(54) Title: INTERACTIVE REALITY AUGMENTATION FOR NATURAL INTERACTION
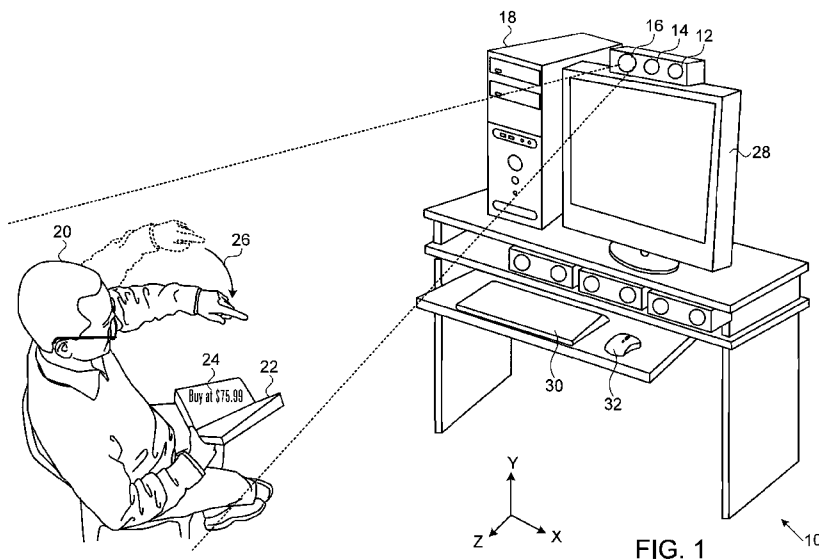
FIG. 1

(57) Abstract: Embodiments of the invention provide apparatus and methods for interactive reality augmentation, including a 2-dimensional camera (36) and a 3-dimensional camera (38), associated depth projector and content projector (48), and a processor (40) linked to the 3-dimensional camera and the 2-dimensional camera. A depth map of the scene is produced using an output of the 3-dimensional camera, and coordinated with a 2-dimensional image captured by the 2-dimensional camera to identify a 3-dimensional object in the scene that meets predetermined criteria for projection of images thereon. The content projector projects a content image onto the 3-dimensional object responsively to instructions of the processor, which can be mediated by automatic recognition of user gestures

Interactive Reality Augmentation for Natural Interaction

CROSS-REFERENCE TO RELATED APPLICATIONS

This Application claims the benefit of U.S. Provisional Application No. 61/365788, filed 20 July 2010, which is herein incorporated by reference.

5   BACKGROUND OF THE INVENTION

1. Field of the Invention.

This invention relates generally to natural interaction systems. More particularly this invention relates to adaptive reality augmentation and 3-dimensional input interfaces.

2. Description of the Related Art.

10   Natural user interfaces are gaining momentum in the entertainment and computer industry. Gesture controls are supplementing or replacing more conventional and less natural interfaces such as keyboard and mouse, game controller, and remote control. The user interactions, however, continue to relate largely to the computer monitor, thus limiting applicability and ease of use of such interfaces. Some of the gesture controls rely on optical 3-dimensional

15   mapping.

Various methods are known in the art for optical 3-D mapping, i.e., generating a 3-dimensional profile of the surface of an object by processing an optical image of the object. This sort of profile is also referred to as a depth map or depth image, and 3-D mapping is also referred to as depth mapping.

20   Some methods are based on projecting a laser speckle pattern onto the object, and then analyzing an image of the pattern on the object. For example, PCT International Publication WO 2007/043036, whose disclosure is incorporated herein by reference, describes a system and method for object reconstruction in which a coherent light source and a generator of a random speckle pattern project onto the object a coherent random speckle pattern. An imaging

25   unit detects the light response of the illuminated region and generates image data. Shifts of the pattern in the image of the object relative to a reference image of the pattern are used in real time reconstruction of a 3-D map of the object. Further methods for 3-D mapping using speckle patterns are described, for example, in PCT International Publication WO 2007/105205, whose disclosure is incorporated herein by reference.

SUMMARY OF THE INVENTION

The present invention, in certain embodiments thereof seeks to provide an improved content projection device, which is aware of objects in its field of view, recognizing such objects as suitable for projection of content thereon. The projection device may adapt to the geometry and character of the objects by controlling scale, distortion, focus of the projected content, and varying the projected content itself. Additionally or alternatively, the projection device may adapt the projected content according to the relationship of the viewer to the projected content, such as its gaze vector, distance from the surface onto which content is projected, and other similar parameters. The 2D/3D input device used to analyze the geometry for projection can also be used to interact with the projected content.

According to disclosed embodiments of the invention, methods and apparatus are provided for the projection of content, such as the input device interface, using a 3-dimensional input device as means of determining the optimal objects to serve as substrate for such content projection.

There is provided according to embodiments of the invention an apparatus for processing data, including a sensing element for acquiring a scene including a 2-dimensional camera and a 3-dimensional camera, a processor linked to the 3-dimensional camera and the 2-dimensional camera and programmed to produce a depth map of the scene using an output of the 3-dimensional camera, and to coordinate the depth map with a 2-dimensional image captured by the 2-dimensional camera to identify a 3-dimensional object in the scene that meets predetermined criteria for projection of images thereon, and a content projector for establishing a projected image onto the 3-dimensional object responsively to instructions of the processor.

According to an aspect of the apparatus, coordinating the depth map includes identifying a position of the 3-dimensional object with six degrees of freedom with respect to a reference system of coordinates, wherein the content projector is operative to compensate for scale, pitch, yaw and angular rotation of the 3-dimensional object.

According to a further aspect of the apparatus, coordinating the depth map includes referencing a database of 3-dimensional object definitions and comparing the 3-dimensional object with the definitions in the database.

An aspect of the apparatus includes a wearable monitor, wherein the content projector is operative to establish the projected image as a virtual image in the wearable monitor or in a

virtual space. The sensing element, the processor and the content projector may be incorporated in the wearable monitor.

According to a further aspect of the apparatus, the content projector is operative to establish the projected image onto a virtual surface for user interaction therewith.

According to yet another aspect of the apparatus, the processor is operative for controlling a computer application responsively to a gesture and wherein the projected image includes a user interface for control of the computer application.

According to aspect of the apparatus, the projected image includes written content.

Other embodiments of the invention provide methods for carrying out the function of the above-described apparatus.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

For a better understanding of the present invention, reference is made to the detailed description of the invention, by way of example, which is to be read in conjunction with the following drawings, wherein like elements are given like reference numerals, and wherein:

Fig. 1 is a schematic pictorial illustration of an interactive three-dimensional video display system, which is constructed and operative in accordance with a disclosed embodiment of the invention;

Fig. 2 is a block diagram of the system shown in Fig. 1, which is constructed and operative in accordance with an embodiment of the invention;

Fig. 3 is a block diagram that shows functional elements of a portion of an exemplary processing device, which is constructed and operative in accordance with an embodiment of the invention;

Fig. 4 is an exemplary flow chart of a method of identifying 3-dimensional objects in a scene in accordance with an embodiment of the invention;

Fig. 5 illustrates a screen of a mobile device that is projected onto a virtual surface in accordance with an embodiment of the invention; and

Fig. 6 illustrates an interactive three-dimensional video display system that includes a wearable monitor in accordance with an embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

In the following description, numerous specific details are set forth in order to provide a thorough understanding of the various principles of the present invention. It will be apparent

to one skilled in the art, however, that not all these details are necessarily always needed for practicing the present invention. In this instance, well-known circuits, control logic, and the details of computer program instructions for conventional algorithms and processes have not been shown in detail in order not to obscure the general concepts unnecessarily.

