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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 December 2003 (24.12.2003)

PCT

(10) International Publication Number
WO 03/107039 A2

(51) International Patent Classification[7]: G02B

(21) International Application Number: PCT/IL03/00496

(22) International Filing Date: 12 June 2003 (12.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/387,901 13 June 2002 (13.06.2002) US

(71) Applicant *(for all designated States except US)*: **I SEE TECH LTD.** [IL/IL]; 33 Raziel St., 46812 Hertzliya (IL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SOFER, Eli** [IL/IL]; 33 Raziel St., 46812 Hertzliya (IL). **TIDHAR, Amnon** [IL/IL]; 44 Tzahal St., 56265 Yahud (IL).

(74) Agent: **LANGER, Edward**; Shiboleth, Yisraeli, Roberts, Zisman & Co., 46 Motefiore St., 65201 Tel-Aviv (IL).

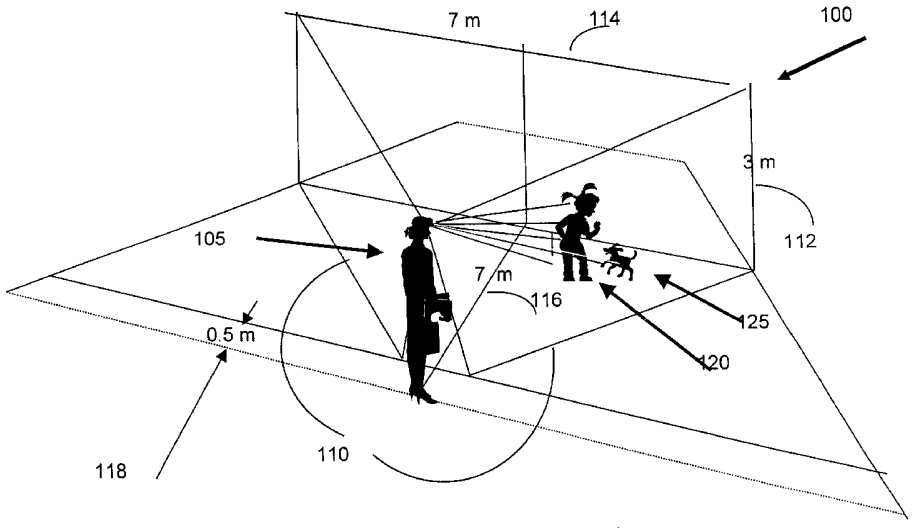
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHOD AND APPARATUS FOR A MULTISENSOR IMAGING AND SCENE INTERPRETATION SYSTEM TO AID THE VISUALLY IMPAIRED



(57) **Abstract:** An apparatus and method for aiding a visually impaired or blind person to detect, identify and avoid objects in the field of vision of the surroundings. The apparatus includes electro-optical devices to detect and identify objects, a control unit, which receives and processes information from the devices and a vocal representation unit to receive instructions from the control unit for purpose of audibly describing the objects to the person in the field of vision, thereby enabling the person to cope and proceed safely in the surroundings. The method for aiding a visually impaired person to detect, identify and objects in the field of vision surroundings includes the steps of receiving visual images of the objects, receiving range information concerning the objects, processing the image information, and audibly describing to the person the objects in the field of vision, thereby enabling the person to cope with, and proceed in his surroundings.

WO 03/107039 A2

METHOD AND APPARATUS FOR A MULTISENSOR IMAGING AND SCENE INTERPRETATION SYSTEM TO AID THE VISUALLY IMPAIRED

Field Of The Invention

The present invention relates generally to aiding a visually impaired person. More specifically, the present invention relates to a multisensor system designed to aid a blind, or otherwise visually impaired, person to detect, identify and avoid obstacles, and to assist him in finding his way to a specific target or to aid in negotiating spaces where one's view is partially or completely obstructed.

Background Of The Invention

Electronic and computerized systems for helping visually impaired and blind people are well known in the art. For example, US Patent No. 5,687,136 to Borenstein describesan ultrasound based system which gives a proximity warning and is embodied in a cane. The cane is equipped with sensors to calculate the distance to nearby obstacles and guide the user away from potential collisions. The cane is pushed by the user ahead of himself and it maneuvers itself around obstacles by means of two steerable wheels. However, when the user approaches an object, like a closed door through which he wishes to pass, the cane tends to guide the user away from the door. In an effort to solve the problem, the robotic cane's guidance can be overridden or modified via a joystick incorporated into the handle. Where there is direct conflict between the commands of the computer and the human - such as walking up to a door - the cane is set to respond to the human's commands. In situations that are less clear, the computer's and human's commands are combined. For example, where a person directs the cane to turn right but the cane senses an object in the way, the cane would turn slightly to the right, accommodating both the user and sensor inputs. A task which appears beyond the capabilities of the disclosed cane is that it cannot prevent a user from stepping off a curb into traffic.

US Patent No. 5,936,528 to Kubayashi, teaches a guidance and support apparatus for a visually handicapped person, comprising a resonance tag, antenna and portable scanner. The antenna transmits microwaves from 5 to 60 MHertz. The system provides voice messages, such as “go,” “stop” and”turn to left.”

Arkenstone is a provider of reading systems for people with visual and reading disabilities, and is used to find out exactly how to get from here to there, and what is encountered along the way. A product called The Talking Map adds software to a user's talking PC, with which the user can explore his neighborhood or city. Once the route is

determined in the privacy of the home, the directions are saved to a tape recorder. However, product provides no assistance in physical step-by-step navigation.

In general, the limiting factor on the amount and quality of information that can be generated by systems that sense and communicate via the electromagnetic frequency spectrum, which includes at the low end sound waves and ultrasonic frequencies just beyond audible sound at about 30 KHz, up through microwave frequencies and then lasers in the trillions of Hertz, or cycles per second, is roughly proportional to the frequency itself. Thus, laser-based communications has enormous potential for communicating detailed information.

State-of-the-art laser-based sensing systems include 3D laser imaging sensors and computerized vision sensors. A 3D laser imaging sensor mmay consist of a laser-based triangulation scanner designed to produce realistic 3D models rapidly. The resulting models are integrated with 3D editing tools. Laser triangulation is one of the most common techniques of 3D data acquisition. It is an active stereoscopic technique where the distance to the object is computed by means of a directional light source and a video camera.

A laser beam is deflected from a mirror onto an object being scanned. The complete object is scanned by incrementally changing first the horizontal angle, α, across a line, and then adding lines by incrementally changing the vertical angle, β. The process is identical to that used to scan an image on to a standard television receiver. For example, there may be 100 by 100 pixels, or scanned points per frame, and five frames are scanned per second. The object scatters the light, which is then collected by a video camera located at a known triangulation distance from the laser. For facilitation purposes the camera is modeled as an ideal lens and the charged coupling device (CCD) detector is modeled as flat. It can be observed that the angle and the pixel position of the scattered light are related. Since the focal length of the camera lens is known, the analysis of the resulting video image can determine the angle of the scattered light.

