



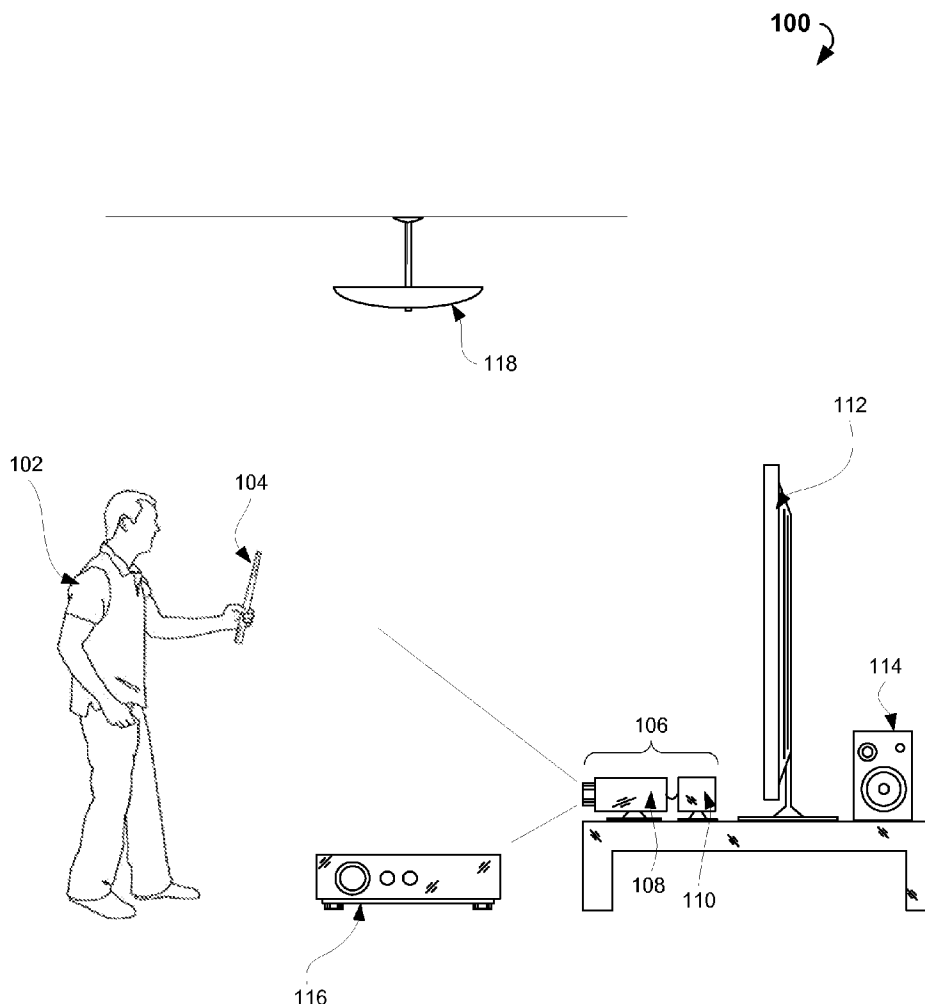
US 20130010207A1

(19) **United States**(12) **Patent Application Publication**
Valik et al.(10) **Pub. No.: US 2013/0010207 A1**(43) **Pub. Date: Jan. 10, 2013**(54) **GESTURE BASED INTERACTIVE CONTROL
OF ELECTRONIC EQUIPMENT***H04N 5/44* (2011.01)*G09G 5/00* (2006.01)(75) Inventors: **Andrey Valik**, Miass (RU); **Pavel
Zaitsev**, Miass (RU); **Dmitry Morozov**,
Miass (RU)(52) **U.S. Cl. 348/734; 345/156; 345/158; 704/275;
348/E05.096; 704/E21.001**(73) Assignee: **3DIVI**, Miass (RU)(21) Appl. No.: **13/478,457**(22) Filed: **May 23, 2012**(30) **Foreign Application Priority Data**

Jul. 4, 2011 (RU) RU2011127116

Publication Classification(51) **Int. Cl.***G06F 3/033* (2006.01)*G10L 21/00* (2006.01)(57) **ABSTRACT**

A computer-implemented method for controlling one or more electronic devices by recognition of gestures made by a three-dimensional object (3D). In one example embodiment, the method comprises capturing a series of successive 3D images in real time, identifying that the object has a predetermined elongated shape, identifying that the object is oriented substantially towards a predetermined direction, determining at least one qualifying action being performed by a user and/or the object, comparing the at least one qualifying action to one or more pre-determined actions associated with the direction towards which the object is oriented, and, based on the comparison, selectively issuing to the one or more electronic devices a command associated with the at least one qualifying action.