5   As used herein, the term "content projection" may encompass establishment of an image of the content onto a wearable transparent monitor, such as see-through eyeglasses, and thus invisible to anyone other than the person wearing the glasses, or onto a physical object that is visible to anyone interacting with the object. The term is not limited to the above examples. It may encompass forming an image by many means, including retinal projection, projection

10   onto see-through glasses, projection of the image into a virtual space, for example as a holo-gram, and other techniques for creating augmented reality.

System Architecture.

         Turning now to the drawings, reference is initially made to Fig. 1, which is a schematic pictorial illustration of an interactive three-dimensional video display system 10, which is

15   constructed and operative in accordance with a disclosed embodiment of the invention. The system 10 incorporates a 3-dimensional (3-D) camera 12, which may include an infra-red (IR) projector and corresponding CMOS/CCD camera open for the projector band. The 3-D cam-era 12 captures 3-D information that may includes the body (or at least parts of the body) of the user, tangible entities wielded or operated by the user for controlling a computer applica-

20   tion, and other objects in the field of view of the 3-D camera 12. Details of a 3-D imaging as-sembly of this sort are described, for example, in PCT International Publication WO 2010/004542 and U.S. Patent Application Publication No. 2009/0183125, which are here-in incorporated by reference. The 3-D camera 12 typically operates in the near infra-red spec-trum. However the principles of the invention are equally applicable to modifications that en-

25   able the 3-D camera 12 to capture electromagnetic energy outside the near infra-red spectrum, for example far infrared or ultraviolet energy. The system 10 may also include a 2-dimensional (2-D) camera 14, which operates in the visible spectrum, and can acquire a scene with sufficient resolution to allow automatic interpretation of written information in the scene and typically produces a Red-Green-Blue (RGB) output signal.

30   The 3-D camera 12 and the 2-D camera 14 are cooperative with a content projector 16, all under the control of a processor, such as a computer 18.

A suitable unit for use in the system 10 that bundles the 3-D camera 12 and the 2-D camera 14 is the PrimeSensor™ Reference Design, available from PrimeSense Corporation, 104 Cambay Ct, Cary North Carolina, 27513, U.S.A. The content projector 16 may be the Pi-coP® display engine, available from MicroVision, Inc., 6222 185th Ave NE Redmond WA, 98052. In some embodiments, the 3-D camera 12 and the 2-D camera 14 may be integral with the content projector 16 as a modification of the PrimeSensor Reference Design. In one embodiment, the 3-D camera 12 is an integrated module that includes an IR projector, which projects a pattern of spots onto the object and captures an image of the projected pattern. Alternatively, the IR projector, may be embodied as a separate module (not shown). The IR projector may be realized according to the teachings of U.S. Provisional Applications 61/372,729 (filed August 11, 2010) and 61/425,788 (filed December 22, 2010), which are herein incorporated by reference. This provisional application teaches also how to reuse the scanning hardware to project both the IR required for depth mapping and the visible content.

The processor may analyze the scene using the teachings of commonly assigned co-pending U.S. Patent Application No. 12/854,187, entitled "Analysis of Three-Dimensional Scenes", which is herein incorporated by reference.

The computer 18 may comprise a general-purpose computer processor, which is programmed in software to carry out the functions described hereinbelow. The software may be downloaded to the processor in electronic form, over a network, for example, or it may alternatively be provided on non-transitory tangible storage media, such as optical, magnetic, or electronic memory media. Alternatively or additionally, some or all of the image functions may be implemented in dedicated hardware, such as a custom or semi-custom integrated circuit or a programmable digital signal processor (DSP). Although the computer 18 is shown in Fig. 1, by way of example, as a separate unit from the 3-D camera 12, some or all of the processing functions of the computer may be performed by suitable dedicated circuitry associated with or within the housing of the 3-D camera 12 and the 2-D camera 14. As will be seen from the discussion below, elements of the system 10 may be miniaturized and incorporated in a wearable monitor to enable the user to move about and more freely interact with the scene in near real-time. In any case the 3-D camera 12 and the 2-D camera 14 function as a sensor component, which observes a scene (users and their surroundings). The computer 18 functions as a perception component, which comprehends the scene and user interaction with-

in these surroundings as mediated or stimulated by information provided by the content projector 16.

The computer 18 may execute programs such as Nite^{TM} Middleware, available from PrimeSense, in cooperation with the PrimeSensor Reference Design. For example, the PrimeSensor Reference Design supplies an application layer in the computer 18 with control widgets, thereby providing an application programming interface (API) that translates user gestures or postures into known deterministic application inputs. The Middleware performs image processing operations on data generated by the components of the system 10, including the 3-D camera 12 with its IR projector, and the 2-D camera 14 in order to reconstruct 3-dimensional maps of a user 20 and acquired scenes. The term "3-dimensional map" refers to a set of 3-dimensional coordinates representing the surface of a given object. One form of 3-dimensional map is referred to as a depth image or depth map, in which each pixel has a value indicating the distance from the camera to the corresponding point in the scene, rather than the brightness and color of the point as in a 2-dimensional image. The computer 18 then computes the three-dimensional coordinates of points on the surface of the control entity by triangulation, based on transverse shifts of the spots in the pattern.

In typical applications, information captured by the 3-D camera 12 is processed by the computer 18, which drives the content projector 16. The computer 18 may operate according to a program that is designed to create a natural or contrived experience for the user. As shown in Fig. 1, the system 10 has recognized a book 22 in the scene, and has projected a sale offer 24 onto the book 22: "Buy at $75.99". The user 20 is reacting to the offer by a hand gesture 26, which acts as an input to the computer 18. Gesture control of a computing device is known, for example, from commonly assigned U.S. Patent Application Publication No. 2009/0183125, which is herein incorporated by reference, and which also teaches methods of projection of scenes into a virtual image space. Gesture control is included in the functionality of the Nite^{TM} Middleware, which may interpret gestures of the user 20, for example in response to the sale offer 24 that are acquired by the 3-D camera 12 and the 2-D camera 14.

Furthermore, as the interaction of the user 20 with the book 22 and the sale offer 24 evolves, for example, by the user 20 grasping the book 22, a gaze identification module executing in the computer 18 may recognize that the user 20 is looking at the book 22. By processing the acquired 2-D images, the book title may be recognized and interpreted in the

system 10. Then, computing optimal projection parameters, a book review may be projected onto the book 22. The user 20 could scroll and interact with the projected book review as if he were viewing it on a display screen. In this way, the system 10, cooperatively with the user 20, converts the book 22 in an *ad hoc* fashion into a virtual information screen for the benefit of
5   the user a20.

The system 10 optionally includes a display screen 28 and conventional input devices such as a keyboard 30 and mouse 32, which may present a user interface for administrative use, e.g., system configuration, and for operational control of the system 10 by the user 20.

Reference is now made to Fig. 2, which is a block diagram of the system 10 (Fig. 1), in
10   accordance with an embodiment of the invention. A scene 34 is acquired concurrently by two cameras, a 2-D camera 36 and a 3-D camera 38, which may be separate units or integral as a combined unit. Alternatively, the scene can be captured by the 3-D camera 38 only or by the 2-D camera 36 only, image analysis performed on the images acquired in any case. As noted above these cameras may be realized as the PrimeSensor Reference Design. Data output by
15   the 2-D camera 36 and a 3-D camera 38 are input to a processor 40, which executes middle-ware, for example, the above-mentioned Nite Middleware. The Middleware places the scenes captured by the two cameras in registration. The middleware includes an object analysis mod-ule 42, which identifies objects in the scene 34 and determines their suitability for content pro-jection thereon. A projector control module 44, another component of the Middleware, con-
20   verts coordinates and characteristics of objects in the scene 34, for example an object 46, and prepares an image for projection. The module 44 issues suitable instructions for a projector 48 such that the image, typically containing information content, is projected onto the object 46. The instructions may contain corrections for distortion attributable to the scale, attitude and configuration of the object 46. Additionally or alternatively, the projector 48 may include its
25   own mechanisms to compensate for such distortion.

