



(86) Date de dépôt PCT/PCT Filing Date: 2008/09/15
(87) Date publication PCT/PCT Publication Date: 2009/03/19
(85) Entrée phase nationale/National Entry: 2010/03/15
(86) N° demande PCT/PCT Application No.: US 2008/010750
(87) N° publication PCT/PCT Publication No.: 2009/035705
(30) Priorité/Priority: 2007/09/14 (US60/993,907)

(51) Cl.Int./Int.Cl. *G06F 3/01* (2006.01),
G06F 3/042 (2006.01)
(71) Demandeur/Applicant:
INTELLECTUAL VENTURES HOLDING 67 LLC, US
(72) Inventeurs/Inventors:
BELL, MATTHEW, US;
CHENNAVASIN, TIPATAT, US;
CLANTON, CHARLES H., US;
HULME, MICHAEL, US;
OPHIR, EYAL, US;
VIETA, MATTHEW, US
(74) Agent: SMART & BIGGAR

(54) Titre : TRAITEMENT D'INTERACTIONS D'UTILISATEUR BASEES SUR DES GESTES
(54) Title: PROCESSING OF GESTURE-BASED USER INTERACTIONS

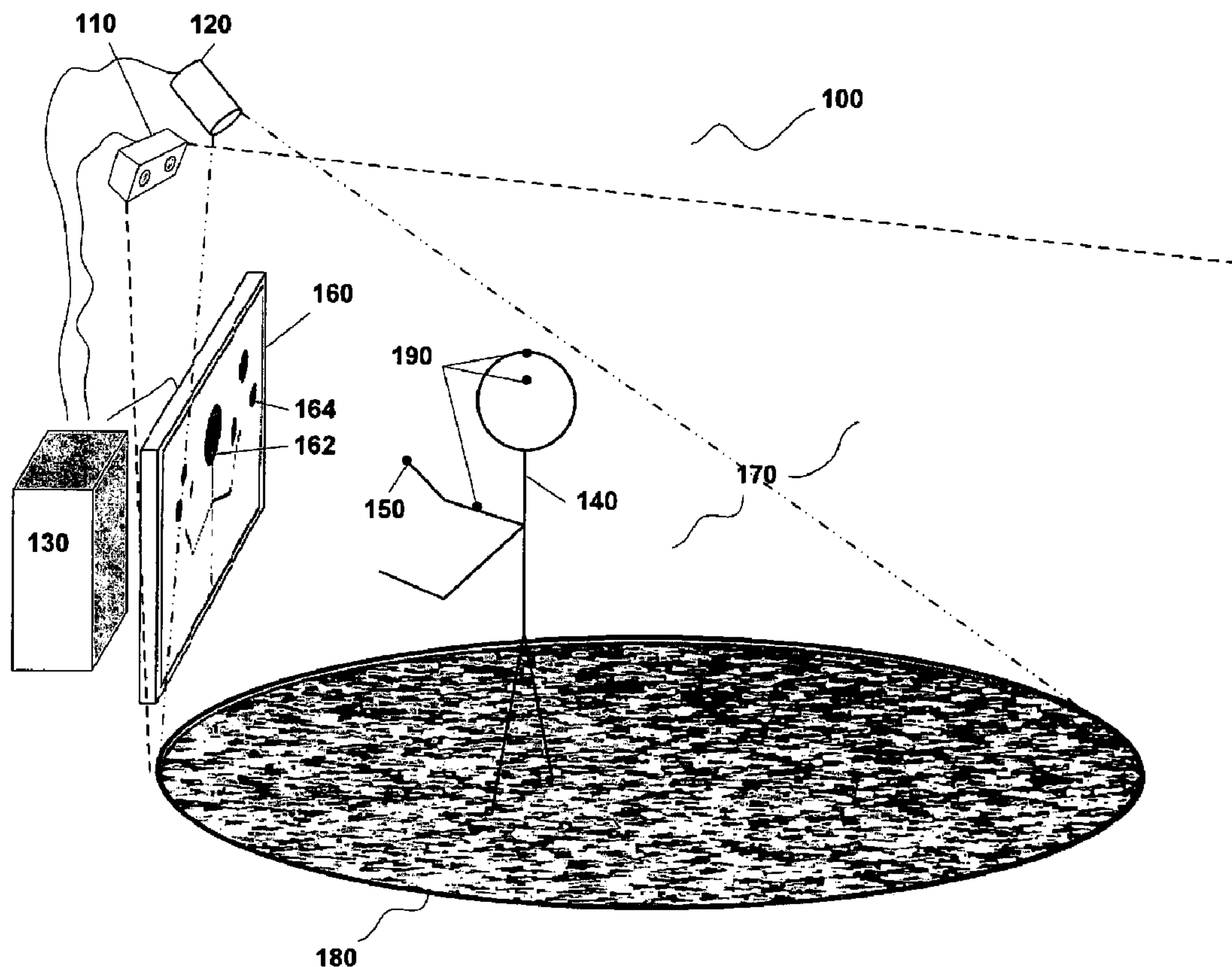


Figure 1

(57) Abrégé/Abstract:
Systems and methods for processing gesture-based user interactions with an interactive display.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
19 March 2009 (19.03.2009)

PCT

(10) International Publication Number
WO 2009/035705 A1(51) International Patent Classification:
G06F 17/00 (2006.01)(21) International Application Number:
PCT/US2008/010750(22) International Filing Date:
15 September 2008 (15.09.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/993,907 14 September 2007 (14.09.2007) US(71) Applicant (for all designated States except US): **REACTRIX SYSTEMS, INC.** [US/US]; 301 Chesapeake Dr., Redwood City, CA 94063 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BELL, Matthew** [US/US]; c/o Reactrix Systems, Inc., 301 Chesapeake Drive, Redwood City, CA 94063 (US). **CHENNAVASIN, Tipatat** [US/US]; c/o Reactrix Systems, Inc., 301 Chesapeake Drive, Redwood City, CA 94063 (US). **CLANTON, Charles, H.** [US/US]; c/o Reactrix Systems, Inc., 301 Chesapeake Drive, Redwood City, CA 94063 (US). **HULME, Michael** [US/US]; c/o Reactrix Systems, Inc., 301 Chesapeake Drive, Redwood City, CA 94063 (US).**OPHIR, Eyal** [US/US]; c/o Reactrix Systems, Inc., 301 Chesapeake Drive, Redwood City, CA 94063 (US). **VI-ETA, Matthew** [US/US]; c/o Reactrix Systems, Inc., 301 Chesapeake Drive, Redwood City, CA 94063 (US).(74) Agents: **PHAM, Tam Thanh** et al.; 2200 Geng Road, Palo Alto, CA 94303 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, IS, JP, KE, KG, KM, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(54) Title: PROCESSING OF GESTURE-BASED USER INTERACTIONS

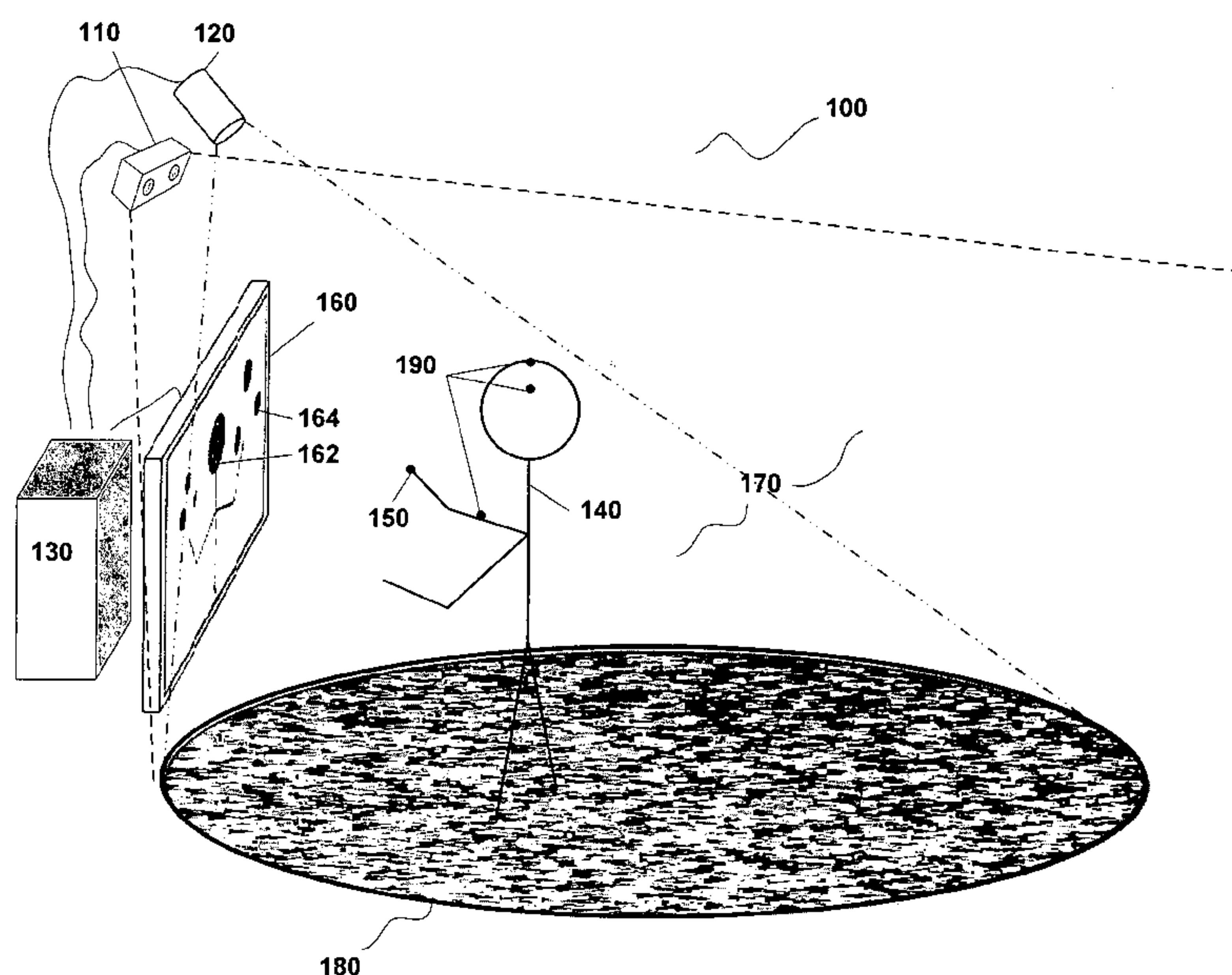


Figure 1

(57) Abstract: Systems and methods for processing gesture-based user interactions with an interactive display.

WO 2009/035705 A1

PROCESSING OF GESTURE-BASED USER INTERACTIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[1] The present application claims the priority benefit of U.S. provisional patent application number 60/993,907 filed September 14, 2007, the disclosure of which is incorporated herein by reference.