A pixel is defined as a 2D (2-dimensional) picture element, the smallest resolvable rectangular area of an image, either on a screen or stored in memory. Each pixel in a color image has its own brightness and color, usually represented as a triple of red, green and blue intensities. By contrast a voxel the smallest distinguishable box-shaped part of a three-dimensional space. A particular voxel is identified by the x, y and z coordinates of one of its eight corners, or its center. The term is used in 3D modelling.

The angle is also known, since it is the projection angle of the laser beam, with two angular coordinate components, α and , β. Thus, using simple trigonometry, the 3D spatial (XYZ) coordinates of a surface point can be determined. Hence the name triangulation.

Meta Exhibit 1008
Meta v. Mullen - Page 003

Simple 3D scanning produces single projection scans from a single viewpoint only. Thus, for 360° model creation, a scan fusion, or "gluing" algorithm is critical.

The first step in gluing is to manually select three corresponding points on the two scans to be glued. This provides a starting translation and rotation vector for the scan fusion algorithm. Since the two scans must have corresponding points, the scans, which are to be glued, must overlap at least ½". Once the points are selected the scan fusion algorithm provides a preview mode allowing the user to see a rough alignment of the fusion. Before starting the fusion process the user may select the scan's geometry or texture to be used as the overriding factor for fusion. Once activated, the fusion algorithm finds the best rotation and translation matrices of the models, "glues" them together creating a single geometry for the object, and finally, re-triangulates. The result is a new seamless geometric model with properly blended textures. This approach permits gluing of almost any object with distinguishing characteristics in either texture or geometry.

Computerized vision sensors with 3D sensing software should include:

- automated digitization with best next view and maybe path planning algorithms, coupled to automated motion when the sensor does have the possibility to move around the digitized object in an autonomous way;
- automated or a least semi-automated transformation and modelization of the raw data, in the form of:
  + cloud points,
  + triangular meshes,
  + parametric surfaces, and
  + high-level geometric primitives such as cylinders, planes or spheres; and
- visualization and editing of the raw and processed data.

Today this perfect piece of software does not exist, but subsets of it do exist. They are often specialized in a particular application and prove efficient enough in that context. Movies are among the first to use 3D geometry capture and modeling techniques.

The general principle of 3D sensing is that sensors receive light, or other forms of energy such as ultrasound or x-rays, from their environment and are supposed to construct a model of this environment. The model depends on the application:

Meta Exhibit 1008
Meta v. Mullen - Page 004

- In 3D sensing, the goal is usually a precise 3D geometrical model of the surface of objects. It is attempted to detect the reflection of light, etc. on the surface of objects;
- Sometimes the 'color' properties of a surface must also be digitized; and
- For inspection purposes, it is sought to map the physical properties inside an object, and to interpret changes in the incoming signal due to the changing physical properties inside an object.

A few years ago, mobile robotics were the main field of application for 3D sensing. Computer vision scientists were mainly interested in passive sensing techniques that were supposed to reflect the way the human eye is working. Passive, means that no energy is emitted for the purpose of sensing, it is only received. Such passive techniques include stereo vision and monocular shape-from-X techniques like shape-from-contour or shape-from-shading. The problem is that recovering 3D information from a single 2D image is an ill-posed problem. Thus, monocular techniques have to add a priori information such as surface smoothness to recover 3D data, a process known as regularization. Regularization is most often context dependent and is difficult to apply outside of the lab.

The majority of automated video identification is done with 2 dimensional modeling. Identification may be greatly enhanced by 3 dimensional recognition of contour through depth perception.

To overcome the above problems of passive sensing, active sensing techniques have been developed. Active means that properly formatted light, etc. is emitted, and then received, once it has been interacted with the object to digitize. Typically, light is emitted in the direction of an object, reflected on its surface and recovered by the sensor; and then distance to the surface is calculated using triangulation or time-of-flight.

Volume digitization techniques like computerized tomography (CT), magnetic resonance imaging (MRI) or ultrasound imaging also fall in this category.

Effective television mini-cameras also are generally expensive, costing $50,000. To be useful for helping the visually impaired and blind person, it is important to be able manufacture an effective camera for approximately $400.

Thus, the trend is to integration or fusion of 3D techniques, and there is therefore a need in the art for a system to provide increased efficiency, adaptability and versatility for the fusion of 3D and computer vision sensing for the benefit of visually impaired and blind people.

Meta Exhibit 1008
Meta v. Mullen - Page 005

Summary Of The Invention

It is an object of the present invention to provide methods and apparatus for increased efficiency, adaptability, economy and versatility in a multi-sensor system for the benefit of visually impaired and blind people in finding their way to a specific target.

It is a further object of the present invention to provide methods and apparatus for an improved 3 dimensional algorithm for sensing and identifying objects in the field of vision sensor confronting visually impaired and blind people.

It is a still further object of the present invention to provide methods and apparatus to enable the creation of an overall mental picture of the surrounding physical environment of a visually impaired or blind person, and to enhance communication concerning the environment.

An apparatus and method are described for aiding a visually impaired or blind person to detect, identify and avoid objects in the field of vision of the surroundings. The apparatus includes electro-optical devices to detect and identify objects, a control unit, which receives and processes information from the devices and a vocal representation unit to receive instructions from the control unit for the purpose of audibly describing the objects to the person in the field of vision, thereby enabling the person to cope and proceed safely in the surroundings. The method describes the aid for a visually impaired person to detect, identify and avoid objects in the field of vision surroundings, including the steps of receiving visual images of the objects, receiving range information concerning the objects, processing the image information and audibly describing to the person the objects in the field of vision, thereby enabling the person to cope with, and proceed in his surroundings.

These objects, and others not specified hereinabove, are achieved by an exemplary embodiment of the present invention, wherein the system of the present invention is designed to aid visually impaired or blind pople to detect, identify and avoid obstacles on their way and assist in finding their way to a specific target. The integration of two sensors, a 3D laser imaging sensor and a computerized vision sensor, create an overall mental image of the surrounding physical environment. The system of the present invention is compact and light weight, and can be mounted on head of the blind person.

The system of the present invention provides detection and identification of fixed or moving obstacles, and can provide estimates of the relative range, speed and bearings of

Meta Exhibit 1008
Meta v. Mullen - Page 006

several simultaneous obstacles and objects. The system of the present invention alerts the blind person of immediate danger and provides the needed information by synthesized voice.

The system of the present invention has small physical dimensions, allowing the user free movement. Object identification and description is provided for a range up to 20 meters, as well as alerts for potential close range obstacles between one and two meters away, including stationary and moving objects, and changes in ground level, for example, steps or a ramp. The system gives estimates of the relative speed of objects, such as "moving fast towards you," or "moving away", etc. A continuous verbal scene description is provided according to the relevance and level of threat. The system of the present invention has the ability to "concentrate" on a specific object in order to achieve a higher level of details, and has the ability to read signs. A capability to "learn" is built in, thereby continually adding to the stored environment model with new objects, relevant details and description methods.

The present invention will be more fully understood from the following detailed description of the preferred embodiments.