The position and attitude of the user may be taken into consideration when computing projection parameters. For example, as noted above, the gaze vector toward the projected con-tent may vary as the user moves about in the scene. The projection parameters may be accor-dingly adjusted to compensate for such variations, e.g., by adjusting for scale, parallax, and
30   similar distortions, so as to simulate a realistic experience for the user. One example of such adjustment is a correction for the fact that 3-dimensional objects appear differently when viewed from different directions, i.e., different sides of the object or different 2-D projections of the object become apparent to the observer. The projection content can be adjusted as a

function of the gaze vector and user position relative to virtual object, thus creating a realistic experience of the object actually being in the presence of the observer. Gaze direction can be determined by methods known in art. For example, in the case of a device embedded in see-through glasses, head position orientation is obtainable by rigid registration of the world rela-

5   tive to the device. Gaze direction can also be measured, for example, using eye-tracking products available from Tobii Technology, Inc., 510 N, Washington Street, Suite 200, Falls Church, VA 22046. Gaze may then be translated into object coordinates using 3D information obtained by the sensor.

Object Awareness.

10   Techniques for identifying and tracking body parts are known from commonly assigned U.S. Patent Application Publication No. 2011/0052006, entitled "Extraction of Skeletons from 3-D Maps", which is herein incorporated by reference. Essentially this is accomplished by receiving a temporal sequence of depth maps of a scene containing a humanoid form. A digital processor processes at least one of the depth maps so as to find a location of a designated body

15   part, such as the head or hand estimates dimensions of the humanoid form based on the location. The processor tracks movements of the humanoid form over the sequence using the estimated dimensions. These teachings are employed in the above-mentioned Nite Middleware, and may be enhanced by linking other known recognition routines by those skilled in the art.

For example, in the case of identifying the head of the body, the processor may segment

20   and analyzes a 3-dimensional form to identify right and left arms, and then search the space between the arms in order to find the head. Additionally or alternatively recognition techniques may be used. The depth maps may be registered with 2-dimensional images of the head or other object. The processor may apply a pattern or face recognition technique to identify the face of a humanoid form in a 2-dimensional image. The face location in the 2-dimensional

25   image is then correlated with the location of the head of the 3-dimensional form. Using the same techniques, an entire scene may be analyzed, segmented, and known categories of objects identified as candidates for projection of images thereon.

Object Processor.

Reference is now made to Fig. 3, which is a block diagram that schematically shows

30   functional elements of a portion of an exemplary processing device 50, which is a component of the processor 40 (Fig. 2), and which is constructed and operative in accordance with an embodiment of the invention. The processing device 50 may be fabricated as a dedicated inte-

grated circuit, on a single semiconductor substrate, with a USB port 52 to an optional host computer 54. Device 50 may include other interfaces, as well, including an object analyzer 56. The object analyzer 56 is linked to a database 58, which holds a library containing descriptions of objects to be recognized and evaluated by the object analyzer 56. It will be appreciated that alternative configurations of the processing device 50 can be constructed by those skilled in the art. As noted above, the operations of the processing device 50 may be controlled by middleware residing in instruction memory 60 and data memory 62

A depth processor 64 processes the information captured by the 3-D camera 12 (Fig. 1) in order to generate a depth map. Depth processor 64 uses dedicated memory space in a memory 66. This memory can also be accessed by a controller 68, which is described hereinbelow, but typically not by the host computer 54. Rather, depth processor 64 may be programmed by the host computer 54 via an application program interface (API).

Depth processor 64 receives input IR data from 3-D camera 12 (Fig. 1) via a depth CMOS interface 70. The depth processor 64 processes the video data in order to generate successive depth maps, i.e., frames of depth data. The depth processor 64 loads these data into a depth first-in-first-out (FIFO) memory 72 in a USB FIFO unit 74.

In parallel with the depth input and processing operations, a color processing block 76 receives input color video data from the 2-D camera 14 (Fig. 1) via a color CMOS sensor interface 78. The block 76 converts the raw input data into output frames of RGB video data, and loads these data into a RGB FIFO memory 80 74 in the unit 74. Alternatively, the block 76 may output the video data in other formats, such as YUV or Bayer mosaic format.

The unit 74 acts as a buffer level between the various data suppliers and a USB controller 82. The unit 74 packs and formats the various data types according to different classes (such as a USB video class and a USB audio class), and also serves to prevent data loss due to USB bandwidth glitches. It arranges the data into USB packets according to the USB protocol and format prior to transferring them to the USB controller.

A high-bandwidth bus, such as an Advanced High-performance Bus (AHB) matrix 84, is used to carry data between the components of the processing device 50, and specifically for conveying data from the unit 74 to the USB controller 82 for transfer to the host computer 54. (AHB is a bus protocol promulgated by ARM Ltd., of Cambridge, England.) When there are packets ready in the unit 74 and space available in the internal memory of USB controller 82, the USB controller 82 uses direct memory access (DMA) to read data from memory 72, memory 80, and an audio FIFO memory 86 via an AHB slave module 88 and the matrix 84. The

USB controller 82 multiplexes the color, depth and audio data into a single data stream for output via the USB port 52 to the host computer 54.

For the purpose of USB communications, they processing device 50 comprises a USB physical layer interface, PHY 90, which may be operated by the USB controller 82 to com-

5   municate via a suitable USB cable with a USB port of the host computer 54. The timing of the USB PHY is controlled by a crystal oscillator 92 and a phase-locked loop 94 (PLL), as is known in the art.

Alternatively, USB controller 86 may optionally communicate with the host computer via a USB 2.0 Transceiver Macrocell Interface (UTMI) and an external PHY 96.

10   Various external devices may connect with the processing device 50 cooperatively with the host computer 54, including a projector control module 98, which accepts instructions from the processing device 50 and the host computer 54 to effect a desired image projection onto specified coordinates in space.

The controller 68 is responsible for managing the functions of the processing device 50,

15   including boot-up, self-test, configuration, power and interface management, and parameter adjustment.

The controller 68 may comprise a digital signal processor (DSP) core 100 and an AHB master 102 for controlling data movement on the matrix 84. Typically, controller 68 boots from a boot read-only memory 104, and then loads program code from a flash memory (not

20   shown) via a flash memory interface 106 into instruction random-access memory 60 and data memory 62. The controller 68 may, in addition, have a test interface 108, such as a Joint Test Action Group (JTAG) interface, for purposes of debugging by an external computer 110.

The controller 68 distributes configuration data and parameters to other components of the processing device 50 via a register configuration interface 112, such as an Advanced Peri-

25   pheral Bus (APB), to which the controller is connected through the matrix 84 and an APB bridge 114.

Further details of the processing device 50 are disclosed in the above-noted PCT International Publication WO 2010/004542.

Object Analysis.

30   Continuing to refer to Fig. 3, the object analyzer evaluates data developed by the depth processor 64 in cooperation with the block 76 and the unit 74 to evaluate a scene captured by the 3-D camera 12 (Fig. 1).

