



- (51) International Patent Classification: G06F 3/01 (2006.01)
- (21) International Application Number: PCT/IL2014/050660
- (22) International Filing Date: 21 July 2014 (21.07.2014)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:

61/856,724	21 July 2013 (21.07.2013)	US
61/896,692	29 October 2013 (29.10.2013)	US
- (71) Applicant: POINTGRAB LTD. [IL/IL]; 4 HaCharash Street, 4524075 Hod Hasharon (IL).
- (72) Inventors: EILAT, Eran; 16 Gansin St., 5341916 Givatayim (IL). GAD, Assaf; 1 Amir Drori St., 5845650

Holon (IL). PERSKI, Haim; 27 Mazal Arieh Street, 4536056 Hod-HaSharon (IL).

- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR TOUCHLESS ACTIVATION OF A DEVICE

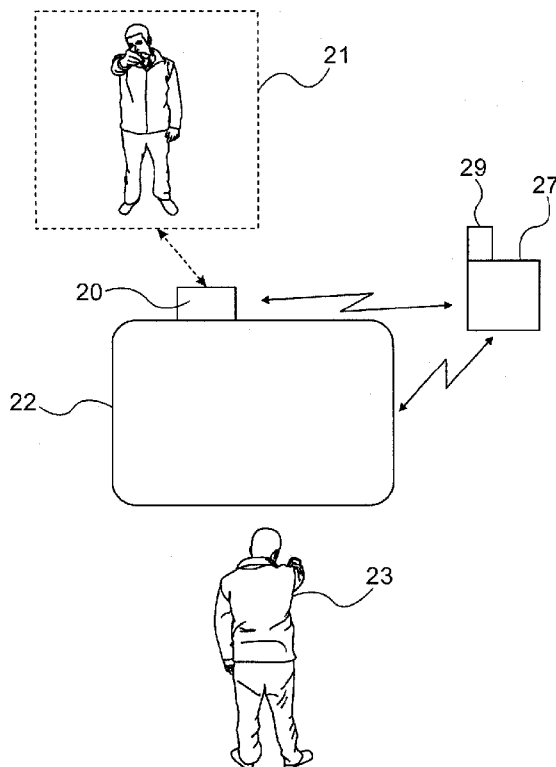


FIGURE 2A

(57) Abstract: A method and system are provided for computer vision based control of a device by obtaining an image via a camera, the camera in communication with a device; detecting in the image a user pointing at the camera; and controlling the device based on the detection of the user pointing at the camera.



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))*

METHOD AND SYSTEM FOR TOUCHLESS ACTIVATION OF A DEVICE

FIELD OF THE INVENTION

[0001] The present invention relates to the field of hand recognition based control of electronic devices. Specifically, the invention relates to touchless activation and other control of a device.

BACKGROUND

[0002] The need for more convenient, intuitive and portable input devices increases, as computers and other electronic devices become more prevalent in our everyday life.

[0003] Recently, human gesturing, such as hand gesturing, has been suggested as a user interface input tool in which a hand gesture is detected by a camera and is translated into a specific command. Gesture recognition enables humans to interface with machines and interact naturally without any mechanical appliances. The development of alternative computer interfaces (forgoing the traditional keyboard and mouse), video games and remote controlling are only some of the fields that may implement human gesturing techniques.

[0004] Recognition of a hand gesture may require identification of an object as a hand and tracking the identified hand to detect a posture or gesture that is being performed.

[0005] Currently, personal computer devices and other mobile devices may include software or dedicated hardware to enable hand gesture control of the device however due to the significant resources needed for hand gesture control, this control mode is typically not part of the basic device operation but must be specifically triggered. A device must typically be already operating in some basic mode in order to enter hand gesture control mode.

[0006] Typically, a device being controlled by gestures includes a user interface, such as a display, allowing the user to interact with the device through the interface and to get feedback regarding his operations. However, only a limited number of devices and home appliances include displays or other user interfaces that allow a user to interact with them.

[0007] Additionally, in a home environment there is usually more than one device. Currently, there is no accurate method for selectively activating a device without interacting with a display of that device.

[0008] Thus, touchless control of devices in a typical home setting is still limited.

[0009] Activation of devices using human voice recognition is also known. Voice recognition capabilities can be found in computer operating systems, commercial software for computers, mobile phones, cars, call centers, internet search engines, home appliances and more.

[0010] Some systems offer gesture recognition and voice recognition capabilities, enabling a user to control devices either by voice or by gestures. Both modalities (voice control and gesture control) are enabled simultaneously and a user signals his desire to use one of the modalities by means of an initializing signal. For example, the SamsungTM Smart TVTM product enables voice control options once a specific phrase is said out loud by the user. Gesture control options are enabled once a user raises his hand in front of a camera attached to the TV. In cases where the Smart TVTM microphone does not pick up the user's voice as a signal, the user may talk into a microphone on a remote control device, to reinforce the initiation voice signal.

[0011] The difficulties in picking up a voice signal, on one hand, and the risk of causing unintended activation (e.g., due to users talking in the background), on the other hand, leave voice controlled systems much to be desired.

SUMMARY

[0012] Embodiments of the present invention provide methods and systems for touchless activation and/or other control of a device.

[0013] Activation and/or other control of a device, according to embodiments of the invention, include the user indicating a device (e.g., if there are several devices, indicating which of the several devices) and a system being able to detect which device the user is indicating and is able to control the device accordingly. Detecting which device is being indicated according to embodiments of the invention, and activating the device based on this identification enables activating and otherwise controlling the device without requiring interaction with a user interface.

[0014] For example, methods and systems according to embodiments of the invention provide accurate and simple activation or enablement of a voice control mode. A user may utilize a gesture or posture of his hand to enable voice control of a device, thereby eliminating the risk of unintentionally activating voice control through unintended

talking and eliminating the need to speak up loudly or talk into a special microphone in order to enable voice control in a device.

[0015] According to one embodiment a V-like shaped posture is used to control voice control of a device. This easy and intuitive control of a device is enabled, according to one embodiment, based on detection of a shape of a user's hand.

BRIEF DESCRIPTION OF THE FIGURES

[0016] The invention will now be described in relation to certain examples and embodiments with reference to the following illustrative figures so that it may be more fully understood. In the drawings:

[0017] Fig. 1 is a schematic illustration of a system according to embodiments of the invention;

[0018] Fig 2A is a schematic illustration of a system to identify a pointing user, according to embodiments of the invention;

[0019] Fig. 2B is a schematic illustration of a system controlled by identification of a pointing user, according to embodiments of the invention;

[0020] Fig. 2C is a schematic illustration of a system for control of voice control of a device, according to one embodiment of the invention;

[0021] Fig. 3 is a schematic illustration of a method for detecting a pointing user, according to embodiments of the invention;

[0022] Fig. 4 is a schematic illustration of a method for detecting a pointing user by detecting a combined shape, according to embodiments of the invention;

[0023] Fig. 5 is a schematic illustration of a method for detecting a pointing user by detecting an occluded face, according to embodiments of the invention;

[0024] Fig. 6 is a schematic illustration of a system for controlling a device in a multi-device environment, according to an embodiment of the invention;

[0025] Fig. 7 is a schematic illustration of a method for controlling a device based on location of a hand in an image compared to a reference point in a reference image, according to an embodiment of the invention;

[0026] Fig. 8 is a schematic illustration of a method for controlling a voice controlled mode of a device, according to embodiments of the invention; and

[0027] Fig. 9 schematically illustrates a method for toggling between voice control enable and disable, according to embodiments of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0028] Methods according to embodiments of the invention may be implemented in a system which includes a device to be operated by a user and an image sensor which is in communication with a processor. The image sensor obtains image data (typically of the user) and sends it to the processor to perform image analysis and to generate user commands to the device based on the image analysis, thereby controlling the device based on computer vision.

[0029] In the following description, various aspects of the present invention will be described. For purposes of explanation, specific configurations and details are set forth in order to provide a thorough understanding of the present invention. However, it will also be apparent to one skilled in the art that the present invention may be practiced without the specific details presented herein. Furthermore, well known features may be omitted or simplified in order not to obscure the present invention.

[0030] Unless specifically stated otherwise, as apparent from the following discussions, it is appreciated that throughout the specification discussions utilizing terms such as "processing," "computing," "calculating," "determining," or the like, refer to the action and/or processes of a computer or computing system, or similar electronic computing device, that manipulates and/or transforms data represented as physical, such as electronic, quantities within the computing system's registers and/or memories into other data similarly represented as physical quantities within the computing system's memories, registers or other such information storage, transmission or display devices.

[0031] An exemplary system, according to one embodiment of the invention, is schematically described in **Fig. 1**, however, other systems may carry out embodiments of the present invention.

[0032] The system **100** may include an image sensor **103**, typically associated with a processor **102**, memory **12**, and a device **101**. The image sensor **103** sends the processor **102** image data of field of view (FOV) **104** to be analyzed by processor **102**.

Typically, image signal processing algorithms and/or image acquisition algorithms may be run in processor **102**. According to one embodiment a user command is generated by processor **102** or by another processor, based on the image analysis, and is sent to the device **101**. According to some embodiments the image processing is performed by a first processor which then sends a signal to a second processor in which a user command is generated based on the signal from the first processor.

[0033] Processor **102** may include, for example, one or more processors and may be a central processing unit (CPU), a digital signal processor (DSP), a microprocessor, a controller, a chip, a microchip, an integrated circuit (IC), or any other suitable multi-purpose or specific processor or controller.

[0034] Memory unit(s) **12** may include, for example, a random access memory (RAM), a dynamic RAM (DRAM), a flash memory, a volatile memory, a non-volatile memory, a cache memory, a buffer, a short term memory unit, a long term memory unit, or other suitable memory units or storage units.

[0035] The device **101** may be any electronic device or home appliance that can accept user commands, e.g., TV, DVD player, PC, mobile phone, camera, set top box (STB) or streamer, smart home console or specific home appliances such as an air conditioner, etc. According to one embodiment, device **101** is an electronic device available with an integrated standard 2D camera. The device **101** may include a display or a display may be separate from but in communication with the device **101**.

[0036] The processor **102** may be integral to the image sensor **103** or may be a separate unit. Alternatively, the processor **102** may be integrated within the device **101**. According to other embodiments a first processor may be integrated within the image sensor and a second processor may be integrated within the device.

[0037] The communication between the image sensor **103** and processor **102** and/or between the processor **102** and the device **101** may be through a wired or wireless link, such as through infrared (IR) communication, radio transmission, Bluetooth technology and other suitable communication routes.

