

US 20050031169A1

(19) **United States**

(12) **Patent Application Publication**
Shulman et al.

(10) **Pub. No.: US 2005/0031169 A1**

(43) **Pub. Date: Feb. 10, 2005**

(54) **BIRDS EYE VIEW VIRTUAL IMAGING FOR
REAL TIME COMPOSITED WIDE FIELD OF
VIEW**

Related U.S. Application Data

(60) Provisional application No. 60/493,579, filed on Aug. 9, 2003.

(76) Inventors: **Alan Shulman**, Santa Rosa, CA (US);
Donald R. Snyder III, Crestview, FL
(US)

Publication Classification

(51) **Int. Cl.⁷** **G06K 9/00**; G06K 9/32;
G01C 21/26

(52) **U.S. Cl.** **382/104**; 382/294; 701/200

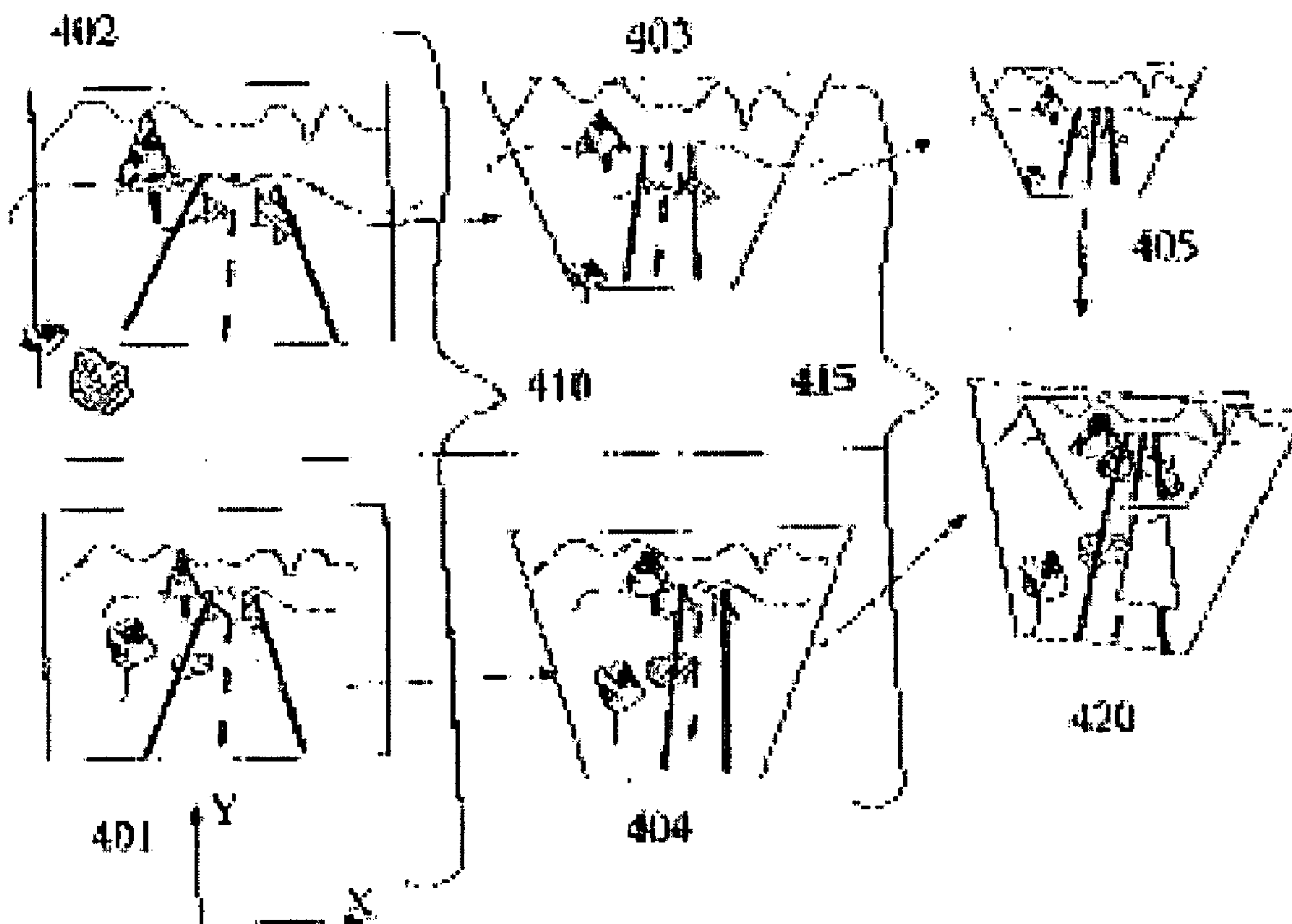
Correspondence Address:
DERGOSITS & NOAH LLP
Attn: **Michael E. Dergosits**
Suite 1450
Four Embarcadero Center
San Francisco, CA 94111 (US)