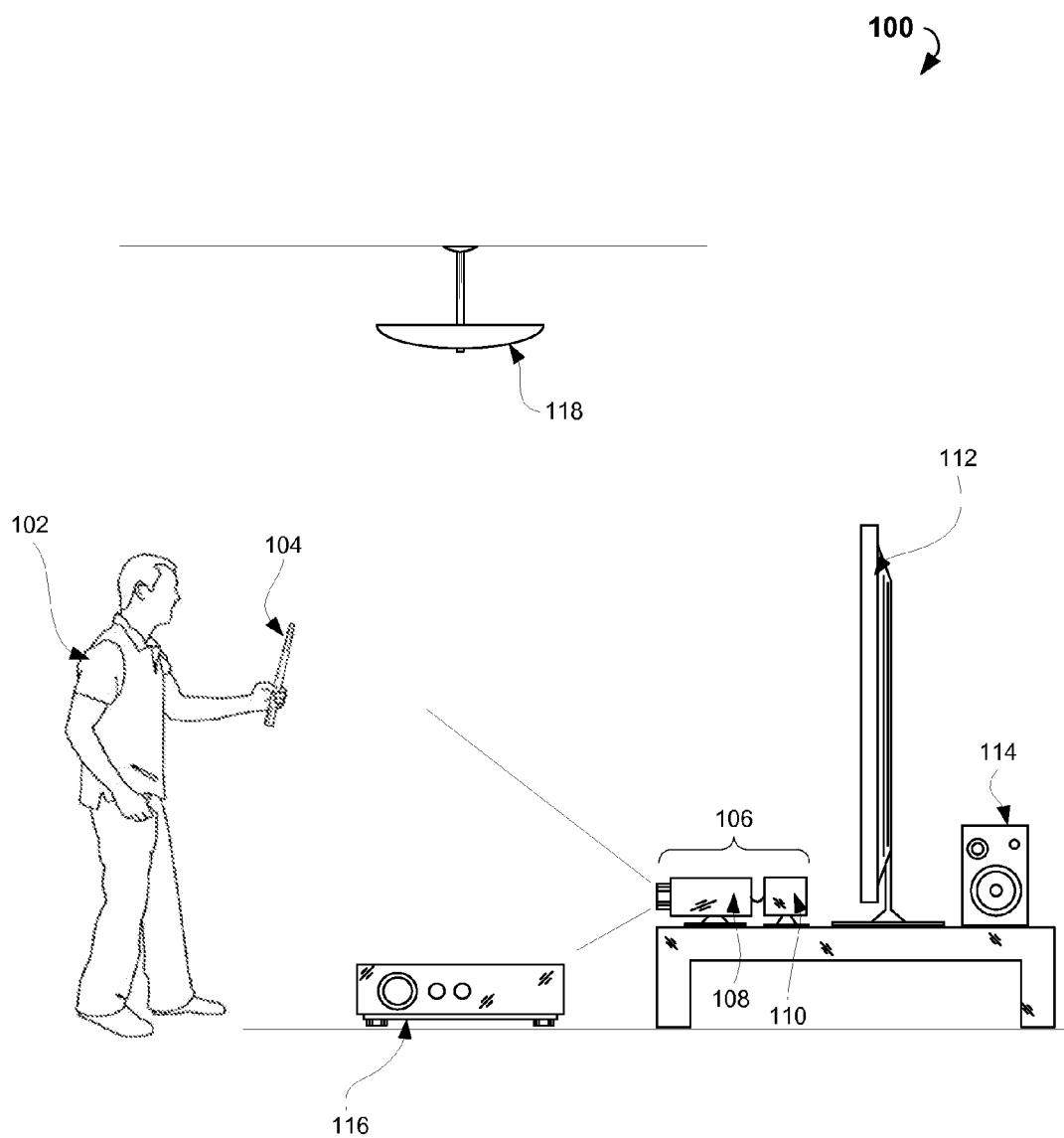


FIG. 1

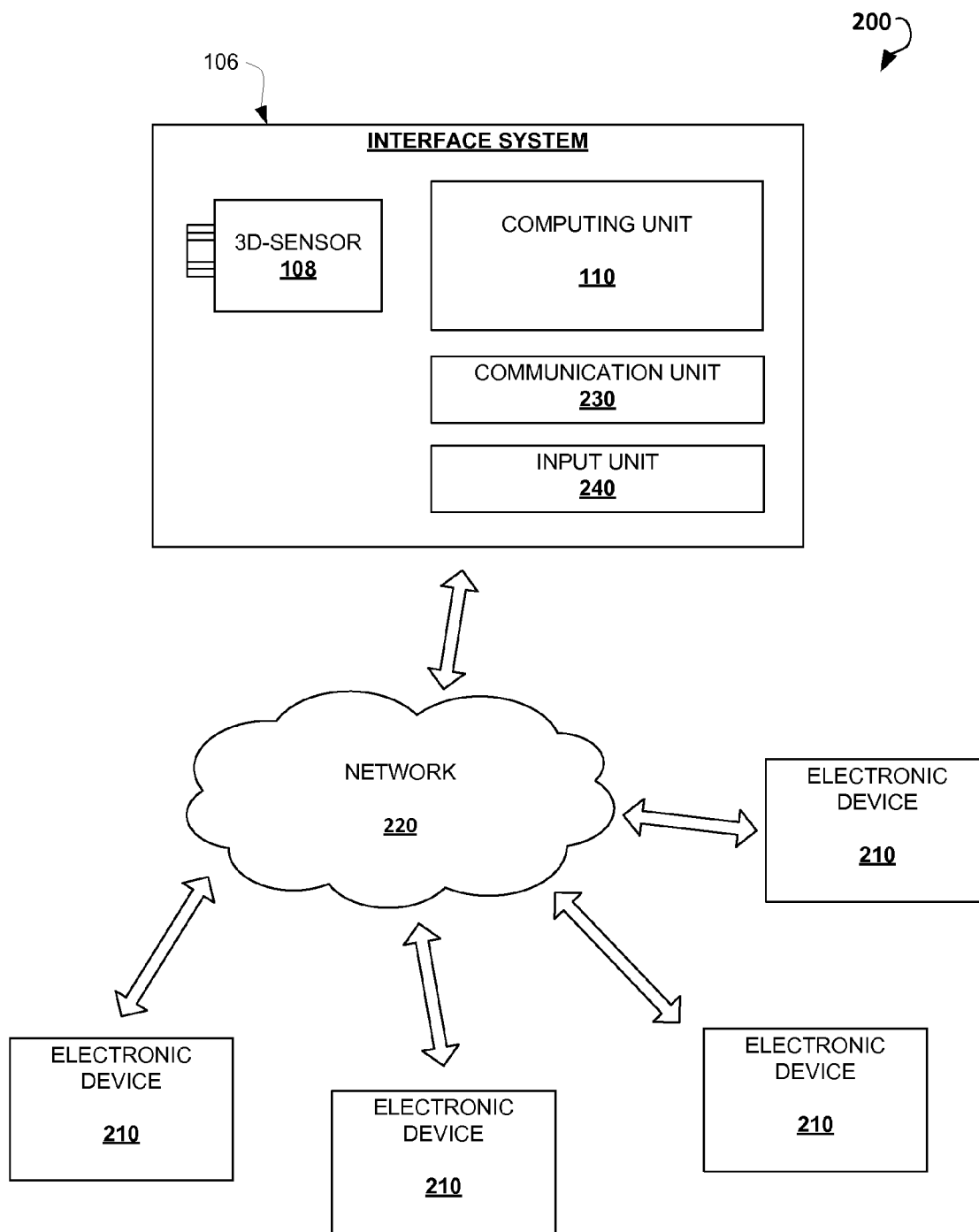


FIG. 2

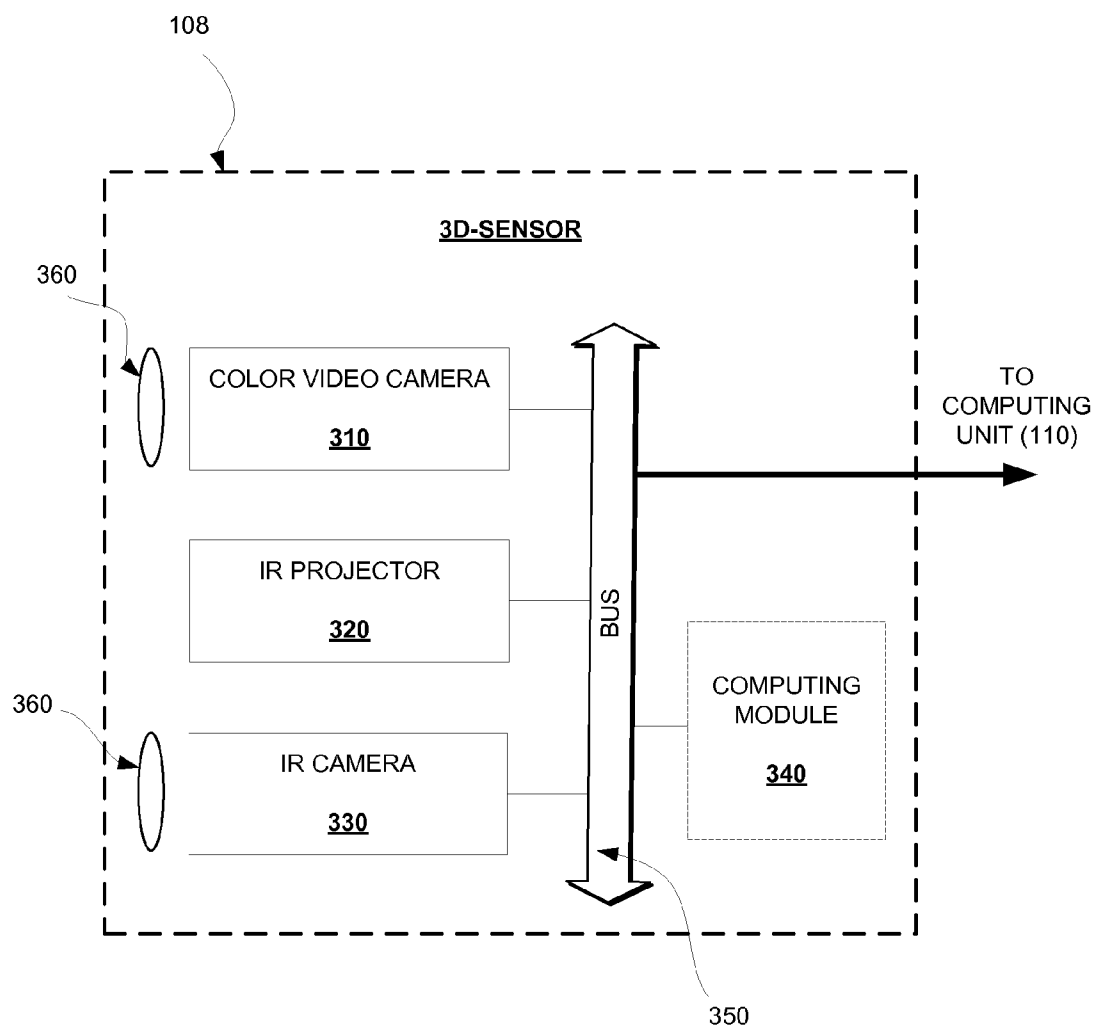


FIG. 3

110
↙

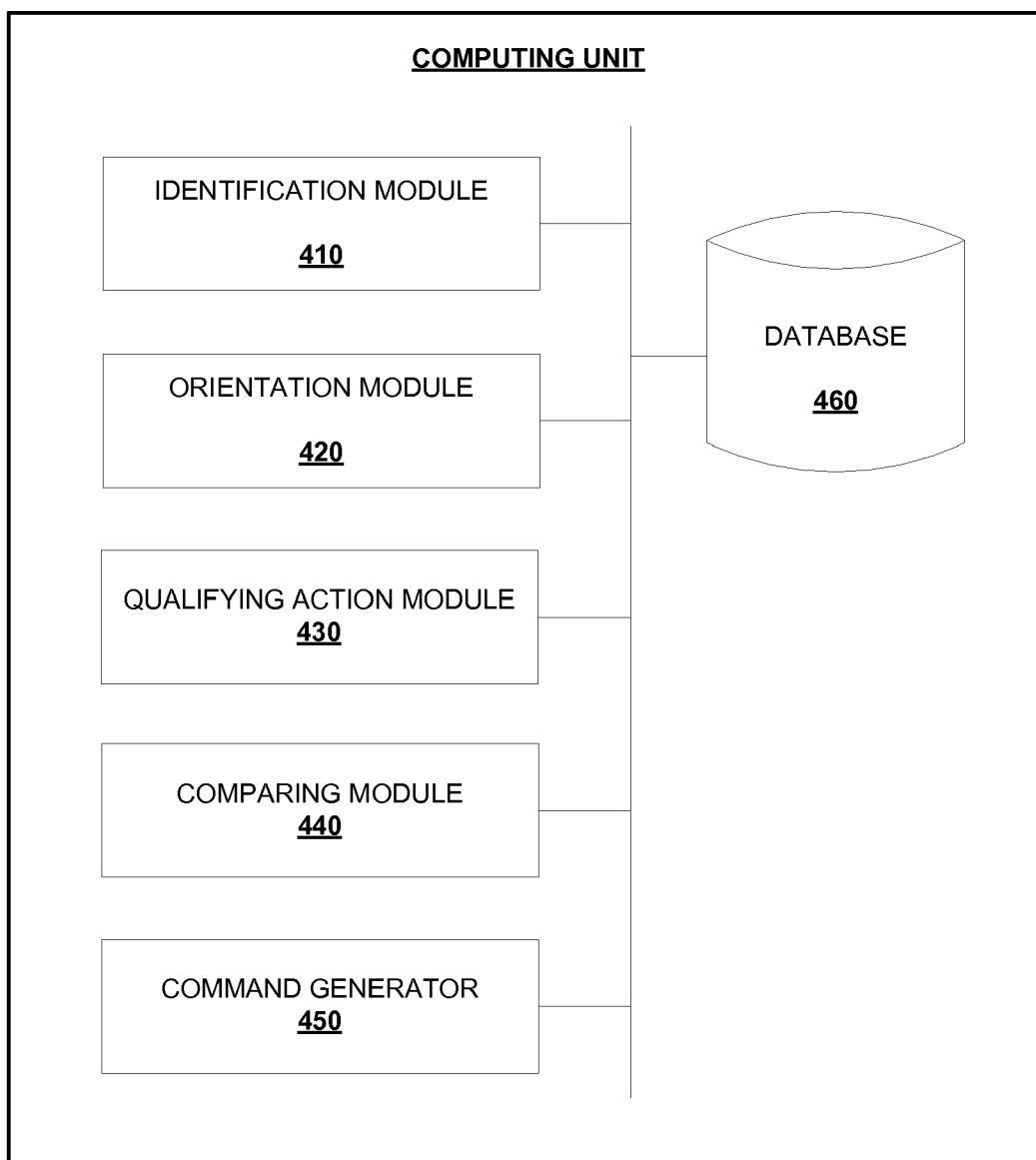


FIG. 4

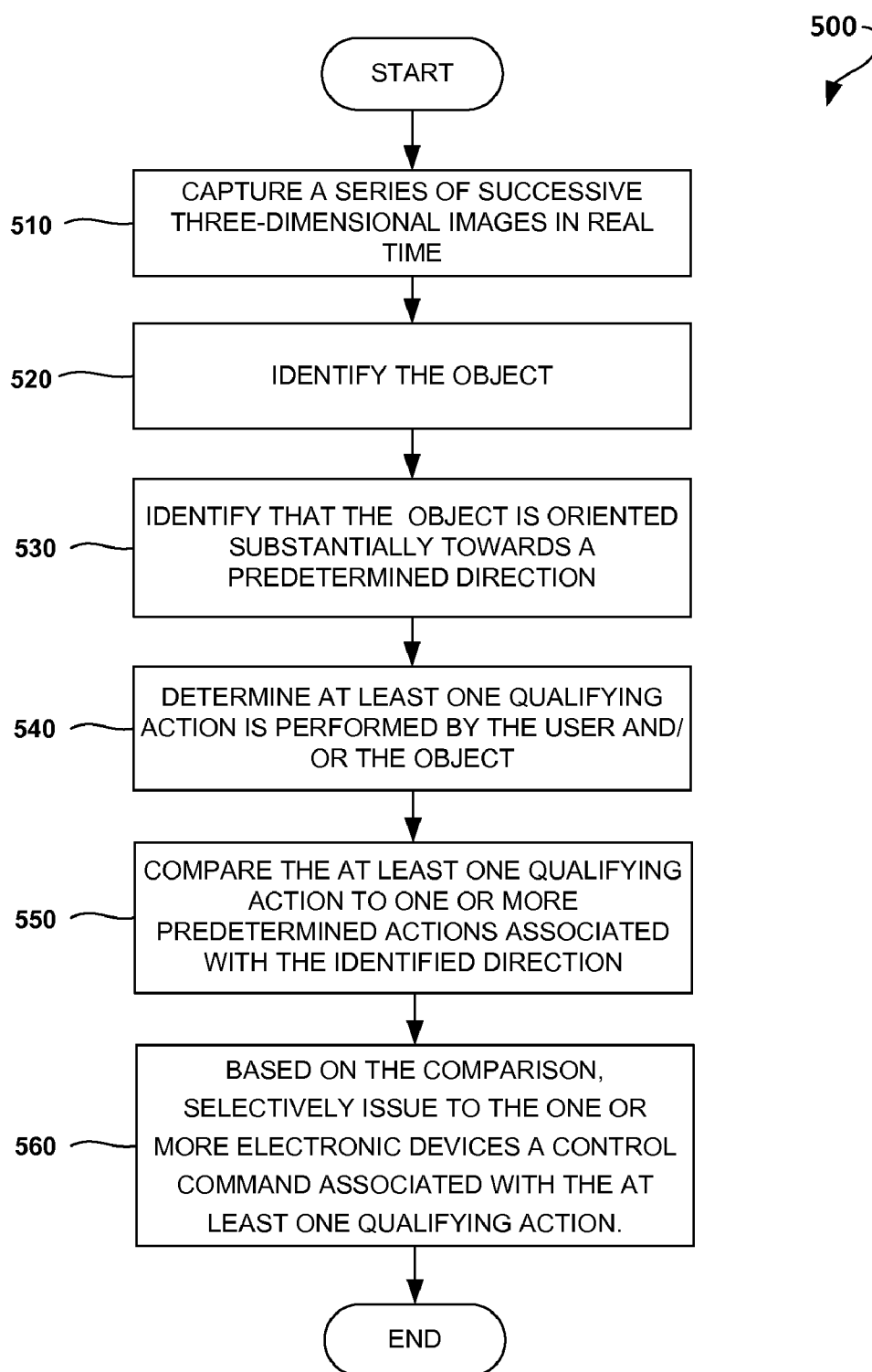


FIG. 5

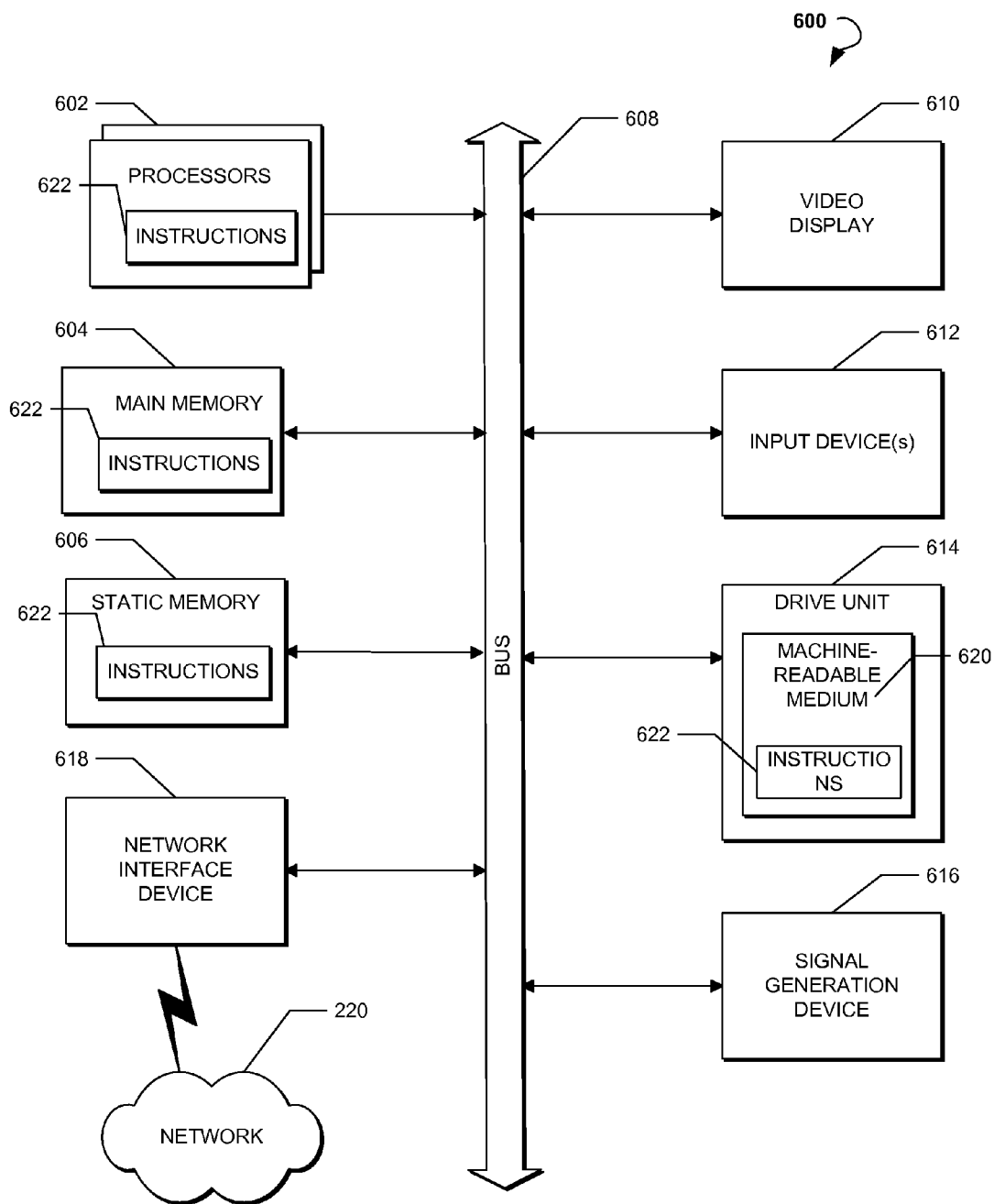


FIG. 6

GESTURE BASED INTERACTIVE CONTROL OF ELECTRONIC EQUIPMENT

RELATED APPLICATIONS

[0001] This application is Continuation-in-Part of Russian Patent Application Serial No. 2011127116, filed Jul. 4, 2011, which is incorporated herein by reference in its entirety for all purposes.

BACKGROUND

[0002] 1. Technical Field

[0003] This disclosure relates generally to computer interfaces and, more particularly, to methods for controlling electronic equipment by recognition of gestures made by an object.

[0004] 2. Description of Related Art

[0005] The approaches described in this section could be pursued but are not necessarily approaches that have previously been conceived or pursued. Therefore, unless otherwise indicated, it should not be assumed that any of the approaches described in this section qualify as prior art, merely by virtue of their inclusion in this section.

[0006] Interactive gesture interface systems are commonly used to interact with various electronic devices including gaming consoles, Television (TV) sets, computers, and so forth. The general principle of such systems is to detect human gestures or motions made by users, and generate commands based thereupon that cause electronic devices to perform certain actions. Gestures can originate from any bodily motion or state, but commonly originate from the face or hand. However, some systems also include emotion recognition features.

[0007] The gesture interface systems may be based on various gesture recognition approaches that involve the utilization of cameras, moving sensors, acceleration sensors, position sensors, electronic handheld controllers, and so forth. Whichever approach is used, human gestures can be captured and recognized, and a particular action can be triggered by an electronic device. Particular examples may include wireless electronic handheld controllers, which enable users to control gaming consoles by detecting motions or gestures made by such controllers. While such systems became very popular, they are still quite complex and require the utilization of various handheld controllers that are typically different for different applications.

