

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
4 June 2009 (04.06.2009)

PCT

(10) International Publication Number
WO 2009/068942 A1

(51) International Patent Classification:
G06T 3/00 (2006.01) *G06T 15/20* (2006.01)
H04N 5/262 (2006.01)

(21) International Application Number:
PCT/IB2008/001369

(22) International Filing Date: 22 January 2008 (22.01.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
GB 0723538.5 30 November 2007 (30.11.2007) GB

(71) Applicant (for all designated States except US):
DOOWORKS FZ CO [AE/AE], PO Box 293783,
Dubai Airport Free Zone (AE).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **JACOB, Stephane,**
Jean-Louis [FR/AE], c/o Dooworks FZ Cores, Dubai Air-
port Free Zone (AE).

(74) Agent: **LLOYD, Patrick, Alexander, Desmond,** Reddie
& Grose, 16 Theobalds Road, London WC1X 8PL (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, **DK**, EE, ES, FI, FR, GB, GR, **HR**, HU, IE, **IS**, **IT**, LT, LU, LV, MC, MT, NL, NO, PL, PT, **RO**, SE, **SI**, SK, TR), OAPI (BF, **BJ**, CF, CG, CI, CM, GA, GN, GQ, GW, ML, **MR**, NE, SN, TD, TG).

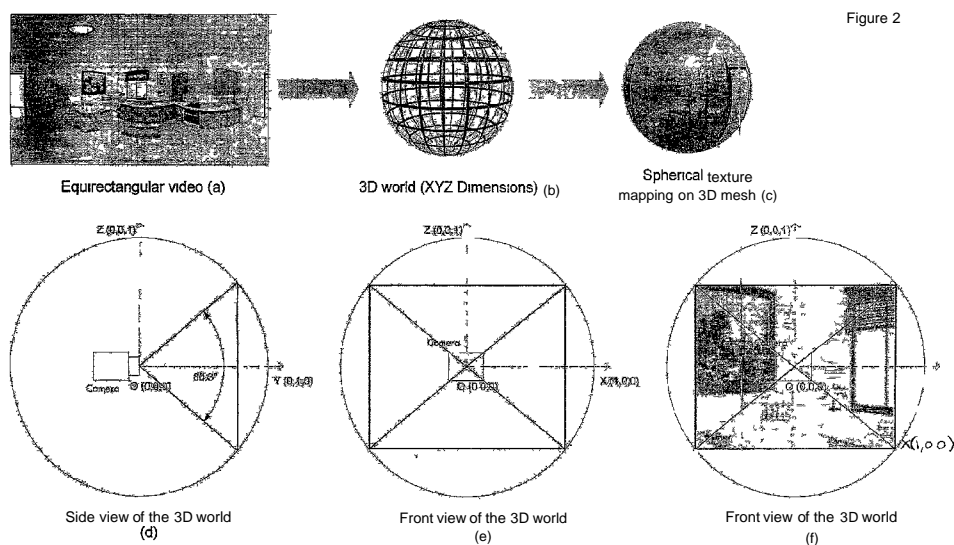
Declaration under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(v))

Published:

— with international search report

(54) Title: METHOD AND SYSTEM FOR PROCESSING OF IMAGES



(57) Abstract: A stream of images is acquired either from a camera and fish eye lens assembly or is computer generated. The image data is then texture mapped onto a 3D mesh and a portion of the mapped data is selected under the control of pan, tilt and zoom functions which define the parameters of a virtual camera within the mesh. The selected data is then converted back into 2D data and displayed.

METHOD AND SYSTEM FOR PROCESSING OF IMAGES

FIELD OF THE INVENTION

This invention relates to processing of images and in particular, to the manipulation and display of images derived from very wide or 360° views.

5 BACKGROUND TO THE INVENTION

There are many situations in which it may be desired to present images to a user in such a manner that the user can control a portion of a 360° image that is viewed on a screen. Such images can be from computer-generated images or from real world images recorded using a video
10 camera.