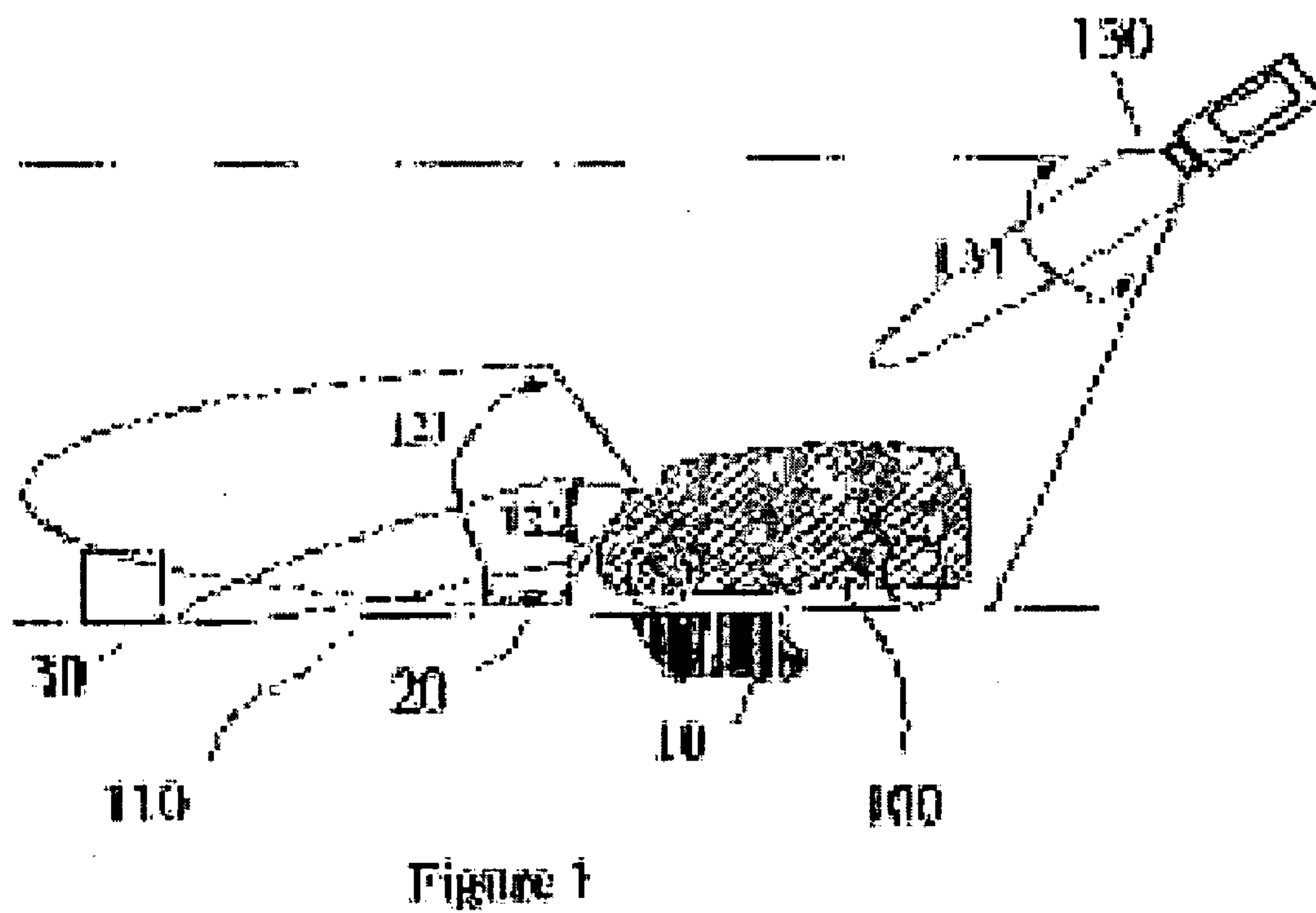
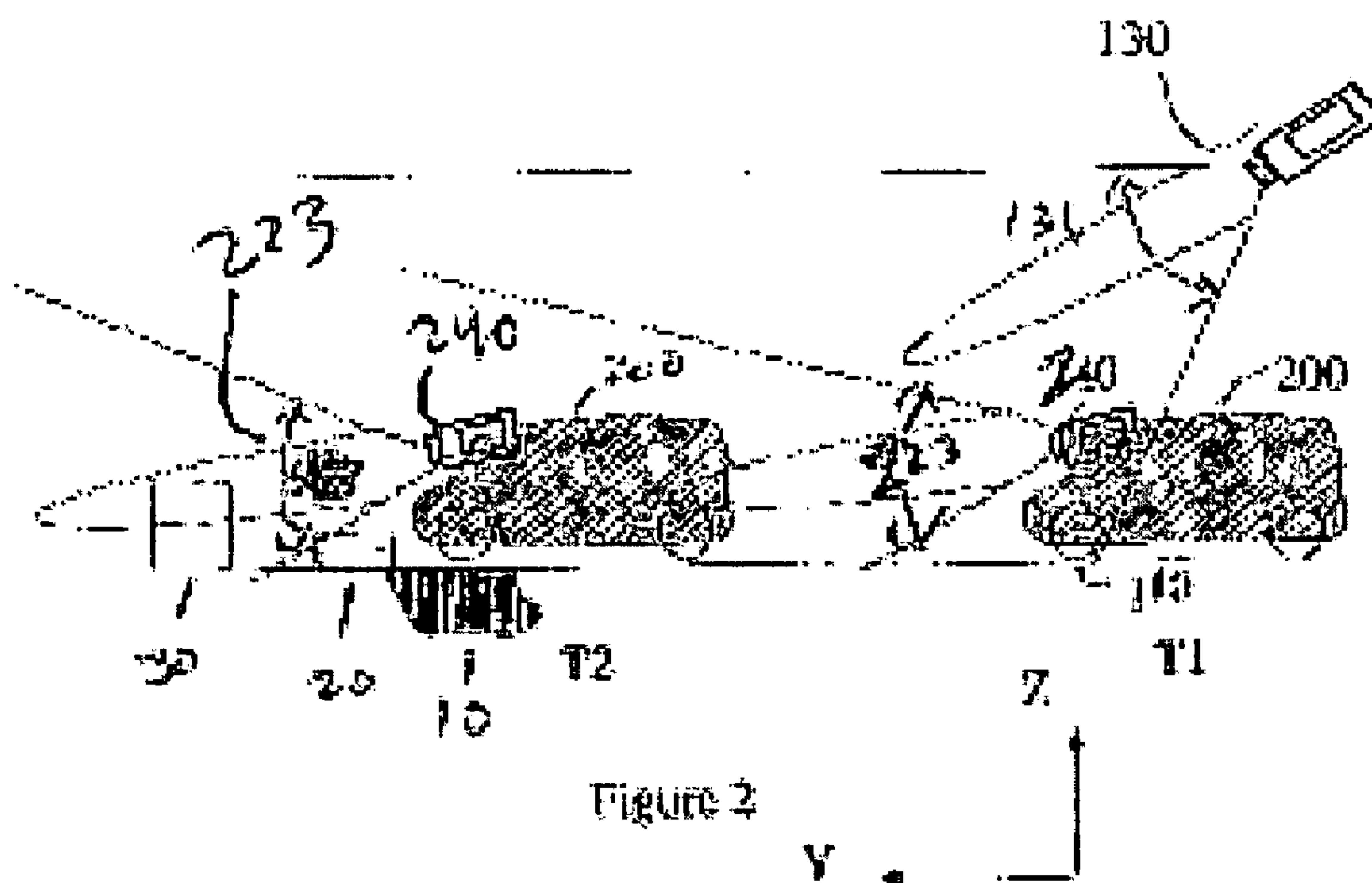
(57) **ABSTRACT**

A live image and a previously acquired or generated image are superimposed or composited to represent a virtual vantage point for flying, driving or navigating a plane, vehicle or vessel.