The algorithm executed by the object analyzer 56 may be dictated by an application program in the host computer 54. For example, the object analyzer 56 may be instructed to search for and report one or more known objects in the scene that are specified in the database 58. The host computer 54 may thereupon instruct the content projector 16 (Fig. 1) to project images on the selected object or objects. Additionally or alternatively, the object analyzer 56 may be instructed to identify and report objects meeting predefined criteria, without resort to the database 58.

The data communicated by the object analyzer 56 with respect to an identified object typically includes the size and location of the object, as well as its orientation, preferably with six degrees of freedom, including scale, pitch, yaw and angular rotation with respect to a reference system of coordinates. This information allows the projector to compensate for distortions by suitably scaling and contorting a projected image so as to be project it onto the selected object such that the viewer sees an image that is substantially distortion-free. Configuration of a projected image is known, e.g., from U.S. Patent Application Publication No. 20110081072, entitled "Image Processing Device, Image Processing Method, and Program". The image may be configured in software in order to avoid the expense of complex optical arrangements and to more easily achieve freedom from such effects as off-axis image distortion Alternatively, As noted above, commercially available projects may provide their own compensation for distortion control.

Reference is now made to Fig. 4, which is an exemplary flow chart of a method of identifying 3-dimensional objects in a scene in accordance with an embodiment of the invention. For convenience of presentation, the method is disclosed in conjunction with the apparatus shown in Fig. 1 and Fig. 3, but it is applicable to apparatus configured differently. The process steps are shown in a particular linear sequence in Fig. 4 for clarity of presentation. However, it will be evident that many of them can be performed in parallel, asynchronously, or in different orders. Those skilled in the art will also appreciate that a process could alternatively be represented as a number of interrelated states or events, e.g., in a state diagram. Moreover, not all illustrated process steps may be required to implement the process. Furthermore, many details may vary according to the dictates of the host computer 54 and the requirements of its application program.

Assume that the viewer is located in a bookshop. At initial step 116 an application program executing in the host computer 54 would like to identify an open book displaying textual information. This is a 3-dimensional object having a known definition in the database 58 that

includes at least one generally light-colored planar surface. The 3-D camera 12 is enabled and a 3-dimensional scene captured in the processing device 50. The object analyzer 56 evaluates the scene, locates and identifies objects in 3-dimensional space.

At decision step 118 it is determined whether a planar surface has been located in the scene.

Control now proceeds to decision step 120, where it is determined if the planar surface meets criteria for a book. The criteria may involve, *inter alia*, size, proximity to certain other objects, and geometric details corresponding to a closed or open book.

If the determination at decision step 120 is affirmative, then control proceeds final step 122. The coordinates and orientation of the book are reported by the object analyzer 56 to the controller 68, which instructs the projector control module 98 cooperatively with the host computer 54 to display an application-determined image (MENU-1) on the identified book. The image may contain, for example, options to purchase the item, or obtain additional details, for example book reviews, and popularity ratings. Indeed, if the 3-D camera 12 was successful in capturing the title of the book, the additional details may be included in the projected image. It is assumed that the host computer 54 has access to a local or distributed database or can make automatic inquiries via the Internet.

If the determination at decision step 120 is negative, then control proceeds to decision step 124. A determination is made if more objects are present in the scene for processing.

If the determination at decision step 124 is affirmative, then control returns to decision step 118.

If the determination at decision step 124 is negative, then a second state of the method commences. It is assumed that the application program falls through to a secondary option, in which an image is projected on the user's hand, if visible to the 3-D camera 12.

Control now proceeds to decision step 126, where it is determined if a body part is present in the scene. This may be accomplished using the teachings of the above-noted U.S. Patent Application Publication No. 2011/0052006.

If the determination at decision step 126 is affirmative, then control proceeds to decision step 128, where it is determined if the body part is a hand.

If the determination at decision step 128 is affirmative, then control proceeds to final step 130, which is similar to final step 122. However, a different menu (MENU-2) is now projected on the hand, which may include, for example, control options for the governing com-

puter application. In both final step 122 and final step 130 the image is configured so as to create a natural feeling on the part of the user when interacting with the content.

If the determination at decision step 128 is negative, then control proceeds to decision step 132. A determination is made if more objects are present in the scene for processing.

5   If the determination at decision step 132 is affirmative, then control returns to decision step 126. Otherwise, control passes to final step 134, in which a conventional menu display is presented on a display screen. Final step 134 represents a failure to identify a suitable external object for projection of an image thereon. It will be appreciated that the method shown in Fig. 4 can be varied, and elaborated as required to comply with the specifications of the go-

10   verning application program. Recognition and prioritization of various objects and images may be programmed so as to accommodate the configuration of a particular scene and the needs of the program itself.

Alternate Embodiment 1.

This embodiment is similar to the first embodiment, except a convenient virtual surface

15   is provided for projection of images and for access by the user. Reference is now made to Fig. 5, which illustrates a screen 136, typically of a mobile information device 138, such as a cellular telephone, e.g., a "smart phone" that is projected onto a virtual surface in accordance with an embodiment of the invention. Such devices are too small for convenient interaction and media consumption. The screen 136 incorporates a miniature projector 140 and sensing

20   device 142, which have the same functions as the 3-D camera 12 and content projector 16 in the embodiment of Fig. 1. Projectors suitable for this purpose are available, for example, from Microvision. In this embodiment, the projector 140 projects an image onto a virtual projection surface 144, which is enlarged relative to the screen 136.

In one mode of operation, the projector 140 may create an enlarged version of informa-

25   tion displayed on the screen 136.

In another mode of operation the sensing device 142 captures an external scene. The mobile information device 138 is configured to perform the method of scene analysis described above with reference to Fig. 4. In this example, an open book 146 was identified in the external scene. An application program executing in the mobile information device 138 has

30   caused the projector 140 to project an image 148 of the book 146 onto the projection surface 144, and to superimpose a menu 150 onto the image 148. The menu 150 invites the user to purchase the book 146 at a sales price of $75.99 or to cancel the display.

Alternate Embodiment 2.

In the first embodiment, images have been described as projections onto a physical object, e.g., a book or a hand. In this embodiment, the projector may be embodied as a device that projects content onto a wearable monitor, such as eye-glasses. In this embodiment final step 122 and final step 130 are modified in the method of Fig. 4.

Reference is now made to Fig. 6, which illustrates an interactive three-dimensional video display system having a wearable monitor in accordance with an embodiment of the invention. The system is configured to project the respective images onto the wearable monitor rather than the object themselves Such devices offer possibilities of allowing a computer-generated image produced by the method described with reference to Fig. 4 to be generated and optionally superimposed on a real-world view. Such devices may operate by projecting the computer-generated image through a partially reflective mirror while viewing an external scene. Alternatively the device may mix the computer-generated image and real-world view electronically.

In the example of Fig. 6, a user 152 employs a wearable monitor 154, which is capable of displaying stereoscopic imagery. The wearable monitor 154 is provided with or interfaced with components similar to those of the system 10 (Fig. 1). Like the system 10, the wearable monitor 154 is adapted to analyze an external scene. In this example, it identifies the book 146, and generates an image 156 containing the same information as the image 148 (Fig. 5). The wearable monitor 154 may be a separate unit or may incorporate other elements of the system 10. In the embodiment of Fig. 6, the wearable monitor 154 includes a miniature projector 158 and a sensing element 160. Additionally or alternatively, the wearable monitor 154 may communicate with an external processor or sensing device via a wireless link. Suitable wearable helmet mounted displays and see-through eyewear displays for use as the wearable monitor 154 are available as the Madison line of Novero (novero.com) or from Lumus Ltd., 2 Bergman Street Rehovot 76705, Israel.