In the field of 3D content display many technologies exist on the market, but can easily be grouped in 3 main families:

3D Real Time

3D Rendering

15 360 Still Navigable

3D Real-time is a 3D engine software that can compute and display in real-time all the elements needed in a 3D world, such as meshes, texture, light, shadow. The rendering unit generates images at the desired FPS (Frame per Second), each frame being computed and rendered. For
20 example, depending on the complexity of the 3D world and its components, 30 FPS means that the scene is rendered 30 times per seconds, giving the end user a smooth illusion of movement in that 3D world.

Many different ways exists to render this 3D world in Real-time. They are
25 all dependent on the hardware regarding the quality and speed of the

render. For example, if the scene is complex (high number of vertices, complicated lights and shadows, high quality of textures) the rendering time will be huge for each frame, and the FPS will drop as a result, losing the impression of movement. But the main constraint is that in 3D Real-time, some very high rendering modes such as Ray-Tracing and Radiosity are not allowed because of computing time. For example, an image rendered in Radiosity can take at least 1 hour to render, but give to the end user a sensation of reality. The user has control of images produced by this method and many apply Pan, Tilt and Zoom to a virtual camera to select the portion of images that is viewed.

The FPS in this case is relative to the rendering process in the 3D engine. This technology provides full interactivity to the 3D environment, but is poor in quality.

3D Rendering relates to video sequences that are preliminarily rendered into a 3D software and are simply displayed using a video player. The 3D video is then composed with sequences of frames (basically 25 Frames Per Seconds FPS). The rendering process is computed in advance using technologies such as Ray-Tracing or Radiosity, not allowed in 3D Real-time. However, in this arrangement, the end user does not have any control of the movie.

The FPS in this case is relative to the final video. This technology provides ultra high quality, but doesn't provide any interactivity control.

360 Still Navigable is a 3D image rendered in any 3D rendering software using Equirectangular settings for virtual camera parameters. The concept of a virtual camera is well understood in the computer graphics field to refer to an image view computed by the software as if it had 'seen' the scene from a certain coordinate location. The image is then displayed using software that corrects and dewarps in Real-time the Equirectangular image to allow the end user to have a corrected image on the screen. The

end user has Pan, Tilt and Zoom (PTZ) control on that static frame enabling a limited degree of control of the static frame that is displayed

This technology combines ultra high quality, but provides a limited interactivity. In particular, only still images can be viewed.

- 5 The following table represents the different interactive possibilities that are allowed in the different technologies:

Technology	Pan	Tilt	Zoom	Time	Space
3D RealTime	X	X	X	X	X
3D Rendering				X	
360 Equirectangular	X	X	X		

- We have appreciated the desirability of providing moving images to a user, which allow the possibility of the user to control the portion of a wide field
 10 image displayed on a screen using parameters such as pan, tilt and zoom. We have further appreciated that hardware constraints mean that it is not computationally feasible to render each frame of such an image sequence in Real-time.

- We have also appreciated a need to process streams of image data
 15 derived from wide-angle views, for example, acquired using a camera fitted with a fish eye or other wide angle lens, so as to allow a user to select only portions of the image for viewing from a particular perspective. We have also appreciated that such processing is computationally intensive. We have further appreciated the advantage of processing
 20 image data in such a way as to allow the computation to be undertaken in a graphics card in a PC rather than requiring processing using a main CPU of a PC. Such processing can make the processing of a wide angle image data commercially viable and not confine it to high-end post production facilities such as are required by existing techniques.

SUMMARY OF THE INVENTION

The invention is defined in the claims to which reference is now directed.