[0008] Another approach involves utilization of 3D sensor devices capable of recognizing users' gestures or motions without dedicated handheld controllers or the like. Gestures are identified by processing users' images obtained by such 3D-sensors, and then they are interpreted to generate a control command. Control commands can be used to trigger particular actions performed by electronic equipment coupled to the 3D-sensor. Such systems are now widely deployed and generally used for gaming consoles.

[0009] One of the major drawbacks of such systems is that they are not flexible and cannot generate control commands for multiple electronic devices concurrently connected to a single 3D-sensor or any other device for capturing human motions or gestures. Thus, the conventional technology fails to provide a technique for improved detection and interpretation of human gestures associated with a particular electronic device among a plurality of devices connected to the common 3D-sensor.

SUMMARY

[0010] This summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0011] In accordance with various embodiments and the corresponding disclosure thereof, methods and systems for controlling one or more electronic devices by recognition of gestures made by an object. The described methodologies enable users to interact with one or a plurality of electronic devices such as gaming consoles, computers, audio systems, video systems, and so forth. The interaction with various electronic devices can be performed with the help of at least one 3D-sensor being configured to recognize not only gestures, but also a particular electronic device among the plurality of electronic devices to which the gestures are dedicated.

[0012] In accordance with one aspect, there is provided a computer-implemented method for controlling one or more electronic devices by recognition of gestures made by an object. The method may comprise capturing a series of successive 3D images in real time and identifying the object. The object may have a predetermined elongated shape. The method may also comprise identifying that the object is oriented substantially towards a predetermined direction, determining at least one qualifying action being performed by an user and/or the object, comparing the at least one qualifying action to one or more predetermined actions associated with the direction to which the object is oriented towards, and, based on the comparison, selectively providing to the one or more electronic devices a command associated with the at least one qualifying action.