BACKGROUND

Field of the Invention

[2] The present invention generally relates to interactive display systems. More specifically, the present invention concerns user interfaces for processing gesture-based user interactions with an interactive display in a three-dimensional environment.

Description of the Related Art

[3] Many existing interactive technologies rely on the likes of two-dimensional cameras. The two-dimensional data generated by these cameras is difficult to reliably parse for the purpose of separating a foreground from a background. As a result, it is difficult to isolate the hands or other appendages of a user and track corresponding movements of the same.

[4] Poor isolation and tracking data leads to difficulties with respect to providing data to an interactive display system. If an interactive display system is unable to gather data for controlling or manipulating information rendered through the same, a user may be forced to rely upon button presses or manipulation of physical objects such as keyboards, joysticks, and actuators. Once a user is forced to rely upon physical instruments for providing input, many of the benefits offered by an interactive display system are lost.

[5] There is a need for processing of gesture-based user interactions with respect to interactive display systems.

SUMMARY

[6] In a first embodiment, a method for processing gesture-based user interactions with an interactive display is disclosed. Through the method, three-dimensional information about an object in a defined three-dimensional space is gathered. A portion of the three-dimensional information as corresponding to a hand of a user in the defined three-dimensional space is identified. An on-screen image of a digital object is displayed as is an on-screen image associated with and approximating the physical shape of the user. The displayed on-screen image approximates that of the user and does not interact with the displayed digital object. An icon corresponding to the hand of the user in the defined three-dimensional space is displayed. The icon is capable of interacting with the displayed digital object.

[7] A second method for processing gesture-based user interactions with an interactive display is disclosed in a further embodiment. In this further embodiment, three-dimensional information about users in a defined three-dimensional space is gathered. The defined three-dimensional space includes volumetric zones. A portion of each of the users as defined by the three-dimensional information is located in one or more of the volumetric zones. A portion of the three-dimensional information corresponds to a hand of one of the users in the defined three-dimensional space. On screen digital objects are displayed as is an indication of whether one of the users is currently capable of interacting with the displayed digital. This indication corresponds to the location of the user in one or more of the volumetric zones.

[8] A third embodiment is for yet another method for processing gesture-based user interactions with an interactive display. Three-dimensional information about an object in a defined three-dimensional space is gathered and a portion of the three-dimensional information is identified as corresponding to a hand of a user in the defined three-dimensional space. An on-screen image of a digital object is displayed and that has a visually perceptible activation state that corresponds to whether the user has exceeded an activation level threshold for the digital object. An icon is displayed for interacting

with the on-screen image of the digital object, the icon corresponding to the hand of the user in the defined three-dimensional space. The activation level corresponds to a position of the hand of the user over time.

[9] In a still further embodiment, a further method for processing gesture-based user interactions with an interactive display is disclosed. Three-dimensional information about an object in a defined three-dimensional space is gathered and a portion of the three-dimensional information is identified as corresponding to a hand of a user in the defined three-dimensional space. An on-screen image of a digital object is displayed, which has an activation level threshold. An icon is displayed for interacting with the on-screen image of the digital object, the icon corresponding to the hand of the user in the defined three-dimensional space. The icon may control an action associated with the digital object only if the activation level threshold of the displayed digital object has been surpassed. The activation level threshold of the digital object is surpassed as a result of the icon being proximate to a region associated with the digital object over time.

BRIEF DESCRIPTION OF THE DRAWINGS

- [10] FIGURE 1 illustrates an exemplary flat panel interface system.
- [11] FIGURE 2 illustrates an interactive display space including volumetric zones.
- [12] FIGURE 3 illustrates an alternative embodiment of an interactive display space including volumetric zones.
- [13] FIGURE 4 illustrates an exemplary method for the processing of gesture-based interactions.
- [14] FIGURE 5 illustrates hand / user representations vis-à-vis a selection level.
- [15] FIGURE 6 illustrates how the physical position of a user can be mapped into a display virtual world of digital objects
- [16] FIGURE 7 illustrates an alternative embodiment of how the physical position of a user can be mapped into a display virtual world of digital objects.

DETAILED DESCRIPTION

[17] FIGURE 1 illustrates an exemplary flat panel interface system 100. The system 100 includes a display 160, a three-dimensional vision system 110, a computing system 130, and an optional illuminator 120. The system 100 may be configured according to any variety of form factors.

[18] Display 160 may assume any variety of configurations including displays capable of showing video images. In one embodiment, the display 160 is a flat panel display, such as a liquid crystal display (LCD), a plasma screen, or an organic light emitting diode display (OLED) screen. The display may alternatively include, without limitation, a cathode ray tube (CRT), an electronic ink screen, a rear projection display, a front projection display, an off-axis front (or rear) projector such as the WT600 projector sold by NEC, a screen that produces a three-dimensional image such as a lenticular three-dimensional video screen, or a fog screen such as the Heliodisplay™ screen from IO2 Technology USA. Multiple screens may be tiled to form a single, larger display. The system 100 may contain more than one interactive display 160 whereby more than one user may concurrently interact with the system 100.

[19] In one configuration (like that of FIGURE 1), various elements of the system 100 may be mounted to a wall. Elements of the system 100 may alternatively be mounted to the ceiling, freestanding, mounted on a table, or in a two-sided configuration. Freestanding systems may contain a large base or may be bolted in to the ground for stability.

[20] Embodiments of the system 100 that incorporate three-dimensional objects may utilize three-dimensional hardware and software systems whereby the display 160, for example, may be cylindrical or spherical. These and other form factor configurations such as those involving the floor, a tabletop, or wall projection system may likewise utilize three-dimensional vision systems. In such embodiments, the three-dimensional vision system 110 may cover areas around the display. Projectors may be used to generate the displays, such as a floor projection, in which a projector mounted on the

ceiling casts an image on the ground, allowing the user 140 to walk on the screen area. The projection may also be cast on a table, a wall or any other surface. Multiple projectors and multiple three-dimensional vision systems may be used to create extended areas of interactivity on a floor, a wall, or another surface.

[21] Embodiments of system 100 may vary in size. For example, three-dimensional vision system 110 may be attached to the top of a laptop computer monitor thereby allowing the system 100 to be portable. The laptop itself may operate as computing system 130 and handle the three-dimensional vision software and generation of visual effects.

[22] System 100 may, as noted above, include an illuminator 120. The position of the three-dimensional vision system 110 and illuminator 120 in FIGURE 1 is illustrative. The three-dimensional vision system 110 and illuminator 120 may be arranged in a variety of locations relative to the display 160 and/or one another. Further, there may be multiple instances of three-dimensional vision system 110 and illuminator 110. The three-dimensional vision system 110 and/or illuminator 120 may use one or mirrors or other reflective surfaces (not shown) to change the position of interactive area 170.

[23] The system 100 may use a microphone or microphone array to capture audio input and may use speakers to create audio output. The microphone(s) may be directional so as to better capture audio from specific users without background noise. The speaker(s) may also be directional so as to focus audio onto specific users and specific areas. Examples of directed speakers include the Maestro™ and the SoloSphere™ from Brown Innovations in addition to those systems manufactured by Dakota Audio, Holosonics, and The American Technology Corporation of San Diego (ATCSD).

[24] Three-dimensional vision system 110 may be implemented using stereo vision. Stereo vision systems use information from two or more cameras to construct a three-dimensional image. Three-dimensional computer vision techniques using algorithms such as those based on the Marr-Poggio algorithm may take as input two or more images of the same scene taken from slightly different angles.

[25] Marr-Poggio-based algorithms are merely one example of stereo algorithms in that they find texture patches from the different cameras' images that correspond to the same part of the same physical object. The disparity between the positions of the patches in the images allows the distance from the camera to that patch to be determined thus providing three-dimensional position data for that patch.

[26] The performance of this algorithm tends to degrade, however, when dealing with objects of uniform color because uniform color makes it difficult to match corresponding patches in different images. Thus, if an illuminator 120 creates light that is textured, the texturing of an otherwise uniform color can improve the distance estimates of some three-dimensional computer vision algorithms.

[27] Exemplary stereo cameras that may be used in the three-dimensional vision system 110 include the Tyzix DeepSea™ and Point Grey Bumblebee™. These cameras may be monochrome or color, and may be sensitive to one or more specific bands of the electromagnetic spectrum including visible light, near-infrared, far infrared, and ultraviolet.

[28] The three-dimensional vision system 110 may also be implemented using time-of-flight cameras. A time-of-flight camera detects distance information for each pixel of an image. The distance information for each pixel may include the distance from the camera to the object seen at that pixel. Time-of-flight cameras obtain this distance data by timing how long it takes an invisible light pulse to travel from a light source that may be located next to the camera to the object being imaged and then back to the camera. The light pulses may be rapidly repeated thereby allowing the time-of-flight camera to have a frame rate similar to that of a video camera. While a time-of-flight camera typically has a range of 1-2 meters at 30 frames per second, the range can be increased by lengthening the exposure time, which, in turn, lowers the frame rate. Manufacturers of time-of-flight cameras include Canesta Inc. of Sunnyvale, California, and 3DV Systems of Israel.

[29] Embodiments of the three-dimensional vision system 110 may also be implemented using laser rangefinders, cameras paired with structured light projectors,

laser scanners, laser line scanners, ultrasonic imagers, systems that derive three-dimensional data through the intersection of foreground images from several cameras, and/or any combination of the above. Multiple three-dimensional vision systems 110 may be simultaneously used whereby their three-dimensional data is merged together to create a unified data set describing objects in the interactive area.

[30] While background illumination may be sufficient in some embodiments of the system 100, using an illuminator 120 may be necessary to improve the texture data to be processed through execution of a stereo algorithm by a processing device (*e.g.*, in three-dimensional vision system 110 or at computing device 130). By lighting objects in the interactive area with a pattern of light, the illuminator 120 improves the amount of texture data that may be used by the stereo algorithm to match patches. The illuminator 120 may create a patterned and/or uneven light pattern to improve the detection of objects using a stereo algorithm. In some instances, the illuminator 120 may simply provide additional illumination of users 140 interacting with display 160.

[31] Illuminator 120 may include concentrated light sources such as high power light-emitting diodes (LEDs), incandescent bulbs, halogen bulbs, metal halide bulbs, and arc lamps. More than one concentrated light source may be used simultaneously, either using the same scattering system as the first light source or a separate one. It should be noted that a substantially collimated light source, such as a lamp with a parabolic reflector or one or more narrow angle LEDs, may be used in place of a concentrated light source to achieve similar results.

[32] A variety of patterns may be used by illuminator 120 such as a random dot pattern, a fractal noise pattern that provides noise on varying length scales, and a pattern comprising a set of parallel lines. The parallel lines may be randomly and varied with respect to their distancing to provide valuable texturing. A variety of methods may be utilized to generate such patterns. For example, the illuminator 120 may include a video projectors designed to project any image that is provided via a video input cable. The image may change over time to optimize the performance of the three-dimensional vision system 110. For example, the pattern may dim in an area where the three-

dimensional vision system's 110 camera images are becoming saturated with light or become higher resolution in areas where the object is close to the camera. More than one video projector may be used to cover the whole interactive area.