[0038] According to one embodiment the image sensor **103** may include a CCD or CMOS or other appropriate chip. The image sensor **103** may be included in a camera such as a forward facing camera, typically, a standard 2D camera such as a webcam or other standard video capture device, typically installed on PCs or other electronic devices. A 3D camera or stereoscopic camera may also be used according to embodiments of the invention.

[0039] The image sensor **103** may obtain frames at varying frame rates. In one embodiment of the invention, image sensor **103** receives image frames at a first frame rate; and when a predetermined shape of an object (e.g., a shape of a user pointing at the image sensor) is detected (e.g., by applying a shape detection algorithm on an image frame(s) received at the first frame rate to detect the predetermined shape of the object, by processor **102**) the frame rate is changed and the image sensor **103** receives image frames at a second frame rate. Typically, the second frame rate is larger than the first frame rate. For example, the first frame rate may be 1 fps (frames per second) and the second frame rate may be 30 fps. The device **101** can then be controlled based on the predetermined shape of the object and/or based on additional shapes detected in images obtained in the second frame rate.

[0040] Detection of the predetermined shape of the object (typically detected in the first frame rate), e.g., a predetermined shape of a user (such as a user using his hand in a specific posture) can generate a command to turn the device **101** on or off. Images obtained in the second frame rate can then be used for tracking the object and for further controlling the device, e.g., based on identification of postures and/or gestures performed by at least part of a user's hand.

[0041] According to one embodiment a first processor, such as a low power image signal processor may be used to identify the predetermined shape of the user whereas a second, possibly higher power processor may be used to track the user's hand and identify further postures and/or shapes of the user's hand or other body parts.

[0042] Gestures or postures performed by a user's hand may be detected by applying shape detection algorithms on the images received at the second frame rate. At least part of a user's hand may be detected in the image frames received at the second frame rate and the device may be controlled based on the shape of the part of the user's hand.

[0043] According to some embodiments different postures are used for turning a device on/off and for further controlling the device. Thus, the shape detected in the image frames received at the first frame rate may be different than the shape detected in the image frames received at the second frame rate.

[0044] According to some embodiments the change from a first frame rate to a second frame rate is to increase the frame rate such that the second frame rate is larger than the first frame rate. Receiving image frames at a larger frame rate can serve to increase speed of reaction of the system in the further control of the device.

[0045] According to some embodiments image data may be stored in processor **102**, for example in a cache memory. Processor **102** can apply image analysis algorithms, such as motion detection and shape recognition algorithms to identify and further track the user's hand. Processor **102** may perform methods according to embodiments discussed herein by for example executing software or instructions stored in memory **12**.

[0046] According to embodiments of the invention shape recognition algorithms may include, for example, an algorithm which calculates Haar-like features in a Viola-Jones object detection framework. Once a shape of a hand is detected the hand shape may be tracked through a series of images using known methods for tracking selected features, such as optical flow techniques. A hand shape may be searched in every image or at a different frequency (e.g., once every 5 images, once every 20 images or other appropriate frequencies) to update the location of the hand to avoid drifting of the tracking of the hand.

[0047] When discussed herein, a processor such as processor **102** which may carry out all or part of a method as discussed herein, may be configured to carry out the method by, for example, being associated with or connected to a memory such as memory **12** storing code or software which, when executed by the processor, carry out the method.

[0048] Optionally, the system **100** may include an electronic display **11**. According to embodiments of the invention, mouse emulation and/or control of a cursor on a display, are based on computer visual identification and tracking of a user's hand, for example, as detailed above.

[0049] Different embodiments are disclosed herein. Features of certain embodiments may be combined with features of other embodiments; thus certain embodiments may be combinations of features of multiple embodiments.

[0050] Embodiments of the invention may include an article such as a computer or processor readable non-transitory storage medium, such as for example a memory, a disk drive, or a USB flash memory encoding, including or storing instructions, e.g., computer-executable instructions, which when executed by a processor or controller, cause the processor or controller to carry out methods disclosed herein.

[0051] Methods according to embodiments of the invention include obtaining an image via a camera, said camera being in communication with a device, and detecting in the image a predetermined shape of an object, e.g., a user pointing at the camera. The device may then be controlled based on the detection of the user pointing at the camera. For

example, as schematically illustrated in **Fig. 2A**, camera **20** which is in communication with device **22** and processor **27** (which may perform methods according to embodiments of the invention by, for example, executing software or instructions stored in memory **29**), obtains an image **21** of a user **23** pointing at the camera **20**. Once a user pointing at the camera is detected, e.g., by processor **27**, a command may be generated to control the device **22**. According to one embodiment the command to control the device **22** is an ON/OFF command. According to another embodiment detection, by a first processor, of the user pointing at the camera may cause a command to be generated to start using a second processor to further detect user gestures and postures and/or to change frame rate of the camera **20** and/or a command to control the device **22** ON/OFF and/or other commands.

[0052] In one embodiment a face recognition algorithm may be applied (e.g., in processor **27** or another processor) to identify the user and generating a command to control the device **22** (e.g., in processor **27** or another processor) may be enabled or not based on the identification of the user.

[0053] In some embodiments the system may include a feedback system which may include a light source, buzzer or sound emitting component or other component to provide an alert to the user of the detection of the user's identity or of the detection of a user pointing at the camera.

[0054] Communication between the camera **20** and the device **22** may be through a wired or wireless link including processor **27** and memory **29**, such as described above.

[0055] According to one embodiment, schematically illustrated in **Fig. 2B**, a system **200** includes camera **203**, typically associated with a processor **202**, memory **222**, and a device **201**.

[0056] According to one embodiment the camera **203** is attached to or integrated in device **201** such that when a user (not shown) indicates at the device **201**, he is essentially indicating at the camera **203**. According to one embodiment the user may indicate at a point relative to the camera. The point relative to the camera may be a point at a predetermined location relative to the camera.

[0057] For example, locations above or below the camera or to the right/left of the camera may be designated for specific controls of an appliance. For example, the device **201**, which may be an electronic device or home appliance that can accept user commands, e.g., TV, DVD player, PC, mobile phone, camera, set top box (STB) or

streamer, smart home console or specific home appliances such as an illumination fixture, an air conditioner, etc., may include a panel **204**, which may include marks **205a** and/or **205b**, which, when placed on the device **201**, are located at predetermined locations relative to the camera **203** (for example, above and below camera **203**).

[0058] According to one embodiment the panel **204** may include a camera view opening **206** which may accommodate the camera **203** or at least the optics of the camera **203**. The camera view opening **206** may include lenses or other optical elements.

[0059] In some embodiments mark **205a** and/or **205b** may be at a predetermined location relative to the camera view opening **206**. If the user is indicating at the mark **205a** or **205b** then the processor **202** may control output of the device **201**. For example, a user may turn on a light source by indicating at camera view opening **206** and then by indicating at mark **205a** the user may make the light brighter and by indicating at mark **205b** the user may dim the light.

[0060] According to one embodiment the panel **204** may include an indicator **207** configured to create an indicator FOV **207'** which correlates with the camera FOV **203'** for providing indication to the user that he is within the camera FOV.

[0061] According to one embodiment the processor **202** may cause a display of control buttons or another display, to be displayed to the user, typically in response to detection of the user indicating at the camera. The control buttons may be arranged in predetermined locations in relation to the camera **203**. For example, the processor **202** may cause marks **205a** and **205b** to be displayed on the panel **204**, for example, based on detection of a user indicating at the camera **203** or based on detection of a predetermined posture or gesture of the user or based on another signal.

[0062] Thus, an image of a user indicating at a camera may be used as a reference image. The location of the user's hand (or part of the hand) in the reference image may be compared to the location of the user's indicating hand (or part of the hand) in a second image and the comparison may enable to calculate the point being indicated at in the second image. For example, when a user activates a light source by indicating at a camera (e.g., at camera view opening **206**), the image of the user indicating at the camera can be used as a reference image. In a next, second, image the user may indicate at mark **205a** which is, for example, located above the camera view opening **206**. The location of the user's hand in the second image can be compared to the location of the user's hand in the reference image and based on this comparison it can be deduced that the user is indicating at a higher point in the second image than in the reference image.

This deduction can then result, for example, in a command to brighten the light, whereas, if the user were indicating a point below the camera view opening **206** (e.g., mark **205b**) then the light would be dimmed.

[0063] A method according to one embodiment, may include determining the location of a point being indicated at by a user in a first image and if the location of the point is determined to be at the location of the camera then controlling the device may include generating an ON/OFF command and/or another command, such as displaying to the user a set of control buttons or other marks arranged in predetermined locations in relation to the camera. Once it is determined that the user is indicating at the camera, the location of the hand in a second image can be determined and it may be determined if the location of the hand in the second image shows that the user is indicating at a predetermined location relative to the camera. For example, determining if the user is indicating at a predetermined location relative to the camera can be done by comparing the location of the hand in the first image to the location of the hand in the second image. If it is determined that the user is indicating at a predetermined location relative to the camera then an output of the device may be controlled, typically, based on the predetermined location

[0064] If the location of the point being indicated at in the first image is not the location of the camera it is determined if the location is a predetermined location relative to the camera. If the location is a predetermined location relative to the camera then an output of the device may be controlled.

[0065] Controlling an output of a device may include modulating the level of the output (e.g., raising or lowering volume of audio output, rewinding or running forward video or audio output, raising or lowering temperature of a heating/cooling device, etc.). Controlling the output of the device may also include controlling a direction of the output (e.g., directing air from an air-conditioner in the direction of the user, directing volume of a TV in the direction of a user, etc.). Other output parameters may be controlled.

[0066] An exemplary system, according to another embodiment of the invention, is schematically described in **Fig. 2C** however other systems may carry out embodiments of the present invention.

[0067] The system **2200** may include an image sensor **2203**, typically associated with a processor **2202**, memory **12**, and a device **2201**. The image sensor **2203** sends the processor **2202** image data of field of view (FOV) **2204** (the FOV including at least a

user's hand or at least a user's fingers **2205**) to be analyzed by processor **2202**. Typically, image signal processing algorithms and/or shape detection or recognition algorithms may be run in processor **2202**.

[0068] The system may also include a voice processor **22022** for running voice recognition algorithms or voice recognition software, typically to control device **2201**. Voice recognition algorithms may include voice activity detection or speech detection or other known techniques used to facilitate speech and voice processing.

[0069] Processor **2202**, which may be an image processor for detecting a shape (e.g., a shape of a user's hand) from an image may communicate with the voice processor **22022** to control voice control of the device **2201** based on the detected shape.

[0070] Processor **2202** and processor **22022** (which may be units of a single processor or may be separate processors) may be part of a central processing unit (CPU), a digital signal processor (DSP), a microprocessor, a controller, a chip, a microchip, an integrated circuit (IC), or any other suitable multi-purpose or specific processor or controller.