In a broad aspect, the invention involves a new two-stage process involving first, pre-rendering pixels of an image using a mesh and second, playing the sequence of frames derived from the mesh. The pre-rendering is arranged such that it is then computationally simple to convert the image data mapped onto the mesh into video images with user selectable parameters such as pan, tilt and zoom. In particular, the mesh is arranged so that the conversion from the image mapped on the mesh into a 2D representation does not require interpolation. To achieve this, the mesh is arranged such that a conceptual virtual camera, if arranged at a central point within the mesh, would see an equal spacing of image pixels for presentation on a 2D display. The computationally intensive step of pre-rendering can therefore be done in non-Real-time and allows a video player to play the images in Real-time whilst allowing user selectable parameters, such as pan, tilt or other degrees of freedom and zoom to also be chosen in Real-time. The image source may be video, for example coming from a real time camera stream, recorded file or computer generated. In the former case, in a preferred embodiment, the mesh is comprised of multiple meshes that have the shape of one, or multiple camera lenses. In the latter case, in a preferred embodiment, the computer generated images are rendered in equirectangular format by rendering software and then mapped onto a spherical mesh.

BRIEF DESCRIPTION OF THE FIGURES

Embodiments of the invention will now be described, by way of example only, with reference to the accompanying figures in which:

Figure 1 is an overview of the whole image creation, pre-rendering and video play out process;

Figure 2 shows the step of producing an image on a mesh for a single image;

Figure 3 shows examples of images resulting from the process;

Figure 4 shows the process steps for an embodiment using wide
5 angle lens capture;

Figures 5a, 5b and 5c show plan, side and perspective views of an 3D or real environment to be derived using the embodiment of Figure 4;

Figure 6 is an example of a fish eye image acquired from the images of Figures 5a - 5c using the embodiment of Figure 4;

10 Figure 7 shows example meshes;

Figures 8a and 8b show the conversion between a hemisphere and a hemi-cube;

Figure 9 shows a resulting image with the scene of Figures 5a - 5c projected onto a hemi-cube lens; and

15 Figure 10 shows examples of a mesh used with a fish eye lens to correct for distortion.

DESCRIPTION OF PREFERRED EMBODIMENT

The invention may be embodied in an image processing system. It is preferred that the invention is embodied into separate systems, namely a
20 pre-rendering system, which operates an embodiment of a first aspect of the invention and a video player, which operates an embodiment of a second aspect of the invention. Each aspect may be embodied in specific hardware designed to operate the process, such as a specific digital signal processor or other hardware arrangement. Alternatively, the embodiments
25 may be implemented in software.

To put the embodiment of both aspects of the invention in context prior to describing the two aspects of the pre-rendering system, on the one hand, and the video player, on the other hand, the advantages of an embodiment of both aspects of the invention, as a whole, will first be described.

- 5 In essence, the embodiment of the invention undertakes the computationally expensive process of converting each frame of an image sequence, in such a way that it is then computationally simple to allow a user to select parameters such as pan, tilt and zoom in Real-time. Accordingly, the end result is an image sequence representing a camera
10 (whether virtual or real) moving through space, which can be manipulated by a user to pan, tilt or zoom in different directions. Accordingly, a user may select different pan, tilt and zoom settings and step forwards and backwards through the image sequence in time. However, unlike the computationally very expensive process of 3D Real-time rendering, the
15 user would not be able to cause the apparent virtual camera to move to different locations in space. However, we have appreciated that there are many environments and applications where this is not a requirement.

- Throughout the description reference is made to pan, tilt and zoom as these are expressions commonly used in the art. Pan and tilt are
20 examples of degrees of freedom which are the set of independent displacements that specify completely the displaced or deformed position of a body or system. In general, a rigid body in d -dimensions has $d(d+1)/2$ degrees of freedom (d translations + $d(d-1)/2$ rotations). It can be reasoned that rotational freedom is the same as fixing a coordinate frame.
25 The first axis of the new frame is unrestricted, except that it has to have the same scale as the original, so it has $(d-1)$ DOFs. The second axis has to be orthogonal to the first, so it has $(d-2)$ DOFs. This leads to $d(d-1)/2$ rotational DOFs in d dimensions. In 1-, 2- and 3- dimensions there are therefore one, three, and six degrees of freedom.