Brief Description Of The Drawings

In order to understand the invention and to see how it may be carried out in practice, a preferred embodiment will now be described, by way of non-limiting example only, with reference to the accompanying drawings, in which:

Figs. 1-1a are, respectively, a schematic spatial relation illustration of a user facing the field of vision, and a coordinate system, in accordance with an exemplary embodiment of the present invention;

Fig. 2 is a schematic system functional block diagram illustrating the aid to a blind person system, in accordance with an exemplary embodiment of the present invention;

Fig. 3 is a schematic illustration of the coordinate system for a user facing the field of vision, in accordance with an exemplary embodiment of the present invention;

Fig. 4 is a schematic illustration of the headset, in accordance with an exemplary embodiment of the present invention; and

Fig. 5 is a schematic illustration of the beltset, in accordance with an exemplary embodiment of the present invention.

Meta Exhibit 1008
Meta v. Mullen - Page 007

Detailed Description Of Preferred Embodiments

Reference is now made to Figs. 1-1a, which are respectively, a schematic spatial relation illustration **100** of a user/blind person **105** facing the field of vision **110**, and a coordinate system, in accordance with an exemplary embodiment of the present invention. The height **112** of field of vision **110** is seen to be 3 meters and the width **114** at the far end is 7 meters at a depth **116** that is also 7 meters, but is no closer than a minimum depth **116**. Two exemplary "objects" **160** are shown, a girl **120** and a dog **125**.

The system analyzes the surroundings according to the locality of anything present in the area within the boundaries, such as girl **120,** dog **125**, dog's tail **127** or the ground, walls, fence, building or treeline present in the area of field of vision **110**, static or moving. The system locates exceptions in the surfaces within the boundaries: a decline or elevation in the ground, a protrusion or recess, an opening in a wall, etc. The system identifies and recognizes movement, and has concern for any living beings, such as girl **120** and dog **125**, in field of vision **110**. The system gains knowledge by analyzing movements derived from objects positioned near user **105**.

Field of vision **110** of the system is defined by coordinate guidelines **150**, which originate from the axes of coordinate **300**, described hereinbelow in Fig. 3, which are located in front of user **105**, and the x-axis **305** in the direction at the start of operation. The system describes the objects in the surrounding area by giving them names.

Fig. 2 is a schematic system functional block diagram illustrating the aid to a blind person system, in accordance with an exemplary embodiment of the present invention.

Descriptions follow the system functional block diagram presented in Fig. 2, and will refer to the other figures in order to enhance the explanation.

The input unit **400**, illustrated in Fig. 4 comprises two sensors, CCD sensor **205** and 3D imager **208**, and an optional third sensor shown as CCD sensor **206**. 3D imager **208** emits sequences of laser beams **204** and measures the time of arrival of the return beams **203** in order to compute the distance to the point on the surface of any object in the field of vision. CCD sensor **205** is a camera that senses the color (as combination of Red, Green, and Blue) of the points on the surfaces of objects. In the optional arrangement, imager **208** is used in order to project a structured light grid on the objects in the field of vision, while both CCD sensors **205** and **206** are used to compute the distance to the object's points in addition to their roles as color sensing cameras.

Meta Exhibit 1008
Meta v. Mullen - Page 008

The sensors processor and control unit (SP&C) **209** performs the necessary integration between the sensed data streams from the sensors, and the computation needed in order to relate distance and color information to each grid point in the field of vision. Control unit **209** also functions to adapt the sensors to any required depth of the field of vision.

The input unit of Fig. 4 is wired or wirelessly linked via port **420** to the belt unit 500 illustrated in Fig. 5. This link enables bi-directional flow of information. The arrow **210** represents the flow of frames of 100X100 grid points, each point with its associated range and color information. Arrow **211** represents the control information flowing into sensors control **209**, requesting a required field of vision **110** setup.

The belt unit, as illustrated in Fig. 5, contains three functional modules: the High level vision processing module **216**; the World Model knowledge base **221**; and the User Interface module **234**.

The World Model (WM) **221** includes the description of known objects, their relations to their components, and the relations among the components. In addition, WM **221** includes the grouping of known objects into classes, the common properties of classes, and the common relations between classes of objects. For example, a specific chair will be described as composed of four column shaped components (legs), arranged as a rectangle with a flat horizontal board on top (seat). It belongs to a class of objects named **Chair**. This class has common physical dimensions, typical ranges of height width and typical behavior in time – it is stationary. The class **Chair** has a typical relationship with the ground: **on top**, and a common relationship with the class of objects **Table**: **near**.

The atom elements (the lowest level elements with no subcomponents) in WM **221** are the set of regular spatial surfaces. i.e. the planes and smooth surfaces of the second order.

Stream of frames **210** is analyzed in order to decompose each frame into regular blocks **212**. The sequence of frames contains range and color information of sensed spatial points. Each block is characterized by its physical dimension, general shape description, color, and/or texture property. Movements of blocks are tracked along a continuous sequence of frames. A moving average is calculated for each sequential group of several frames, thereby smoothing any effect of random head movements. The smoothed movement vector is added to the other regular surface properties, and the complete list of properties is passed on to the Object Synthesis and Recognition module **214**.

In Object Synthesis and Recognition module **214**, relations between surfaces are computed, and a bottom up search into WM **221** is initiated as represented by arrow **223**, in order to find an object class description that matches any subset of the surfaces and the

Meta Exhibit 1008
Meta v. Mullen - Page 009

relations among them. In turn, subsets that were identified as belonging to an object class are going through the same process in order to find a match with a more complex (higher in the hierarchy) object class. The subset of surfaces is then regarded as recognized, it is named as belonging to the object class and its output to the next sub-module is represented by arrow **250**. The process continues until all the surfaces are interpreted as belonging to the highest-level object class possible.

The recognition process makes use of all available information that can be extracted from the sensed data. Internal structure or shape is the way the object is composed of lower level (less complex) components, e.g. the way **a bench** is composed of the legs seat and backrest, and the relations among them. The object's measurable properties are its physical dimensions, dominant color and dominant texture. For example, the typical size of a bench is 1 meter wide and 130-150 cm high. The object's typical relations with other object classes in the surrounding neighborhood include physical relations like **beside**, **on top**, **above** etc.
The recognition process follows the principle that recognition is achieved if the information available is enough to define it as belonging to a unique object class in WM **221**. If the recognition using this principle is not satisfactory a correction is made (see below) the WM **221** is enriched and a learning process takes place. Hence, in a moderately rich WM **221** the following simple examples are feasible:

- A regular recess in the ground – is a curb of the sidewalk.
- An object moving at knee height – is a dog.
- A green object on the ground about 140 cm high and 150 wide – is a bench.
- A hole in a vertical surface 190 cm high and 80 cm wide – is an entrance.

The list of object classes identified, together with their position information, physical dimensions, and relative velocity vectors, constitute an internal scene description **225**. Scene **225** description is then ordered according to distance from the user and filtered according to the mode of operation **215** as described hereinbelow. The filtered scene description is passed, as represented by arrow **229**, to a text to speech translator **228,** located in interface unit **234** where a verbal description is generated and transmitted through the headphones **230,** which are physically a component of the head mounted unit Fig. 4.