(21) Appl. No.: **10/914,375**

(22) Filed: **Aug. 9, 2004**





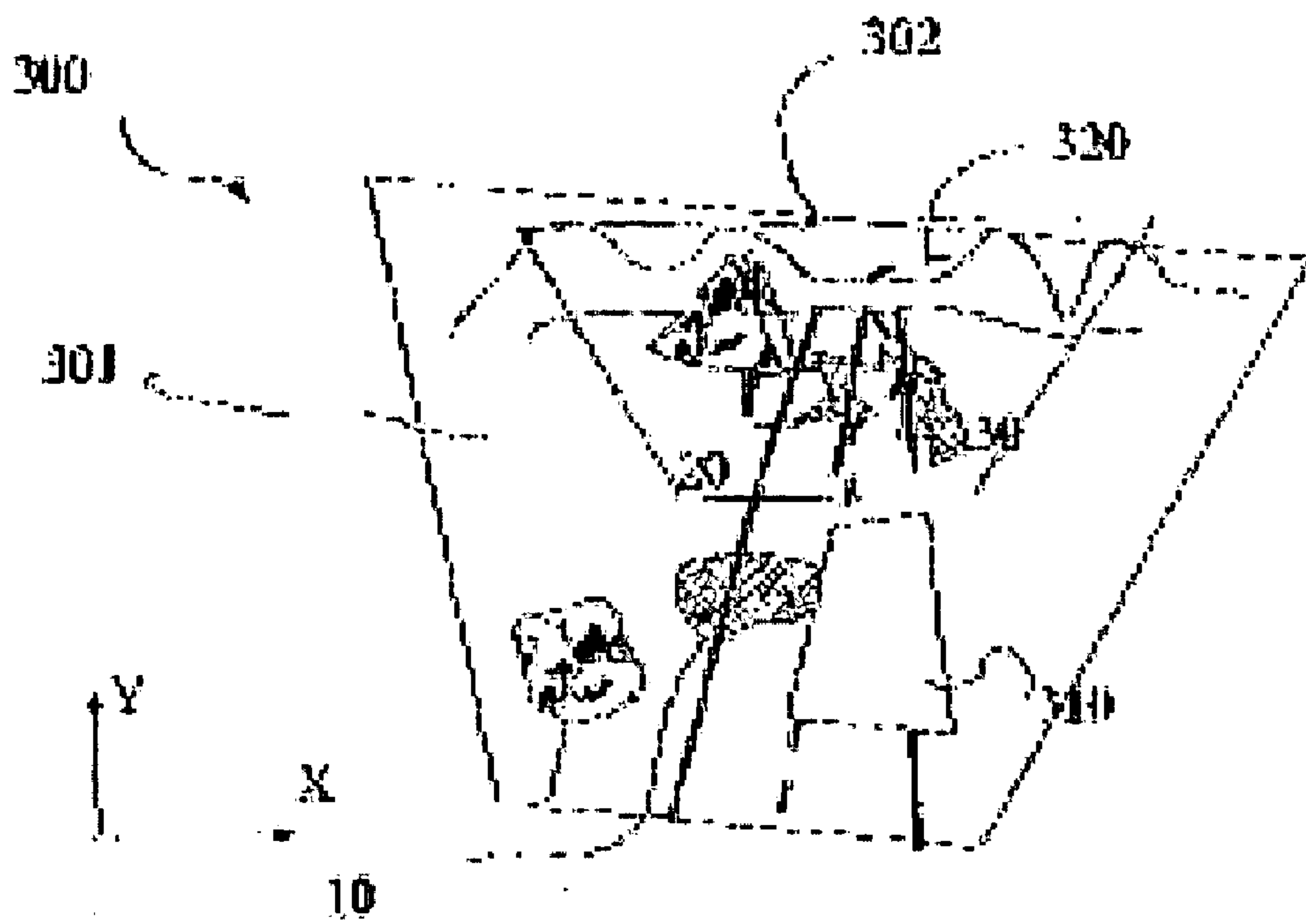


Figure 3

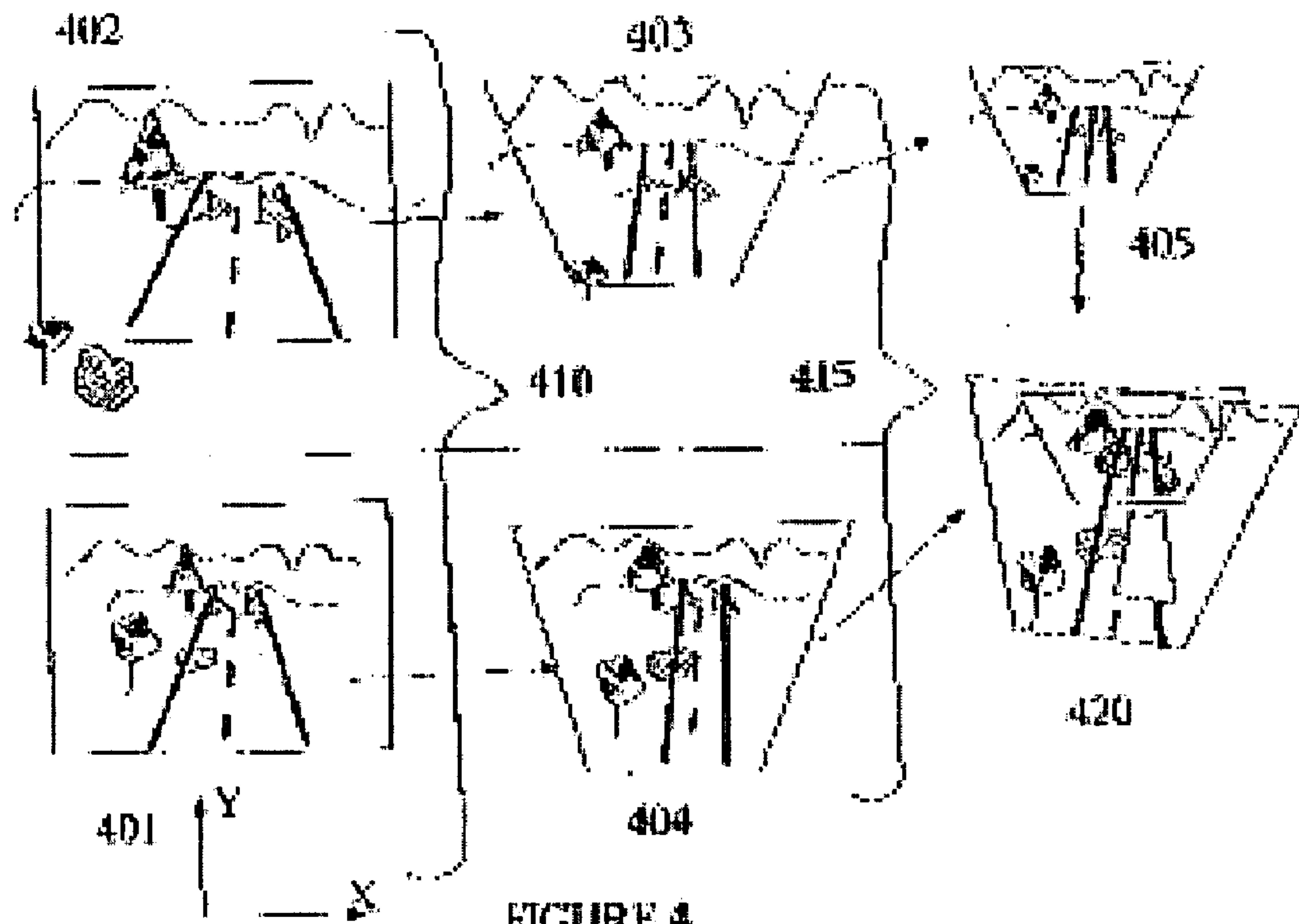


FIGURE 4

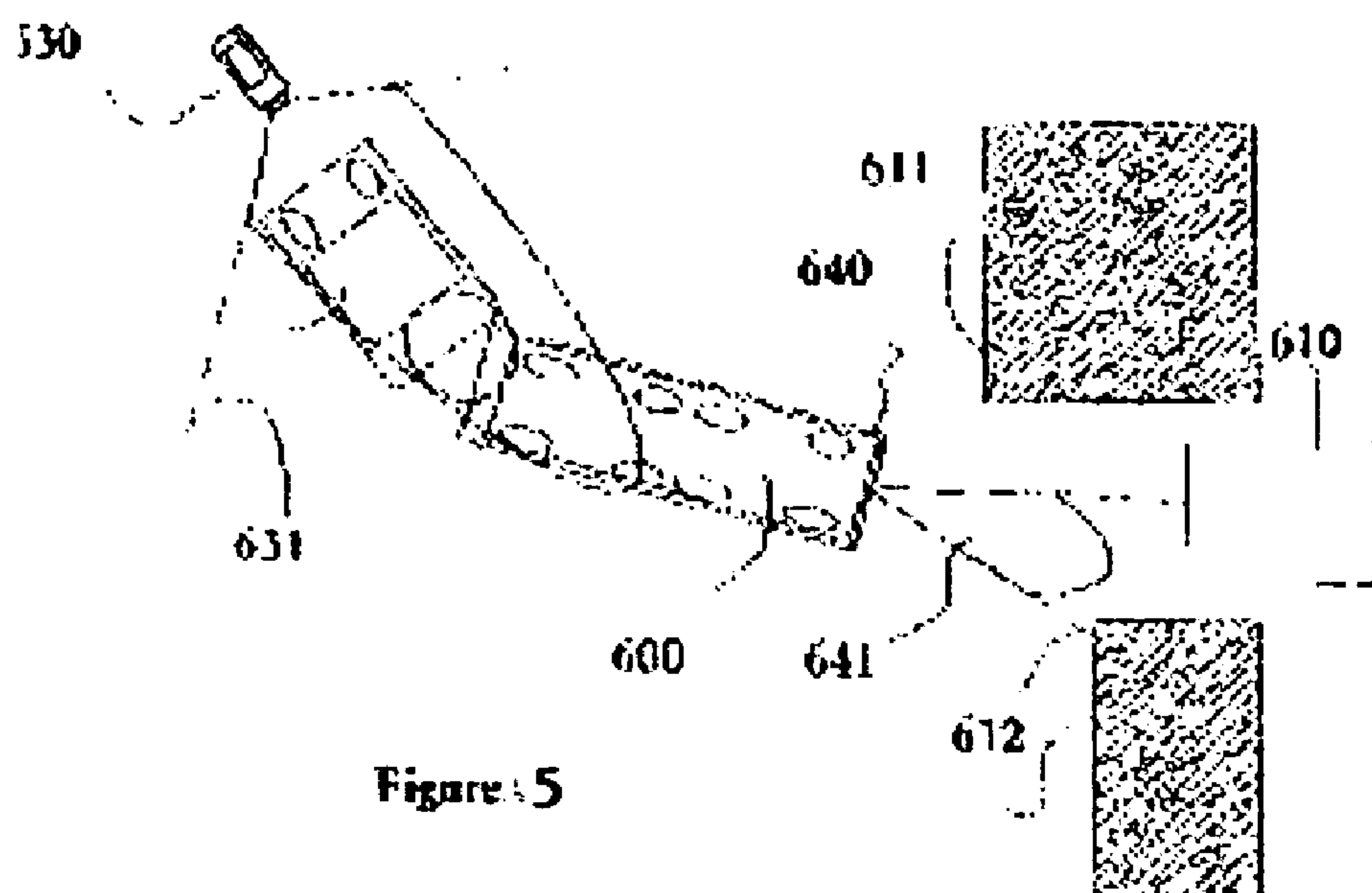


Figure 5

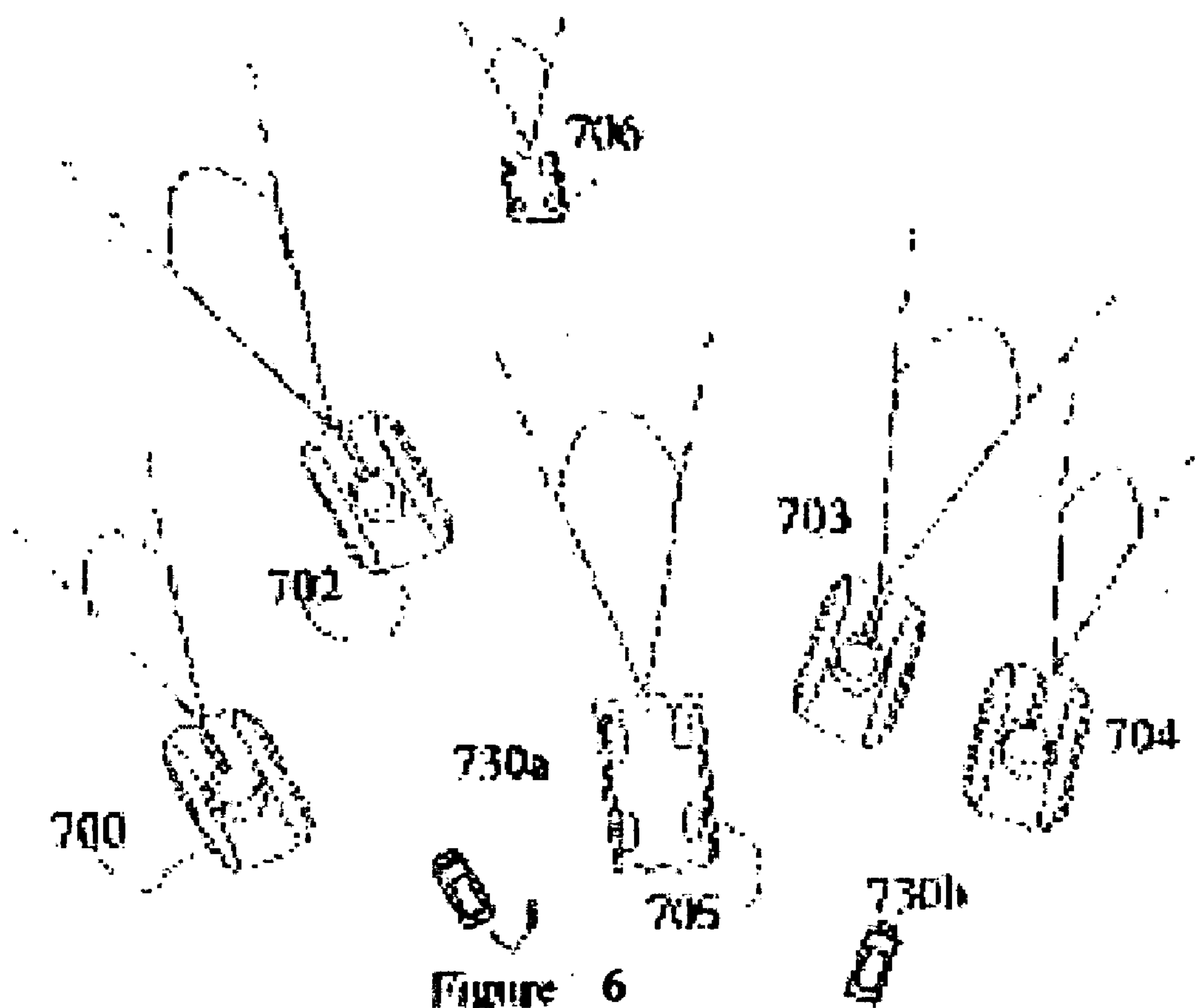


Figure 6

BIRDS EYE VIEW VIRTUAL IMAGING FOR REAL TIME COMPOSITED WIDE FIELD OF VIEW

RELATED APPLICATIONS

[0001] The present application is based upon provisional application Ser. No. 60/493,579 filed on Aug. 9, 2003.

TECHNICAL FIELD

[0002] The present invention relates to methods of acquiring multiple images for display in the form of a composite image from alternative vantage points, and more particularly to the use of such methods for steering transportation vehicles to real-time or providing situational awareness.

BACKGROUND OF THE INVENTION

[0003] The present invention recognizes certain limitations which inherently exist in an attempt to navigate a vehicle. Often times, vehicles provide a driver with a limited view of the driver's surroundings. For example, large trucks and military vehicles such as a tanks position a driver high above a roadway with a somewhat limited viewing angle. By the time potholes and other impediments are closely approached by the vehicle, they are no longer in a driver's field of view. As such, an attempt was made to