While the image 156 is actually established within the wearable monitor 154, in some embodiments it may be perceived by the user 152 as being superimposed in an external region of space as shown in Fig. 6. The wearable monitor 154 in such embodiments may be equipped with positioning, head-tracking and eye-tracking subsystems.

It will be appreciated by persons skilled in the art that the present invention is not limited to what has been particularly shown and described hereinabove. Rather, the scope of the

present invention includes both combinations and sub-combinations of the various features described hereinabove, as well as variations and modifications thereof that are not in the prior art, which would occur to persons skilled in the art upon reading the foregoing description.

Claims.

1. An apparatus for processing data, comprising:

a sensing element for acquiring a scene comprising a 3-dimensional camera;

a processor linked to the 3-dimensional camera and programmed to produce a depth map of the scene using an output of the 3-dimensional camera, and making a scene analysis to identify a 3-dimensional object in the scene that meets predetermined criteria for projection of images thereon; and

a content projector for forming an image responsively to instructions of the processor.

2. The apparatus according to claim 1, wherein forming the image comprises projecting the image onto the 3-dimensional object.

3. The apparatus according to claim 1, wherein forming the image comprises projecting the image onto a retina of a user.

4. The apparatus according to claim 1, wherein forming the image comprises projecting the image onto see-through eyeglasses.

5. The apparatus according to claim 1, wherein forming the image comprises projecting the image into a 3-dimensional virtual space.

6. The apparatus according to claim 1, wherein the instructions of the processor are responsive to the scene analysis, and wherein the processor is cooperative with the content projector for varying at least one of   projection parameters and content   of the image responsively to the scene analysis.

7. The apparatus according to claim 1, wherein the processor is cooperative with the content projector for varying characteristics of the image responsively to an interaction between a user and the scene.

8. The apparatus according to claim 7, wherein the interaction comprises a variation in a gaze vector of the user toward the 3-dimensional object.

16

9. The apparatus according to claim 7, wherein the interaction comprises a gesture of the user relating to the 3-dimensional object.

10. The apparatus according to claim 7, wherein varying characteristics of the image comprises varying at least one of a scale and a compensation for distortion.

11. The apparatus according to claim 1, further comprising a 2-dimensional camera for acquiring a 2-dimensional image of the scene, wherein the processor is operative for recognizing information relating to the 3-dimensional object in the 2-dimensional image, and is further operative and for instructing the content projector to include metadata of the 3-dimensional object in the image responsively to the recognized information.

12. The apparatus according to claim 1, wherein identifying the 3-dimensional object comprises identifying a position of the 3-dimensional object with six degrees of freedom with respect to a reference system of coordinates, wherein the content projector is operative to compensate for scale, pitch, yaw and angular rotation of the 3-dimensional object.

13. The apparatus according to claim 1, wherein identifying the 3-dimensional object comprises referencing a database of 3-dimensional object definitions and comparing the 3-dimensional object with the definitions in the database.

14. The apparatus according to claim 1, further comprising a wearable monitor, wherein the content projector is operative to establish the image as a virtual image in the wearable monitor.

15. The apparatus according to claim 14, wherein the sensing element, the processor and the content projector are incorporated in the wearable monitor.

16. The apparatus according to claim 14, wherein the wearable monitor comprises see-through eyeglasses.

17. The apparatus according to claim 1, wherein the content projector is operative to establish the image onto a virtual surface for user interaction therewith.

18. The apparatus according to claim 1, wherein the processor is operative for controlling a computer application responsively to a gesture and wherein the image comprises a user interface for control of the computer application.

19. The apparatus according to claim 1, wherein the image comprises written content.

20. A method for augmented interaction with a data processing system, comprising the steps of:

capturing a 3-dimensional image of a scene;

using a digital processor, processing the 3-dimensional image to locate 3-dimensional objects therein, and to determine that the 3-dimensional objects satisfy predefined criteria; and

forming a content-containing image for interaction therewith by a user.

21. The method according to claim 20, wherein forming the content-containing image comprises projecting the content-containing image onto one of the 3-dimensional objects.

22. The method according to claim 20, wherein forming the content-containing image comprises projecting the content-containing image onto a retina of the user.

23. The method according to claim 20, wherein forming the content-containing image comprises projecting the content-containing image onto see-through eyeglasses.

24. The method according to claim 20, wherein forming the content-containing image comprises projecting the content-containing image into a into a 3-dimensional virtual space.

25. The method according to claim 20, further comprising the steps of varying characteristics of the content-containing image responsively to an interaction between the user and the scene.

26. The method according to claim 25, wherein the interaction comprises a variation in a gaze vector of the user toward one of the 3-dimensional objects.

27. The method according to claim 25, wherein the interaction comprises a gesture of the user relating to one of the 3-dimensional objects.

28. The method according to claim 25, wherein varying characteristics of the content-containing image comprises varying at least one of a scale and a compensation for distortion.

29. The method according to claim 20, further comprising the steps of:
capturing a 2-dimensional image of the scene in registration with the 3-dimensional image;
recognizing information relating to the 3-dimensional objects in the 2-dimensional image; and
varying the content-containing image responsively to the recognized information.

30. The method according to claim 20, further comprising the steps of:
recognizing a gesture relating to the content-containing image; and
responsively to the gesture controlling a computer application.

31. The method according to claim 30, wherein the content-containing image comprises a user interface for control of the computer application.

32. The method according to claim 20, wherein processing the 3-dimensional image comprises identifying a position of one of the 3-dimensional objects with six degrees of freedom with respect to a reference system of coordinates, and forming a content-containing image comprises compensating for scale, pitch, yaw and angular rotation of the one 3-dimensional object.

33. The method according to claim 20 wherein processing the 3-dimensional image comprises referencing a database of 3-dimensional object definitions and comparing the 3-dimensional objects with the definitions in the database.

34. The method according to claim 20, wherein one of the 3-dimensional objects is a portion of a humanoid form.

35. A method for augmented interaction with a data processing system, comprising the steps of:

receiving a depth map of a scene containing 3-dimensional objects meeting first criteria for projection of images thereon;

receiving a 2-dimensional image of the scene;

using a digital processor, processing the depth map to locate the objects;

executing a computer application to analyze the 2-dimensional image to identify a most suitable one of the 3-dimensional objects according to second criteria;

projecting a user interface to the computer application as a content-containing image onto the most suitable one of the 3-dimensional objects;

using the digital processor, recognizing a gesture;

interpreting the gesture as an interaction with the user interface; and

controlling the computer application responsively to the interaction.

36. The method according to claim 35, wherein processing the depth map comprises identifying a position of the most suitable one of the 3-dimensional objects with six degrees of freedom with respect to a reference system of coordinates, and projecting a user interface comprises compensating for scale, pitch, yaw and angular rotation thereof.

37. The method according to claim 35, wherein processing the depth map comprises referencing a database of 3-dimensional object definitions and comparing the 3-dimensional objects with the definitions in the database.

38. The method according to claim 35, wherein one of the 3-dimensional objects in the depth map is a portion of a humanoid form.

39. A method for augmented interaction with a data processing system, comprising the steps of:

receiving a depth map of a scene containing 3-dimensional objects meeting first criteria for projection of images thereon;

using a digital processor, processing the depth map to locate the objects;

executing a computer application to identify a most suitable one of the objects according to second criteria;

projecting an image of the one object onto a virtual surface;

5   using the digital processor, recognizing a gesture;

interpreting the gesture as an interaction with the image of the one object; and

controlling the computer application responsively to the interaction.