[0013] In some embodiments, the predetermined direction can be associated with the one or more electronic devices. The object can be selected from a group comprising a wand, an elongated pointing device, an arm, a hand, and one or more fingers of the user. The series of successive 3D images can be captured using at least one video camera or a 3D image sensor. In some examples, the object can be identified by performing one or more of: processing the captured series of successive 3D images to generate a depth map, determining geometrical parameters of the object, and identifying the object by matching the geometrical parameters to a predetermined object database. The determination of at least one qualifying action may comprise the determining and acknowledging of one or more of: a predetermined motion of the object, a predetermined gesture of the object, a gaze of the user towards the predetermined direction associated with one or more electronic devices, a predetermined motion of the user, a predetermined gesture of the user, biometric data of the user, and a voice command provided by the user. Biometric data of the user can be determined based on one or more of the following: face recognition, voice recognition, user body recognition, and recognition of a user motion dynamics pattern. The gaze of the user can be determined based on one or more of the following: a position of the eyes of the user, a position of the pupils or a contour of the irises of the eyes of the user, a position of the head of the user, an angle of inclination of the head of the user, and a rotation of the head of the user. The mentioned one or more electronic devices may comprise a computer, a game console, a TV set, a TV adapter, a commu-

nication device, a Personal Digital Assistant (PDA), a lighting device, an audio system, and a video system.

[0014] According to another aspect, there is provided a system for controlling one or more electronic devices by recognition of gestures made by an object. The system may comprise at least one 3D image sensor configured to capture a series of successive 3D images in real time and a computing unit communicatively coupled to the at least one 3D image sensor. The computing unit can be configured to: identify the object; identify that the object is oriented substantially towards a predetermined direction; determine at least one qualifying action being performed by a user and/or the object; compare the at least one qualifying action to one or more predetermined actions associated with the direction to which the object is oriented towards; and, based on the comparison, selectively provide to the one or more electronic devices a command associated with the at least one qualifying action.