In the case where an object is identified as belonging to the object class of **Signs,** control is passed to the OCR **233,** which attempts to interpret words in the sign, and return the words interpreted as part of internal scene description **225**.

The World Model enrichment module (WME) **219** is responsible for the continuous adaptation of WM **221** to new experience (The "learning" process). Adaptation is carried out in several modes:

Meta Exhibit 1008
Meta v. Mullen - Page 0010

- **Direct teaching**: A part of a scene is given a name by the user. This means that a number of recognized objects fulfilling a group of relations among themselves, themselves constitute an object belonging to an object class by the given name. It might be a new object class to WM **221**, in which case the given part of the scene is the first sensed instance of a particular variation of the object class. And, it might be a new development for an existing class, which means that the object class definition has to be modified in order to include a variation of the newly defined object. For example, a part of a total scene is composed of two vertical boards and a group of several horizontal boards between them is named by the user **a cabinet**. If it is the first time this name is used a new object class is generated by this name. If an object class **a cabinet** already exists, grouping some other variations of cabinets, such as having a closed door, the object class definition will be broadened in order to include the new type of cabinet in the definition.
- **Generalization (by positive enforcement or by default)**: A part of the scene is identified as belonging to a specific object class, although some of the relationships, or some of the components, deviate from the definition in WM **221**. Cases occur where this specific object class is still the best match for the given segment of the scene. Either acceptance by default, or by positive confirmation of the user, causes the process to generalize the set of definitions of the object class in order to include the newly sensed segment of the scene as an included variation. For example, a segment of the scene, which consists of four column shaped objects and a flat board on top, is identified as belonging to the object class **a chair**, although the definition of this object class includes the existence of a backrest. No response or confirmation by the user causes the process to generalize the set of definitions of **a chair** to include such stool type chairs.
- **Refinement (by negative enforcement or correction)**: If a part of the scene is identified as belonging to an object class, and the user rejects the identification, the process refines the set of definitions to permanently exclude such identification in the future. However, if the user makes a specific correction by later accepting identification of this segment of the scene,

Meta Exhibit 1008
Meta v. Mullen - Page 0011

it will be treated as an example of the direct teaching case above.

Hence, starting from a given set of object classes, which represent regular surfaces, WM **221** is enriched by presenting objects to the sensor, naming them, confirming right identifications, and correcting wrong identifications throughout the routine usage of the system.

The task & mode manager (T&MM) **213** is the control module that coordinates the various components according to the mode of operation dictated by the user. The following are the various modes of operations made possible by the module:

- **Complete scene description.** The T&MM **213** sets the regular surfaces decomposition (RSD) module **212** to pass on the complete set of surfaces in field of vision **110**, each time adding new surfaces entering field of vision **110** as a result of conscious head movements. And, T&MM **213** sets internal scene description module (ISD) **225** to pass the full scene description, represented by arrow **229**, to the translator **228.**
- **On going description.** T&MM **213** sets ISD **225** to pass on totranslator **228** only incremental changes in field of vision **110**.
- **Obstacle avoidance.** T&MM **213** sets RSD **212** to relate only to surfaces of up to 3 meters in range or moving towards the user, for example.
- **Focus mode.** T&MM **213** sets SP&C **209** to narrow the spatial angle of field of vision **110** around the central axis, and updates RSD **212** and OSR **214** on the change of setting. Hence the sensors acquire more detailed information on the object in the center of field of vision **110**. The various "teaching modes" described above take place while the system is in the focus mode, "concentrating" on a single object.
- **Search mode.** T&MM **213** sets OSR **214** to search for a specific object or object class in field of vision **110**. This means that a top down search is being carried out, and surfaces not relevant to these are being disregarded. T&MM **213** sets ISD **225** to pass on only one object or none at all (search failed).

Meta Exhibit 1008
Meta v. Mullen - Page 0012

T&MM **213** interfaces to the user via the path represented by arrow **226** by means of the control unit **227**, which integrates the inputs from both a control panel **232** and the microphone **231.** The user instructions as to the mode he wishes the system to operate in, are given either by punching keys on the control panel, or by voicing a verbal command. In the same manner, naming a new object or object class, approving, and rejecting are done either by the control panel or by voice. Control unit **227** operates the decoding algorithms and the voice recognition engine. Control unit **227** decodes the instructions or the responses, and passes them on to T&MM **213**, which in turn sets up the various modules, or passes the proper response to WME **219**.

Fig. 3 is a schematic illustration of the coordinate system **300** for a user facing the field of vision, in accordance with an exemplary embodiment of the present invention. The x-axis **305** and the y-axis **310** form a horizontal plane. The z-axis **315** is vertical. By integrating along the time axis, which is a non-visible "fourth dimension," relative velocity vectors are calculated, which are associated with the polygons. The system provides position measurements for the distance to object **155**, i.e., **r** the range **302** and the exact vertical angle **320**, α, and the exact horizontal angle **322**, β, wherein each angle is measured to within 0.3°. Each pixel is expressed as [R, G, B, r](α, β, t), where R, G, B represent primary color components red, green and blue, r is the range the distance, and t is the time. Thus each of the four values for each pixel – R, B, G and r - are a function of α, β and t. Vertical angle **320**, α, varies from +10□ above the horizontal to -80° nearly straight down, and horizontal angle **322**, β varies from -45° to +45°.

Fig. 4 is a schematic illustration of the headset **400 for use with the present invention**, in accordance with an exemplary embodiment of the present invention. Having temples **405** in the manner of glasses, and worn similarly to glasses, headset **400 is preferably constructed so that it** does not inhibit the head movements of user **105**. Video sensor **205**, i.e., a CCD chip or miniature television camera, and scanning range-finder **215** are mounted in the headset carrying compartment **415**. A pair of earphones **410** and a microphone are sub-components of headset **400**. A small processing unit, not specifically shown, carried in the beltset **500** of user **105**, is connected by a link (wired or wireless) to port **420**, which may be a physical input port or a wireless port for example of the kind used for short range infrared transmissions (such as required by for Bluetooth specification), and therefrom is connected to sensors **222**, **224** and **226** and earphones **410**. In a preferred embodiment link **420** is wireless, using infrared (IR) communication, for example, which is quite over the fraction of a meter distance concerned. Preferably, headset **400** does not have external parts detracting from the appearance of user **105**, or parts sensitive to environmental conditions.

Meta Exhibit 1008
Meta v. Mullen - Page 0013

So as not to block audibility of ambient sounds, so important to a blind person, earphones **410** are designed so as not to fit directly into or over the ears, and/or only a single earphone **410** is provided.

Functions embodied by headset **400** are not necessarily embodied specifically by a device worn directly on the head in the manner of eyeglasses.

Fig. 5 is a schematic illustration of a beltset **500**, in accordance with an exemplary embodiment of the present invention. Beltset **500** carrying a compartment **505** includes the Processing and Computation Unit and the power supply. Beltset **500** is coordinated to be worn conveniently by means of the pants belt **510**. Beltset **500** contains a number of operating switches and dials **515**, which are optionally designed to detach and hold conveniently in the hand. The operating switches **520 and 525** on the front of beltset **500** are used to dictate the mode of operation of the device.