[33] The illuminator 120 may also use a structured light projector to cast out static or time-varying patterns. Examples of structured light projectors include the LCD-640™ and the MiniRot-H1™ from ABW. In another embodiment, the pattern may be generated by a slide projector including a light source, a transparent slide having an image, and a lens system or curved mirror system to project the image of the slide. More than one slide projector may be used to cover the whole of the interactive area. In another embodiment, the pattern is generated by a highly concentrated light source (a light that emanates from a small area). Alternately, a reflective or refractive system may be used.

[34] The illuminator 120 may use an invisible or minimally visible light source, for example, a near-infrared light source, so as to avoid creating a visible pattern on the ground that may be noticed by users. Alternately, the illuminator 120 may have a light source that emits both visible and invisible light and may include a filter to reduce the visible light, leaving the invisible light. The illuminator 120 and cameras may be strobed through pulsing of the illuminator 120 during a synchronized camera exposure whereby the effect of ambient light is reduced or eliminated. The three-dimensional vision system 110 may contain optical filters that suppress light from wavelengths not emitted by the illuminator 120 thus allowing light from the illuminator 120 to be featured more prominently in images captured by the vision system 110.

[35] Vision software may be executed at three-dimensional vision system 110 or at a computing device 130 coupled to the vision system 110. Execution of the vision software may improve the usefulness of captured image data. For example, where multiple three-dimensional cameras are used, vision software may aid in the calibration of the cameras such that the data from the cameras can be merged into the same coordinate space (*i.e.*, the two cameras may cover a similar area from two different angles to provide coverage of the area from multiple angles). Such cameras may cover

different areas where one camera may focus on close-up interaction while the other camera focuses on distant interaction. The quality and resolution of stereo processing may also be varied and effectuated through the execution of vision software. For example, the area closest to the display 160 may be processed at a higher resolution in order to resolve a user's 140 individual fingers 150 near the display 160 with more accuracy.

[36] Several methods may be used to remove inaccuracies and noise in the three-dimensional data. For example, background methods may be used to mask out three-dimensional data from areas of the camera's field of view that are known to have not moved for a particular period of time. These background methods (also known as background subtraction methods) may be adaptive, allowing the background methods to adjust to changes in the background over time. These background methods may use luminance, chrominance, and/or distance data from the cameras in order to form the background and determine foreground. Once the foreground is determined, three-dimensional data gathered from outside the foreground region may be removed.

[37] In a further embodiment, a color camera may be used to obtain chrominance data for the three-dimensional data of the user 140 and other objects in front of the display 160. This chrominance data may be used to acquire a color three-dimensional representation of the user 140, which allows their likeness to be recognized and tracked. The likeness of the recognized and tracked user 140 may subsequently be displayed.

[38] Noise filtering may be applied to either a depth image, which is the distance from the camera to each pixel of the camera's image from the camera's point of view, or directly to the three-dimensional data gathered by the vision system 110. Smoothing and averaging techniques such as median filtering may be applied to the camera's depth image in order to reduce depth inaccuracies. Isolated points or small clusters of points may be removed from the three-dimensional data set if they do not correspond to a larger shape thus eliminating noise while leaving users intact.

[39] The three-dimensional data may be analyzed in a variety of ways to produce high level information. For example, a user's fingertips, fingers, and hands may be

detected (collectively represented as element 150 in FIGURE 1). Methods for doing so include various shape recognition and object recognition algorithms. Objects may be segmented using any combination of two-dimensional and/or three-dimensional spatial, temporal, chrominance, or luminance data. Furthermore, objects may be segmented under various linear or non-linear transformations of the aforementioned domains. Examples of object detection algorithms include, but are not limited to deformable template matching, Hough transforms, and the aggregation of spatially contiguous pixels and/or voxels in an appropriately transformed space.

[40] As another example, the three-dimensional points belonging to a user 140 may be clustered and labeled such that the cluster of points belonging to the user is identified. Various body parts, such as the head and arms (190) of a user 140 may be segmented as markers. Points may also be also clustered in three-dimensional space using unsupervised methods such as k-means or hierarchical clustering. The identified clusters may then enter a feature extraction and classification engine. Feature extraction and classification routines are not limited to three-dimensional spatial data but may also be applied to previous feature extractions or classifications in other data domains such as two-dimensional spatial, luminance, chrominance, or transformations thereof.

[41] A skeletal model may also be mapped to the three-dimensional points belonging to a given user 140 via a variety of methods including but not limited to expectation maximization, gradient descent, particle filtering, and feature tracking. In addition, face recognition algorithms, such as *eigenface* or *fisherface* may use data from the vision system such as two-dimensional and/or three-dimensional spatial, temporal, chrominance, and luminance data in order to identify users and their facial expressions. Facial recognition algorithms used may be image based or video based. This information may be used to identify users, especially in situations where they leave and return to the interactive area as well as change interactions with displayed content based on face, gender, identity, race, facial expression, or other characteristics.

[42] Fingertips or other body parts may be tracked over time in order to recognize specific gestures such as pushing, grabbing, dragging, dropping, poking, and/or drawing of shapes using a finger, pinching, and other such movements.

[43] The three-dimensional vision system 110 may be specially configured to detect specific objects other than the user 140. This detection may utilize object recognition algorithms executable to recognize specific aspects of the appearance or shape of the object, radio frequency identification (RFID) tags in the object read by an RFID reader (not shown) to provide identifying information, and/or a light source strobed in a particular pattern on an object to provide identifying information.

[44] The user(s) 140 or other objects detected by the three-dimensional vision system 110 may be represented on the display 160. This representation 162 of the user 140 on the display 160 may be useful in allowing the user to interact with virtual, digital objects 164 (which may be referred to interchangeably as virtual or digital objects) shown on the display 160. Interaction may be facilitated by giving a visual indication of user position (162) relative to the digital objects 164.

[45] This representation may include a digital shadow of the user(s) or other objects. For example, a two-dimensional shape that represents a projection of the three-dimensional data corresponding to the user's body onto a flat surface. The two-dimensional shape may approximate that of the user as well as with respect to size. Depending on a particular embodiment of the present invention, the representation of the user (or portions thereof) may or may not interact with digital objects 164 on the display 160.

[46] A digital outline of the user(s) or other objects may also be used to represent the user 162 on display 160 with respect to interacting with digital objects 164. The outline can be thought of as the edges of the digital shadow. This outline, similar to the digital shadow, may approximate the shape and size of the user. This shape may be colored, highlighted, rendered, or otherwise processed arbitrarily before display.

[47] Various other images, icons, or two-dimensional or three-dimensional renderings representing the users' hands (150) or other body parts (190) or other objects may be

rendered on display 160 for the purpose of interacting with digital objects 164. For example, the user's hand 150 may be represented on the display 160 as an approximation of the user's actual hand or as a 'hand-like' icon corresponding to the same.

[48] Other interaction and representation methodologies include:

[49] The shape of the user(s) rendered in the virtual space may be combined with markers on the user's hands 150 that are displayed when the hands are in a position to interact with on-screen objects. For example, the markers on the hands may only show up when the hands are pointed at the screen.

[50] Points that represent the user(s) or other objects from the point cloud of three-dimensional data from the vision system 110 may be displayed as objects, which may be small and/or semitransparent.

[51] Cursors representing the position of users' fingers (150) may be used. These cursors may be displayed or change appearance when the finger 150 of a user 140 is capable of a specific type of interaction in the virtual space.

[52] Objects that move along with and/or are attached to various parts (190) of the users' bodies may be utilized. For example, a user 140 may have a helmet that moves and rotates with the movement and rotation of the user's head 190.

[53] Digital avatars that match the body position of the user(s) 140 or other objects as they move. In one embodiment, the digital avatars are mapped to a skeletal model of the users' positions.

[54] Any combination of the aforementioned representations.

[55] In some embodiments, the representation may change appearance based on the users' allowed forms of interactions with on-screen objects. For example, a user 140 may be shown as a gray shadow and not be able to interact with objects until they come within a certain distance of the display 160, at which point their shadow changes color and they can begin to interact with on-screen objects.

[56] Given the large number of potential features that can be extracted and gathered by the three-dimensional vision system 110 and the variety of virtual objects 164 that can

be displayed on the screen, there are a large number of potential interactions between the users 140 and the virtual objects 164. Some examples of potential interactions include two-dimensional force-based interactions and influence image based interactions that can be extended to third dimension. Thus, three-dimensional data about the position of a user could be used to generate a three-dimensional influence image to affect the motion of a three-dimensional object. These interactions, in both two- and three-dimensions, allow the strength and direction of the force the user imparts on a virtual object 164 to be computed thereby giving the user control over how they impact the object's motion on the display 160.

[57] Users 140 (or their displayed representation 162) may interact with digital objects 164 by intersecting with them in virtual space. This intersection may be calculated in three-dimensions or three-dimensional data from the user may be projected down to two-dimensions and calculated as a two-dimensional intersection.

[58] Visual effects may be generated based on the three-dimensional data from the user. For example, a glow, warping, emission of particles, flame trail, or other visual effects may be generated using the three-dimensional position data or some portion thereof. Visual effects may be based on the position of specific body parts. For example, a user could create virtual fireballs by bringing their hands together. Users 140 may use specific gestures to pick up, drop, move, rotate, or otherwise modify virtual, digital objects 164 displayed on-screen.

[59] The virtual space depicted on the display 160 may be shown in two- or three-dimensions. In either case, the system 100 merges information about the user 140 with information about the digital objects 164 and other images (*e.g.*, user representation 162) displayed in the virtual space of the display 160. If the user 140 is depicted two-dimensionally in the virtual space, then the three-dimensional data about the user's position may be projected onto a two-dimensional plane.

[60] The mapping between the physical space 180 in front of the display 160 and the virtual space shown on the display 160 can be arbitrarily defined and can even change over time. The actual scene seen by the users 140 on the display 160 may vary based on

the definitions of the physical space 180. Similarly, the virtual space (or just the user's representation 162) may be two-dimensional. In this case, the depth component of the user's virtual representation 162 may be ignored.

[61] Mapping may operate in a manner similar to that of a mirror such that the motions of the user's representation 162 in the virtual space as seen by the user 140 are akin to a mirror image of the user's 140 actual motions. The mapping may be calibrated such that when the user 140 touches or brings a part of their body (150) near to the screen display 160, their virtual representation 162 touches or brings the same part of their body near to the same part of the screen 160. The mapping may also show the user's representation 162 appearing to recede from the surface of the screen 162 as the user 140 approaches the screen 160.

[62] Multiple interactive areas may be created by partitioning the space in front of the vision system into multiple volumetric spaces and sorting the gathered data into these spaces. FIGURE 2 illustrates an interactive display space including volumetric zones. Display space includes vision system 220 and display 210.