40. The method according to claim 39, wherein processing the depth map comprises

10   identifying a position of the one object with six degrees of freedom with respect to a reference system of coordinates, and projecting the image of the one object comprises compensating for scale, pitch, yaw and angular rotation of the one object.

41. The method according to claim 39, wherein processing the depth map comprises

15   referencing a database of 3-dimensional object definitions and comparing the 3-dimensional objects in the depth map with the definitions in the database.

42. The method according to claim 39, wherein the image of the one object further comprises a user interface for control of the computer application.

20

43. A method for augmented interaction with a data processing system, comprising the steps of:

receiving a depth map of a scene containing 3-dimensional objects meeting first criteria for projection of images thereon;

25   using a digital processor, processing the depth map to locate the objects;

executing a computer application to identify a most suitable one of the objects according to second criteria;

projecting an image of the one object onto a wearable monitor;

using the digital processor, recognizing a gesture of a user;

30   interpreting the gesture as an interaction with the image of the one object; and

controlling the computer application responsively to the interaction.

44. The method according to claim 43, wherein processing the depth map comprises identifying a position of the one object with six degrees of freedom with respect to a reference system of coordinates, and projecting the image of the one object comprises compensating for scale, pitch, yaw and angular rotation of the one object.

5

45. The method according to claim 43, wherein processing the depth map comprises referencing a database of 3-dimensional object definitions and comparing the 3-dimensional objects with the definitions in the database.