suggest a means of providing a driver both with obstacle positioning and coordinance together with a broader view of the vehicle's upcoming terrain. To the inventor's understanding, there has been no successful means suggested to date for providing such useful information to a vehicle operator.

SUMMARY OF THE INVENTION

[0004] A first object of the instant invention is to display a virtual image to an individual steering or driving a plane, vessel or transportation vehicle in real time, or other visualization requirement when the image comprising a live image is suitably transformed to reflect a vantage point reflecting the position of the vehicle in reference to obstacles and hazards that are no longer in view.

[0005] Yet another object of the present invention is to display virtual imaging that combines visual and non-visual imaging sensors in real-time.

[0006] It is yet another object of the present invention is to combine live images acquired by multiple vehicles to form composite images reflecting a wider virtual field of view, the field of view optionally combining using previously acquired or generated images superimposed thereon.

[0007] A further object of the invention is to superimpose reference information on the aforementioned composite images illustrating, for example, the relative position of the vehicle, hazards, targets and the desired path or roadway between such objects.

[0008] One aspect of the invention is characterized in that images acquired at times t1 and t2 are optionally superimposed or composited by correlating the relative magnification such that pixels from the distant image are placed with the corresponding pixels of the live image. However, to the extent that the virtual viewpoint is intended to enable navigation around objects that are no longer visible to the live image, this superposition is preferably continuously updated to account for both forward movement and rotation (X, Y, Z) of the image frames.

[0009] The above and other objects, effects, features, and advantages of the present invention will become more apparent from the following description of the embodiments thereof taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE FIGURES

[0010] In **FIG. 1** the elevation of a vehicle, traveling on the road is intended to schematically illustrate the driver's actual viewpoint and the preferred viewpoint according to the teachings of the instant invention.

[0011] **FIG. 2** is an elevation showing the principal of acquiring and utilizing time sequence images corresponding to a vehicle's first position at time t 1 and second position at t 2 while traveling on the road depicted in **FIG. 1**.

