camera module 804 provides scene data to the identification/capturing module 808, and the user interaction module 810 provides user interaction data to the identification/capture module 808. The scene data corresponds to image data of the real world scene visible in the field of view of the HMD module 806 and captured by the camera 804. The user interaction data may be eye gaze data or gesture data. In either case, the user interaction data includes coordinate data that defines a user interaction point, eye-gaze point (E), in terms of the coordinate system corresponding to the HMD screen.



[0048] In the case where the user interaction is an eye gaze, the identification/capture module **808** identifies a portion **508** of an object **504** by tracking the eye gaze of the user to determine a location of the eye gaze, locating a region **512** of the object corresponding to the location of the eye gaze, and detecting the portion **508** within the region. In the case where the user interaction is a gesture, the identification/capture module **808** identifies a portion of an object by track the gesture of the user to determine a location of the gesture, locating a region of the object corresponding to the location of the gesture, and detecting the portion within the region.

[0049] Upon detecting the portion **508**, the identification/capture module **808** captures an image of the portion. To this end, the identification/capture module **808** may determine a boundary around the portion and display the boundary as an augmented reality on the HMD. In one configuration, the image of the portion includes only image data within the boundary **510**. This is beneficial in that it reduces the amount of image data to be transmitted to the search engine and also provides a more focused search by reducing the amount of extraneous searchable content from the search request.

[0050] The communication module **812** transmits the image to a search engine. Subsequently, the communication module **812** receives data corresponding to search results from the search engine. The search results may be provided to the identification/capturing module **808** and displayed as an augmented reality image on the HMD screen in the vicinity of the object.

[0051] The AR devices, as illustrated in FIGS. **4** and **8** may include additional modules that perform each of the steps of the algorithm in the aforementioned flow chart of FIG. **7**. As such, each step in the aforementioned flow chart of FIG. **7** may be performed by a module and the apparatus may include one or more of those modules. The modules may be one or more hardware components specifically configured to carry out the stated processes/algorithm, implemented by a processor configured to perform the stated processes/algorithm, stored within a computer-readable medium for implementation by a processor, or some combination thereof.

[0052] FIG. **9** is a diagram **900** illustrating an example of a hardware implementation for an apparatus **802'** employing a processing system **914**. The processing system **914** may be implemented with a bus architecture, represented generally by the bus **924**. The bus **924** may include any number of interconnecting buses and bridges depending on the specific application of the processing system **914** and the overall design constraints. The bus **924** links together various circuits including one or more processors and/or hardware modules, represented by the processor **904**, the modules **804**, **806**, **808**, **810**, **812** and the computer-readable medium/memory **906**. The bus **924** may also link various other circuits such as timing sources, peripherals, voltage regulators, and power management circuits, which are well known in the art, and therefore, will not be described any further.

[0053] The processing system **914** includes a processor **904** coupled to a computer-readable medium/memory **906**. The processor **904** is responsible for general processing, including the execution of software stored on the computer-readable medium/memory **906**. The software, when executed by the processor **904**, causes the processing system **914** to perform the various functions described supra for any particular apparatus. The computer-readable medium/memory **906** may also be used for storing data that is

manipulated by the processor **904** when executing software. The processing system further includes at least one of the modules **804**, **806**, **808**, **810** and **812**. The modules may be software modules running in the processor **904**, resident/stored in the computer readable medium/memory **906**, one or more hardware modules coupled to the processor **904**, or some combination thereof.

[0054] In one configuration, the apparatus **802/802'** includes means for identifying a portion of an object in a field of view of the HMD based on user interaction with the HMD, means for capturing an image of the portion, wherein the size of the image is less than the size of the object in the field of view, and means for transmitting the image to a search engine. The apparatus **802/802'** also includes means for displaying an augmented reality on the HMD in the vicinity of the portion, means for receiving an indication to initiate an online visual search of the object using the portion, means for receiving search results from the search engine; and means for displaying the search results as an augmented reality image on the HMD screen in the vicinity of the object. The aforementioned means may be one or more of the aforementioned modules of the apparatus **802** and/or the processing system **914** of the apparatus **802'** configured to perform the functions recited by the aforementioned means.

[0055] In summary, methods and apparatus disclosed herein allow a user to select only a subset of scene with an eye gaze that contains an object of interest. With reference to FIGS. **5** and **6**, a visual search using a HMD with augmented reality and eye tracking works as follows: A user looks at the object of interest in a scene visible through the HMD and initiates a visual search via other input techniques such as a gesture. An eye tracking system included in the HMD system determines a location of a gaze point corresponding to the point of the scene at which the user is gazing. The eye tracking system identifies the object of interest based on the gaze point. For example, with reference to FIG. **5**, based on the location of the eye gaze, the eye tracking system identifies the object of interest as the bar code. Such identification may be made by color recognition, wherein the system detects a change in color between the edge of the bar code and the back cover of the book.