[0071] Memory unit(s) **12** may include, for example, a random access memory (RAM), a dynamic RAM (DRAM), a flash memory, a volatile memory, a non-volatile memory, a cache memory, a buffer, a short term memory unit, a long term memory unit, or other suitable memory units or storage units.

[0072] According to one embodiment a command to enable voice control of device **2201** is generated by processor **2202** or by another processor, based on the image analysis. According to some embodiments the image processing is performed by a first processor which then sends a signal to a second processor in which a command is generated based on the signal from the first processor.

[0073] Processor **2202** may run shape recognition algorithms, for example, an algorithm which calculates Haar-like features in a Viola–Jones object detection framework, to detect a hand shape which includes, for example, a V-like component (such as the "component" created by fingers **2205**) or other shapes (such as the shape of the user's face and finger in a "mute" or "silence" posture **2205'**) and to communicate with processor **22022** to activate, disable or otherwise control voice control of the device **2201** based on the detection of the V-like component and/or based on other shapes detected.

[0074] The system may also include an adjustable voice recognition component **2206**, such as an array of microphones or a sound system. According to one embodiment the image processor (e.g., processor **2202**) may generate a command to adjust the voice recognition component **2206** based on the detected shape of the user's hand or based on the detection a V-like shape. For example, a microphone may be rotated or otherwise moved to be directed at a user, once a V-like shape is detected or sound received by an array of microphones may be filtered according to the location/direction of the V-like shape with respect to the array of microphones, or the sensitivity of a sound system may be adjusted or other adjustments may be made to better enable receiving and enhancing voice signals.

[0075] In another embodiment a face recognition algorithm may be applied (e.g., in processor **2202** or another processor) to identify or classify the user according to gender/age/ethnicity, etc. and voice detection and recognition algorithms (e.g., in processor **22022** or another processor) may be more efficiently run based on the classification of the user.

[0076] In some embodiments the system includes a feedback unit **2223** which may include a light source, buzzer or sound emitting component or other component to provide an alert to the user of the detection of the user's fingers in a V-like shape (or other shapes). According to one embodiment the alert is a sound alert, which may be desired in a situation where the user cannot look at the system (e.g., while driving) to get confirmation that voice control is now enabled/disabled, etc.

[0077] The device **2201** may be any electronic device or home appliance or appliance in a vehicle that can accept user commands, e.g., TV, DVD player, PC, mobile phone, camera, set top box (STB) or streamer, etc. According to one embodiment, device **2201** is an electronic device available with an integrated 2D camera. The device **2201** may include a display **22211** or a display may be separate from but in communication with the device **2201**.

[0078] The processors **2202** and **22022** may be integral to the image sensor **2203** or may be in separate units. Alternatively, the processors may be integrated within the device **2201**. According to other embodiments a first processor may be integrated within the image sensor and a second processor may be integrated within the device.

[0079] The communication between the image sensor **2203** (or other sensors) and processors **2202** and **22022** (or other processors) and/or between the processors **2202** and **22022** and the device **2201** (or other devices) may be through a wired or wireless

link, such as through infrared (IR) communication, radio transmission, Bluetooth technology and other suitable communication routes.

[0080] According to one embodiment the image sensor **2203** may be a 2D camera including a CCD or CMOS or other appropriate chip. A 3D camera or stereoscopic camera may also be used according to embodiments of the invention.

[0081] According to some embodiments image data may be stored in processor **2202**, for example in a cache memory. Processor **2202** can apply image analysis algorithms, such as motion detection and shape recognition algorithms to identify a user's hand and/or to detect specific shapes of the user's hand and/or shapes of a hand in combination with a user's face or other shapes. Processor **2202** may perform methods according to embodiments discussed herein by for example executing software or instructions stored in memory **12**.

[0082] When discussed herein, a processor such as processors **2202** and **22022** which may carry out all or part of a method as discussed herein, may be configured to carry out the method by, for example, being associated with or connected to a memory such as memory **12** storing code or software which, when executed by the processor, carry out the method.

[0083] According to one embodiment which is schematically illustrated in **Fig. 3**, the method includes obtaining an image via a camera (**310**), said camera being in communication with a device. In the image a shape of a user pointing at the camera (or at a different location related to a device) is detected (**320**) and based on the detection of the shape of the user pointing at the camera (or other location), generating a command to control the device (**330**). According to one embodiment a detector trained to recognize a shape of a pointing person is used to detect the shape of the user pointing at the camera or at a different location related to a device. Shape detection algorithms, such as described above, may be used.

[0084] A shape of a user pointing at the camera can be detected in a single image, unlike detecting gestures which involve motion, which cannot be detected from a single image but requires checking at least two images.

[0085] According to one embodiment the camera is a 2D camera and the detector's training input includes 2D images.

[0086] When pointing at a camera, the user is typically looking at the camera and is holding his pointing finger in the line of sight between his eyes and the camera. Thus, a "shape of a pointing user", according to one embodiment, will typically include at least

part of the user's face. According to some embodiments a "shape of a pointing user" includes a combined shape of the user's face and the user's hand in a pointing posture (for example **21** in **Fig. 2A**).

[0087] Thus, a method for computer vision based control of a device according to one embodiment, which is schematically illustrated in **Fig. 4**, includes the steps of obtaining an image of a field of view, the field of view including a user (**410**) and detecting a combined shape of the user's face (or part of the user's face) and the user's hand in a pointing posture (**420**). A device may then be controlled based on the detection of the combined shape (**430**).

[0088] According to another embodiment the device may be controlled based on detecting a combined shape of the user's face and the user's hand, the user's hand being held away from the user's face. Thus, a user does not necessarily have to point in order to indicate a desired device. The user may be looking at a desired device (or at the camera attached to the device) and may raise his arm in the direction he is looking at, thus indicating that device.

[0089] For example, detection of a combined shape of the user's face (or part of the user's face) and the user's hand held at a distance from the face (but in the line of sight between his eyes and the camera), for example, in a pointing posture, may generate a command to change a first (slow) frame rate of the camera obtaining images of the user to a second (quicker) frame rate. In addition, or alternatively, the detection of the combined shape may generate a command to turn a device ON/OFF or any other command, for example as described above.

[0090] According to one embodiment one or more detectors may be used to detect a combined shape. For example, one detector may identify a partially obscured face whereas another detector may identify a hand or part of a hand on a background of a face. One or both detectors may be used in identifying a user pointing at a camera.

[0091] A face or facial landmarks may be continuously or periodically searched for in the images and may be detected, for example, using known face detection algorithms (e.g., using Intel's OpenCV). According to some embodiments a shape can be detected or identified in an image, as the combined shape, only if a face was detected in that image. In some embodiments the search for facial landmarks and/or for the combined shape may be limited to a certain area in the image (thereby reducing computing power) based for example, on size (limiting the size of the searched area

based on an estimated or average face size), on location (e.g., based on the expected location of the face) and/or on other suitable parameters.

[0092] According to another embodiment detection of a user pointing at the camera or at a different location related to a device may be done by identifying a partially occluded face. For example, as schematically illustrated in **Fig. 5**, a method according to one embodiment of the invention may include the steps of obtaining an image via a camera (**502**); detecting in the image a user's face partially occluded around an area of the user's eyes (**504**); and controlling the device based on the detection of the partially occluded user's face (**506**).

[0093] The area of the eyes may be detected within a face by detecting a face (e.g., as described above) and then detecting an area of the eyes within the face. According to some embodiments an eye detector may be used to detect at least one of the user's eyes. Eye detection using OpenCV's boosted cascade of Haar-like features may be applied. Other methods may be used. The method may further include tracking at least one of the user's eyes (e.g., by using known eye trackers).

[0094] According to one embodiment the user's dominant eye is detected, or the location in the image of the dominant eye is detected, and is used to detect a pointing user. Eye dominance (also known as ocular dominance) is the tendency to prefer visual input from one eye to the other. In normal human vision there is an effect of parallax, and therefore the dominant eye is the one that is primarily relied on for precise positional information. Thus, detecting the user's dominant eye and using the dominant eye as a reference point for detecting a pointing user, may assist in more accurate control of a device.

[0095] According to one embodiment the method includes detecting a shape of a partially occluded user's face. According to one embodiment the face is partially occluded by a hand or part of a hand.

[0096] The partially occluded face may be detected in a single image by using one or more detectors, for example, as described above.

[0097] According to one embodiment, for example in a multi-device environment, the system identifies an "indication posture" and can thus determine which device (of several devices) is being indicated by the user. The "indication posture" may be a static posture (such as the user pointing at the device or at the camera associated with the device). According to one embodiment a system includes a camera operating at a low

frame rate and/or having a long exposure time such that motion causes blurriness and is easily detected and discarded, facilitating detection of the static "indication posture".

[0098] For example, as schematically illustrated in **Fig. 6**, a single room **600** may include several home appliances or devices that need to be turned on or off by a user, such as an audio system **61**, an air conditioner **62** and a light fixture **63**. Cameras **614**, **624** and **634** attached at each of these devices may be operating at low energy such as at low frame rate. Each camera may be in communication with a processor (such as processor **102** in **Fig. 1**) to identify a user indicating at it and to turn the device on or off based on the detection of the indication posture. For example, if a user **611** is standing in the room **600** pointing at air conditioner **62**, the image **625** of the user which is obtained by camera **624** which is located at or near the air conditioner will be different than the images **615** and **635** of that same user **611** obtained by the other cameras **614** and **634**. Typically, the image **625** obtained by camera **624** will include a combined shape of a face and hand or a partially occluded face because the user is looking at and pointing at or near the camera **624**, whereas the other images will not include a combined shape of a face and hand or a partially occluded face. Upon detection of a combined shape or partially occluded face (or other sign that the user is pointing at or near the camera), the device (e.g. air conditioner **62**) may be turned on or off or may be otherwise controlled.

[0099] Some known devices can be activated based on detected motion or sound however, this type of activation is not specific and would not enable activating a specific device in a multi-device environment since movement or a sound performed by the user will be received at all the devices indiscriminately and will activate all the device instead of just one. Interacting with a display of a device may enable more specificity however typical home appliances, such as audio system **61**, air conditioner **62** and light fixture **63**, do not include a display. Embodiments of the current invention do not require interacting with a display and enable touchlessly activating a specific device even in a multi-device environment.

[00100] A method according to another embodiment of the invention is schematically illustrated in **Fig. 7**. The method includes using a processor to detect, in an image, a location of a hand (or part of a hand) of a user, the hand indicating at a point relative to the camera used to obtain the image (**702**), comparing the location of the hand in the image to a location of the hand in a reference image (**704**); and controlling the device based on the comparison (**706**).

[00101] According to one embodiment the reference image includes the user indicating at the camera.