[63] In FIGURE 2, a "touch" space 250 primarily captures finger pointing interactions, a "primary users" space 240 captures the bodies of the people nearest the screen, and a "distant users" space 230 captures the bodies of the people who are more distant. The user representation on the screen may be different depending on what region in which they are present. For example, fingers in the "touch" space 250 may be represented by cursors, bodies 260 in the "primary users" space 240 may be represented by colored outlines, and bodies 270 in the "distant users" area 230 may be represented by gray outlines. The boundaries of the spaces may change. For example, if there is no one in the "primary users" space 250, the boundary of that space may expand to cover the "distant users" space 270.

[64] Users who are beyond a certain distance from the screen or in a "distant users" space 270 may have their ability to interact with on-screen objects reduced or eliminated. This allows users close to the screen to interact with virtual objects without interference from more distant users. In some instances, a user may be too close to the display 210

and because of that proximity, likewise unable to interact with objects on the screen.

Various indicia may be represented on the display 210 indicating the present locale of a user and/or their ability to interact with objects on the display 210.

[65] Providing indicia concerning the ability for a present interaction may be related to one or more reasons. For example, the display may be out of reach or a particular situation may call for large numbers of users interacting simultaneously and to discourage any one user from coming up close and dominating the interaction. Other situations may include those where the display is behind a window and close-up interaction is impossible. In this case, a "too close" interaction space may be created.

[66] In some embodiments, various effects may be displayed or audio emissions made in order to attract a user outside of the aforementioned volumetric areas into coming into the same. In one embodiment, the display shows a very large, highly visible flash or other reaction when no one is currently in the interactive space and a first user moves in to the interactive space. This feature is designed to draw a user's attention to the display when no one is currently interacting and the user passing into the interactive space may only see the display in their peripheral vision.

[67] As addressed previously, the vision systems of the present invention may detect the hands and bodies of users. Hand and body detection may use the range, luminance, color, and/or other data. Hand and body detection methods may be applied to the raw color/luminance image, range image, three-dimensional point cloud, and/or other data. Hand and body detection methods may also be applied to any transformed or projected data based on the aforementioned data sets.

[68] Bodies may be detected as spatially separated clusters of points in the three-dimensional cloud. Bodies may also be detected by projecting the point cloud onto the floor plane, and identifying regions of high density. Bodies may also be identified by looking for large connected components in the projected depth image after the removal of quickly varying regions of depth. Bodies may also be identified by performing template matching or skeletal matching on either the raw three-dimensional data or on a suitable projection. Preliminary identification may occur in one projection and then

further refined in a separate projection. For instance, the vision system may use a floor plan view to pick out rough bounding boxes for users bodies. A frontal projection of the points within the bounding box may then be used to separate very close users.

[69] Hands may be detected as isolated objects or as part of the user. In the former case, hands may be associated with a user. The position and motion of detected hands may be defined in relation to a user, the display, or some region in space either predefined or dynamically created in the course of interaction. Many possible coordinate systems may be used to specify this relationship, depending on the desired effect. For example, a polar coordinate system centered on the shoulder of the user from which the hand comes may be used. Hand detection methods may include identification through a skeletal fitting of the body; identification through template matching; identification through color based skin segmentation; identification as a body relative object; and identification based on connected component properties.

[70] Once users' bodies have been identified, a bounding region such as a box, sphere or cylinder, may be centered at the user. Objects connected to the body, but outside of this region may become candidates for hands. These connected objects may be further segmented to identify the 'tip' or hand. Hands may be identified by taking the extreme regions of these objects such as the most distant point from the bounding box or some point on the body as could be defined, and any point that is within some threshold value below the maximum distance could be considered part of a hand. The hand may also be extracted from these regions through template matching. Hands may also be segmented from this region based on various features including, but not limited to curvature, distance from bounding box or body, and color properties. A skeleton model of an arm may also be fit to the region using expectation maximization.

[71] FIGURE 3 illustrates an alternative embodiment of an interactive display space including volumetric zones. In FIGURE 3, interactive display 310 and camera/illuminator system 320 create an interactive space consisting of "too close" space 350, "hands" space 340, and "distant users" space 330. Users in "distant users" space 330 can only interact with their whole bodies and may see full-body user representations

of themselves on the display. Users in “hands” space 340 can interact with their hands as well as their bodies and may see hand representations as well as full-body representations on the display.

[72] Users who put their hand in “too close” space 350 may not be able to interact with either their hands or bodies and a visual on-screen representation such as a disappearing puff of smoke at the position of their hands or a fading image of their bodies may clue them in that they are getting too close to interact. For example, the user 360 may see the visual representation of their hands turn into a puff of smoke and may lose the ability to interact with on-screen objects using their hands. In contrast, user 370, who is in the “hands” space 340, has full ability to interact with on-screen objects using their hands and sees a different representation of their hands than user 360. The inner edge of the “too close” space may coincide with the boundary of the camera’s field of view or the illuminator’s field of illumination.

[73] The display may display a representation of the user’s body or specific body parts in order to provide visual feedback to the user as to their location within the virtual space shown on the display, and facilitate accurate interaction with the objects shown on the display. The display may also display a representation of a user’s hand. Since many interactions between the user and the display may be via a user’s hand, it may be advantageous to specifically highlight the hands of users. Hand representations may only be shown when the hand is sufficiently away from the body to convey some kind of gestural use. The representation used for the hand may depend on the position of the hand or how it is moving. It may also reflect whether the hand is capable of selecting or interacting with objects. It may be an anthropomorphic hand, a different type of object, or a combination of the two.

[74] In one embodiment, the hand representation may only appear when the user, or their hand, is within a specific space in front of the display. Alternatively, the user’s full-body representation fades away or transforms when a user’s hand representation begins to be displayed and returns when the user’s hand representation is removed.

[75] FIGURE 4 illustrates an exemplary method for the processing of gesture-based interactions. In step 410, a camera capable of gathering real time three-dimensional data about its environment, such as a stereo or time-of-flight camera, captures the motion of one or more human users or other objects. In step 420, the data from the three-dimensional camera, which may take the form of a three-dimensional point cloud, is analyzed by a computer vision system, and potentially coordinate-transformed to determine the position of the data relative to the display and virtual objects shown on the display. In step 430, a computer vision system analyzes this data to determine the location of users' bodies and hands, and potentially other body parts. These body parts may be grouped by person, and their positions may be tracked over time to determine velocity, acceleration, and other characteristics.

[76] In step 440, the degree to which each hand is making an action associated with selection of a virtual object, or some other interaction with a virtual object, is determined. In step 450, a visual representation of each hand, which may be different depending on its level of selection or other action, is determined. These visual representations are then rendered on the display in step 480, along with a representation of the users' bodies, which is created in step 470, and the virtual objects that the users have been interacting with, which are controlled by simulation 490. By rendering the representations on the display, the users can receive visual feedback about the position of their bodies and hands relative to virtual objects on the screen.

[77] Separately, the selection level of each hand, which is determined by step 450, is processed in step 460 to determine whether the hands have interacted with any virtual objects from simulation 490. The user interaction process shown in FIGURE 4 can repeatedly happen so quickly that it generates the impression of real-time interaction.

[78] The process shown in FIGURE 4 illustrates just one possible embodiment. Alternative embodiments may be implemented; for example, objects in simulation 490 may analyze the vision signals around them in order to determine if there is a nearby hand or selection behavior occurring thereby bypassing the need for steps 440 and 430.

[79] FIGURE 5 illustrates hand / user representations vis-à-vis a selection level. In step 510, the degree to which the hand is selecting an object is determined. In the illustrated embodiment, a simple distance along the Z axis from the torso to the hand is determined. As the user extends their arm forward, this value increases. In step 520, a visual representation for the hand on the display is determined based on the distance value computed in step 510. The hand representation gradually closes from an open hand to a single pointing finger as the distance increases. This gives users feedback as to their level of partial selection—how far their hands are from reaching the threshold at which they gain the ability to select objects. When the hand has crossed the threshold into selection mode, where it has the ability to select objects, a crosshairs appears. If no hand is raised, no hand representation is shown. Separately, the distance computed in step 510 serves as an “activation level” in step 530, where the hand becomes able to select objects if the activation level exceeds, for example, 0.5. The example shown in FIGURE 5 is merely illustrative as one possible embodiment. Numerous other selection methods can be used to compute an activation level.

[80] In one embodiment, users’ hands or bodies can interact with virtual objects via a two-dimensional or three-dimensional simulation of physical collisions. In one embodiment, the position of each hand in the virtual space shown on the display is used as input to the physics simulation that governs the behavior of the virtual objects. The hand may be modeled with a specific shape that can be either predefined, such as a sphere of specific radius, or derived from the two-dimensional or three-dimensional shape of the user’s hand itself. In one embodiment, the hand’s visual representation may serve as the model by the physics simulation.

[81] The model of the hand in the physics simulation could be used to detect overlap and collisions with virtual objects shown on the display. The position of the hand over time can be used to derive velocity and acceleration, which would allow realistic collisions to be computed with virtual objects on the display. Collisions may be inelastic, partially elastic, or fully elastic, and may be computed using standard collision formulas known to those skilled in the art. Since the hand cannot physically be moved

by the virtual object, the hand may be treated as if it has infinite mass for the purpose of calculating post-collision velocities.

[82] Other body parts, such as arms, heads, or the entire body, may be input to the physics simulation in a similar fashion. The motion information about the hands or other body parts may be used to compute interactions other than collisions, such as fluid or ripple simulations, or particle emissions that are dependent on velocity or acceleration.

[83] A virtual object attached to the hand or another body part may also be used to interact with virtual objects. For example, a virtual sword may be attached to the user's hand. Orientation of the object may be computed by analyzing the orientation of the user's hand, forearm, or arm, allowing the user to easily control the object's orientation. This object may be analyzed by the physics engine to determine interactions with other objects.

[84] FIGURE 6 illustrates how the physical position of a user can be mapped into a display virtual world of digital objects. In FIGURE 6, user 620 is interacting in front of display 630. User 620 is within interactive space 610, which is the space within which their interactions can affect the display. Axes 660 define a coordinate system. The X and Y axes define horizontal and vertical position on the display 630, while the Z axis defines a perpendicular distance from the display 630. In this interactive space 610, user 620 is raising their hand 670 to interact with virtual object 640 displayed on the screen 630.

[85] In one embodiment, overlap or proximity between hand 670 and virtual object 640 is computed by projecting the hand's position on the Z axis onto the (X,Y) plane of the display surface and then analyzing the relative (X,Y) positions of the hand 670 and virtual object 640. In addition, a representation 650 of hand 640 is displayed on display 630 at the position computed by projecting the hand's position along the Z axis. This representation provides user 620 with visual feedback as to the location of their hand 670 on the virtual world of display 630.

[86] FIGURE 7 illustrates an alternative embodiment of how the physical position of a user can be mapped into a display virtual world of digital objects. In one embodiment,

display 730 displays images on a two-dimensional surface and depth cues such as shadows or mist are used to convey depth. In another embodiment, display 730 is a display capable of producing an image that appears three-dimensional to user 720, giving the illusion that displayed images are at a particular distance in front of or behind the screen. Numerous three-dimensional display mechanisms exist and are commercially available, including, but not limited to, stereoscopic projections and lenticular video screens.