[0012] **FIG. 3** is schematic illustration of a vehicle is the actual position as display to the driver of a vehicle from the preferred viewpoint on an information display.

[0013] **FIG. 4** a schematic diagram **FIG. 4** a systematic diagram illustrating the principles underlying one embodiment of a method of image processing to transform the images to equivalent virtual view point above and behind the vehicle at the same magnification, including discrete steps in transforming, aligning and superimposed a real-time image with a corresponding synthetic image for display.

[0014] **FIG. 5** is a plain view illustrating the use of the instant invention for assisting a truck driver to backup a truck rig into a loading ramp.

[0015] **FIG. 6** is a plan view illustrating the use of the instant invention for the command and control of a variety of combat vehicle executing a mission was spreading out over the terrain

DETAILED DESCRIPTION OF THE INVENTION

[0016] **FIG. 1** is an elevation view to illustrate the deficiencies of the vantage point of a driver in vehicle **100**. Objects having reference numerals **10**, **20** and **30** represent normally visible obstacles in or adjacent to the road **110**. More specifically, the driver has just passed by pothole **10**, which is no longer visible from the front window, whereas the farthest obstacles **30** are still within the field of view, being subtended by angle **121** to represent the driver's vantage point. However, as object **20** is about to leave the driver's field of view as the vehicle progresses forward, it is very difficult to pass such road obstacles at a close distance, or drive between them as the spacing approaches the width of the vehicle. Thus, to the extent a road, bridge or rough terrain is extremely narrow with obstacles or borders that represent significant safety hazards, or the driver is required to navigate in close proximity to such hazards, the disappearance at close range from the driver's field of vision is undesirable. Further, to the extent the navigation vehicle depends on non-visual imaging system, for instance an infrared imaging system for use during night driving or fog, the viewing angle of such imaging systems, represented by reference numerals **122**, can be a further limitation. Under these and other conditions it would be preferable if the driver could simultaneously have a sufficiently distant view in front of the vehicle, while at the same time perceive the vehicles

position with respect to close objects and hazards they are attempting or forced to avoid at a close distance.

[0017] More specifically, in particularly hazardous situations it would be desirable if the driver could control the vehicle with a virtual viewpoint situated slightly behind a vehicle, as indicated by the camera icon **130**, having a virtual viewing area within the arc subtended by angle **131**. Thus, a display of the virtual viewing angle shows the driver the vehicle location with respect to road hazards **10**, **20** and **30**, which might be located either just in front or to either side of the vehicle.

[0018] Accordingly, **FIG. 2** illustrates the operative principles for a first embodiment of the instant invention, which includes acquiring and displaying the desirable image of virtual camera **130** of **FIG. 1**. Vehicle **200** in **FIG. 2** includes a video or digital imaging camera **240** that continually acquires at least a forward image as vehicle **200** progresses from the right side of the Figure at time t1 to the left side of the Figure at time t2. Camera **240** has an actual viewing area within the arc subtended by angle **223**, which may be the same, narrower or wider than the driver's actual field of view. Two or more images acquired by camera **240** between t1 and t2 are used to generate a composite, for example, an actual image acquired at time t2 as if acquired from virtual camera position **130**, but overlaid in correspondence with image data acquired at time t1. Thus, the driver would be able to view and steer around or close to the edge of the road but avoid hazards **10**, **20** and **30**.

[0019] Accordingly, **FIG. 3** illustrates a display **300** of the driver's view of composite image corresponding to time t2, as would be seen from virtual camera position **130**. The display **300** has a first frame region **302** and a surrounding region **301**. The first or inner frame region **301** preferably is used to display the current, or live image accorded time t2, whereas the surrounding region **301** was acquired earlier, that is at time t1, or between time t2 and time t1, and thus includes pothole type road **10**, which is adjacent to vehicle, and out of the driver's current actual view. The image of the vehicle **310** is synthesized, as it is never in actual view. In yet another embodiment, Display **300** may also include various indicia, such as a guideline **320** to follow to drive between other hazard objects **20** and **30**.

[0020] It should be appreciated that since it is very difficult to position a camera for recording images that correspond with the virtual camera **130** position each of the images acquired at time t1 and time t2 are generally transformed prior to display of the composite image representing the virtual viewpoint at time t2. However, the image at time t1 in the above example can be generated from another image source not on the vehicle, including an image database, and may in fact have been acquired at the reference viewpoint.

[0021] Generating the Birds Eye View (BEV) image of **FIG. 3** utilizes one or more images acquired at a different viewpoint, which transformed to reflect a different vantage point above vehicle. Those of ordinary skill in the art of computer graphics will recognize that there are numerous schemes for performing such transformations. Aerial video is rapidly emerging as a low cost, widely used source of imagery for mapping, surveillance and monitoring applications. The disclosure of U.S. Pat. No. 5,259,037, which is incorporated herein by reference, discloses a method of acquiring individual images from an aerial video that can be

aligned with one another and merged to form an image mosaic. In surveillance applications, such a video map provides the basis for estimating motion of objects within a scene. U.S. Pat. No. 5,590,037, which is incorporated herein by reference, discloses a method for converting forward-looking video or motion picture imagery into a down looking database suitable for use in an image generation system to generate real-time perspective images for simulation purposes.

[0022] Further, U.S. Pat. No. 5,649,032, which is incorporated herein by reference, discloses methods for automatically generating a mosaic from a plurality of input images. The inventor's of the '032 patent teach a mosaic construction system that sequentially executes an image alignment process and a mosaic composition process such that, from a sequence of images, the system automatically produces a mosaic for utilization by various applications. The invention is described as being capable of constructing both dynamic and static mosaics. A dynamic mosaic includes imagery that is time variant, e.g., the mosaic is updated with new content over time, while the content of a static mosaic is time invariant.