10

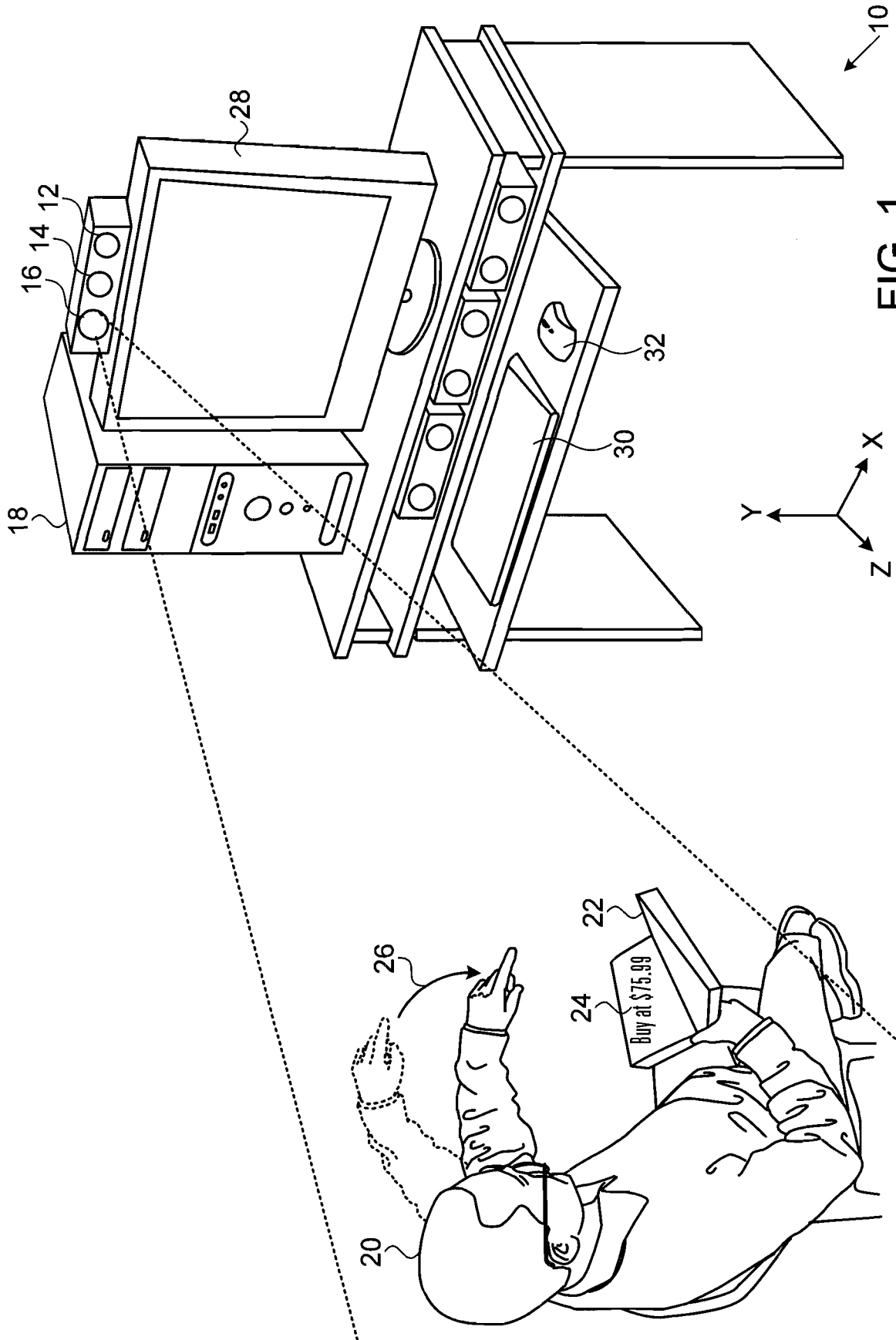46. The method according to claim 43, wherein the image of the one object further comprises a user interface for control of the computer application.
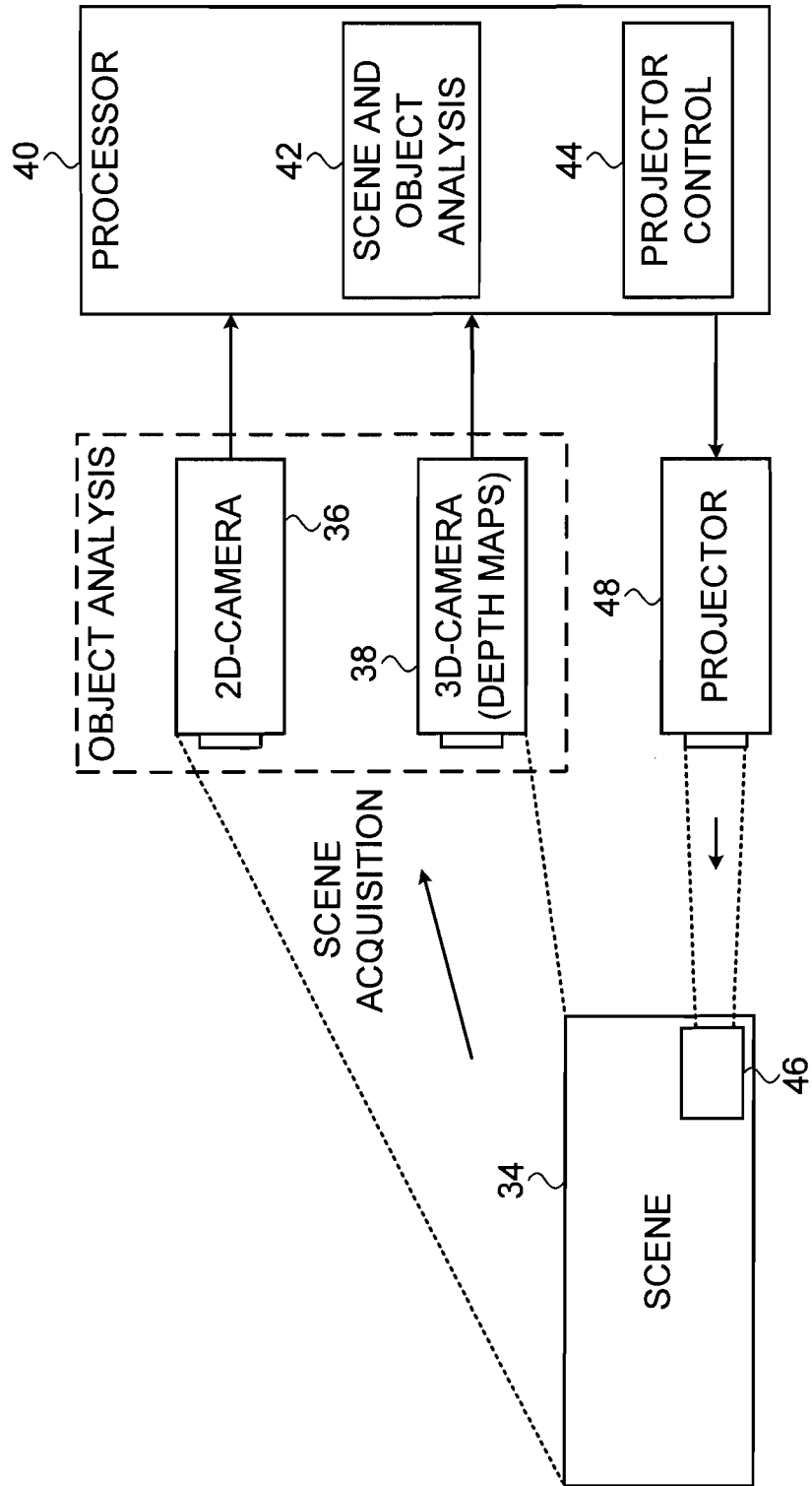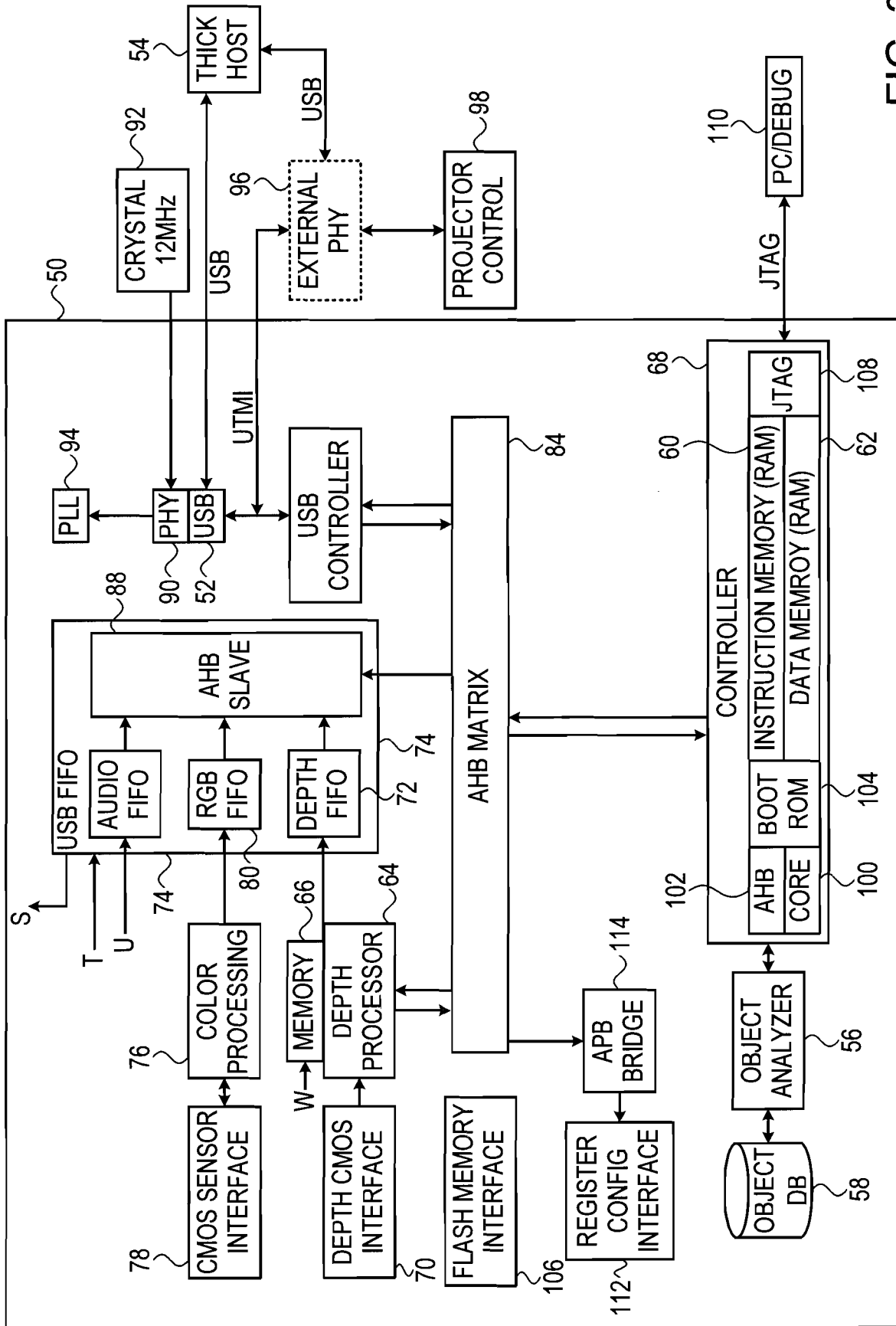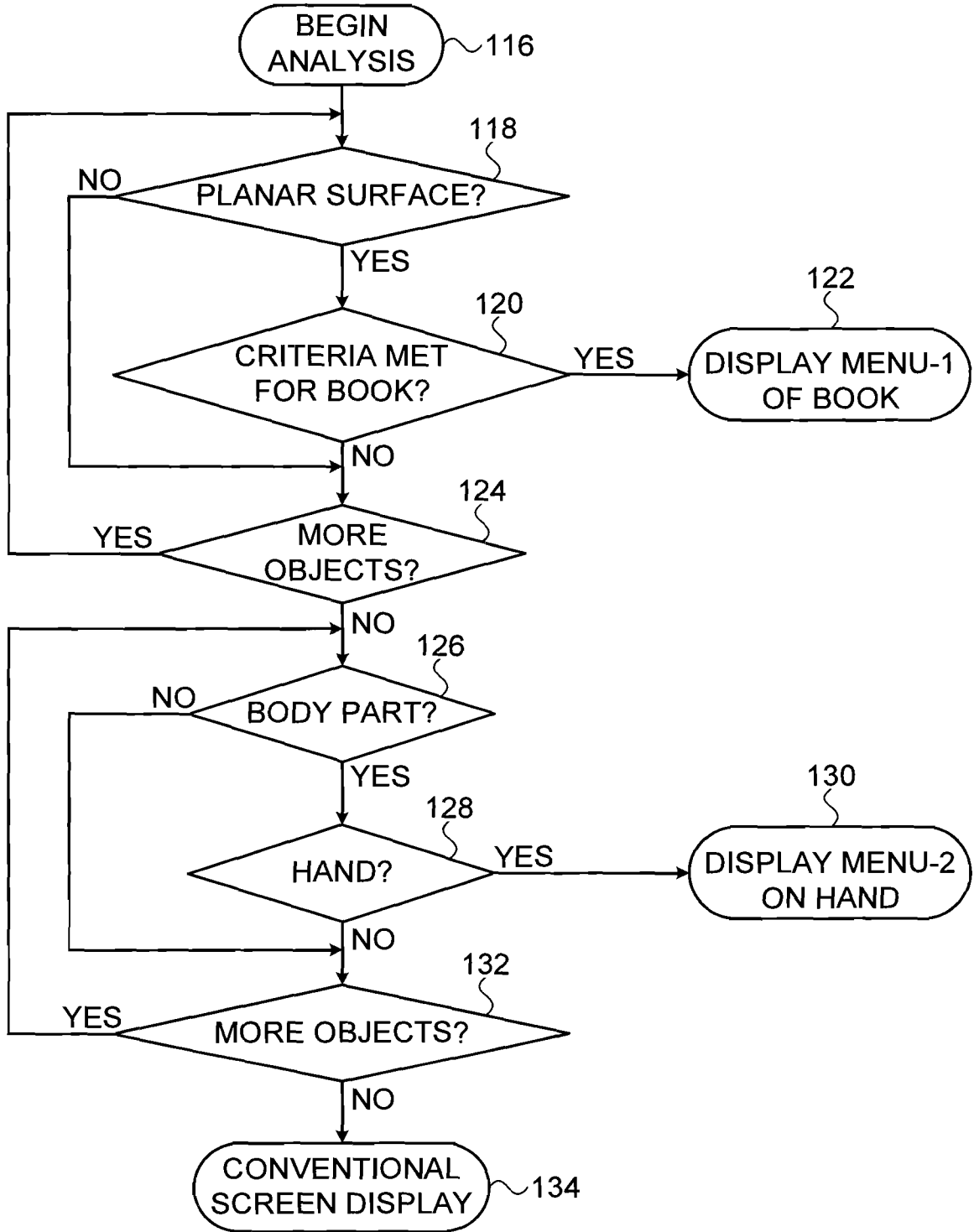
FIG. 1