[87] User 720 interacts in interactive space 710. When user 710 reaches their hand 770 out, a virtual representation 750 of their hand may be displayed by display 730. As the user's hand 770 enters within touch space 780 around virtual object 740 displayed by display 730, an interaction occurs. Coordinate axes 760 are illustrated for reference.

[88] There are many potential methods that the system can use to select an on-screen object. It is important to minimize both false negatives and false positives thereby preventing the user from accidentally selecting objects that they do not want to select, and being unable to select objects they want to select. The information used to compute these user behaviors may use various inputs from the vision system, including but not limited to, current and past hand position and velocity, arm angle, current and past head position, and current and past torso position.

[89] Selection can occur by holding a hand still over an object or region to be selected. The act of holding a hand still can be detected in several ways. For example, activation level may be tracked over time, with the object or region being selected if the activation level passes a threshold. Partial selection may be shown based on activation level. The hand position over time may contribute to the activation level. For example, the activation level may be decreased by an amount determined by a formula if the hand is moving quickly over the object or region or is outside the object or region, while the activation level may be increased by an amount determined by a formula if the hand is over the object or region and is moving slowly or stopped. The formula may take the velocity of the hand's motion into account such that the slower the hand moves, the faster the activation level will rise.

[90] The average speed of the hand over a specified range of time below a specific value may also operate as a selection mechanism. Speed may be defined over a two-dimensional plane defining the interactive zone on the display, over the three-dimensional space in front of the display, or some other method; for example, the X and Y axes shown in FIGURED 6 and 7.

[91] Other selection mechanisms include: the maximum speed of the hand over a specified range of time being below a specific value; the position of the hand in some two-dimensional (e.g., the X and Y axes shown in FIGURES 6 and 7) or three-dimensional coordinate system stays within a specified region for a specified period of time; the variance of the position of the hand in some two-dimensional (e.g., the X and Y axes shown in FIGURES 6 and 7) or three-dimensional coordinate system stays below a specific value for a specified period of time; the position data of the hand is averaged over a time interval, then the averaged position data is examined over a second time interval using any of the techniques described herein or another technique.

[92] Selection can also occur by thrusting the hand toward the screen or virtual object in a poking motion. This act of poking can be detected in a variety of ways. For example: the distance of the hand from the screen decreases by more than a threshold amount within a specific period of time; the speed of the hand relative to the screen along the Z axis exceeds a particular value, and the hand is moving toward the screen; the time-averaged speed of the hand relative to the screen along the Z axis exceeds a particular value for a specified period of time, and the hand is moving towards the screen; any of the above examples, with the additional criterion that the X and Y speed (or time-averaged X and Y speed) of the hand remain below threshold values for a specified period of time. This prevents accidental selection when the users swing their hands; the velocity of the user is in a direction within a specified range of directions, and the speed exceeds a given amount. In one such embodiment, this range of directions consists of directions within a specified angle of a vector pointing directly along the Z axis toward the screen.

[93] Selection may also occur based on the distance of the user's hand from the screen or an arbitrary three-dimensional region. Thus, when the user extends their hand into a "selection zone," the object can be selected. Selection can also occur based on the position of the user's hand relative to their body. This act can be detected in a variety of ways such as a user's hand is greater than a threshold distance from that user's head, torso, shoulder, or center of gravity; the ratio of the user's height to the distance from the user's hand to that user's head, torso, shoulder, center of gravity goes below a certain amount, which ensures that shorter people and taller people can make similar movements to select an object; distance may be defined in several ways; for example, it may consist of distance along the Z axis (toward the screen), or in three-dimensional space (X,Y,Z), or in a weighted coordinate system.

[94] Selection can occur based on the user waving their hand rapidly over a specified location. Rapid waving can be detected by the average speed of the user's hand exceeds a threshold amount over a particular time interval but the position remains within a specific region. Selection can also occur based on the average speed of the user's hand exceeding a threshold amount over a particular time interval, but the variance of the position remains below a specific amount.

[95] Selection may also occur based on the execution of a specific gesture.

[96] Because different people may intuitively use different methods to select an object, it may be advantageous to allow multiple methods of selection. Thus, for example, a user may select an object if they hold their hand still over it for more than a specific period of time, or if they make a rapid poking motion at it.

[97] In order to prevent accidental selection, multiple methods of selection may be required to be invoked simultaneously to trigger selection. For example, a user may need to hold their hand still for a specified period of time and keep their hand at least a minimum distance from their body.

[98] Feedback may be used in the selection process to indicate the fact that a selection is taking place. In one embodiment, the on-screen representation for the user's hand changes when the hand is selecting an object. In another embodiment, the on-screen

representation for the user's hand changes when the hand would be selecting an object based on its position or movement, even if no object is present in that area. In a still further embodiment, the on-screen object changes as a user selects, or has partially selected, the object.

[99] The on-screen representation of the hand may alternatively begin to change as it starts to select an object. This provides visual feedback to the user about the nature of the selection process, and helps the user prevent accidental selections. In one embodiment, the change in user representation is chosen from a sequence of images based on a particular value of the selection detection algorithm. For example, if the selection detection algorithm is based on the distance of the hand from the user's body, the hand can gradually change appearance along an animation sequence, where the frame from the animation sequence is chosen using the distance value.

[100] Multiple users can simultaneously interact with the display. The camera system is capable of sensing multiple bodies, and the vision system may be capable of disambiguating data from multiple users' bodies. This disambiguation enables the vision system to assign each hand to a specific user's body. As a result, different bodies and hands can have different user representations as well as different interactions with the virtual objects displayed on the display. In addition, the display can ignore specific users, or allow one user or a group of users to control a specific set of interactions.

[101] The above-described components and functions can be comprised of instructions that are stored on a computer-readable storage medium. The instructions can be retrieved and executed by a processor. Some examples of instructions are software, program code, and firmware. Some examples of storage medium are memory devices, tape, disks, integrated circuits, and servers. The instructions are operational when executed by the processor to direct the processor to operate in accord with the invention. Those skilled in the art are familiar with instructions, processor(s), and storage medium.

[102] While various embodiments have been described above, it should be understood that they have been presented by way of example only, and not limitation. For example, any of the elements associated with the user interface may employ any of the desired

functionality set forth hereinabove. Thus, the breadth and scope of a preferred embodiment should not be limited by any of the above-described exemplary embodiments.

[103] For example, information from one or more display systems may be shared over a high speed data connection allowing for a variety of applications. For example, users of the display may be able to navigate and interact with an online virtual world such as Second Life, There, or World of Warcraft, and objects and avatars within that world using a gesture-based interface. In addition, the three-dimensional data of the position and movement of the users can be transmitted and displayed in the online virtual world thereby allowing other users of the online virtual world to see and interact with the users of the display. These other users may be using a variety of interfaces (e.g., a traditional keyboard-and-mouse interface) or another display to interact in the virtual world. The users may be near to or distant from each other.

[104] The interface of the presently described embodiments may provide for a variety of potential uses. For example, with respect to sports, users may box, play tennis (with a virtual racket), throw virtual balls, or engage in other sports activity with a computer or human opponent shown on the screen. In the context of virtual world navigation, users may use natural body motions such as leaning to move around a virtual world, and use their hands to interact with objects in the virtual world. Virtual characters may be controlled through the aforementioned interface system where a digital character on the screen may talk, play, and otherwise interact with people in front of the display as they pass by it. This digital character may be computer controlled or may be controlled by a human being at a remote location.

[105] Embodiments of the present invention allow for new implementations of advertising including interactive product demos and interactive brand experiences as well as multiuser workspaces where groups of users can move and manipulate data represented on the screen in a collaborative manner. Video games may also implement the aforementioned interface system such that users can play games and control their onscreen characters via gestures and natural body movements. Modeling and 'trying

on' of clothes may also be utilized through this system where clothes are placed on the image of the user on the display thus allowing them to virtually try on clothes.

CLAIMS

WHAT IS CLAIMED IS:

1. A method for processing gesture-based user interactions with an interactive display, the method comprising:

gathering three-dimensional information about an object in a defined three-dimensional space;

identifying a portion of the three-dimensional information as corresponding to a hand of a user in the defined three-dimensional space;

displaying an on-screen image of a digital object;

displaying an on-screen image associated with and approximating the physical shape of the user, wherein the displayed on-screen image approximating that of the user does not interact with the displayed digital object; and

displaying an icon corresponding to the hand of the user in the defined three-dimensional space, wherein the displayed icon is capable of interacting with the displayed digital object.

2. A method for processing gesture-based user interactions with an interactive display, the method comprising:

gathering three-dimensional information about a plurality of users in a defined three-dimensional space, wherein the defined three-dimensional space includes a plurality of volumetric zones, wherein a portion of each of the plurality of users as defined by the three-dimensional information is located in one or more of the plurality of volumetric zones;

identifying a portion of the three-dimensional information as corresponding to a hand of one of the plurality of users in the defined three-dimensional space;

displaying on-screen digital objects; and

displaying an indication of whether one of the plurality of users is currently capable of interacting with the displayed digital objects based on the location of the one of the plurality of users in one or more of the plurality of volumetric zones.

3. The method of claim 2, wherein the plurality of volumetric zones includes a volumetric zone furthest from the interactive display relative to the other volumetric zones.

4. The method of claim 3, further comprising displaying an on-screen image that does not interact with the displayed digital objects and that is associated with and approximating the physical shape of one of the plurality of users located in the volumetric zone furthest from the interactive display relative to the other volumetric zones.

5. The method of claim 4, wherein the display of the on-screen image of one of the plurality of users in the volumetric zone furthest from the interactive display relative to the other volumetric zones differs from the display of an on-screen image associated with and approximating the physical shape of one of the plurality of users located in a volumetric zone for interacting with the displayed on-screen digital objects.

6. The method of claim 2, wherein the plurality of volumetric zones includes a volumetric zone closest to the interactive display relative to the other volumetric zones.
7. The method of claim 6, wherein the indication of whether one of the plurality of users is currently capable of interacting with the displayed digital objects corresponds to the proximity of the one or more of the plurality of volumetric zones with respect to the interactive display.
8. The method of claim 2, wherein the plurality of volumetric zones includes a volumetric zone for interacting with the displayed digital objects, and an icon corresponding to the hand of the one of the plurality of users is displayed only when the corresponding hand of the one of the plurality of users is in the volumetric zone for interacting with the display objects.
9. A method for processing gesture-based user interactions with an interactive display, the method comprising:
 - gathering three-dimensional information about an object in a defined three-dimensional space;
 - identifying a portion of the three-dimensional information as corresponding to a hand of a user in the defined three-dimensional space;
 - displaying an on-screen image of a digital object, the digital object having a visually perceptible activation state corresponding to whether the user has exceeded an activation level threshold for the digital object; and
 - displaying an icon for interacting with the on-screen image of the digital object, the icon corresponding to the hand of the user in the defined three-dimensional space, and wherein the activation level corresponds to a position of the hand of the user over time.