[00102] Detecting the user indicating at the camera may be done, for example, by detecting the user's face partially occluded around an area of the user's eyes, as described above.

[00103] Detecting a location of a hand of a user indicating at the camera or at a point relative to the camera may include detecting the location the user's hand relative to the user's face, or part of face, for example relative to an area of the user's eyes.

[00104] According to one embodiment detecting a location of a hand of a user indicating at a camera or at a point relative to the camera involves detecting the shape of the user. The shape detected may be a combined shape of the user's face and the user's hand, the user's hand being held away from the user's face. According to one embodiment detecting the user indicating at the camera and/or at a point relative to the camera is done by detecting a combined shape of the user's face and the user's hand in a pointing posture.

[00105] According to embodiments of the invention detection of a user indicating at the camera or at a point relative to the camera may be done based on detecting a part of a hand and may include detecting specific parts of the hand. For example, detection of an indicating user may involve detection a finger or tip of a finger. A finger may be identified by identifying, for example, the longest line that can be constructed by both connecting two pixels of a contour of a detected hand and crossing a calculated center of mass of the area defined by the contour of the hand. A tip of a finger may be identified as the extreme most point in a contour of a detected hand or the point closest to the camera.

[00106] According to one embodiment a user's hand (e.g., a shape of a hand or part of hand) may be searched for in a location in an image where a face (e.g., a shape of a face) has been previously detected, thereby reducing computing power.

[00107] Detecting the user indicating at the camera may involve detecting a predetermined shape of the user's hand (e.g., the hand in a pointing posture or in another posture).

[00108] According to one embodiment, for example in a multi-device environment, the system identifies an "indication posture" and can thus determine which device (of several devices) is being indicated by the user. The "indication posture" may be a static posture (such as the user pointing at the device or at the camera associated

with the device). According to one embodiment a system includes a camera operating at a low frame rate and/or having a long exposure time such that motion causes blurriness and is easily detected and discarded, facilitating detection of the static "indication posture".

[00109] A method according to another embodiment of the invention may include using a processor to detect a reference point in an image (e.g., a first image), the reference point related to the user's face (for example, an area of the user's eyes) or the reference point being the location of a hand indicating at a camera used to obtain the image; detect in another image (e.g., a second image) a location of a hand of a user; compare the location of the hand in the second image to the location of the reference point; and control the device based on the comparison.

[00110] As described above, when a user indicates at a camera, the user is typically looking at the camera and is holding his arm/ hand in the line of sight between his eyes and the camera. Accordingly, an image of a user indicating at the camera will typically include at least part of the user's face. Thus, comparing the location of a user's hand (or part of hand) in an image to a reference point (which is related to the user's face) in that image enables to deduce the location relative to the camera at which the user is indicating and a device can be controlled based on the comparison, as described above.

[00111] A method for computer vision based control of a device according to another embodiment of the invention is schematically illustrated in **Fig. 8**.

[00112] According to one embodiment the method includes obtaining an image of a field of view, which includes a user's fingers (**802**) and detecting in the image the user's fingers in a V-like shape (**804**). Based on the detection of the V-like shape voice control of a device is controlled (**806**).

[00113] Detecting the user's fingers in a V-like shape may be done by applying a shape detection or shape recognition algorithm to detect the user's fingers (e.g., index and middle finger) in a V-like shape. In some embodiments motion may be detected in a set of images and the shape detection algorithm can be applied based on the detection of motion. In some embodiments the shape detection algorithm may be applied only when motion is detected and/or the shape detection algorithm may be applied at a location in the images where the motion was detected.

[00114] According to one embodiment controlling voice control includes enabling or disabling voice control. Enabling voice control may include running known voice recognition algorithms or applying known voice activity detection or speech detection techniques. The step of controlling voice control may also include a step of adjusting sensitivity of voice recognition components. For example, a voice recognition component may include a microphone or array of microphones or a sound system that can be adjusted for better receiving and enhancing voice signals.

[00115] According to one embodiment the method may include generating an alert to the user based on detection of the user's fingers in a V-like shape. The alert may include a sound component, such as a buzz, click, jingle etc.

[00116] According to one embodiment, which is schematically illustrated in **Fig. 9**, the method includes obtaining an image of a field of view, which includes a user (**902**) and detecting in the image a first V-like shape (**904**). Based on the detection of the first V-like shape voice control of a device is enabled (**906**). The method further includes detecting in the image a second shape (**908**), which may be a second V-like shape or a different shape, typically a shape which includes the user's fingers, and disabling voice control based on the detection of the second shape (**910**).

[00117] In one embodiment the detection of a second V-like shape is confirmed to be the second detection (and cause a change in the status of the voice control (e.g., enabled/disabled)) only if it occurs after (e.g., within a predetermined time period) the detection of the first V-like shape.

[00118] According to one embodiment the method may include generating an alert to the user based on detection of the second shape.

[00119] According to one embodiment the second shape may be a combination of a portion of the user's face and at least a portion of the user's hand, for example, the shape of a finger positioned over or near the user's lips.

[00120] Thus, a user may toggle between voice control and other control modalities by posturing, either by using the same posture or by using different postures.

Claims

1. A method for computer vision based control of a device, the method comprising

obtaining an image via a camera, said camera in communication with a device;

detecting in the image a user pointing at the camera; and

controlling the device based on the detection of the user pointing at the camera.
2. The method of claim 1 wherein controlling the device comprises generating an ON/OFF command.
3. The method of claim 1 wherein detecting a user pointing at the camera comprises detecting a shape of a user pointing at the camera.
4. The method of claim 1 wherein detecting a user pointing at the camera comprises detecting a user's face partially occluded around an area of the user's eyes.
5. The method of claim 1 wherein detecting a user pointing at the camera comprises detecting a combined shape of the user's face and the user's hand, the user's hand being held away from the user's face.
6. The method of claim 1 wherein detecting a user pointing at the camera comprises detecting a combined shape of the user's face and the user's hand in a pointing posture.
7. The method of claim 1 wherein detecting in the image a user pointing at the camera comprises detecting a static posture of the user.
8. A method for computer vision based control of a device, the method comprising

obtaining an image via a camera;

detecting in the image a user's face partially occluded around an area of the user's eyes; and

controlling the device based on the detection of the partially occluded user's face.

9. The method of claim 8 comprising detecting a shape of a partially occluded user's face.
10. The method of claim 8 comprising detecting a partially occluded face in a single image.
11. The method of claim 8 wherein detecting a partially occluded user's face comprises applying a detector that identifies a partially occluded face.
12. The method of claim 11 wherein the detector identifies user's face partially occluded by at least part of a hand.
13. The method of claim 11 wherein detecting a partially occluded user's face comprises applying another detector that identifies at least part of a user's hand with a background comprising part of a face.
14. A method for computer vision based control of a device, the method comprising
obtaining an image of a field of view, the field of view comprising a user;
detecting a combined shape of the user's face and the user's hand in a pointing posture; and
controlling the device based on the detection of the combined shape.
15. A method for computer vision based control of a device, the method comprising
detecting in a first image a location of a hand indicating at a camera used to obtain the image, the location of the hand being a reference point;
comparing the location of the hand in a second image to a location of the reference point in the first image, and
controlling the device based on the comparison.
16. The method of all the preceding claims wherein the device is selected from the group consisting of: a TV, DVD player, PC, mobile phone or tablet, camera, STB (Set Top Box) or streamer, smart home console or specific home appliances.
17. A method for computer vision based control of a device, the method comprising

obtaining an image of a field of view, the field of view comprising a user's fingers;

detecting in the image the user's fingers in a V-like shape; and

controlling voice control of a device based on the detection of the V-like shape.

18. The method of claim 17 comprising applying a shape detection algorithm to detect the V-like shape.
19. The method of claim 18 comprising detecting motion in a set of images and applying the shape detection algorithm based on the detection of motion.
20. The method of claim 17 wherein the user's fingers comprise the index and middle finger.
21. The method of claim 17 wherein controlling voice control comprises enabling or disabling voice control or adjusting sensitivity of voice recognition components.
22. The method of claim 21 comprising detecting in the image a second shape comprising the user's fingers and wherein disabling voice control is based on the detection of the second shape.
23. The method of claim 22 wherein the second shape comprises a portion of the user's face and at least a portion of the user's hand.
24. The method of claim 23 wherein the second shape comprises a finger positioned over or near the user's lips.
25. The method of claim 17 comprising generating an alert to the user based on detection of the user's fingers in a V-like shape.
26. The method of claim 22 comprising generating an alert to the user based on detection of the second shape.
27. The method of claim 26 wherein the alert comprises a sound component.
28. A system for controlling a device, the system comprising
 - a voice processor for running voice recognition algorithms;
 - a device configured to be controlled by the voice recognition algorithms;
 - an image processor for detecting a shape from an image and for communicating with the voice processor to control voice control of the device based on the detected shape.

29. The system of claim 28 wherein the image processor is for detecting a hand shape comprising a V-like component and to control voice control of the device based on the detected of the V-like component.
30. The system of claim 28 comprising a camera in communication with the image processor, the camera to obtain an image of the user's hand.
31. The system of claim 30 wherein the camera is a 2D camera.
32. The system of claim 28 further comprising an adjustable voice recognition component, wherein the image processor is to adjust the voice recognition component based on the detected shape.
33. The system of claim 32 wherein the voice recognition component comprises a microphone.
34. The system of claim 28 further comprising a detector to detect in the image a second shape comprising the user's fingers and wherein the image processor is for disabling voice control based on the detection of the second shape.
35. The system of claim 34 wherein the second shape comprises a portion of the user's face and at least a portion of the user's hand.
36. The system of claim 35 wherein the second shape comprises a finger positioned over or near the user's lips.
37. The system of claim 28 comprising a feedback unit configured to notify the user of the detection of the user's fingers in a V-like shape.
38. The system of claim 37 comprising a feedback unit configured to notify the user of the detection of the second shape.
39. The system of claim 37 wherein the feedback unit is configured to provide a sound alert.
40. The system of claim 38 wherein the feedback unit is configured to provide a sound alert.
41. A system comprising
 - a camera for obtaining an image of a user's fingers;
 - a processor for running voice recognition software;
 - a processor for detecting a V-like shape in the image of the user's fingers and for activating the voice recognition software when the V-like shape is detected.
42. The system of claim 41 comprising a device to be controlled by the voice recognition software.

43. The system of claim 41 wherein the processor for detecting a V-like shape is configured to detect a shape of the user's hand and to control the device based on the detection of the shape of the user's hand.
44. The system of claim 41 further comprising an adjustable voice recognition component, wherein the processor for detecting a V-like shape is to adjust the voice recognition component based on the detection of the V-like shape.
45. The system of claim 41 wherein the processor for detecting the V-like shape is to activate the voice recognition software when a first V-like shape is detected and to disable voice recognition when a second V-like shape is detected.