[0015] In some example embodiments, the at least one 3D image sensor may comprise one or more of an infrared (IR) projector to generate modulated light, an IR camera to capture 3D images associated with the object or the user, and a color video camera. The IR projector, color video camera, and IR camera can be installed in a common housing. The color video camera and/or IR camera can be equipped with liquid lenses. The mentioned predetermined direction can be associated with the one or more electronic devices. The object can be selected from a group comprising a wand, an elongated pointing device, an arm, a hand, and one or more fingers of the user. The computing unit can be configured to identify the object by performing the acts of: processing the captured series of successive 3D images to generate a depth map, determining geometrical parameters of the object, and identifying the object by matching the geometrical parameters to a predetermined object database. Furthermore, the computing unit can be configured to determine at least one qualifying action by performing the acts of determining and acknowledging of one or more of: a predetermined motion of the object, a predetermined gesture of the object, a gaze of the user towards the predetermined direction associated with one or more electronic devices, a predetermined motion of the user, a predetermined gesture of the user, biometric data of the user, and a voice command provided by the user. Biometric data of the user can be determined based on one or more of the following: face recognition, voice recognition, user body recognition, and recognition of user motion dynamics pattern. The gaze of the user can be determined based on one or more of the following: a position of the eyes of the user, a position of the pupils or a contour of the irises of the eyes of the user, a position of the head of the user, an angle of inclination of the head of the user, and a rotation of the head of the user.

[0016] According to yet another aspect, there is provided a processor-readable medium. The medium may store instructions, which when executed by one or more processors, cause the one or more processors to: capture a series of successive 3D images in real time; identify the object; identify that the object is oriented substantially towards a predetermined direction; determine at least one qualifying action being performed by a user and/or the object; compare the at least one qualifying action to one or more predetermined actions associated with the direction to which the object is oriented towards; and, based on the comparison, selectively provide to the one or more electronic devices a command associated with the at least one qualifying action.

[0017] To the accomplishment of the foregoing and related ends, the one or more aspects comprise the features herein-after fully described and particularly pointed out in the claims. The following description and the drawings set forth in detail certain illustrative features of the one or more aspects. These features are indicative, however, of but a few of the various ways in which the principles of various aspects may be employed, and this description is intended to include all such aspects and their equivalents.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] Embodiments are illustrated by way of example, and not limitation, in the figures of the accompanying drawings, in which like references indicate similar elements and in which:

[0019] FIG. 1 is a general illustration of a scene suitable for implementing methods for controlling one or more electronic devices by recognition of gestures made by an object.

[0020] FIG. 2 shows an example system environment suitable for implementing methods for controlling one or more electronic devices by recognition of gestures made by an object.

[0021] FIG. 3 shows an example embodiment of the 3D-sensor.

[0022] FIG. 4 is a diagram of the computing unit, according to an example embodiment.

[0023] FIG. 5 is a process flow diagram showing a method for controlling one or more electronic devices by recognition of gestures made by the object, according to an example embodiment.

[0024] FIG. 6 is a diagrammatic representation of an example machine in the form of a computer system within which a set of instructions for the machine to perform any one or more of the methodologies discussed herein is executed.

DETAILED DESCRIPTION

[0025] The following detailed description includes references to the accompanying drawings, which form a part of the detailed description. The drawings show illustrations in accordance with example embodiments. These example embodiments, which are also referred to herein as “examples,” are described in enough detail to enable those skilled in the art to practice the present subject matter. The embodiments can be combined, other embodiments can be utilized, or structural, logical and electrical changes can be made without departing from the scope of what is claimed. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope is defined by the appended claims and their equivalents. In this document, the terms “a” and “an” are used, as is common in patent documents, to include one or more than one. In this document, the term “or” is used to refer to a nonexclusive “or,” such that “A or B” includes “A but not B,” “B but not A,” and “A and B,” unless otherwise indicated.

[0026] The techniques of the embodiments disclosed herein may be implemented using a variety of technologies. For example, the methods described herein may be implemented in software executing on a computer system or in hardware utilizing either a combination of microprocessors or other specially designed application-specific integrated circuits (ASICs), programmable logic devices, or various combinations thereof. In particular, the methods described herein may be implemented by a series of computer-execut-

able instructions residing on a storage medium such as a carrier wave, disk drive, or computer-readable medium. Exemplary forms of carrier waves may take the form of electrical, electromagnetic, or optical signals conveying digital data streams along a local network or a publicly accessible network such as the Internet.

[0027] The embodiments described herein relate to computer-implemented methods and systems for controlling one or more electronic devices by recognition of gestures made by an object. The object, as used herein, may refer to any elongated object having prolonged shape and may include, for example, a wand, an elongated handheld pointing device, an arm, a hand, or one or more fingers of the user. Thus, according to inventive approaches described herein, gestures can be made by users with either their hands (arms or fingers) or handheld, elongated objects. In some embodiments, gestures can be made by handheld objects in combination with motions of arms, fingers, or other body parts of the user.

[0028] In general, one or more 3D-sensors or video cameras can be used to recognize gestures. In the context of this document, various techniques for gesture identification and recognition can be used, and accordingly, various devices can be utilized. In one example embodiment, a single 3D-sensor can be used and may include an IR projector, an IR camera, and an optional color video camera all embedded within a single housing. Image processing and interpretation can be performed by any computing device coupled to or embedding the 3D-sensor. Some examples may include a tabletop computer, laptop, tablet computer, gaming console, audio system, video system, phone, smart phone, PDA, or any other wired or wireless electronic device. Based on image processing and interpretation, a particular control command can be generated and outputted by the computing device. For example, the computing device may recognize a particular gesture associated with a predetermined command and generate such command for further input into a particular electronic device selected from a plurality of electronic devices. For instance, one command generated by the computing device and associated with a first gesture can be inputted to a gaming console, while another command, being associated with a second gesture, can be inputted to an audio-system. In other words, the computing device can be coupled to multiple electronic devices of same or various types, and such electronic devices can be selectively controlled by the user.

[0029] In some example embodiments, the computing device may be integrated with one or more controlled electronic device(s). For instance, the computing device and optional 3D-sensor can be integrated with a gaming console. This gaming console can be configured to be coupled to other electronic devices such as a lighting device, audio system, video system, TV set, and so forth. Those skilled in the art would appreciate that the 3D-sensor, the computing device, and various controlled electronic devices can be integrated with each other or interconnected numerous different ways. It should also be understood that such systems may construe at least some parts of “intelligent house” and may be used as part of home automation systems.

[0030] To select a particular electronic device and generate a control command for such electronic device, the user should perform two actions either concurrently or in series. The first action includes pointing the object towards the particular device. This may include posing the elongated object such that it is substantially oriented towards the particular device to be controlled. For example, the user may show the device

with the pointer finger. Alternatively, the user may orient an arm or hand towards the device. In some other examples, the user may orient the handheld object (e.g., a wand) towards the electronic device. In general, it should be understood that any elongated object can be used to designate a particular electronic device for further action. Such elongated object may or may not include electronic components.

[0031] To generate a particular control command to the selected electronic device, the user should perform the second action, which as used herein is referred to as a “qualifying action.” Once the interface system identifies that the user conducts the first and the second action, a predetermined control command is generated for a desired electronic device. The qualifying action may include one or more different actions. In some embodiments, the qualifying action may refer to a predetermined motion or gesture made by the object. For example, the user may first point to an electronic device to “select” it, and then make a certain gesture (e.g., make a circle motion, a nodding motion, or any other predetermined motion or gesture) to trigger the generation and output of a control command associated with the recognized gesture and “selected” electronic device. In some other embodiments, the qualifying action may include a predetermined motion or gesture of the user. For example, the user may at first point to an electronic device, and then make a certain gesture by hand or head (e.g., the user may knock a pointer finger onto a wand held in the arm).

[0032] Furthermore, in some other embodiments, the qualifying action may include a gaze of the user towards the predetermined direction associated with one or more electronic devices. For example, the user may point to an electronic device while also looking at it. Such a combination of actions can be unequivocally interpreted by the interface system to mean that a certain command is to be performed. The qualifying action may refer to a voice command generated by the user. For example, the user may point to a TV set and say, “turn on,” to generate a turn on command. In some embodiments, the qualifying action may include receipt and identification of biometric data associated with the user. The biometric data may include a face, a voice, motion dynamics pattern, and so forth. For example, face recognition or voice recognition can be used to authorize the user to control certain electronic devices.