10. The method of claim 9, wherein the visually perceptible activation state of the digital object indicates if the activation level threshold for the digital object has been surpassed.
11. The method of claim 10, wherein the visually perceptible activation state of the digital object is indicated by the icon corresponding to the hand of the user.
12. The method of claim 10, wherein the visually perceptible activation state of the digital object is indicated by the on-screen image of the digital object.
13. The method of claim 10, wherein the visually perceptible activation state of the displayed digital object is selected from the group consisting of not active, partially active, and active.
14. A method for processing gesture-based user interactions with an interactive display, the method comprising:
 - gathering three-dimensional information about an object in a defined three-dimensional space;
 - identifying a portion of the three-dimensional information as corresponding to a hand of a user in the defined three-dimensional space;
 - displaying an on-screen image of a digital object, the digital object having a visually perceptible activation state corresponding to whether the user has exceeded an activation level threshold for the digital object; and
 - displaying an icon for interacting with the on-screen image of the digital object, the icon corresponding to the hand of the user in the defined three-dimensional space, and wherein the activation level threshold corresponds to a position of the displayed icon corresponding to the hand of the user versus the displayed digital object over time.

15. A method for processing gesture-based user interactions with an interactive display, the method comprising:

gathering three-dimensional information about an object in a defined three-dimensional space;

identifying a portion of the three-dimensional information as corresponding to a hand of a user in the defined three-dimensional space;

displaying an on-screen image of a digital object, the digital object having an activation level threshold; and

displaying an icon for interacting with the on-screen image of the digital object, the icon corresponding to the hand of the user in the defined three-dimensional space, and wherein the icon may control an action associated with the digital object only if the activation level threshold of the displayed digital object has been surpassed, and wherein the activation level threshold of the digital object is surpassed as a result of the icon being proximate to a region associated with the digital object over time.

16. The method of claim 15, wherein the time to surpass the activation level threshold of digital object corresponds to the amount of motion of the icon with respect to the region associated with the digital object.

17. The method of claim 15, wherein the time to surpass the activation level threshold of the digital object is inversely proportional to the amount of motion of the icon with respect to the region associated with the digital object.

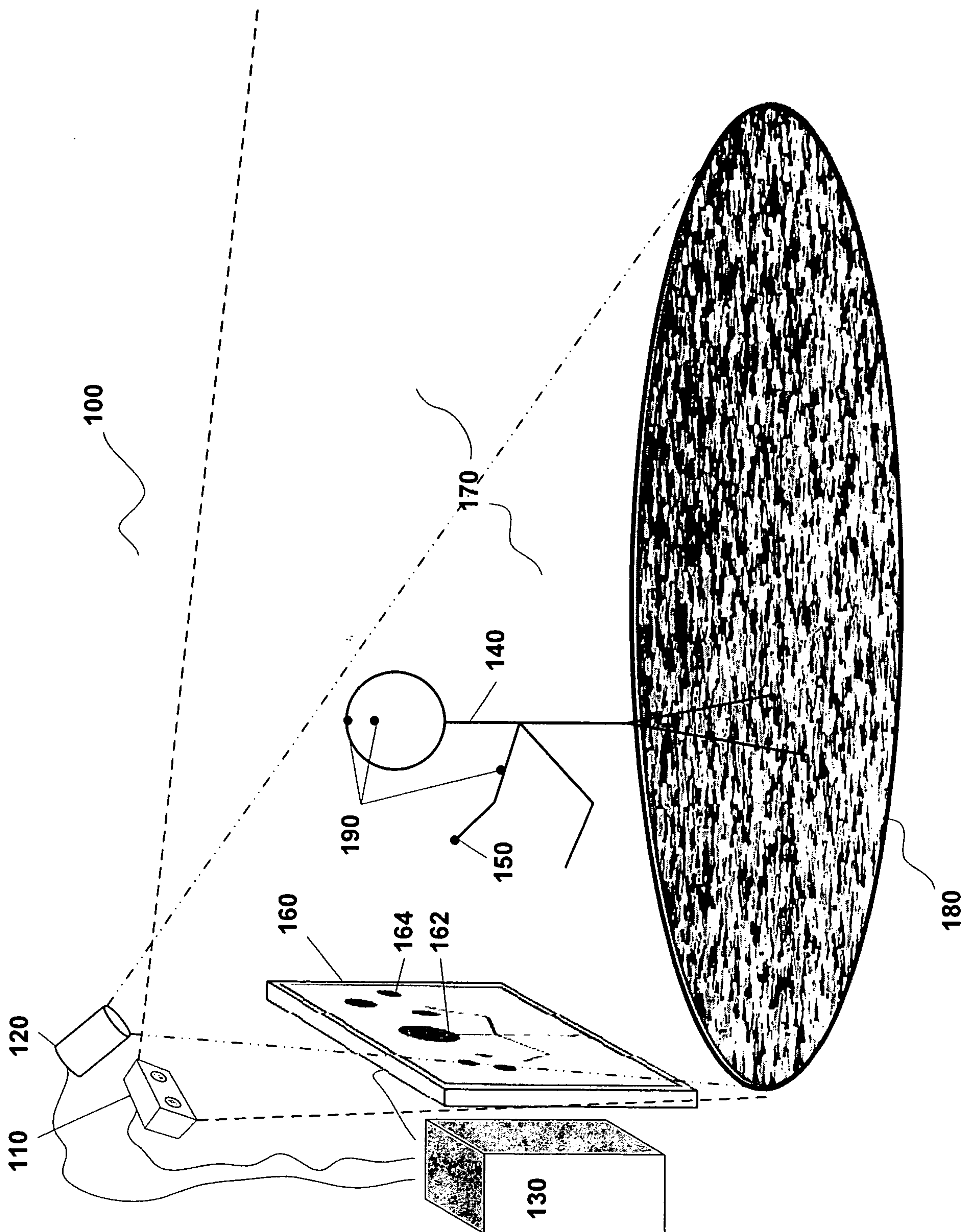
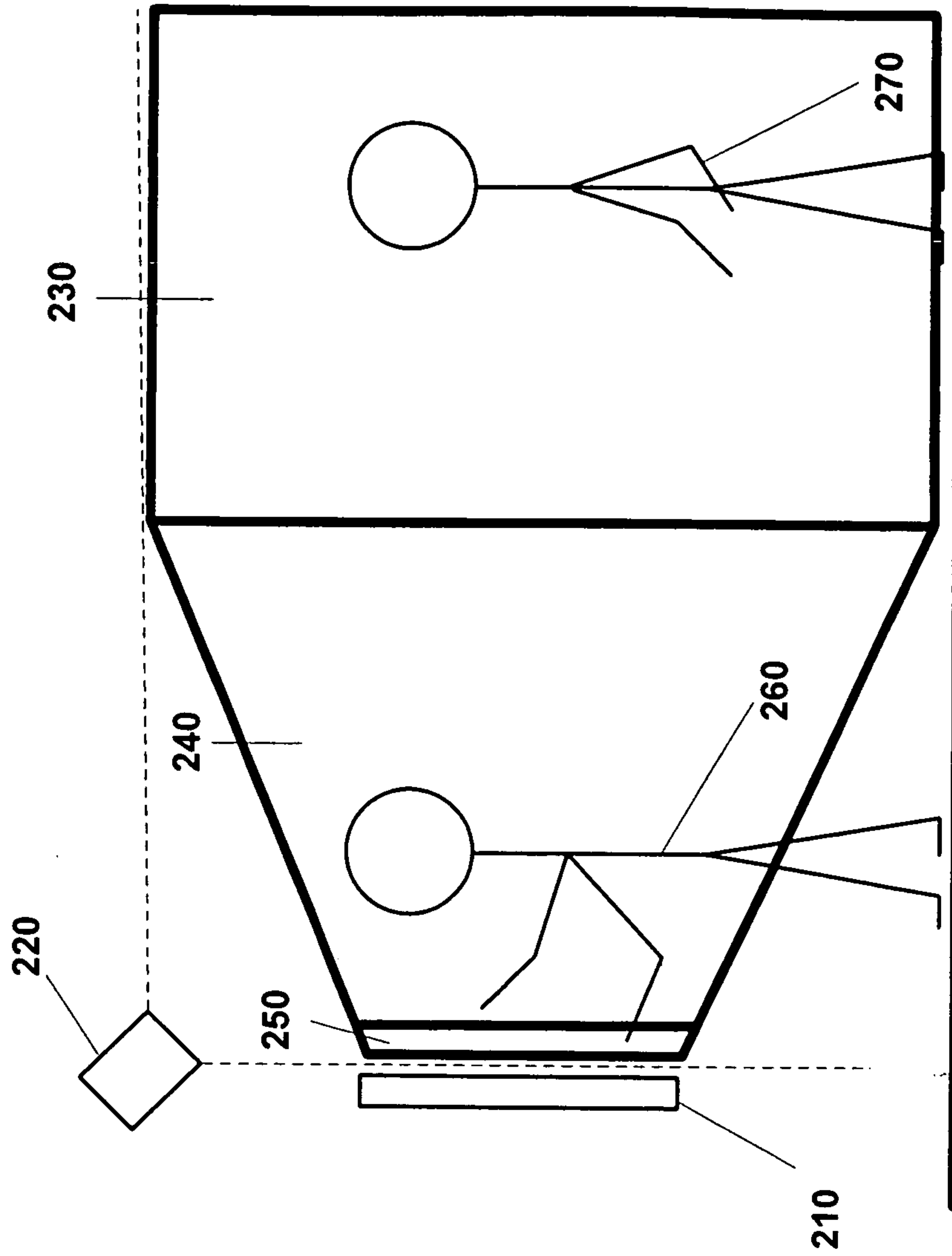