[0023] U.S. Pat. No. 6,512,857, which is incorporated herein by reference, discloses a system and method for accurately mapping between camera coordinates and geo-coordinates, called geo-spatial registration. The method utilizes the imagery and terrain information contained in the geo-spatial database to precisely align the reference imagery with input imagery, such as dynamically generated video images or video mosaics, and thus achieve a high accuracy identification of locations within the scene. The geo-spatial reference database generally contains a substantial amount of reference imagery as well as scene annotation information and object identification information. When a sensor, such as a video camera, images a scene contained in the geo-spatial database, the system recalls a reference image pertaining to the imaged scene. This reference image is aligned very accurately with the sensor's images using a parametric transformation. Thereafter, other information (annotation, sound, and the like) that is associated with the reference image can easily be overlaid upon or otherwise associated with the sensor imagery. Applications of geo-spatial registration include text/graphical/audio annotations of objects of interest in the current video using the stored annotations in the reference database to augment and add meaning to the current video

[0024] Commercial software is available for performing the manipulations disclosed in **FIG. 4**, or alternative methods of combining adjacent images having some overlap, known as "mosaic tiling" may be deployed. For example, Observera Inc. of 4451 Brookfield Corporate Drive, Suite 107, Chantilly, Va. 20151-1693 provides software that has a range of features allowing modification to serve a variety of applications. In addition, Samoff Corporation and Pyramid Vision Technologies, both of 201 Washington Road, CN 5300 Princeton, N.J. 08543-5300, supply commercial software and hardware for performing Birds-Eye-View transformations to create fly by images.

[0025] The methods for generating the useful displayed image in **FIG. 3** can be deconstructed into a sequence of steps, although they need not all occur in a discrete manner, depending on the method of implementation. Thus, **FIG. 4**

illustrates one embodiment of operative principles for acquiring, transforming and aligning the image data used to generate the real-time display of **FIG. 3**. Accordingly, the following description should not be construed as limiting the scope of the patent.

[0026] **FIG. 4** illustrates an alternative embodiment for generating the image display **FIG. 3** image **401** is acquired the time t_1 whereas image **400** and to is acquired a time t_2 , accordingly the pair of image is represented by bracket **410** correspond to live actual images recorded for the moving vehicle. In process of generating image **420** for display, the first image frame **402** is acquired in either video or digital format. Then after movement or displacement of the camera on the vehicle, a second or live image **402** is similarly acquired. Bracket **410** contains a digital representation of the actual images acquired at time t_1 and t_2 for further transformation and merger to form a composite image **420**. Although the Bird's Eye View can be created before or after merger of the images, images **403** and **404**, in bracket **420** represent bird's eye view transformations of the corresponding adjacent images within bracket **410**. The rectangular image frames in bracket **410** distorted to trapezoidal shapes in generating the higher elevation or Birds-eye-view. One embodiment for generating such bird's eye view images optionally includes performing scaled transformation of the rectangular image frame to a trapezoid to simulate the loss of perspective as the birds eye view camera increases in azimuth angle from the actual viewpoint of the camera mounted on the vehicle. The trapezoid results from transforming each row of the x-axis gradually with increased compression starting from the upper edge of the picture frame of the actual view **410**, with increasing compression towards the bottom of the frame. As shown in image **403**, a trapezoidal transformation decreases the divergence of the lines representing the highway traffic lanes in the images of bracket **410**.

[0027] Although images **401** is preferably modified by digital processing to image **403** to correspond to the expected appearance from camera position **130** in **FIG. 2**, alternative viewpoints are possible, including a position forward of the vehicle at the time t_1 at which image **401** is recorded, provided images **403** and **404** are generated with substantially the same virtual camera position.

[0028] Once the transformed image **403** and **404** are generated, the near image acquired at time t_1 is appropriately scaled and overlapped with respect to the earlier acquired image at time t_2 . Thus in forming the composite image **420**, the scaling factor to convert image **403** to image **405** must be determined, as well as any x and y displacement for overlay of image **405** on image **404**.

[0029] Further, virtual features, such as the image of the vehicle, frame separating the image regions, optional projection of any intended driving path, and the like, are preferably overlaid on the penultimate composite images to form the final composite image. Further, the composite image is most preferably refreshed in real time to reflect the forward progress of the vehicle. Thus, image **405** is represented as de-magnified from image **403** to represent its scaling prior to merger over virtual image **404** to create display image **420**. However, equivalent operations can be performed on image **401**, such that detail in the real time image is more fully preserved.

[0030] The displayed image **420** is optionally generated by merging de-magnified image **405** with image **404**, taking into account lateral translation and rotation of the actual cameras viewpoint between the acquisition of frames **401** at time t_1 and **402** at time t_2 . The overlay can be determined by mapping the displacement of pixels from image frames **405** to **404**, such that a selected sub group of pixels in image **404** is replaced with image frame **405** pixels prior to display. As the image acquired at time t_1 and t_2 are mapped to the same magnification, a relative movement or rotation is optionally determined by first searching each image field to identify high contrast features, and then comparing the relative orientation of these features to generate the appropriate correction factors. That is, when the correction factors are applied the high contrast features must coincide to compose an accurate virtual image for display.

[0031] Image **403** is scaled to generate image **405** by a de-magnification factor based on Z-axis displacement of the vehicle between image frames **401** and **402**. The factor used to generate image **405** from image **403** can be determined by several alternative methods. Thus, depending on the method of forming and generating the composite image **420**, it is desirable to know the absolute movement of the actual camera position between times t_1 and t_2 , as this determines the scaling factor for converting image **403** to **405** by geometric calculations, or magnifying images **404** with respect to **403**. The Z-axis distance can be determined by several methods, including but not limited to global satellite positioning, or calculating the change in position by integrating the speedometer output overtime to synchronize time period between t_1 and t_2 .

[0032] To the extent that the initial images are readily acquired in digital format by converting an analog video feed into a JPEG or MPEG format data stream, the correction factors can be generated from selected parameters of the digital data streams. Briefly, JPEG and MPEG data format transmits full images, or I frames, infrequently to conserve bandwidth, using a sequence of intervening frames (B and P type in MPEG) to communicate the changes to portions of the image between I frames. In forming the MPEG/JPEG data stream the image is broken down into macro blocks, which is collections of pixels, and analyzed to identify macro blocks that change location between successive image frames, which are then used to reconstruct the full image for final display. In the MPEG format both B and P frames identify and track macro blocks that change location between I frames. Specifically, the translation and rotation of image **404** with respect image **405** necessary for merging these images may be determined from the movement of macro block represented in the JPEG and MPEG formats by extracting an average macro block translation to represent the relative movement between consecutive I frames. Thus, the vector sum of the individual translation factors can be applied to align images **404** and **405**. To the extent rotation and translation have been limited, the magnification factor is alternatively computed from the y-axis components of the macro block translation between I frames.