[0056] With continued reference to FIG. **5**, once the eye tracking system determines the object of interest, it identifies an area of the scene containing the object and highlights the area with an augmented reality, such as an augmented reality boundary surrounding the object. The selected part of the scene corresponding to the area within the boundary is extracted and sent to the cloud (online visual search engine e.g. Google Goggles) for further analysis. As shown in FIG. **6**, the user's view is augmented with the retrieved search result.

[0057] The benefits of this approach are: The user is not required to move to the object of interest. Only relevant part of the image is sent to the cloud; thus saves redundant data transfer. Minimum workload on the search engine since it does not need to analyze all objects presented in the user's scene, only a subset which contains object of interest.

[0058] In an aspect of the disclosure, a method, an apparatus, and a computer program product provide for conducting an online visual search of an object present in a scene visible through a head mounted display worn by a user. A method includes tracking an eye gaze of the user to locate the object, capturing an image of the object, and transmitting



the image of the object to a search engine. The size of the image is less than the size of the scene. The method may also include receiving an indication to initiate an online visual search of an object. The method may further include receiving search results from the search engine, and displaying the search results as an augmented reality image on the HMD screen in the vicinity of the object.

**[0059]** A corresponding apparatus for conducting an online visual search of an object present in a scene visible through a head mounted display worn by a user, includes means for tracking an eye gaze of the user to locate the object, means for capturing an image of the object, and means for transmitting the image of the object to a search engine. The size of the image is less than the size of the scene. The apparatus may also include means for receiving an indication to initiate an online visual search of an object. The apparatus may further include means for receiving search results from the search engine, and means for displaying the search results as an augmented reality image on the HMD screen in the vicinity of the object.

**[0060]** Another corresponding apparatus for conducting an online visual search of an object present in a scene visible through a head mounted display worn by a user, includes a memory, and at least one processor coupled to the memory and configured to track an eye gaze of the user to locate the object, capture an image of the object, and transmit the image of the object to a search engine. The size of the image is less than the size of the scene. The processor may be configured to receive an indication to initiate an online visual search of an object. The processor may be further configured to receive search results from the search engine, and display the search results as an augmented reality image on the HMD screen in the vicinity of the object.

**[0061]** A computer program product for conducting an online visual search of an object present in a scene visible through a head mounted display worn by a user, includes a computer-readable medium comprising code for tracking an eye gaze of the user to locate the object, for capturing an image of the object, and for transmitting the image of the object to a search engine. The size of the image is less than the size of the scene. The product may include code for receiving an indication to initiate an online visual search of an object. The product may further include code for receiving search results from the search engine, and displaying the search results as an augmented reality image on the HMD screen in the vicinity of the object.

**[0062]** It is understood that the specific order or hierarchy of steps in the processes disclosed is an illustration of exemplary approaches. Based upon design preferences, it is understood that the specific order or hierarchy of steps in the processes may be rearranged. Further, some steps may be combined or omitted. The accompanying method claims present elements of the various steps in a sample order, and are not meant to be limited to the specific order or hierarchy presented.

**[0063]** The previous description is provided to enable any person skilled in the art to practice the various aspects described herein. Various modifications to these aspects will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other aspects. Thus, the claims are not intended to be limited to the aspects shown herein, but is to be accorded the full scope consistent with the language claims, wherein reference to an element in the singular is not intended to mean “one and only one”

unless specifically so stated, but rather “one or more.” The word “exemplary” is used herein to mean “serving as an example, instance, or illustration.” Any aspect described herein as “exemplary” is not necessarily to be construed as preferred or advantageous over other aspects.” Unless specifically stated otherwise, the term “some” refers to one or more. Combinations such as “at least one of A, B, or C,” “at least one of A, B, and C,” and “A, B, C, or any combination thereof” include any combination of A, B, and/or C, and may include multiples of A, multiples of B, or multiples of C. Specifically, combinations such as “at least one of A, B, or C,” “at least one of A, B, and C,” and “A, B, C, or any combination thereof” may be A only, B only, C only, A and B, A and C, B and C, or A and B and C, where any such combinations may contain one or more member or members of A, B, or C. All structural and functional equivalents to the elements of the various aspects described throughout this disclosure that are known or later come to be known to those of ordinary skill in the art are expressly incorporated herein by reference and are intended to be encompassed by the claims. Moreover, nothing disclosed herein is intended to be dedicated to the public regardless of whether such disclosure is explicitly recited in the claims. No claim element is to be construed as a means plus function unless the element is expressly recited using the phrase “means for.”