Figure 1

**Figure 2**

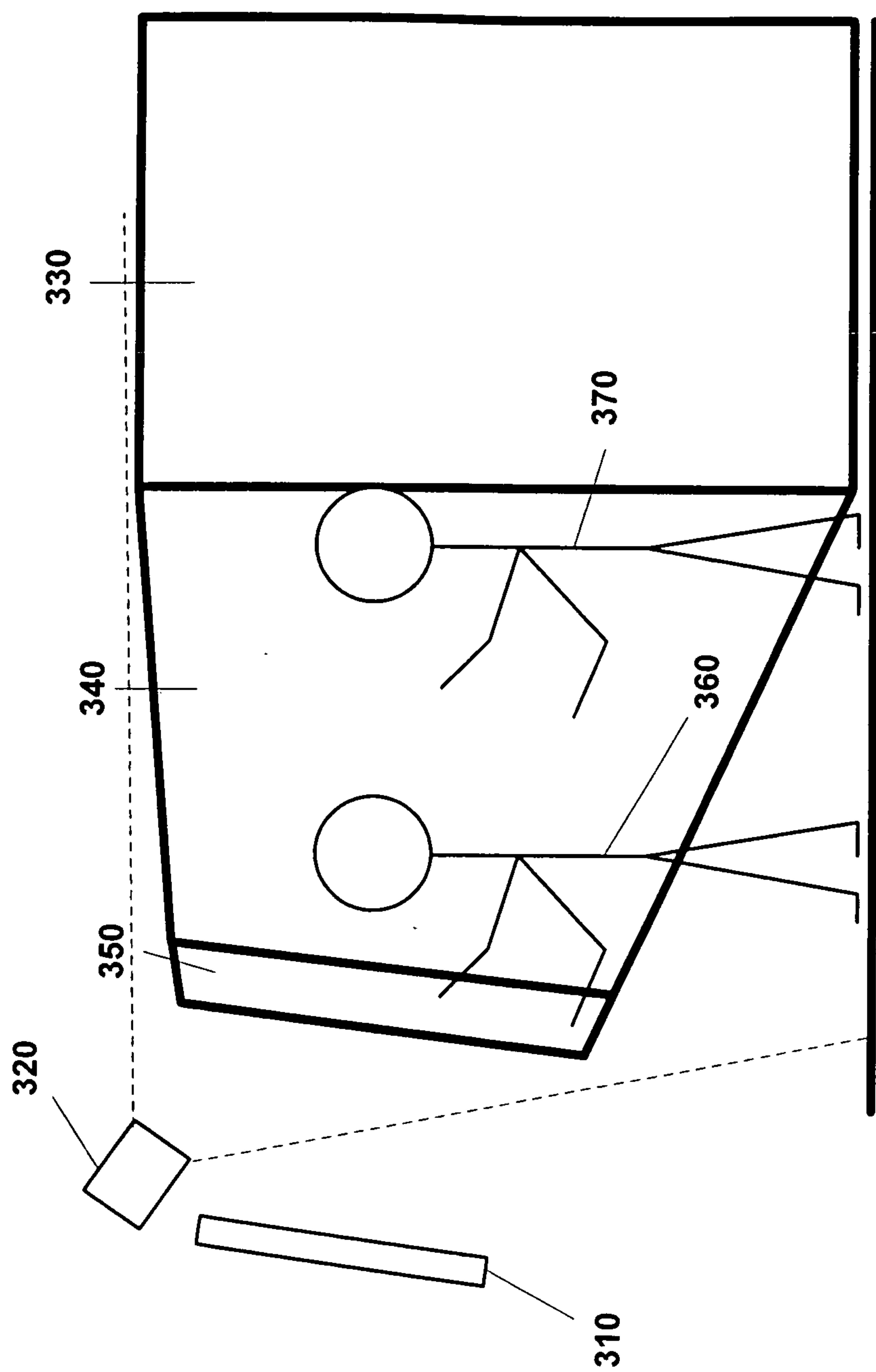


FIGURE 3

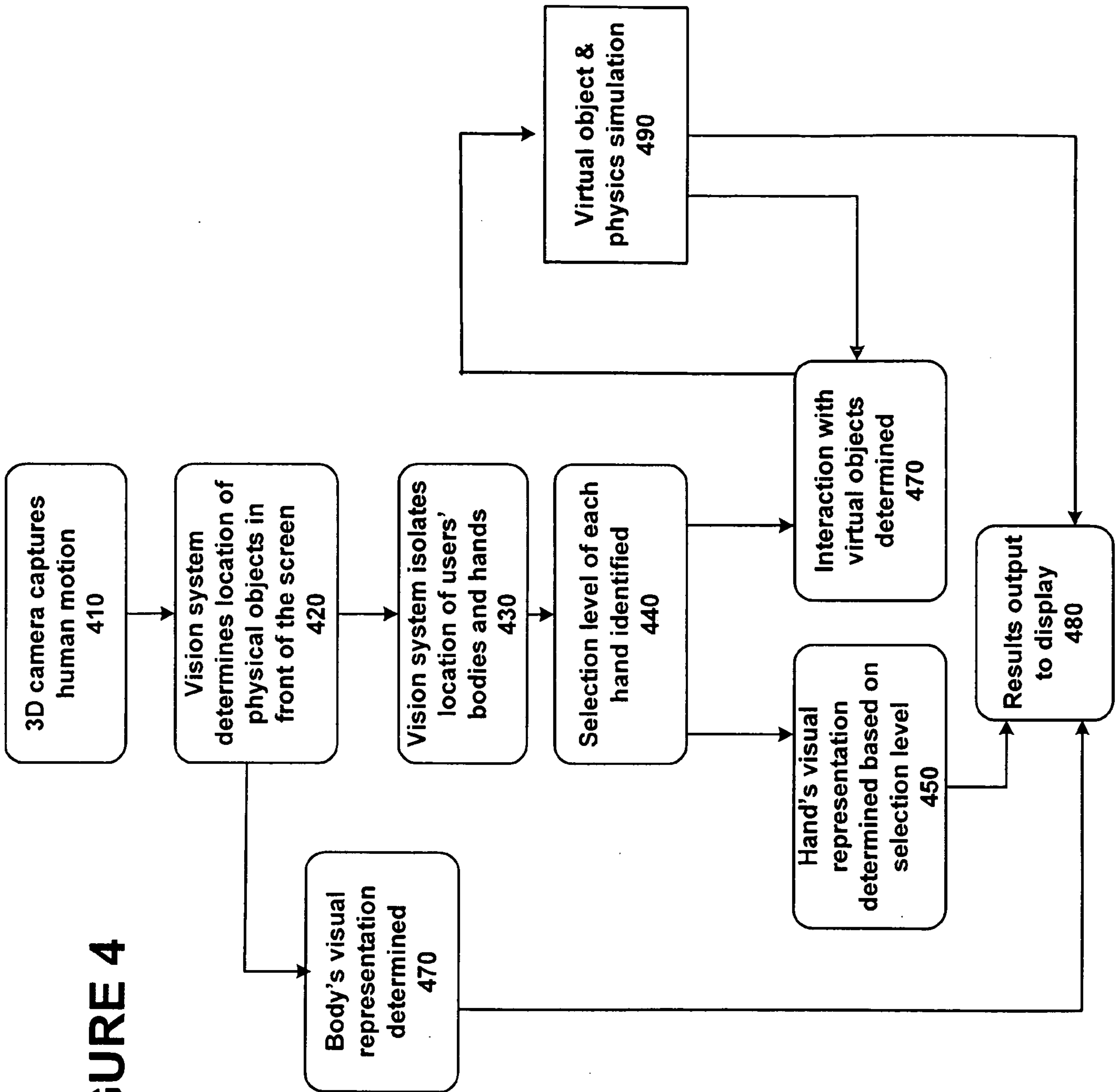
FIGURE 4

FIGURE 5

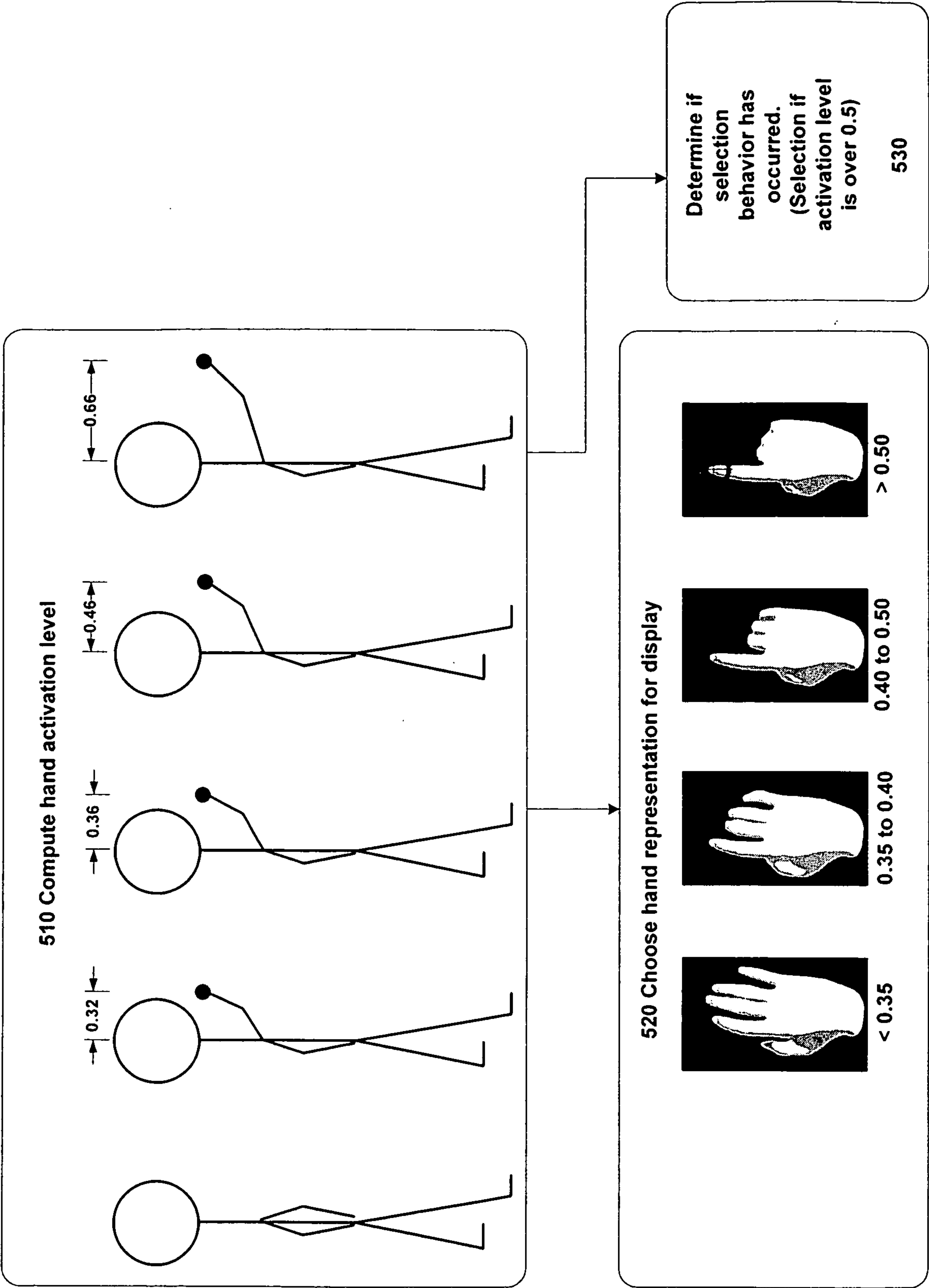