[0033] In some additional embodiments, the interface system may require the user to perform at least two or more qualifying actions. For example, to generate a particular control command for an electronic device, the user shall first use the object to point out the electronic device, then make a predetermined gesture using the object, and then provide a voice command. In another example, the user may point towards the electronic device, make a gesture, and turn the face towards the 3D sensor for further face recognition and authentication. It should be understood that various combinations of qualifying actions can be performed and predetermined for generation of a particular command.

[0034] The interface system may include a database of predetermined gestures, objects and related information. Once a gesture is captured by the 3D-sensor, the computing device may compare the captured gesture with the list of predetermined gestures to find the match. Based on such comparison, a predetermined command can be generated. Accordingly, the database may store and populate a list of predetermined commands, each of which is associated with a particular device and a particular qualifying action (or com-

bination of qualifying actions). It should also be understood that locations of various electronic devices can be pre-programmed in the system, or alternatively, they can be identified by the 3D sensor in real time. For this purpose, the electronic devices can be provided with tags to be attached to their surfaces. Those skilled in the art would appreciate that various techniques can be used to identify electronic devices for the interface system.

[0035] Referring now to the drawings, FIG. 1 is a general illustration of scene 100 suitable for implementing methods for controlling one or more electronic device(s) by recognition of gestures made by an object. In particular, FIG. 1 shows a user 102 holding a handheld elongated object 104, which can be used for interaction with an interface system 106. The interface system 106 may include both a 3D-sensor 108 and a computing unit 110, which can be stand-alone devices or can be embedded within a single housing.

[0036] The 3D-sensor 108 can be configured to capture a series of 3D images, which can be further transmitted to and processed by the computing unit 110. As a result of the image processing, the computing unit 110 may first identify the object 104 and its relative orientation in a certain direction, and second, identify one or more “qualifying actions” as discussed above (e.g., identify a gesture made by the user 102 or the object 104).

[0037] The interface system 106 may be operatively connected with various electronic devices 112-118. The electronic devices 112-118 may include any device capable of receiving electronic control commands and performing one or more certain actions upon receipt of such commands. For example, the electronic devices 112-118 may include desktop computers, laptops, tabletop computers, tablet computers, cellular phones, smart phones, PDAs, gaming consoles, TV sets, TV adapters, displays, audio systems, video systems, lighting devices, home appliances, or any combination or part thereof. According to the example shown in FIG. 1, there is a TV set 112, an audio system 114, a gaming console 116, and a lighting device 118. The electronic devices 112-118 are all operatively coupled to the interface system 106, as further depicted in FIG. 2. In some example embodiments, the interface system 106 may integrate one or more electronic devices (not shown). For example, the interface system 106 may be embedded in a gaming console 116 or desktop computer. Those skilled in the art should understand that various interconnections may be deployed for the devices 112-118.

[0038] The user 102 may interact with the interface system 106 by making gestures or various motions with his or her hands, arms, fingers, legs, head, or other body parts; by making gestures or motions using the object 104; or by making voice commands; by looking in a certain direction; or any combination thereof. All of these motions, gestures, and voice commands can be predetermined so that the interface system 106 is able to identify them, match them to the list of pre-stored user commands, and generate a particular command for electronic devices 112-118. In other words, the interface system 106 may be “taught” to identify and differentiate one or more motions or gestures.

[0039] The object 104 may be any device of elongated shape and design. One example of the object 104 may include a wand or elongated pointing device. It is important to note that the object 104 may be free of any electronics. It could be any article of prolonged shape. Although it is not described in this document (so as not to take away from the general principles), the interface system 106 may be trained to identify

and differentiate the object 104 as used by the user 102. The electronics-free object 104 may have a different design and may imitate various sporting equipment (e.g., a baseball bat, racket, machete, sword, steering wheel, and so forth). In some embodiments, the object 104 may have a specific color design or color tags. Such color tags or colored areas may have various designs and shapes, and in general, they may help facilitate better identification of the object 104 by the interface system 106.

[0040] FIG. 2 shows an example system environment 200 suitable for implementing methods for controlling one or more electronic device(s) by recognition of gestures made by an object. The system environment 200 comprises the interface system 106, one or more electronic devices 210, and a network 220.

[0041] The interface system 106 may include at least one 3D-sensor 108, the computing unit 110, a communication unit 230, and an optional input unit 240. All of these units 108, 110, 230, and 240 can be operatively interconnected. The 3D-sensor 108 may be implemented differently and may include an image capture device. Further details about the 3D-sensor 108 are documented below, with reference to FIG. 3. It should also be appreciated that the interface system 106 may include two or more 3D-sensors 108 spaced apart from each other.

[0042] The aforementioned one or more electronic device(s) 210 are, in general, any device configured to trigger one or more predefined action(s) upon receipt of a certain control command. Some examples of electronic devices 210 include, but are not limited to computers, displays, audio systems, video systems, gaming consoles, and lighting devices. In one embodiment, the system environment 200 may comprise multiple electronic devices 210 of different types, while in another embodiment, the multiple electronic devices 210 may be of the same type (e.g., two or more interconnected gaming consoles are used).

[0043] The communication unit 230 may be configured to transfer data between the interface system 106 and one or more electronic device(s) 210. The communication unit 230 may include any wireless or wired network interface controller, including, for example, a Local Area Network (LAN) adapter, Wide Area Network (WAN) adapter, Wireless Transmit Receiving Unit (WTRU), WiFi adapter, Bluetooth adapter, GSM/CDMA adapter, and so forth.

[0044] The input unit 240 may be configured to enable users to input data of any nature. In one example, the input unit 240 may include a keyboard or ad hoc buttons allowing the users to input commands, program an interface, customize settings, and so forth. According to another example, the input unit 240 includes a microphone to capture user voice commands, which can then be processed by the computing unit 110. Various different input technologies can be used in the input unit 240, including touch screen technologies, pointing devices, and so forth.

[0045] The network 220 may couple the interface system 106 and one or more electronic device(s) 210. The network 220 is a network of data processing nodes interconnected for the purpose of data communication and may be utilized to communicatively couple various components of the environment 200. The network 220 may include the Internet or any other network capable of communicating data between devices. Suitable networks may include or interface with any one or more of the following: local intranet, PAN (Personal Area Network), LAN, WAN, MAN (Metropolitan Area Net-

work), virtual private network (VPN), storage area network (SAN), frame relay connection, Advanced Intelligent Network (AIN) connection, synchronous optical network (SONET) connection, digital T1, T3, E1 or E3 line, Digital Data Service (DDS) connection, DSL (Digital Subscriber Line) connection, Ethernet connection, ISDN (Integrated Services Digital Network) line, dial-up port such as a V.90, V.34 or V.34b is analog modem connection, cable modem, ATM (Asynchronous Transfer Mode) connection, or an FDDI (Fiber Distributed Data Interface), or CDDI (Copper Distributed Data Interface) connection. Furthermore, communications may also include links to any of a variety of wireless networks including, WAP (Wireless Application Protocol), GPRS (General Packet Radio Service), GSM, CDMA or TDMA (Time Division Multiple Access), cellular phone networks, GPS, CDPD (cellular digital packet data), RIM (Research in Motion, Limited) duplex paging network, Bluetooth radio, or an IEEE 802.11-based radio frequency network. The network 220 can further include or interface with any one or more of the following: RS-232 serial connection, IEEE-1394 (Firewire) connection, Fiber Channel connection, IrDA (infrared) port, SCSI (Small Computer Systems Interface) connection, USB (Universal Serial Bus) connection, or other wired or wireless, digital or analog interface or connection, mesh or Digi® networking.

[0046] FIG. 3 shows an example embodiment of the 3D-sensor 108. In some embodiments, the 3D-sensor 108 may comprise at least a color video camera 310 configured to capture images. In some other embodiments, the 3D-sensor 108 may include an IR projector 320 to generate modulated light and also an IR camera 330 to capture 3D images associated with the object 104 or the user 102. In yet more exemplary embodiments, the 3D-sensor 108 may comprise the color video camera 310, IR projector 320, and IR camera 330. In an example, the color video camera 310, IR projector 320, and IR camera 330 are all encased within a single housing.