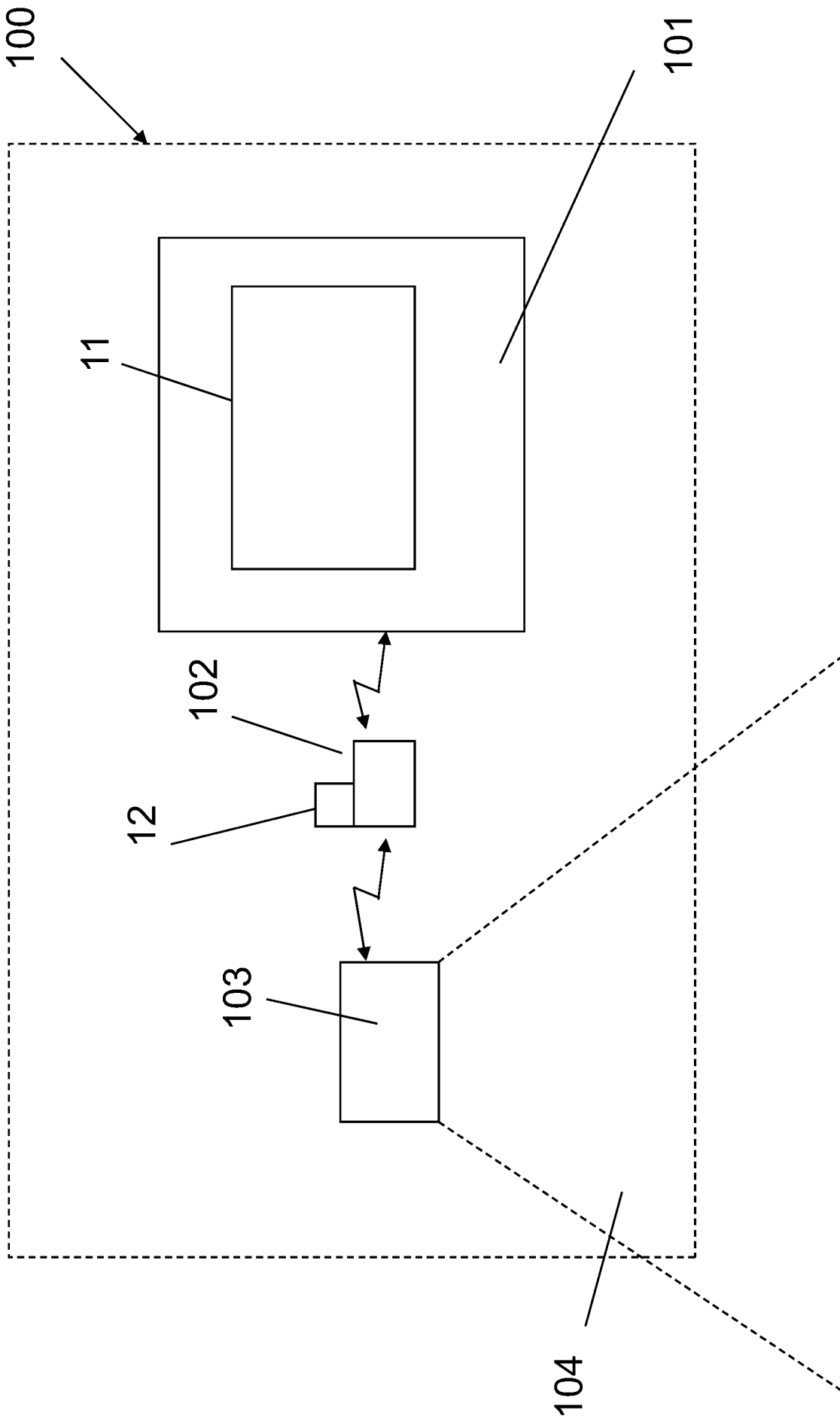


FIGURE 1

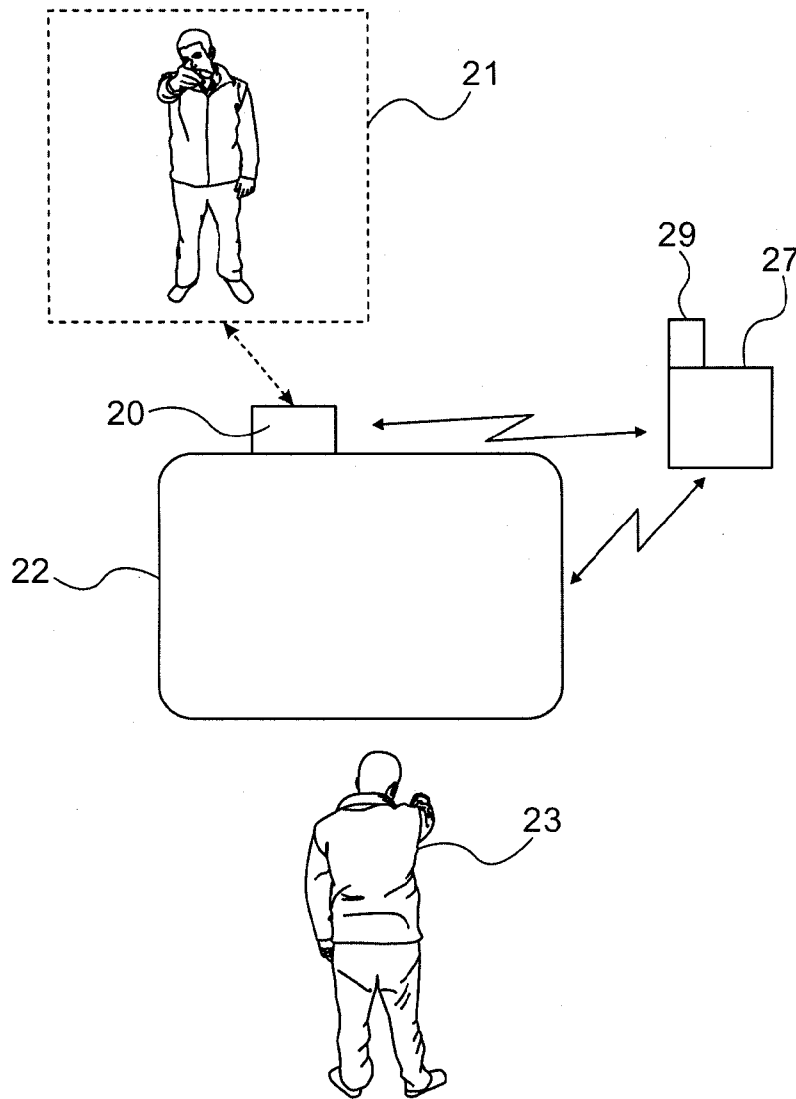


FIGURE 2A

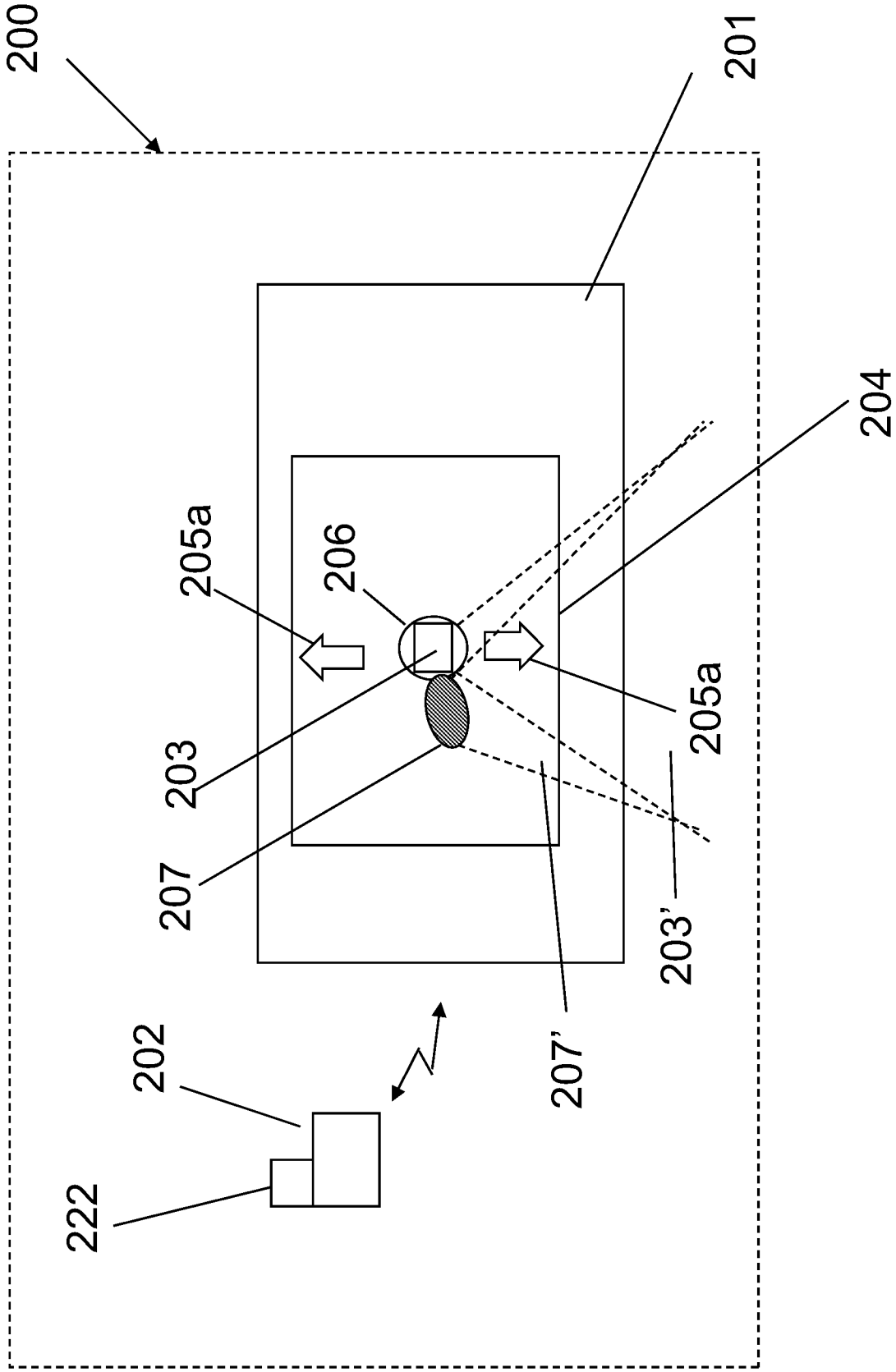


FIGURE 2B

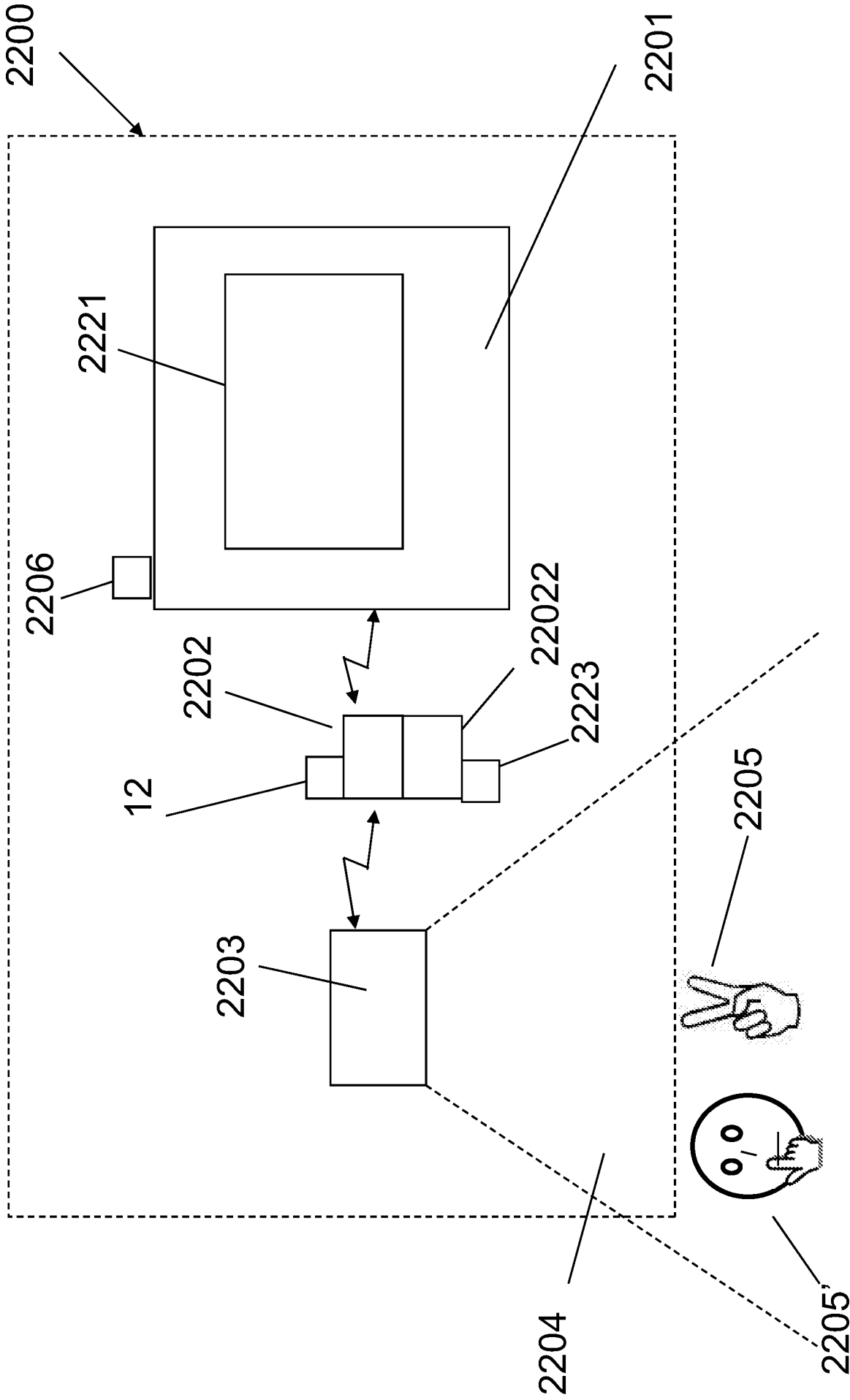


FIGURE 2C

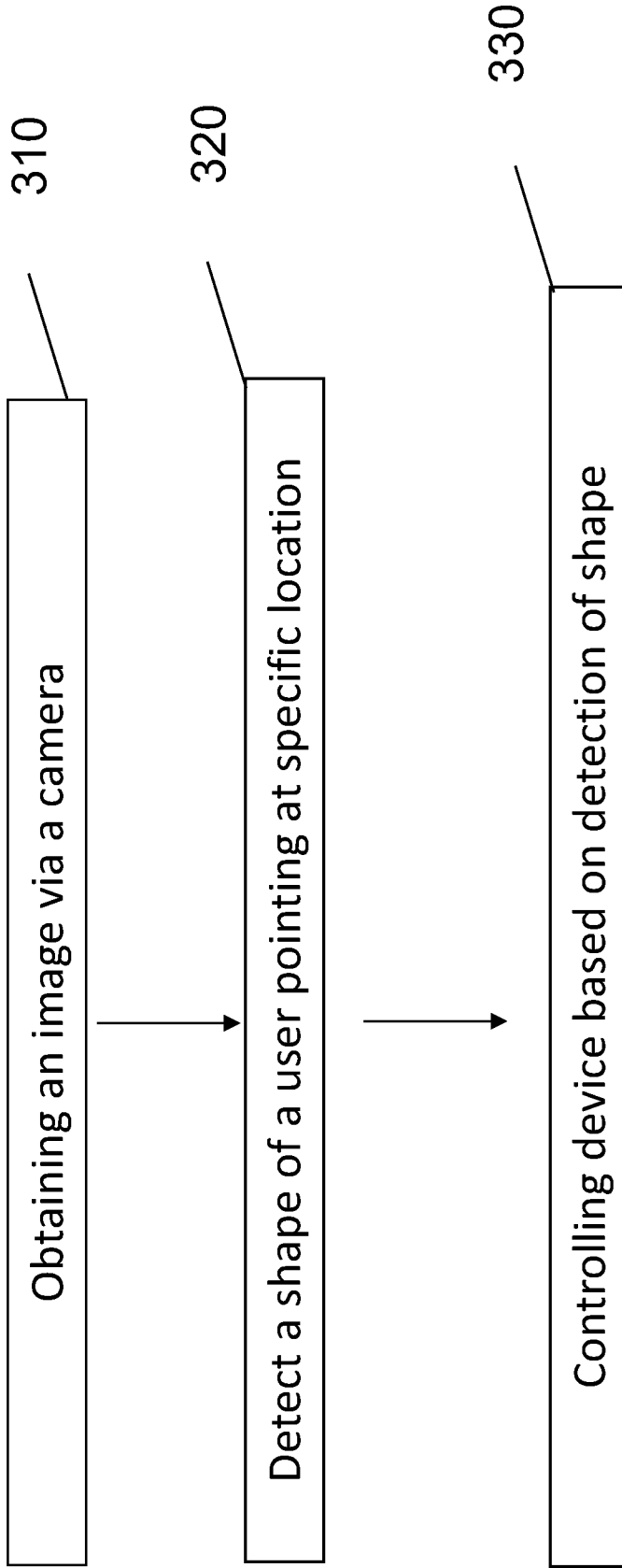


FIGURE 3

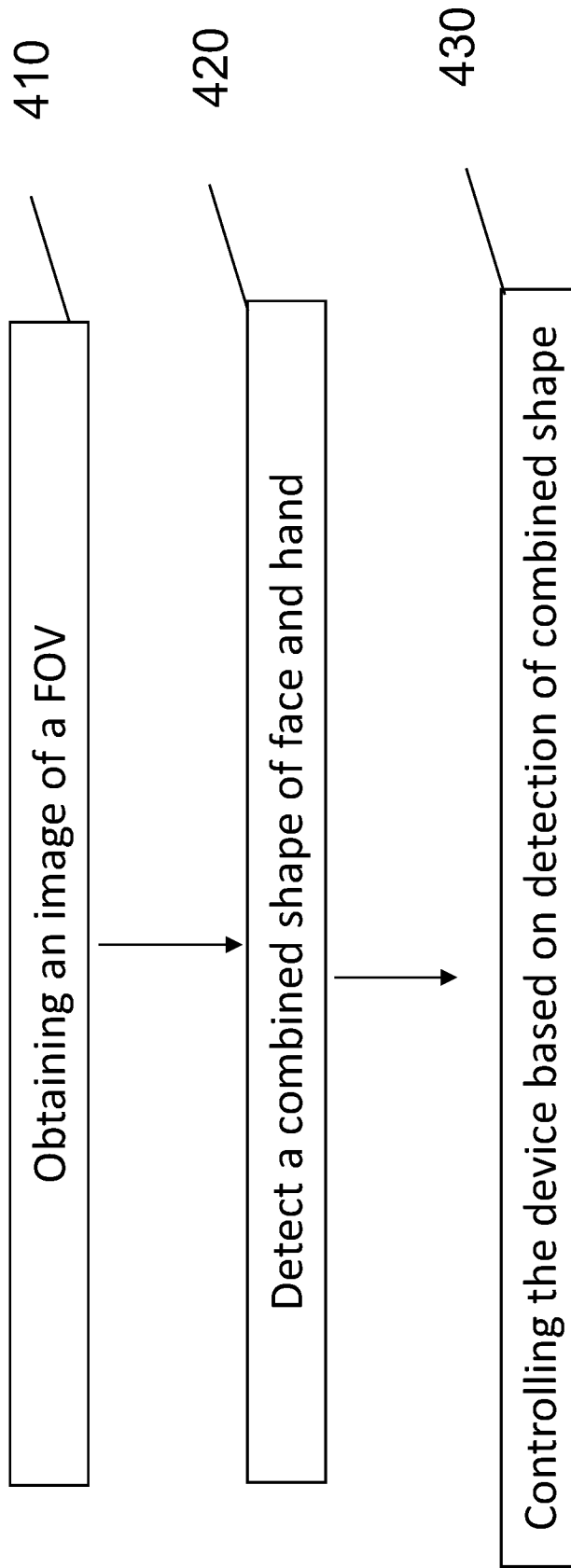


FIGURE 4

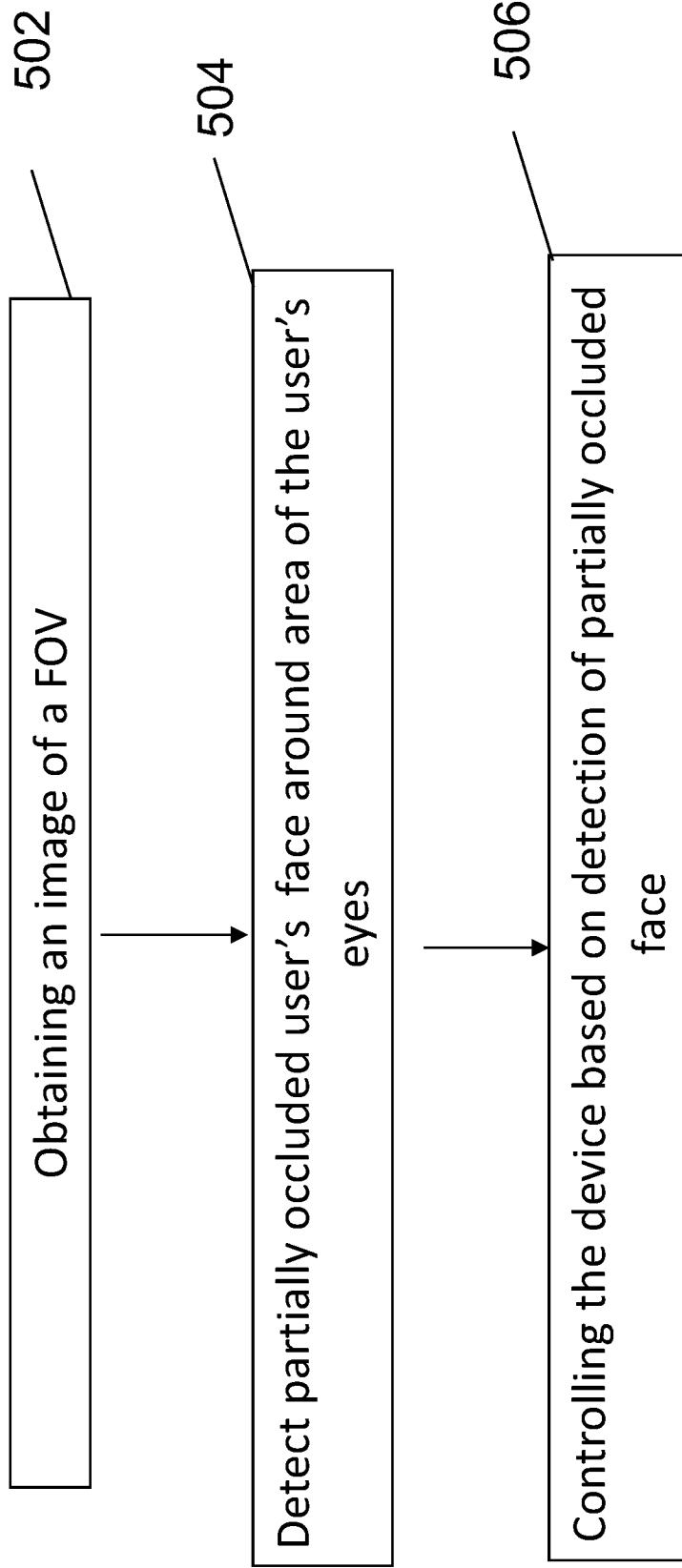


FIGURE 5

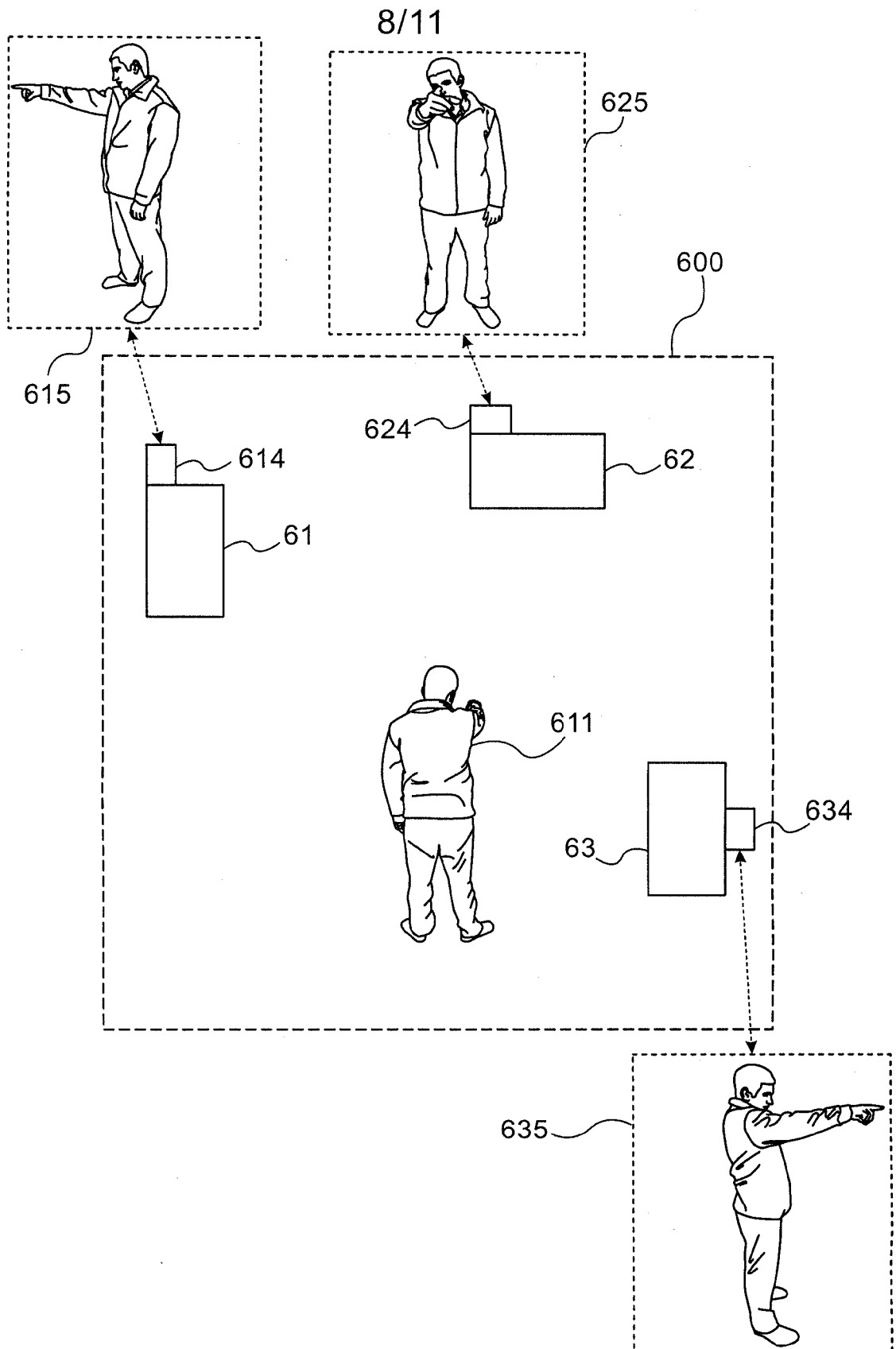


FIGURE 6

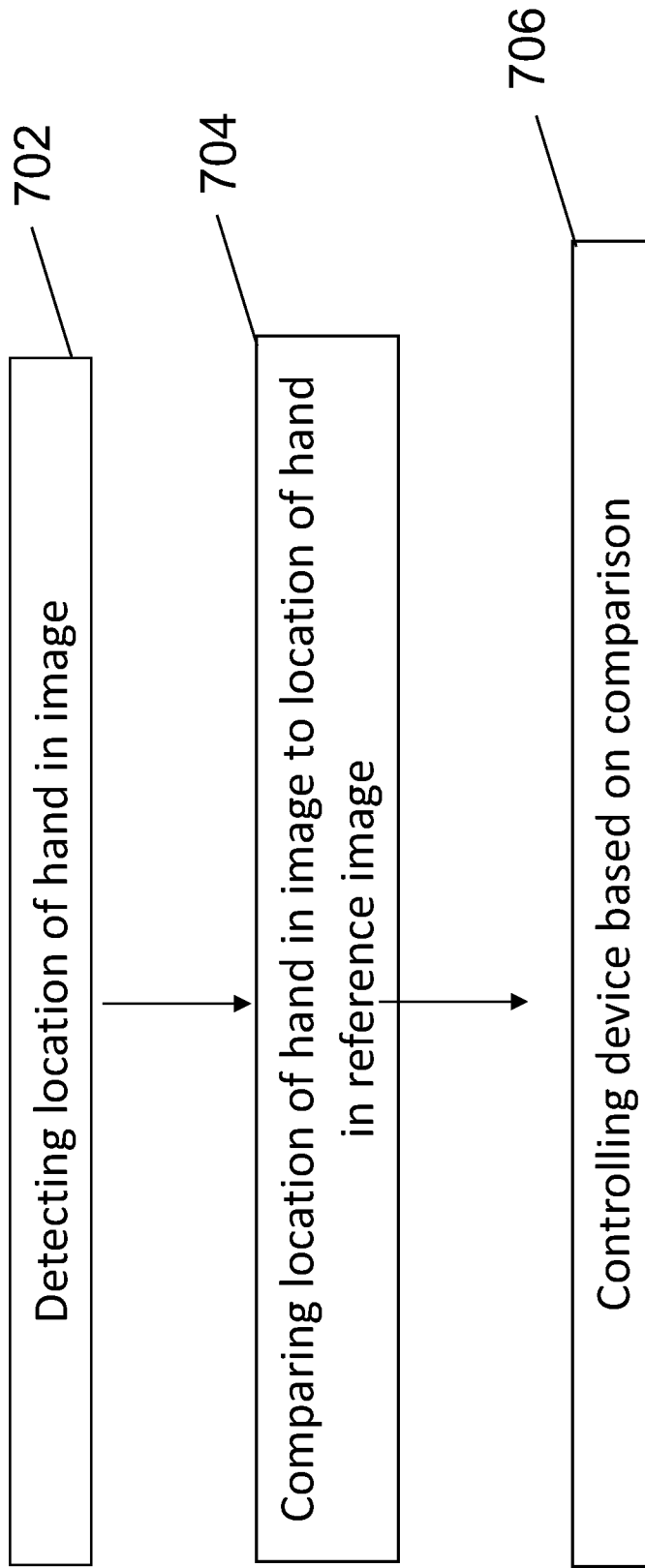


FIGURE 7

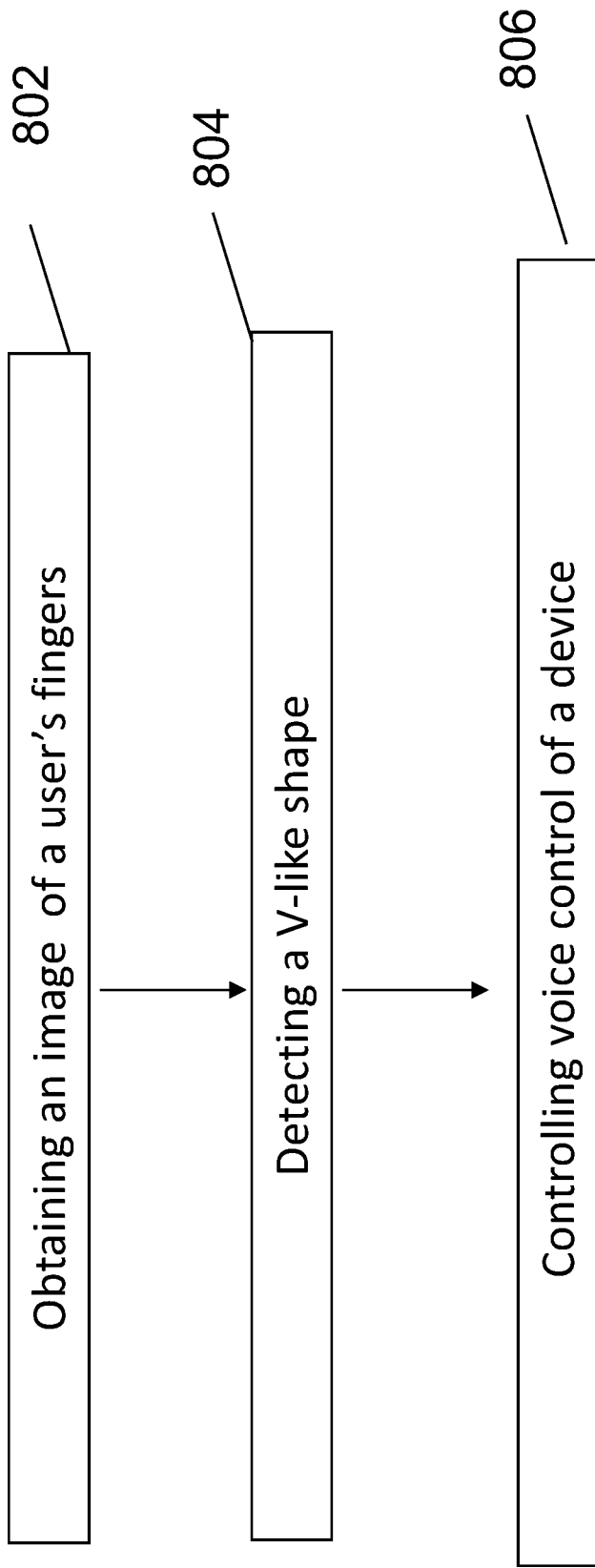


FIGURE 8

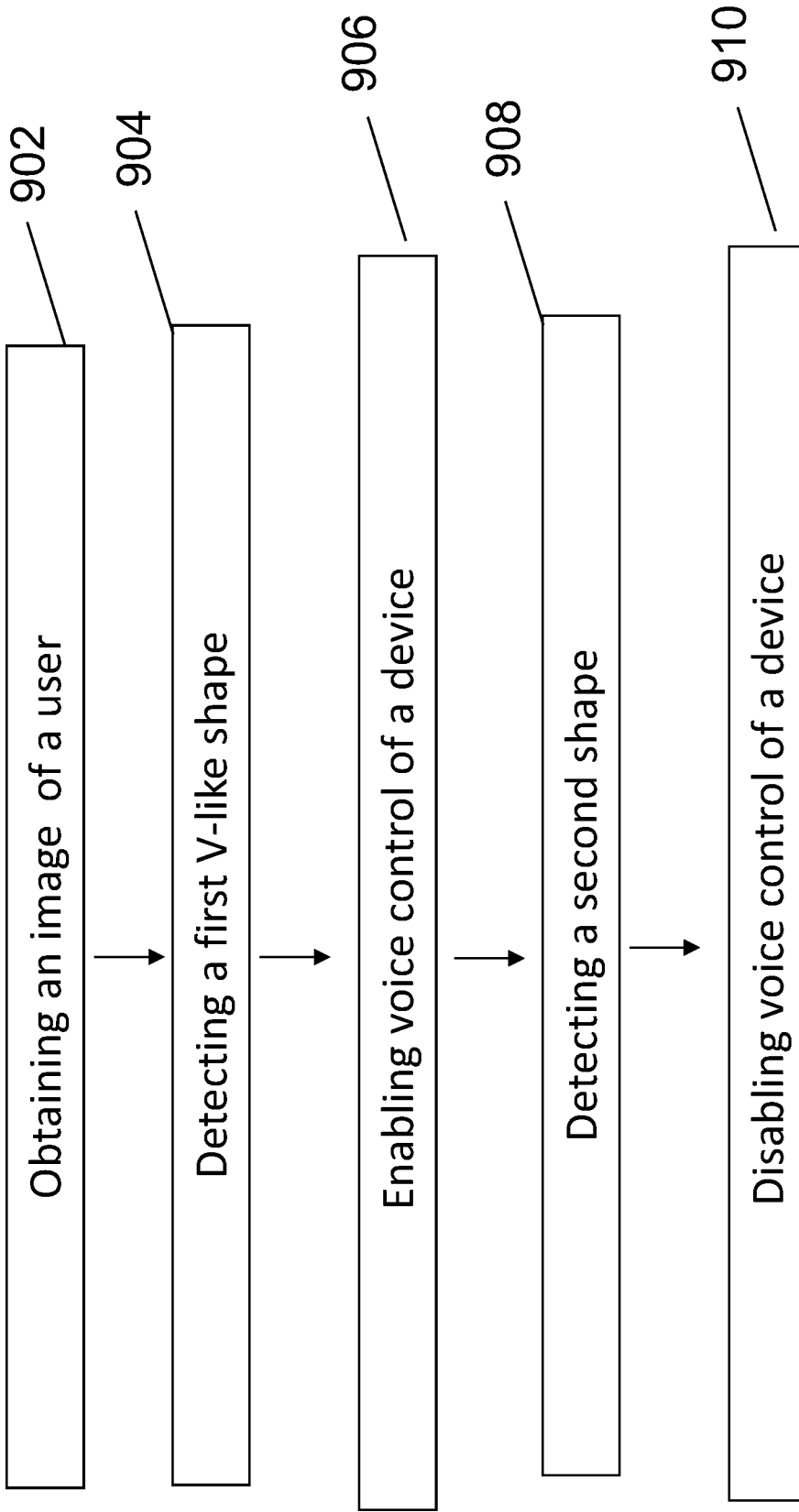


FIGURE 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL2014/050660

A. CLASSIFICATION OF SUBJECT MATTER IPC (2014.01) G06F 3/01		
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) IPC (2014.01) G06F 3/01		
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) Databases consulted: Eponline Search terms used: pointing camera gesture plurality ("select camera" OR "particular device")		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2012099584 A1 HEWLETT-PACKARD DEVELOPMENT COMPANY 27 Jul 2012 (2012/07/27) abstract, ¶¶10,11,12,14,16,20-23, page 11 claim 4;figs. 1,3a-3i,4a-4e	1-3,15-45