[0033] It should be appreciated by one of ordinary skill in the art that the position of car **310** in **FIG. 3** is synthetically generated based on the data set representing the actual car's dimensions, and the selection of the birds eye view position, that is, Z, Y position in **FIG. 1**. For example by specifying

two or more parameters which may include, the azimuth angle, vertical height above the vehicle, horizontal distance behind the vehicle (at time t2), viewing angle and like parameters define the Z and Y positions necessary to determine the appropriate transformation factors used to generate the pair transformed images in bracket **415** from the corresponding actual images in bracket **410** in **FIG. 4**.

[0034] In another alternative embodiment, MPEG conversion can be limited to images **401** and **402**, rather than the entire video frame sequence thus simplifying the computational complexity. However, in a preferred embodiment, the entire bit stream representing each individual frame recorded by the video camera between images **401** and **402** is utilized.

[0035] Further, in yet another embodiment the aforementioned method of macro block tracking can be extended to determine the factor used to compute the magnification of image **403** for generating image **405** while simultaneously correcting for what has been initially described as discrete steps of image rotation and translations determination, i.e. the steps used to place image **405** in image **404**. Although it may be possible in some instances to identify one or more macro blocks that correspond to distinct objects or edges of the vision field it is unlikely that the same macro blocks can be uniquely identified for each frame of the video source from t1 to t2. However, this is not necessarily, provided a refreshed or updated subset of macro blocks is used at each I-frame. The updated subset would correspond to the same x-y coordinate range of the macro blocks in the previous I frame, as updated to reflect the most recent preceding image. More specifically, using the MPEG bit stream to trace the displacement of macro blocks at corners of image frame **402** includes correction for magnification, translation and rotation. Reverse tracking the relative positions of macro blocks that correspond to the corner regions of frame **402** define a relative position for edges of the frame acquired at time t1 frame with respect to the frame acquired at time t2. Thus applying a linear scaling between the compression and distortion ratios necessary to re-map the frame corner also accomplishes translation and rotation. It is anticipated that either of frames **402** or **401** can be modified, either before or after the trapezoidal distortion, that represents the "bird-eye-view". Alternatively, if the distance traversed between image frames **401** and **402** is significant, or the bit stream has been interrupted, the magnification factor can also be determined by computation from the integrated speedometer readings, alternatively larger gaps can also be accommodated by calculation based on the GPS coordinates recorded at the time of acquisition for images **401** and **402**.

[0036] Alternatively, if the driver is concerned about a particular image feature or region of the live image, the translation and rotation factors are preferably acquired by selecting the corresponding macro blocks that represent such features. The previous calculation, likewise carried out by starting with the last live frame **402**, may be carried out by averaging (before accumulating the sum thereof) a limited number of macro block translation factors depending on the area selected. Although the identical macro blocks used as the starting point for the reverse computation (from a particular region of the live image **402**) may only maintain the same identity between I Frames in the MPEG bit stream, the accuracy is the likely to be improved by selecting a

newer subset of macro blocks that correspond to the same x-y coordinate range of the macro blocks in the previous I frame.

[0037] Thus, the steps in executing the aforementioned method of macro block tracking include; 1) identifying the first set of nearest neighbor macro blocks corresponding to corners of frame **402** or selected portions of the live image, 2) record in the average translation to the previous I frame, 3) record the average x-y coordinate position corresponding to average translation to previous I frame, 4) identifying a second set of nearest neighbor macro blocks corresponding to average x-y coordinate positions, 5) recording the average translation to the next previous I frame, 6) compute the sum of the first average translation and each subsequent average translation for each of the corresponding four corners of frame **402** (or selected regions therein), 7) repeating the previous steps of identifying the second set of nearest neighbor's until the subsequent I frame corresponds to the closest I frame from image frame **401**, 8) linearly distorting one of image frame **401** to **402** to aligned the corresponding corners according to their respective translation factors, 9) either before or after merging the distorted and undistorted image frame from the previous step generating a second composite image by distorting a first composite image to correspond with the position of the virtual camera, 10) calculating a second pixel subset corresponding to the profile of the vehicle as determined by the position of the virtual camera, 11) replacing selected pixels in the second composite image with the second subset of pixels to form a third composite image, and 12) displaying the third composite image.

[0038] **FIG. 5** is a plain view illustrating the use of the instant invention for assisting a truck driver to backup a truck rig **600** into a loading ramp **610**. The driver views a display that provides a composite of a live and recorded image from camera **640** (with viewing angle **641**) according to the teachings of the invention with virtual camera position **630** (having viewing angle **631**) such that the drivers "sees" the corners or other obstacles **611** and **612** in close proximity as they back up the truck to the loading dock.

[0039] **FIG. 6** is a plan view illustrating the use of the instant invention for the command and control (optionally from vehicle **705**) of a variety of combat vehicles **700-704** executing a mission was spreading out over the terrain. Alternative virtual camera positions **730a** and **730b** allow a wide-angle view of the battlefield from any vehicle, with the images being acquired from manned vehicles **700-705** or drones **706**. The image preferably shows the actual view from each vehicle, identifies "friend" and "foe" with additional icons thus avoiding friendly fire accidents.

[0040] It should be appreciated that the images described and combined need not be solely from visual sources, but include IR, NIR and other non-visual sources, and may combine visual images with non-visual or enhanced images in either monoscopic or stereoscopic views in the final composite images.

[0041] While the invention has been described in connection with a preferred embodiment, it is not intended to limit the scope of the invention to the particular form set forth, but on the contrary, it is intended to cover such alternatives, modifications, and equivalents as may be within the spirit and scope of the invention as defined by the appended

claims. For example, it should be appreciated that the alternative methods of forming a composite image disclosed herein can be combined with any of the prior art methods of digital image processing provided the real-time images are either acquired in digital format or converted to digital format from an analog video recorder or camera.

1. A method for generating a dynamic composite image by combining a live image and a previously acquired or generated input image, the method comprising the steps of aligning said live image with said previous image, wherein the alignment is performed before or after the input image

and the live image are converted to represent a common virtual vantage point.

2. A method of illustrating relative position of a vehicle with respect to its surroundings comprising taking a first image at a first time and taking a second image at a second time, superimposing the first and second images and correcting the first and second images to account for different image sizes due to the difference in the first and second times.

* * * * *