It will be appreciated that the preferred embodiments described above are cited by way of example, and 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 which would occur to persons skilled in the art upon reading the foregoing description, and which are not specifically disclosed in the prior art.

Meta Exhibit 1008
Meta v. Mullen - Page 0014

What is claimed is:

1. A system for aiding a user with a visual impairment or visual obstruction, comprising:

   means for sensing time/space characteristics and physical characteristics information about an object in a field of view;

   means for interpreting said time/space characteristics and physical characteristics information and for characterizing (recognition) the object; and

   audible information delivery device (Means for communicating) for audibly describing to the user the characterization of the object and said interpretation about the object.

2. The system according to claim 1, wherein said means for sensing further comprises devices selected from the group consisting of: CCD sensors, laser imaging sensors, radar sensors, electromagnetic radiation sensors.

3. The system according to claim 1, wherein said time/space characteristics and physical characteristics information about an object includes any of: physical dimensions, general shape description, texture, color, the distance and position of said user from said object, motion of said object, spatial relationships between objects.

4. The system according to claim 1, wherein said means for interpreting said time/space and physical characteristics information comprises: a sensors processor, a control unit, and an Object Synthesis and recognition system.

5. The system according to claim 1, wherein said characterizing(recognition) is performed by a World Model.

6. The system according to claim 5, further comprising means for correcting and learning when said recognition by said World Model fails.

7. The system according to claim 5, wherein said learning process by a World Model enrichment learning process comprises:

   direct teaching.

   generalization teaching; and

   refinement teaching.

8. The system according to claim 1, wherein said audible information delivery device includes a text to speech translator, and headphones or a speaker.

9. The system according to claim 5, wherein CCD data derived from an object identified as belonging to the object class of signs is processed by OCR software.

Meta Exhibit 1008
Meta v. Mullen - Page 0015

10. Apparatus for aiding in obstacles avoidence , comprising:

at least one electro-optical device to detect and identify said at least one object;

a processing unit, which receives and processes information from said devices; and

a user communication module to receive instructions from said processing unit for the purpose of audibly describing to said person said at least one object in said field of vision, thereby enabling said person to cope with and proceed in said surroundings.

11. The apparatus according to claim 10, wherein at least one of said at least one electro-optical device is a computerized vision device.

12. The apparatus according to claim 11, wherein at least one of said at least one computerized vision device is a CCD sensor.

13. The apparatus according to claim 10, wherein said at least one electro-optical device is a laser imaging sensor.

14. The control unit according to claim 10, wherein said processed information further comprises images.

15. The processed information according to claim 14, further comprising image understanding.

16. The apparatus according to claim 10, further comprising a knowledge database.

17. The apparatus according to claim 10, further comprising a knowledge editor.

18. The apparatus according to claim 10, further comprising a communications and learning subsystem.

19. The apparatus according to claim 10, wherein said apparatus is worn by said person.

20. The apparatus according to claim 19, wherein said apparatus comprises a headset.

21. The apparatus according to claim 20, wherein said headset comprises said at least one electro-optical device.

Meta Exhibit 1008
Meta v. Mullen - Page 0016

22. The apparatus according to claim 20, wherein said headset comprises said vocal representation unit.

23. The apparatus according to claim 20, further comprising a beltset.

24. The apparatus according to claim 23, wherein said beltset comprises said control unit.

25. The apparatus according to claim 24, further comprising manual controls which are detachable from said beltset.

26. The apparatus according to claim 10, further comprising an automatic speech recognition unit for verbal input from said person to said information processing.

27. A method for detecting and identifying at least one object in a field of vision, and notifying the user, comprising the steps of:
acquiring a plurality of visual images of said at least one object;
processing said plurality of images and said information; and
audibly describing to said person said at least one object in said field of vision.

28. The method according to claim 27, and further comprising controlling the adaptive sensitivity of the brightness of illumination.

29. The method according to claim 27, and further comprising range and features extraction.

30. The method according to claim 27, and further comprising analyzing said plurality of visual images into convex polygons.

31. The method according to claim 27, and further comprising 3D object rendering.

32. The method according to claim 27, and further comprising identifying said at least one object.

33. The method according to claim 27, and further comprising incorporation of a knowledge database to interactively help said person identify said at least one object.

34. The method according to claim 27, and further comprising the automated layout of a virtual world image.

35. The method according to claim 27, and further comprising the verbal description of said virtual world image to said person.

Meta Exhibit 1008
Meta v. Mullen - Page 0017

36. A system for aiding a user with a visual impairment or visual obstruction, comprising:

means for sensing time/space characteristics and physical characteristics information about an environment in a field of view;

meansfor interpreting said time/space characteristics and physical characteristics information and for characterizing (recognition) the environment; and

audible information delivery device (Means for communicating) for audibly describing to the user the characterization of the environment and said interpretation about the environment.