FIG. 2

FIG. 3

FIG. 4

5/6

buy-$75.99

cancel
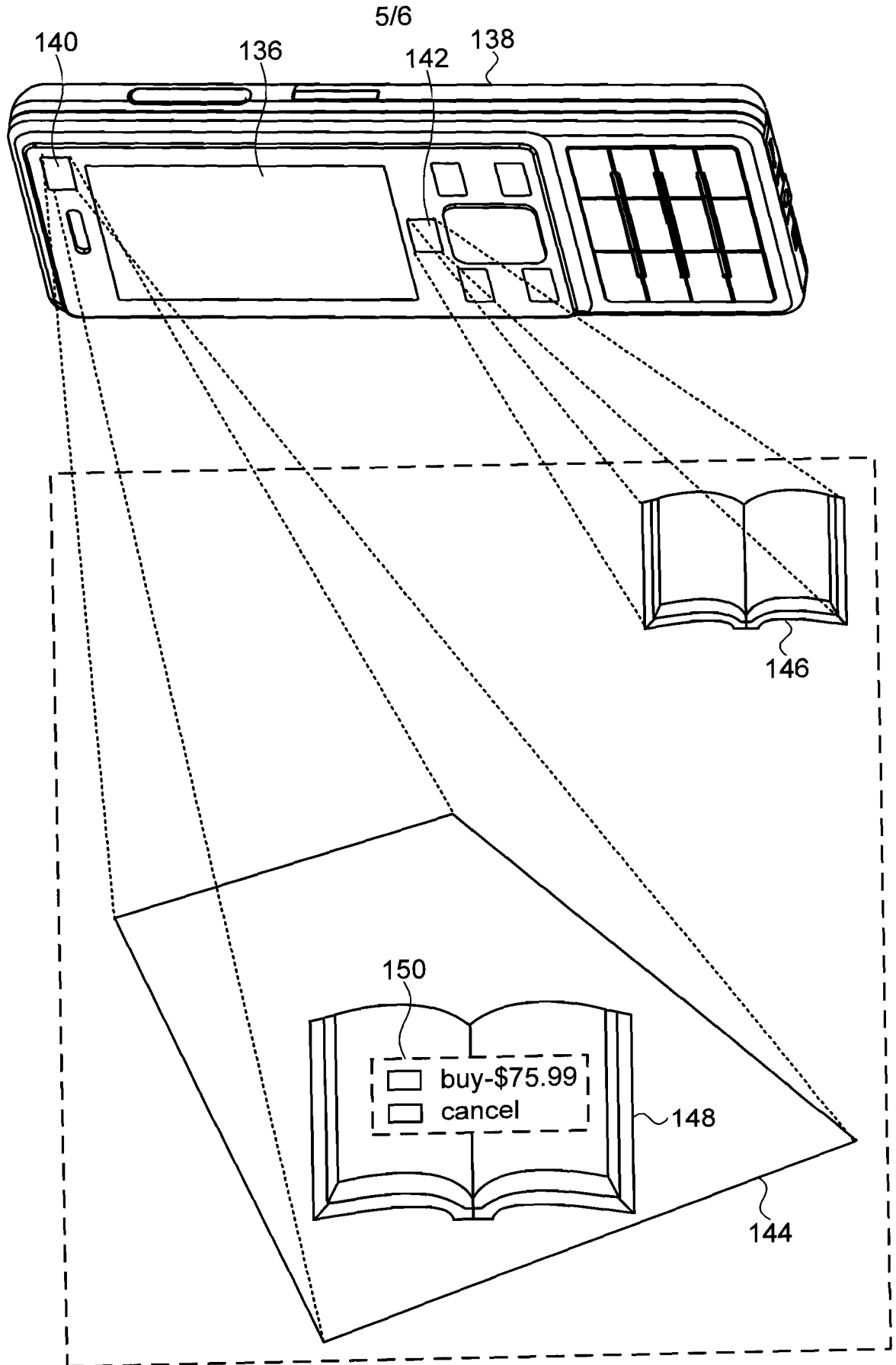
FIG. 5

FIG. 6

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G09G 5/00 (2011.01)
USPC - 345/633

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC: G09G 5/00 (2011.01)
USPC: 345/633

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
IPC: G09G 5/00 (2011.01)
USPC: 345/633; 345/418, 619, 629, 632; 382/103; 725/34, 37, 42, 62, 88, 89; 715/773 (keyword limited; terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
pubWEST(USPT,PGPB,EPAB,JPAB,USOCR); Google(Web); Search terms used: augment interactive virtual reality image object superimpose overlay GUI interface advertisement keyboard gaze perspective view angle direction projection compensation retina eyeglasses HUD HMD stereo mapping depth align pose 3D degrees of freedom convex non-planar gesture surface

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	US 2004/0258314 A1 (HASHIMOTO) 23 December 2004 (23.12.2004) entire document especially Abstract; para [0066]-[0070], [0110]-[0112]	1-33, 35-37, 39-46 ----------- 34, 38
Y	US 2004/0135744 A1 (BIMBER et al.) 15 July 2004 (15.07.2004) entire document especially Abstract; para [0136], [0150]	34, 38
X	US 2005/0254726 A1 (FUCHS et al.) 17 November 2005 (17.11.2005) entire document especially Abstract; para [0040], [0064], [0085], [0087], [0102]	1, 20

☐ Further documents are listed in the continuation of Box C.   ☐

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier application or patent but published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 November 2011 (17.11.2011)	**0 6 DEC 2011**

Name and mailing address of the ISA/US	Authorized officer:
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 · Facsimile No. 571-273-3201	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)