A non-rigid or deformable body may be thought of as a collection of many minute particles (infinite number of DOFs); this is often approximated by a finite DOF system. When motion involving large displacements is the main objective of study a deformable body may be approximated as a rigid body
5 (or even a particle) in order to simplify the analysis.

The 6 degrees of freedom may be described as:

1. Moving up and down (heaving);
2. Moving left and right (swaying);
3. Moving forward and backward (surging);
- 10 4. Tilting up and down (pitching);
5. Turning left and right (yawing);
6. Tilting side to side (rolling).

It can be seen that Pan equates to yawing and that tilt equates to pitching. The zoom function is not a DoF. In the 3D realtime engine described
15 below, the system also has the moving DoF 1-3. In the following description, references to pan and tilt should be treated as references also to other possible degrees of freedom. The embodiment allows interactive ultra high 3D Rendering video sequences to be visualised using standard computer processing power available in the market. In addition, the
20 embodiment combines pan, tilt zoom and time interactive functions on the content, keeping an original frame rate, as generated with the 3D Rendering application or captured from the real world by camera. Thus sequences may be rendered using the graphics card of a commercially available PC instead of requiring use of the main PC processor or a
25 dedicated processor.

The following table comparisons between the functionality of embodiments of the present invention and the existing technologies described above:

The pre-rendering and video play out may be performed on image data acquired from any source. However, it is particularly suited, but not limited, to data acquired by a camera using a wide or ultra wide angle lens such as a fish eye lens. Such lenses can produce a view of up to 360° and it is desirable to be able to select a portion of such an image for further viewing. Thus, the ability to pan, tilt and zoom is most important. Moreover, in many applications it is desirable to be able to view such images as a series of images in real time, in which case the PTZ capability available in existing 3D Real-time and 360 Equirectangular systems are not suitable, bearing in mind, particularly, that the equirectangular system is only suitable for still images. More than one camera may be used as is described below.

Technology	Pan	Tilt	Zoom	Time	Space
3D RealTime	X	X	^	X	X
3D Rendering				X	
360 Equirectangular	X	X	X		
Embodiment	X	X	X	X	

15 Overview of the System

The overall steps in the system will first be described with respect to Figures 1 to 3. Subsequently, the pre-rendering system and video player system will be described separately.

As shown in Figure 1, in an embodiment using computer-generated images, the principal idea is to create a 3D world into a 3D software at 10 and to pre-render it at 20 using particular parameters, and to display them into a 3D world at 30 under the control of user defined pan, tilt and zoom parameters input into the system via a graphical user interface at 40.

The process is able to match the final frame rate displayed at 50 to the frame rate created 10.

The 3D world 10 may be created using any known 3D graphic design software. The 3D world 10 is composed with multiple elements to create the illusion of volume and reality. Each object is composed with meshes, textures, lights, movement. Then, a virtual camera is located in that environment.

The virtual camera parameters (focal length, aperture, ratio, projection, movement) are then set up and sent to the renderer 20. The renderer 20 has parameters set up, such as resolution, frame rate with number of frames, and the type of renderer algorithm. Any suitable algorithm may be used. Known algorithms include Phong, Gouraud, Ray Tracing and Radiosity. For the avoidance of doubt, the term rendering applies to the processing of computer generated images. The renderer operates on either CG images or images acquired from other sources such as a video camera.

The main camera and renderer common parameter is to set the view projection as equirectangular, with an image ratio to 2:1. This parameter may be changed according to user preference.

Then the image sequence is rendered and saved as an images file sequence or a video file. The resulting file or files are then converted as a texture file or files. This is a pre-rendering step.

The texture file is mapped onto a mesh in the form of a 3D sphere (normal orientation=inside / minimum number of vertex = 100) using texture mapping parameters (global illumination - 100%). A virtual camera is then located in the centre of that sphere. The mapping onto the mesh is the key step that saves computation time. This is a rendering step. The virtual camera parameters can be modified in real time. For example, the focal

length can be very short, providing a wide angle view. The mesh sphere, the texture and the virtual camera are combined into software giving the end user control the pan, tilt and zoom function of the virtual camera.