[0047] Furthermore, in some embodiments, the 3D-sensor 108 may also comprise a computing module 340 for image analysis, pre-processing, processing, or generation of commands for the color video camera 310, IR projector 320, or IR camera 330. In some other examples, such operations can be done by the computing unit 110. The 3D-sensor 108 may also include a bus 350 interconnecting the color video camera 310, IR projector 320, and/or IR camera 330, depending on which devices are used.

[0048] The 3D-sensor 108 may also include one or more liquid lenses 360, which can be used for the color video camera 310, IR camera 330, or both. In general, liquid lenses 360 can be used to adaptively focus cameras onto a certain object or objects. The liquid lens 360 may use one or more fluids to create an infinitely-variable lens without any moving parts, by controlling the meniscus (the surface of the liquid.) The control of the liquid lens 360 may be performed by the computing module 340 or the computing unit 110.

[0049] Additional details of the 3D-sensor 108 and how captured image data can be processed are disclosed in the Russian patent application serial number 2011127116, which is incorporated herein by reference in its entirety.

[0050] FIG. 4 is a diagram of the computing unit 110, according to an example embodiment. As shown in the figure, the computing unit 110 may comprise an identification module 410, orientation module 420, qualifying action module 430, comparing module 440, command generator 450, and database 460. In other embodiments, the computing unit 110

may include additional, fewer, or different modules for various applications. Furthermore, all modules can be integrated within a single system, or alternatively, can be remotely located and optionally accessed via a third party.

[0051] The identification module 410 can be configured to identify the object 104 and/or the user 102. The identification process may include processing the series of successive 3D images as captured by the 3D-sensor 108. A depth map is generated as the result of such processing. Further processing of the depth map enables the determination of geometrical parameters of the object 104 or the user 102. For example, a virtual hull or skeleton can be created. Once geometrical parameters are defined, the object 104 can be identified by matching these geometrical parameters to predetermined objects as stored in the database 460.

[0052] The orientation module 420 can be configured to identify that the object 104 is oriented substantially towards a predetermined direction. More specifically, the orientation module 420 can track movements of the object 104 so as to identify that the object 104 is oriented towards a certain direction for a predetermined period of time. Such certain directions can be associated with the electronic devices 210 or the interface system 106. It should be understood that the position of the electronic devices 210 can be preliminarily stored in the database 460, or the interface system 106 can be trained to identify and store locations associated with the electronic devices 210. In some embodiments, the interface system 106 can obtain and store images of various electronic devices 210 such that in future, they can be easily identified. In some further embodiments, the electronic devices 210 can be provided with tags (e.g., color tags, RFID tags, bar code tags, and so forth). Once they are identified, the interface system 106 can associate the tags with certain locations in a 3D space. Those skilled in the art would appreciate that various approaches can be used to identify the electronic devices 210 and their associated locations, so that the orientation of the object 104 towards such locations can be easily identified by the interface system 106.

[0053] The qualifying action module 430 can be configured to track motions of the user 102 or the object 104, and determine at least one qualifying action being performed by the user 102 or the object 104. As mentioned, the qualifying action may include one or more different actions. In some embodiments, the qualifying action may refer to a predetermined motion or gesture made by the object 104. For example, a nodding motion or a circle motion can be considered as a qualifying action. In some other embodiments, the qualifying action may include a predetermined motion or gesture of the user 102. For example, the user 102 may perform a gesture by the hand or head. There are no restrictions to such gestures. The only requirement is that the interface system 106 should be able to differentiate and identify them. Accordingly, it should be understood that the interface system 106 can store previously stored reference motions or reference gestures in the database 460. In some embodiments, the interface system 106 can be trained by performing various motions and gestures, such that the sample motions and gestures can be stored in the database 460 for further comparison with gestures captured in real time.

[0054] In some other embodiments, the qualifying action may include a gaze of the user 102 towards the predetermined direction associated with one or more electronic devices 210. The gaze of the user 102 can be determined based on one or more of the following: position of the eyes of the user 102,

position of the pupils or a contour of the irises of the eyes of the user **102**, position of the head of the user **102**, angle of inclination of the head of the user **102**, and a rotation of the head of the user **102**.

[0055] In some additional embodiments, the qualifying action may include a voice command generated by the user **102**. Voice commands can be captured by the input unit **240** and processed by the computing unit **110** in order to recognize the command and compare it to a predetermined list of voice commands to find a match.

[0056] In some additional embodiments, the qualifying action may include receipt and identification of biometric data associated with the user **102**. The biometric data may include a face, voice, motion dynamics patterns, and so forth. Based on the captured biometrics data, the computing unit **110** may authenticate a particular user **102** to control one or another electronic device **210**. Such a feature may prevent, for example, children from operating dangerous or unwanted electronic devices **210**.

[0057] In some embodiments, the interface system **106** may require that the user **102** perform at least two or more qualifying actions. For example, to generate a particular control command for an electronic device **210**, the user **102** shall first point to a certain electronic device **210** using the object **104**, then make a predetermined gesture using the object **104**, and thirdly, provide a voice command or perform another gesture. Given that various qualifying actions are provided, it should be understood that multiple combinations can be used to generate different control commands.

[0058] The comparing module **440** can be configured to compare the captured qualifying action to one or more predetermined actions being associated with the direction to which the object **104** was oriented (as defined by the orientation module **420**). The aforementioned one or more predetermined actions can be stored in the database **460**.

[0059] The command generator **450** can be configured to selectively provide to one or more electronic devices **210**, based on the comparison performed by the comparing module **440**, a command associated with at least one qualifying action that is identified by the qualifying action module **430**. Accordingly, each control command, among a plurality of control commands stored in the database **460**, is predetermined for a certain electronic device **210** and is associated with one or more certain gestures or motions and the location of the electronic device **210**.

[0060] The database **460** can be configured to store predetermined gestures, motions, qualifying actions, voice commands, control commands, electronic device location data, visual hulls or representations of users and objects, and so forth.

[0061] FIG. **5** is a process flow diagram showing a method **500** for controlling one or more electronic devices **210** by recognition of gestures made by the object **104**, according to an example embodiment. The method **500** may be performed by processing logic that may comprise hardware (e.g., dedicated logic, programmable logic, and microcode), software (such as software run on a general-purpose computer system or a dedicated machine), or a combination of both. In one example embodiment, the processing logic resides at the interface system **106**.

[0062] The method **500** can be performed by the various modules discussed above with reference to FIGS. **2** and **4**. Each of these modules can comprise processing logic. It will be appreciated by one of ordinary skill in the art that examples

of the foregoing modules may be virtual, and instructions said to be executed by a module may, in fact, be retrieved and executed by a processor. The foregoing modules may also include memory cards, servers, and/or computer discs. Although various modules may be configured to perform some or all of the various steps described herein, fewer or more modules may be provided and still fall within the scope of example embodiments.

[0063] As shown in FIG. **5**, the method **500** may commence at operation **510**, with the one or more 3D-sensors **108** capturing a series of successive 3D images in real time. At operation **520**, the identification module **410** identifies the object **104**.

[0064] At operation **530**, the orientation module **420** identifies the orientation of the object **104** and that the object **104** is oriented substantially towards a predetermined direction associated with a particular electronic device **210**. Accordingly, the orientation module **420** may track motion of the object **104** in real time.

[0065] At operation **540**, the qualifying action module **430** determines that at least one qualifying action is performed by the user **102** and/or the object **104**. For this purpose, the qualifying action module **430** may track motions of the user **102** and/or the object **104** in real time.

[0066] At operation **550**, the comparing module **440** compares the qualifying action, as identified at operation **540**, to one or more predetermined actions associated with the direction that was identified at operation **530** as the direction towards which the object is sustainably oriented.

[0067] At operation **560**, the command generator **450** selectively provides, to the one or more electronic devices **210**, a control command associated with at least one qualifying action and based on the comparison performed at operation **550**.

[0068] FIG. **6** shows a diagrammatic representation of a computing device for a machine in the example electronic form of a computer system **600**, within which a set of instructions for causing the machine to perform any one or more of the methodologies discussed herein can be executed. In example embodiments, the machine operates as a standalone device or can be connected (e.g., networked) to other machines. In a networked deployment, the machine can operate in the capacity of a server, a client machine in a server-client network environment, or as a peer machine in a peer-to-peer (or distributed) network environment. The machine can be a personal computer (PC), tablet PC, set-top box (STB), PDA, cellular telephone, portable music player (e.g., a portable hard drive audio device, such as a Moving Picture Experts Group Audio Layer 3 (MP3) player), web appliance, network router, switch, bridge, or any machine capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken by that machine. Further, while only a single machine is illustrated, the term "machine" shall also be taken to include any collection of machines that use or jointly execute a set (or multiple sets) of instructions to perform any one or more of the methodologies discussed herein.