What is claimed is:

1. An apparatus for processing one or more images, the apparatus comprising:
  - at least one memory; and
  - at least one processor coupled to the at least one memory and configured to:
    - receive a first input from a user of the apparatus, wherein the first input comprises an eye gaze of the user;
    - determine a portion of interest within a scene based on the first input;
    - in response to the first input, cause content associated with the portion of interest to be displayed on a display;
    - receive a second input from the user of the apparatus, wherein the second input comprises a gesture, and wherein the second input is associated with the displayed content; and
    - initiate an action based on the second input.
2. The apparatus of claim 1, wherein the gesture includes a hand gesture.
3. The apparatus of claim 2, wherein the hand gesture comprises a physical interaction with the apparatus.
4. The apparatus of claim 3, wherein the physical interaction with the apparatus comprises an interaction with a physical button of the apparatus.
5. The apparatus of claim 2, wherein the at least one processor is configured to:
  - receive a third input from the user of the apparatus; and
  - in response to the third input, cause additional content to be displayed.
6. The apparatus of claim 5, wherein the third input comprises an interaction with a physical interface of the apparatus.
7. The apparatus of claim 6, wherein the physical interface is a physical button.
8. The apparatus of claim 6, wherein, to cause the content associated with the portion of interest to be displayed, the at



least one processor is configured to cause displayed content to be adjusted based on the first input.

9. The apparatus of claim 2, wherein, to initiate the action, the at least one processor is configured to launch an application.

10. The apparatus of claim 2, wherein the first input includes an interaction with a search icon.

11. The apparatus of claim 10, wherein, based on the interaction with the search icon, a search is performed based on a boundary around a portion in the scene.

12. The apparatus of claim 2, wherein, to receive the first input, the at least one processor is configured to track the eye gaze of the user to determine a location of the eye gaze relative to the display, wherein the portion of interest within the scene corresponds to the location of the eye gaze.

13. The apparatus of claim 1, wherein, to receive the second input, the at least one processor is configured to track a gesture of the user to determine a location of the gesture.

14. The apparatus of claim 1, wherein the content comprises a boundary.

15. The apparatus of claim 1, wherein the action comprises a visual search process associated with the portion of interest.

16. The apparatus of claim 2, wherein the at least one processor is configured to obtain, from an image capture apparatus, at least one image of the scene.

17. The apparatus of claim 16, further comprising the image capture apparatus.

18. The apparatus of claim 16, wherein the at least one processor is configured to process the at least one image of the scene to generate data associated with the portion of interest.

19. The apparatus of claim 18, wherein the at least one processor is configured to cause the data to be transmitted to a remote device configured to generate visual search results.

20. The apparatus of claim 19, wherein the at least one processor is configured to:

receive the visual search results from the remote device;  
and

cause the visual search results to be displayed on the display.

21. The apparatus of claim 16, wherein the apparatus is a head mounted device, and wherein the at least one image of the scene corresponds to a field of view observed by the user through a head mounted device.

22. The apparatus of claim 1, wherein the apparatus is a head mounted device.

23. The apparatus of claim 1, further comprising the display configured to display the content.

24. An apparatus for processing one or more images, the apparatus comprising:

at least one memory;  
at least one eye tracking camera;  
at least one scene camera;  
at least one infrared (IR) camera;  
at least one display;  
at least one battery;  
at least one transceiver; and

at least one processor coupled to the at least one memory and configured to:

receive a first input from a user of the apparatus,  
wherein the first input comprises an eye gaze of the user;

determine a portion of interest within a scene based on the first input;

in response to the first input, cause content associated with the portion of interest to be displayed on the at least one display;

receive a second input from the user of the apparatus, wherein the second input comprises a gesture, and wherein the second input is associated with the displayed content; and

initiate an action based on the second input.

25. The apparatus of claim 24, wherein the gesture includes a hand gesture.

26. The apparatus of claim 25, wherein the hand gesture comprises a physical interaction with the apparatus.

27. The apparatus of claim 26, wherein the physical interaction with the apparatus comprises an interaction with a physical button of the apparatus.

28. The apparatus of claim 25, wherein the at least one processor is configured to:

receive a third input from the user of the apparatus; and  
in response to the third input, cause additional content to be displayed on the at least one display.