Meta Exhibit 1008
Meta v. Mullen - Page 0018

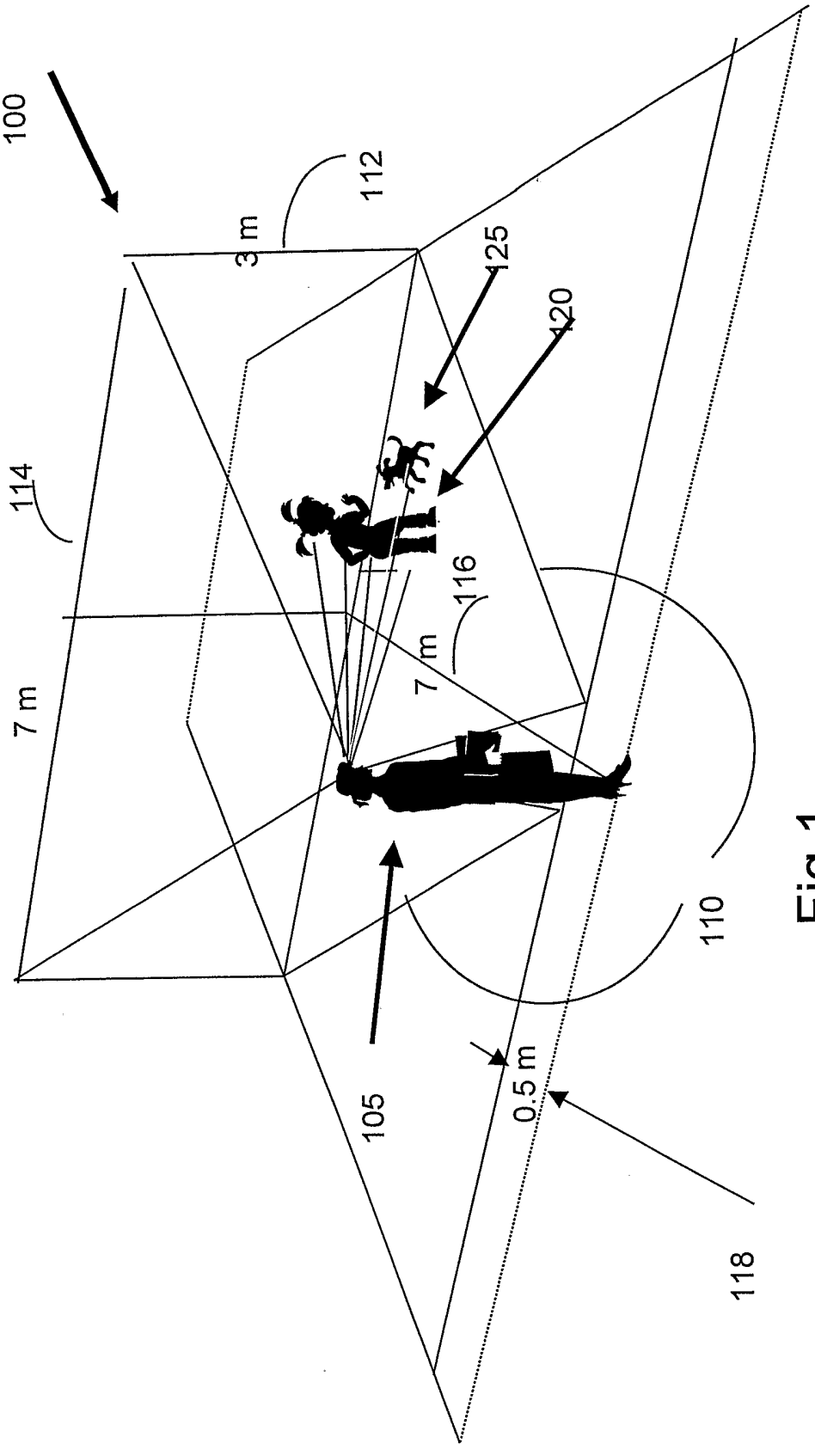


Fig 1

Meta Exhibit 1008
Meta v. Mullen - Page 0019

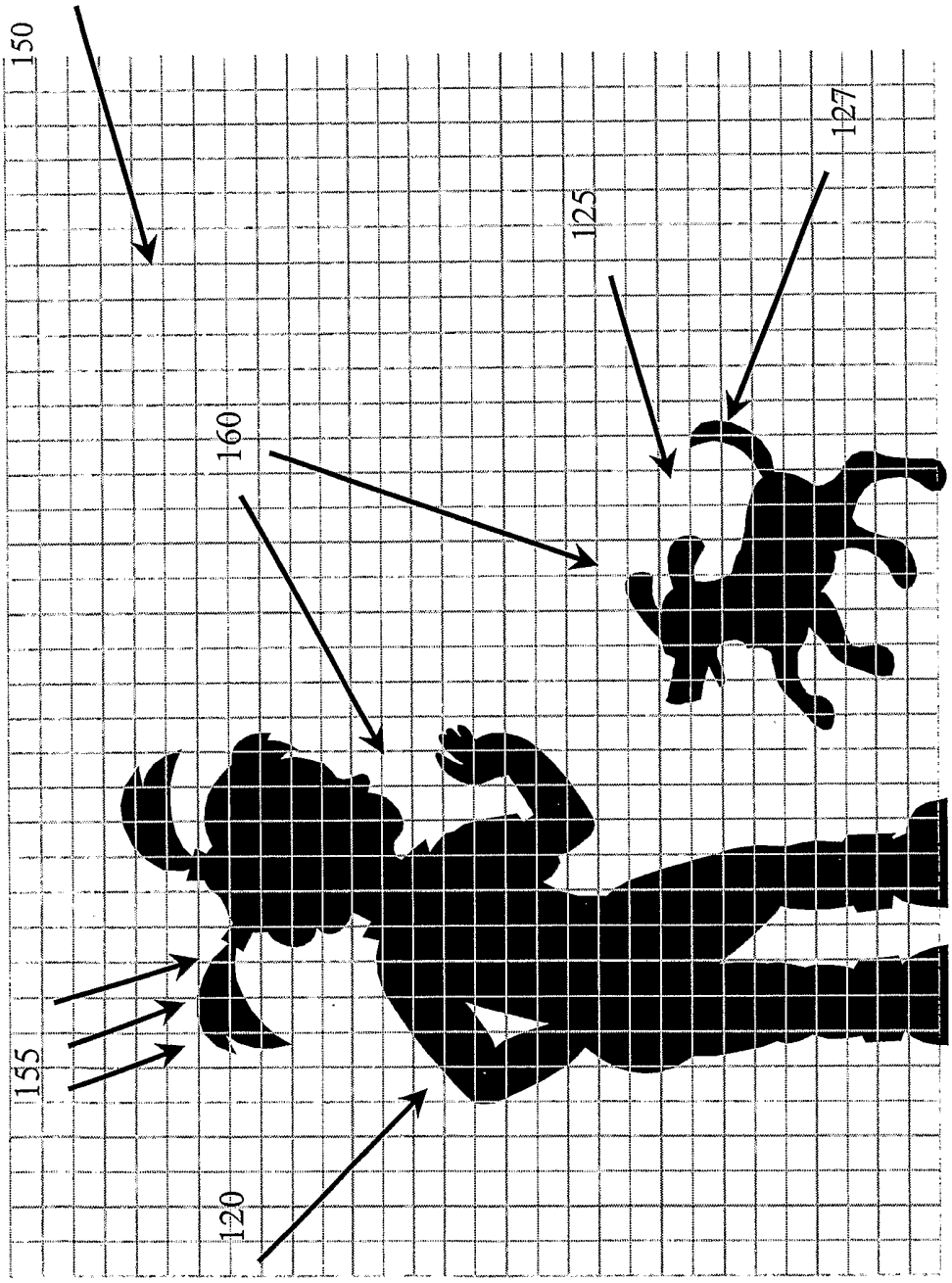


Fig 1a

Meta Exhibit 1008