This combination has a 3D frame rate (rendered in real-time). The texture
5 has a 2D frame rate (pre-rendered).

The graphic user interface 40 generates and sends the pan, tilt and zoom parameters to the software virtual camera 30 and then allows the view to be generated and sent to the general display at step 50. Basically, all this sequence is determined into a loop. That loop determines the 3D frame
10 rate.

The 2D frame rate is still continuing during all the sequence. The 3D frame rate continues until the end user stops the application. The software application can embed normal movie player's functions such as pause, rewind, forward and stop. These functions act on the 2D frame rate.

15 Figure 2 shows the key steps in the process and the available view as a result. As previously described, a 3D view of the world as first created by any known 3D graphic techniques and rendered into a 2D equirectangular view. The equirectangular views are part of a sequence of 2D representations with a 2D frame rate being a sequence representing a
20 view as a camera tracks through a scene. The camera is, of course, a virtual camera in this example as the image is computer-generated. The key step in software is then taking the equirectangular image frames and mapping these onto a mesh, in this case a spherical mesh, in such a way that a virtual camera located at the centre of that mesh views a non-
25 distorted view of the image in any direction. Consequently, as shown in Figure 2, a camera, with given pan, tilt and zoom parameters, will view a portion of the scene represented on the mesh substantially non-distorted. The user can alter the pan, tilt and zoom parameters of the virtual camera, in any direction, to select a portion of the scene to view. As the user also

has control of the whole sequence of frames, the user is able to step forwards or backwards, in the manner of re-winding or playing, as a virtual video player.

Thus, in Figure 2 at (a) an equirectangular image is provided. For
5 simplicity this is shown as a single image although in practice it can be a sequence of images at a real time frame rate. At (b) the 3-D mesh is produced and at (c) texture mapping techniques are used to map the equirectangular image onto the mesh. The remaining parts of the figure show how a portion of the spherical image may be selected for viewing
10 substantially without distortion. Figure 2 (d) is a side view of the 3D world, that is the image mapped onto the sphere. A virtual camera is positioned at the centre of the sphere, that is, in X Y Z co-ordinates (0, 0, 0). The camera can 'see' a portion of the 3D image, with the aspect ratio being set in advance and the actual portion of the 3D image seen being determined
15 by user selected controls including pan, tilt and zoom. Figure 2 (d) shows a front view of the same 3D world, this time showing the image selected by the camera in the X Y directions. Figure 2 (f) shows the same view as Figure 2 (e), but with the selected image portion shown within the camera frame.

20 It will be appreciated that the virtual camera is a tool for selecting a defined portion of the image rendered onto the 3D mesh for viewing.

Figure 3 shows the same example scene, showing how the user can pan around each image of an image frame sequence, to select portions of the image.

25 Thus, in Figures 3a and 3b, the full equirectangular image is shown with a different user window selected. In Figure 3a, this window is to the left of centre. In figure 3b the window is to the right of centre. The window is the portion of the 3D image that is selected for viewing by the virtual camera. The window is moved around the equirectangular image by the pan and tilt

function. Of course, this function is operating to select an area of the image projected onto the 3D mesh, but the user will see the equirectangular image on his display and move the window around the displayed equirectangular image. In figure 3(c) the zoom function is
5 illustrated. Here, the user window occupies a considerably smaller portion of the equirectangular image resulting in a small part of the image filling the display than in the Figure 3(a) and 3(b) examples, giving a zoom function.