Y	abstract	4-13
X	US 2009102788 A1 Nishida et al. 23 Apr 2009 (2009/04/23) abstract, ¶0034	14
Y	abstract, ¶0034	4-13
A	US 2009324008 A1 Kongqiao et al. 31 Dec 2009 (2009/12/31) figs. 1, 6, 10; ¶¶4,28,45	1-16
A	US 2011134250 A1 Kim et al. 09 Jun 2011 (2011/06/09) The entire document	1-45
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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	"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	
"E" earlier application or patent but published on or after the international filing date	"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	
"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)	"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	
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family	
"P" document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search 10 Nov 2014	Date of mailing of the international search report 12 Nov 2014	
Name and mailing address of the ISA: Israel Patent Office Technology Park, Bldg.5, Malcha, Jerusalem, 9695101, Israel Facsimile No. 972-2-5651616	Authorized officer MAUDA Nissim Telephone No. 972-2-5651733	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL2014/050660

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2013066526 A1 Mondragon et al. 14 Mar 2013 (2013/03/14) The entire document	1-45
A	US 2009303176 A1 CHEN et al. 10 Dec 2013 (2013/12/10) The entire document	1-45
A	US 2011107216 A1 BI 05 May 2011 (2011/05/05) The entire document	1-16

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IL2014/050660

Patent document cited search report	Publication date	Patent family member(s)	Publication Date
WO 2012099584 A1	27 Jul 2012	WO 2012099584 A1	26 Jul 2012
		CN 103329066 A	25 Sep 2013
		EP 2666070 A1	27 Nov 2013
		JP 2014507714 A	27 Mar 2014
		KR 20140014129 A	05 Feb 2014
		US 2013290911 A1	31 Oct 2013
US 2009324008 A1	31 Dec 2009	US 2009324008 A1	31 Dec 2009
		US 8194921 B2	05 Jun 2012
		EP 2291722 A1	09 Mar 2011
		EP 2291722 A4	02 Apr 2014
		TW 201003539 A	16 Jan 2010
		WO 2009156565 A1	30 Dec 2009
US 2011134250 A1	09 Jun 2011	US 2011134250 A1	09 Jun 2011
		US 8599265 B2	03 Dec 2013