[0069] The example computer system **600** includes a processor or multiple processors **602** (e.g., a central processing unit (CPU), graphics processing unit (GPU), or both), main memory **604** and static memory **606**, which communicate with each other via a bus **608**. The computer system **600** can further include a video display unit **610** (e.g., a liquid crystal display (LCD) or cathode ray tube (CRT)). The computer system **700** also includes at least one input device **612**, such as

an alphanumeric input device (e.g., a keyboard), cursor control device (e.g., a mouse), microphone, digital camera, video camera, and so forth. The computer system 600 also includes a disk drive unit 614, signal generation device 616 (e.g., a speaker), and network interface device 618.

[0070] The disk drive unit 614 includes a computer-readable medium 620, which stores one or more sets of instructions and data structures (e.g., instructions 622) embodying or utilized by any one or more of the methodologies or functions described herein. The instructions 622 can also reside, completely or at least partially, within the main memory 604 and/or within the processors 602 during execution by the computer system 600. The main memory 604 and the processors 602 also constitute machine-readable media.

[0071] The instructions 622 can further be transmitted or received over the network 220 via the network interface device 618 utilizing any one of a number of well-known transfer protocols (e.g., Hyper Text Transfer Protocol (HTTP), CAN, Serial, and Modbus).

[0072] While the computer-readable medium 620 is shown in an example embodiment to be a single medium, the term “computer-readable medium” should be taken to include a single medium or multiple media (e.g., a centralized or distributed database, and/or associated caches and servers) that store the one or more sets of instructions. The term “computer-readable medium” shall also be taken to include any medium that is capable of storing, encoding, or carrying a set of instructions for execution by the machine, and that causes the machine to perform any one or more of the methodologies of the present application, or that is capable of storing, encoding, or carrying data structures utilized by or associated with such a set of instructions. The term “computer-readable medium” shall accordingly be taken to include, but not be limited to, solid-state memories, optical and magnetic media. Such media can also include, without limitation, hard disks, floppy disks, flash memory cards, digital video disks, random access memory (RAM), read only memory (ROM), and the like.

[0073] The example embodiments described herein can be implemented in an operating environment comprising computer-executable instructions (e.g., software) installed on a computer, in hardware, or in a combination of software and hardware. The computer-executable instructions can be written in a computer programming language or can be embodied in firmware logic. If written in a programming language conforming to a recognized standard, such instructions can be executed on a variety of hardware platforms and for interfaces to a variety of operating systems. Although not limited thereto, computer software programs for implementing the present method can be written in any number of suitable programming languages such as, for example, C, C++, C#, Cobol, Eiffel, Haskell, Visual Basic, Java, JavaScript, Python, or other compilers, assemblers, interpreters or other computer languages or platforms.

[0074] Thus, methods and systems for controlling one or more electronic device(s) by recognition of gestures made by an object have been described. The disclosed technique provides a useful tool to enable people to interact with various electronic devices based on gestures, motions, voice commands, and gaze information.

[0075] Although embodiments have been described with reference to specific example embodiments, it will be evident that various modifications and changes can be made to these example embodiments without departing from the broader

spirit and scope of the present application. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense.

What is claimed is:

1. A computer-implemented method for controlling one or more electronic devices by recognition of gestures made by a three-dimensional object, the method comprising:

- capturing a series of successive 3D images in real time;
- identifying that the object has a predetermined elongated shape;
- identifying that the object is oriented substantially towards a predetermined direction;
- determining at least one qualifying action being performed by a user and/or the object;
- comparing the at least one qualifying action to one or more pre-determined actions associated with the direction towards which the object is oriented; and
- based on the comparison, selectively issuing to the one or more electronic devices a command associated with the at least one qualifying action.

2. The method of claim 1, wherein the predetermined direction is associated with the one or more electronic devices.

3. The method of claim 1, wherein the object is selected from a group comprising a wand, a elongated pointing device, an arm, a hand, and one or more fingers of the user.

4. The method of claim 1, wherein the series of successive 3D images are captured using at least one video camera or a 3D image sensor.

5. The method of claim 1, wherein identifying the object comprises one or more of:

- processing the captured series of successive 3D images to generate a depth map;
- determining geometrical parameters of the object; and
- identifying the object by matching the geometrical parameters to a predetermined object database.

6. The method of claim 1, wherein determining at least one qualifying action comprises determining and acknowledging one or more of:

- a predetermined motion of the object;
- a predetermined gesture of the object;
- a gaze of the user towards the predetermined direction associated with one or more electronic devices;
- a predetermined motion of the user;
- a predetermined gesture of the user;
- biometric data of the user; and
- a voice command provided by the user.

7. The method of claim 6, wherein biometric data of the user is determined based on one or more of the following: face recognition, voice recognition, user body recognition, and recognition of a user motion dynamics pattern.

8. The method of claim 6, wherein the gaze of the user is determined based on one or more of the following: a position of eyes of the user, a position of pupils or a contour of irises of the eyes, a position of a head of the user, an angle of inclination of the head, and a rotation of the head.

9. The method of claim 1, wherein the one or more electronic devices comprising a computer, a game console, a TV set, a TV adapter, a communication device, a Personal Digital Assistant (PDA), a lighting device, an audio system, and a video system.

10. A system for controlling one or more electronic devices by optical recognition of gestures made by an object, the system comprising:

at least one three-dimensional image sensor configured to capture a series of successive 3D images in real time; and

a computing unit communicatively coupled to the at least one 3D image sensor, the computing unit being configured to:

identify that the object has a predetermined elongated shape;

identify that the object is oriented substantially towards a predetermined direction;

determine at least one qualifying action being performed by a user and/or the object;

compare the at least one qualifying action to one or more predetermined actions associated with the direction towards which the object is oriented; and

based on the comparison, selectively issue to the one or more electronic devices a command associated with the at least one qualifying action.

11. The system of claim **10**, wherein the at least one 3D image sensor comprises one or more of an infrared (IR) projector to generate modulated light, an IR camera to capture 3D images associated with the object or the user, and a color video camera.

12. The system of claim **11**, wherein the IR projector, the color video camera, and IR camera are installed in a common housing.

13. The system of claim **11**, wherein the color video camera and/or the IR camera are equipped with liquid lenses.

14. The system of claim **10**, wherein the predetermined direction is associated with the one or more electronic devices.

15. The system of claim **10**, wherein the object is selected from a group comprising a wand, a elongated pointing device, an arm, a hand, and one or more fingers of the user.

16. The system of claim **10**, wherein the computing unit is configured to identify the object by performing the acts of:
processing the captured series of successive 3D images to generate a depth map;

determining geometrical parameters of the object; and identifying the object by matching the geometrical parameters to a predetermined object database.

17. The system of claim **10**, wherein the computing unit is configured to determine at least one qualifying action by performing the acts of determining and acknowledging one or more of:

a predetermined motion of the object;

a predetermined gesture of the object;

a gaze of the user towards the predetermined direction associated with one or more electronic devices;

a predetermined motion of the user;

a predetermined gesture of the user;

biometric data of the user; and

a voice command provided by the user.

18. The system of claim **17**, wherein biometric data of the user is determined based on one or more of the following: face recognition, voice recognition, user body recognition, and recognition of a user motion dynamics pattern.

19. The system of claim **17**, wherein the gaze of the user is determined based on the one or more of the following: a position of eyes of the user, a position of pupils or a contour of irises of the eyes, a position of a head of the user, an angle of inclination of the head, and a rotation of the head.

20. A processor-readable medium having instructions stored thereon, which when executed by one or more processors, cause the one or more processors to:

capture a series of successive 3D images in real time;

identify that an object has a predetermined elongated shape;

identify that the object is oriented substantially towards a predetermined direction;

determine at least one qualifying action being performed by a user and/or the object;

compare the at least one qualifying action to one or more predetermined actions associated with the direction towards which the object is oriented; and

based on the comparison, selectively provide to the one or more electronic devices a command associated with the at least one qualifying action.

* * * * *