In the examples of Figures 1 to 3, the 3D world onto which the equirectangular image is mapped is a sphere. This is a convenient shape
10 and particularly suited to mapping computer generated images. However, the choice of 3D shape is defined by the system provider. In the more detailed example which follows, the images are not computer generated but are acquired using a fish eye lens and camera. In this case it is
15 appropriate to use a mesh which approximates the fish eye lens. Thus, a 180° fish eye will use a hemispherical mesh. In practice, the mesh may be adjusted to compensate for optical distortions in the lens as shown later with respect to Figure 10. The present invention is not limited to any particular mesh shape, although for a given input, the correct choice of
20 mesh is key to outputting good distortion free images. Where the image source is video, the equirectangular images of the computer graphics example are replaced by mapped texture images.

The pre-rendering system will first be described followed by the video player system.

25 **Pre-Rendering System**

The pre-rendering process takes frames of image data and maps these onto a mesh in such a manner that the subsequent process for displaying the video is computationally simple. The key feature of the pre-rendering process is the choice of mesh. The mesh arrangement is chosen so that if

a theoretical virtual camera were located at the centre of the mesh, the view of the camera in any direction would give a substantially non-distorted two-dimensional view of the image. If the original image representing a wide-angle or 360° view is computer generated, then the mesh itself can be calculated. On the other hand, if the input image data is captured in the real world, using a camera system involving one or more fish eye lenses, the mesh may be empirically determined by calculation from captured test images, as described later.

In the present example, the images are captured from a real environment using cameras. The camera arrangement could be a single wide-angled lens camera, but is preferably at least a pair of cameras with each having a lens with a 180° view. In addition, an optional zoom, pan, tilt and zoom camera, may also be included.

The pre-rendering system will now be described in greater detail with respect to Figures 4 to 10.

In the embodiment of Figure 4, a real environment is captured using a camera system 60. The camera system includes two XHD cameras with ultra wide angle lens. These two cameras are disposed back-to-back, preferably with their focal axes parallel, or ideally identical. This system is one of the possible ways to provide a full 360 vision. In Figure 4 this camera is shown as dual fish eye capture 62.

Optionally, these cameras are combined with one other camera 64 with a lens having the capability to provide mechanical zoom. That camera is controlled with standard PTZ functions. The protocol used to control that camera could be the PELCO protocol which aims to reduce time delays in IP based system which is desirable to improve operator control. The PTZ controls are managed and sent through a software application 90 which governs behaviour in the 3D world.

Optionally, further cameras could be linked to that system. For example, a night vision camera. As a minimum, a single camera is used, but the preferred embodiment uses two 180° cameras. The option to include a further camera is shown at 66.

- 5 The multiple streams of image data from the camera system may be transmitted to a converter 30. The converter operates on the XHD format data from the camera and may be any suitable digital video codec. However, it is preferred to use the systems described in our co-pending applications GB 070971 1 and GB 071 801 5.
- 10 Once the multiple streams have been processed by the Codec 70 they are converted into texture mapping data by converter 80.

The texture files are mapped onto a 3D mesh, for example, normal orientation=inside / minimum number of vertex = 100 and using texture mapping parameters for example with global illumination = 100%. For
15 playback, a virtual camera is then located in the centre of that mesh.

The 3D mesh used depends on what kind of focal length used in the cameras and the characteristics of the lenses. A method of empirically determining the mesh is described later.

- For example, to have a full 360 view, we combine two hemispherical
20 meshes. The resulting mesh is a sphere. In Figure 4, the mesh generation and the mapping of the image represented as a texture map onto the mesh, followed by the selection of a portion of the 3D image by the virtual camera is performed at 90 under the control of the virtual camera PTZ controls 92 input from the user GUI 96. As shown at 90, the
25 3D world may store a number of pre-generated meshes including a sphere, a hemisphere, a cube, a hemi-cube and a spline. This is not an exhaustive list and, as mentioned above, the shape may be chosen according to the input characteristics such as the camera lens. However,

this mesh is pre-calculated and does not need to be generated on the fly as the system is operating in real time.

The portion of the image from the 3D world is then passed to a user display 100.