		CN 102640080 A	15 Aug 2012
		DE 112010004671 T5	17 Jan 2013
		KR 20110062475 A	10 Jun 2011
		WO 2011068362 A2	09 Jun 2011
		WO 2011068362 A3	10 Nov 2011
US 2013066526 A1	14 Mar 2013	US 2013066526 A1	14 Mar 2013
		CN 103782255 A	07 May 2014
		CN 103842941 A	04 Jun 2014
		US 2013063340 A1	14 Mar 2013
		WO 2013036621 A1	14 Mar 2013
		WO 2013036632 A1	14 Mar 2013
US 2009303176 A1	10 Dec 2013	US 2009303176 A1	10 Dec 2009
		US 8599132 B2	03 Dec 2013

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IL2014/050660

Patent document cited search report	Publication date	Patent family member(s)	Publication Date
		CN 101604205 A	16 Dec 2009
		CN 101604205 B	27 Jun 2012
		CN 102778947 A	14 Nov 2012
		US 2014049466 A1	20 Feb 2014

US 2011107216 A1	05 May 2011	US 2011107216 A1	05 May 2011

US 2009102788 A1	23 Apr 2009	US 2009102788 A1	23 Apr 2009
		US 8378970 B2	19 Feb 2013
		CN 101419498 A	29 Apr 2009
		CN 101419498 B	04 May 2011
		DE 102008048325 A1	23 Apr 2009
		DE 102008048325 B4	08 Dec 2011
		JP 2009104297 A	14 May 2009
		JP 5228439 B2	03 Jul 2013
		US 2013162535 A1	27 Jun 2013
		US 8681099 B2	25 Mar 2014