Meta v. Mullen - Page 0020

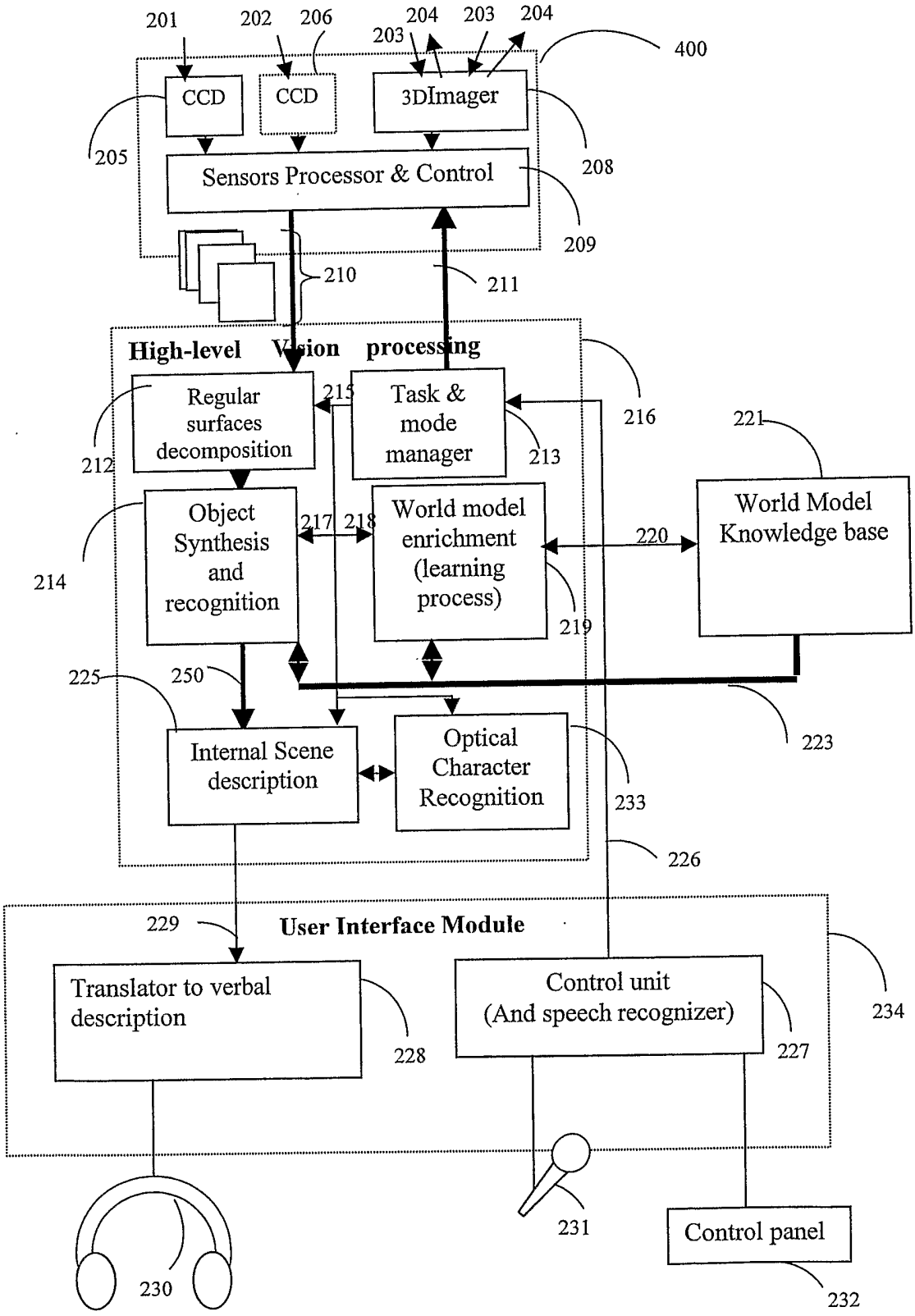


Fig. 2

Meta Exhibit 1008
Meta v. Mullen - Page 0021

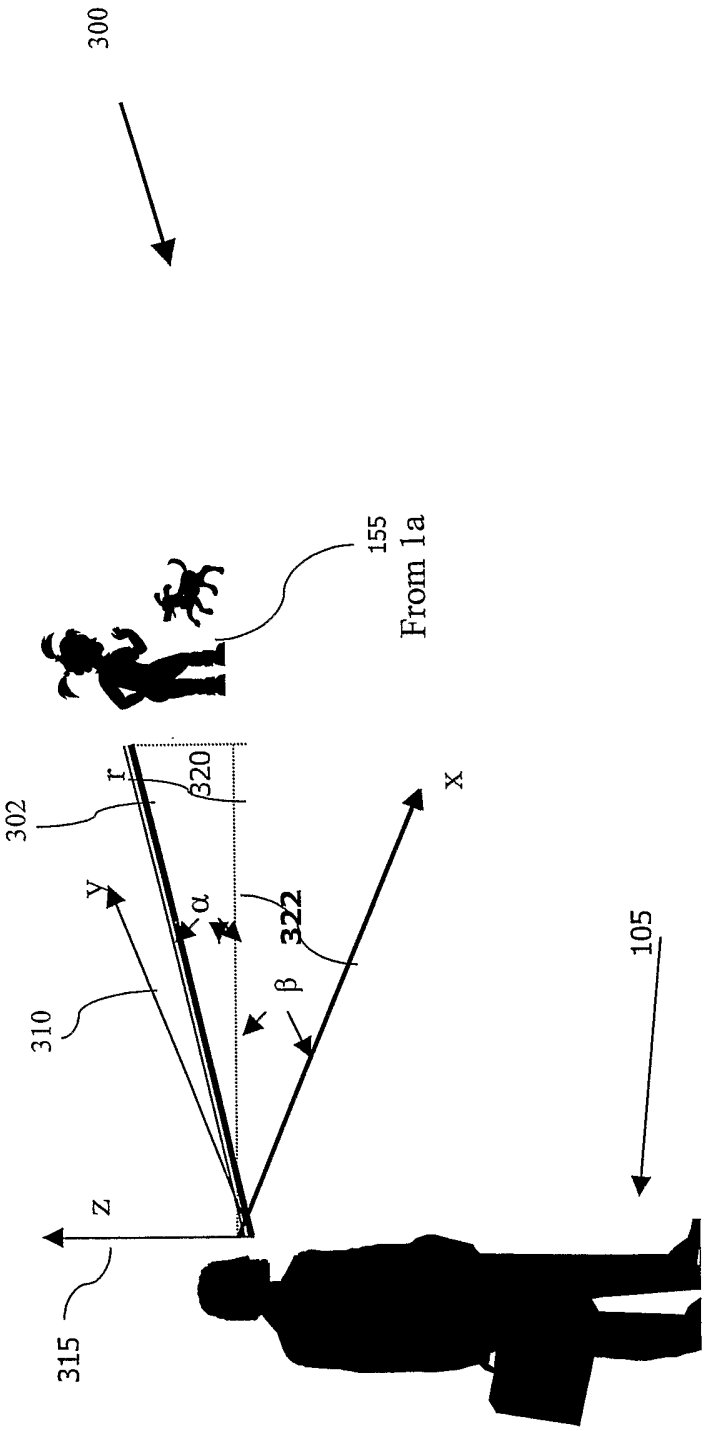


Fig 3

Meta Exhibit 1008