29. The apparatus of claim 28, wherein the third input comprises an interaction with a physical interface of the apparatus.

30. The apparatus of claim 29, wherein the physical interface is a physical button.

31. The apparatus of claim 29, wherein, to cause the content associated with the portion of interest to be displayed, the at least one processor is configured to cause displayed content to be adjusted based on the first input.

32. The apparatus of claim 25, wherein, to initiate the action, the at least one processor is configured to launch an application.

33. The apparatus of claim 25, wherein the first input includes an interaction with a search icon.

34. The apparatus of claim 33, wherein, based on the interaction with the search icon, a search is performed based on a boundary around a portion in the scene.

35. The apparatus of claim 25, wherein, to receive the first input, the at least one processor is configured to track the eye gaze of the user to determine a location of the eye gaze relative to the at least one display, wherein the portion of interest within the scene corresponds to the location of the eye gaze.

36. The apparatus of claim 24, wherein, to receive the second input, the at least one processor is configured to track a gesture of the user to determine a location of the gesture.

37. The apparatus of claim 24, wherein the content comprises a boundary.

38. The apparatus of claim 24, wherein the action comprises a visual search process associated with the portion of interest.

39. The apparatus of claim 25, wherein the at least one processor is configured to obtain, from the at least one scene camera, at least one image of the scene.

40. The apparatus of claim 39, wherein the at least one processor is configured to process the at least one image of the scene to generate data associated with the portion of interest.

41. The apparatus of claim 40, wherein the at least one processor is configured to cause the data to be transmitted, via the at least one transceiver, to a remote device configured to generate visual search results.



**42.** The apparatus of claim **41**, wherein the at least one processor is configured to:

- receive the visual search results from the remote device;
- and
- cause the visual search results to be displayed on the at least one display.

**43.** The apparatus of claim **39**, wherein the apparatus is a head mounted device, and wherein the at least one image of the scene corresponds to a field of view observed by the user through a head mounted device.

**44.** The apparatus of claim **24**, wherein the apparatus is a head mounted device.

**45.** A method of processing one or more images, the method comprising:

- receiving a first input from a user of a device, wherein the first input comprises an eye gaze of the user;
- determining a portion of interest within a scene based on the first input;
- displaying, in response to the first input, content associated with the portion of interest on a display;
- receiving a second input from the user of the device, wherein the second input comprises a gesture, and wherein the second input is associated with the displayed content; and
- initiating an action based on the second input.

**46.** The method of claim **45**, wherein the gesture includes a hand gesture.

**47.** The method of claim **46**, wherein the hand gesture comprises a physical interaction with the device.

**48.** The method of claim **47**, wherein the physical interaction with the device comprises an interaction with a physical button of the device.

- 49.** The method of claim **46**, further comprising:
- receiving a third input from the user of the device; and
  - displaying, in response to the third input, additional content.

**50.** The method of claim **49**, wherein the third input comprises an interaction with a physical interface of the device.

**51.** The method of claim **50**, wherein the physical interface is a physical button.

**52.** The method of claim **50**, wherein displaying the content associated with the portion of interest comprises adjusting displayed content based on the first input.

**53.** The method of claim **46**, wherein initiating the action comprises launching an application.

**54.** The method of claim **46**, wherein the first input includes an interaction with a search icon.

**55.** The method of claim **54**, wherein, based on the interaction with the search icon, a search is performed based on a boundary around a portion in the scene.

**56.** The method of claim **46**, wherein receiving the first input comprises tracking the eye gaze of the user to determine a location of the eye gaze relative to the display, wherein the portion of interest within the scene corresponds to the location of the eye gaze.

**57.** The method of claim **45**, wherein receiving the second input comprises tracking a gesture of the user to determine a location of the gesture.

**58.** The method of claim **45**, wherein the content comprises a boundary.

**59.** The method of claim **45**, wherein the action comprises a visual search process associated with the portion of interest.

**60.** The method of claim **46**, further comprising obtaining, from an image capture device, at least one image of the scene.

**61.** The method of claim **60**, further comprising processing the at least one image of the scene to generate data associated with the portion of interest.

**62.** The method of claim **61**, further comprising transmitting the data to a remote device configured to generate visual search results.

- 63.** The method of claim **62**, further comprising:
- receiving the visual search results from the remote device;
  - and
  - displaying the visual search results on the display.

**64.** The method of claim **60**, wherein the device is a head mounted device, and wherein the at least one image of the scene corresponds to a field of view observed by the user through a head mounted device.

\* \* \* \* \*