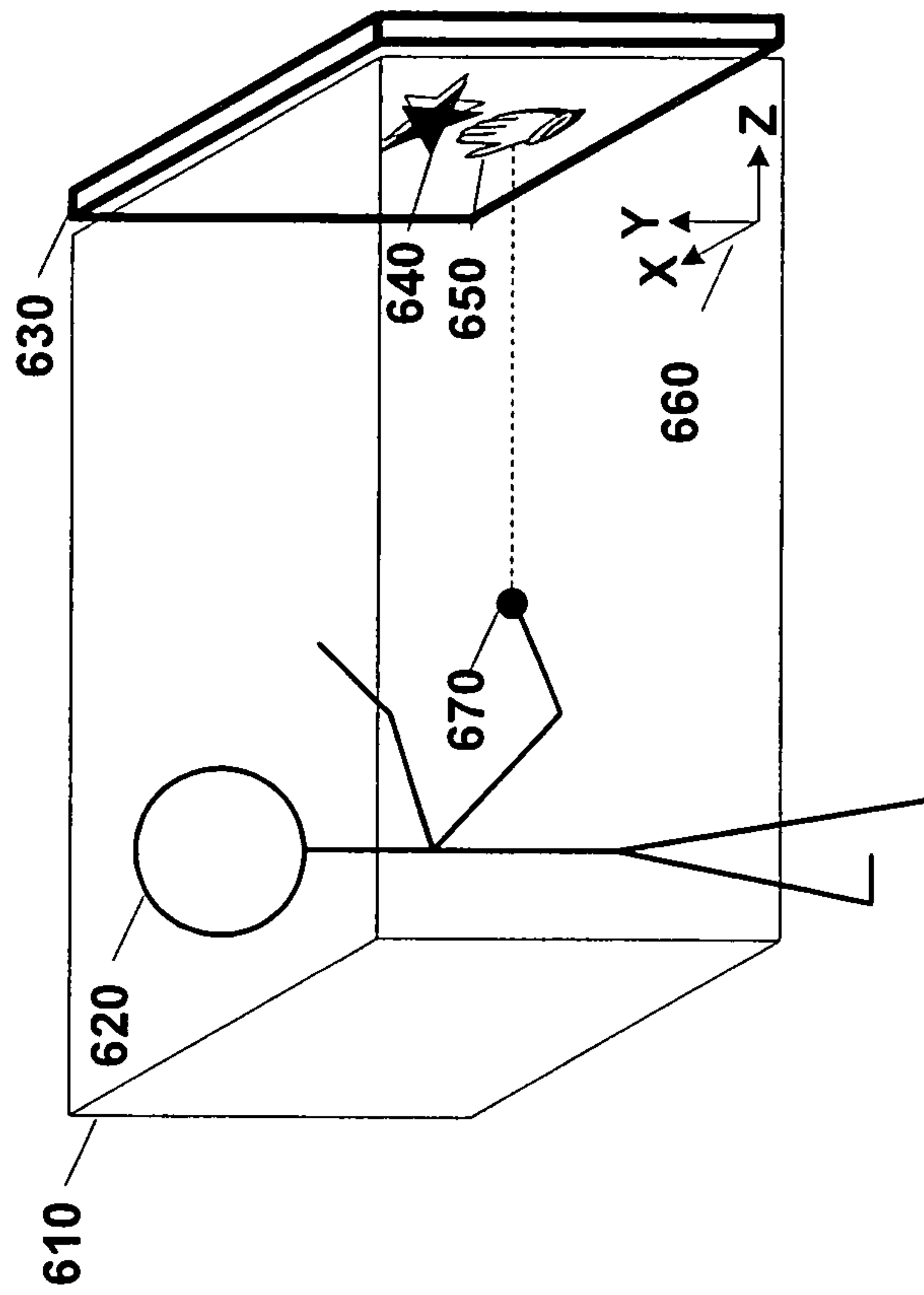
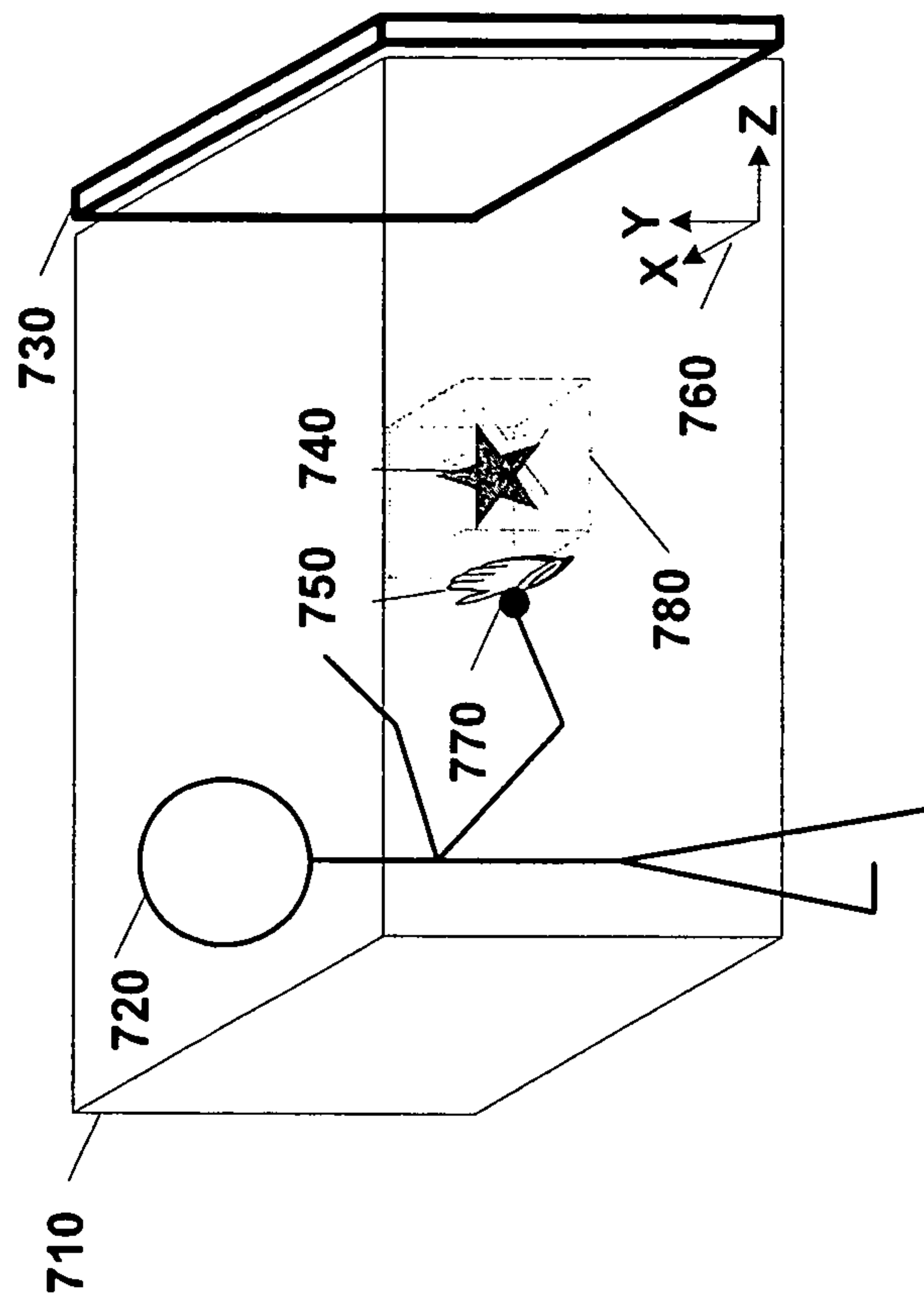
FIGURE 6

FIGURE 7



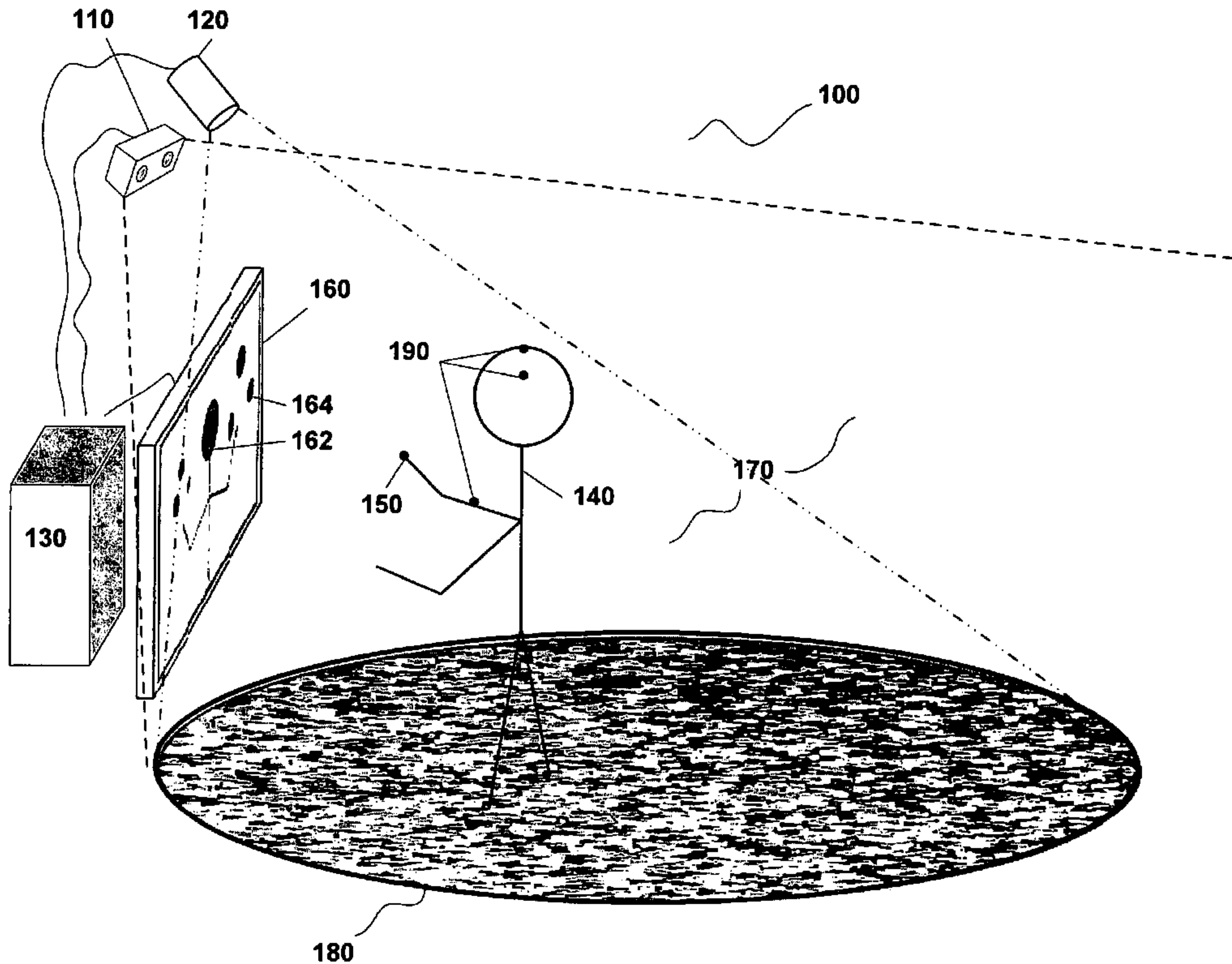


Figure 1