Meta v. Mullen - Page 0022

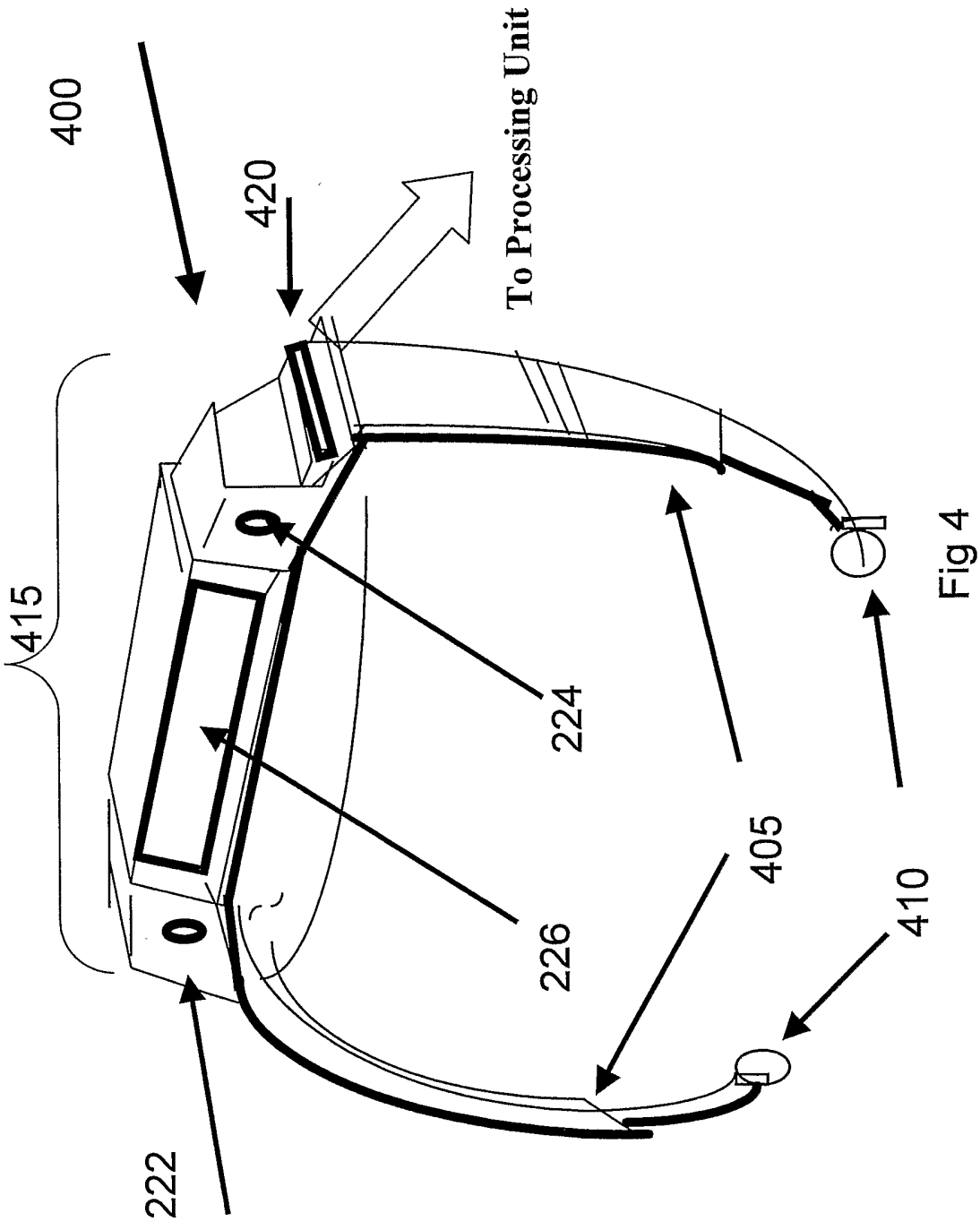


Fig 4

Meta Exhibit 1008
Meta v. Mullen - Page 0023

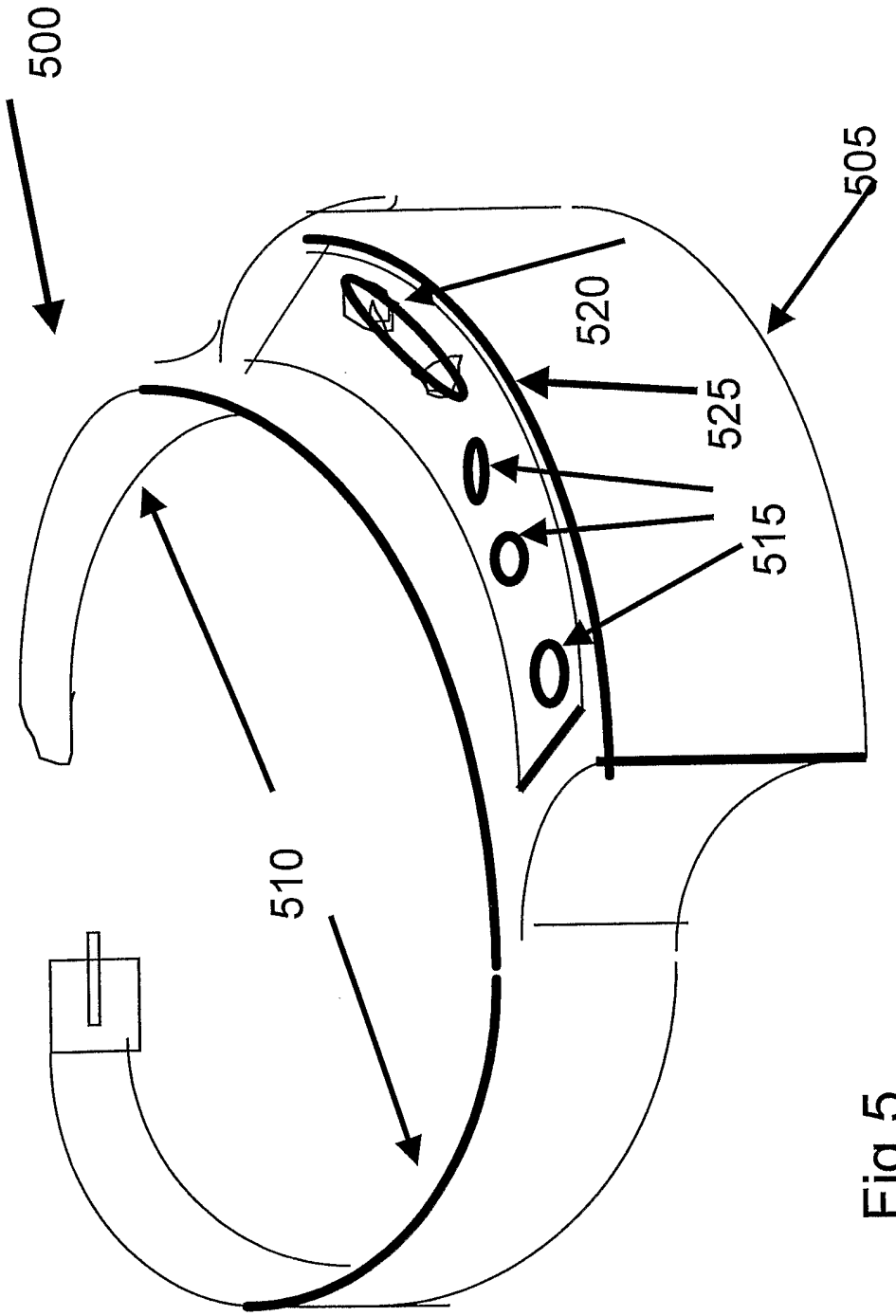


Fig 5

Meta Exhibit 1008
Meta v. Mullen - Page 0024