5 Figures 5a - 5c show schematically how the system captures and processes images for display. The camera and fish eye lens may be positioned to form an image for mapping onto a mesh. The image here, includes a number of components, namely a car 110, a flower 112, an aeroplane 114, a house 116, the sun 118 and a tree 120. Figure 5a
10 shows a plan of the camera, the fish eye lens and image components, Figure 5b shows a side view of the camera, the fish eye lens and the image components and Figure 5c is an isometric view. An image of the components captured in any resolution is taken with a fish eye lens. A mapping 80 is performed such that the image is mapped onto a 3D mesh.
15 This allows a virtual camera, located at the focal point of the 3D mesh, to view a non-distorted view in any direction. In figure 5, the camera represents the real position of the camera in the real world.

Figure 6 shows the first stage of this process. An image of the components is captured using wide-angle lens camera. This may be
20 referred to as the input image and, as already explained, could be an image from a wide-angle lens such as a fish eye lens, including a 360° lens arrangement or could be a computer generated image. In either case, the image represents a wide field of view. The embodiment of the invention is particularly advantageous in that a sequence of frames
25 representing a moving image can easily be pre-rendered using the embodiment of the invention. Thus, Figure 6 shows a fish eye view of the image contents of Figure 5 with each component distorted by an amount which is dependent on the distance from the axis of the camera as is usual with a fish eye lens.

Figure 7 shows a mesh in the form of a hemisphere. As is known, a mesh in graphics processing is a group of polygons connected together by shared vertices. The mesh is used to map co-ordinates of the pixels of the input to conceptual 3D locations. A mesh is thus a conceptual
5 arrangement describing the arrangement of pixels of an image on a 3D surface. As such, a mesh is a theoretical arrangement, rather than a physical arrangement, and represents the manner in which data is stored for each frame. The example on Figure 7 is a theoretical perfect hemispherical mesh, which would be used assuming a perfect 180° lens.
10 Generation of meshes is well known on the computer graphics industry. The mesh may be defined using one of a number of known techniques such as calculating and storing a list of the vertices that make up each polygon in the mesh, or by using an edge array where each edge consists of its vertices. This technique may be enhanced by also storing the
15 polygons that share a given edge. All these techniques describes a 3D object as a polygon mesh such as that shown in Figure 7.

Figures 8a, b and 9 illustrate the problem of distortion which is present particularly when using fish eye lenses. Here, the fish eye lens is represented as a hemi cube conversion shown in Figure 8b. A hemi-cube
20 is a well known representation of a fish eye image. As can be seen from Figures 8a and 8b, the front of the fish eye image becomes the front of the hemi-cube and the remaining annulus of the fish eye images is divided into four equal rectangles shown as the top, bottom, left and right sides of the hemi-cube.

25 Figure 9 shows how the image of Figure 6 may be represented using the cube conversion of Figure 8b. It can be seen that there is distortion at the edges of the shape, see for example the aeroplane which appears curved. This distortion needs to be compensated for in the mesh, as is described later. As mentioned above, texture mapping is used to map the image
30 data onto the chosen 3D mesh. Surface mapping techniques such as

- texture mapping are well known in the computer graphics industry. Texture mapping techniques take into account the fact that real life surfaces are not just coloured but have textures and patterns. Many also have small or large surface displacements such as bumps or gouges.
- 5 Texture mapping simulates these surfaces to make images more realistic. There are three well known techniques for creating the impression of natural colours, textures and appearances : texture mapping, bump mapping and displacement mapping. Texture mapping adds a separately defined texture or pattern to a surface, for example, wall paper on a wall.
- 10 This does not affect the smoothness of the surface, but only changes its colour patterns. Texture mapping therefore adds a separately defined texture or pattern to a surface. In the embodiment of the invention, this pattern is an image that has been generated by a camera with a fish eye lens. The process involves the mathematical mapping of the image from
- 15 one domain to another and requires knowledge of the 3D World Device Co-ordinates (WDC), that is the X, Y, Z value of the surface of the image to be mapped. When ray tracing is used, this would be the point at which the ray intersects the object. Where scan line graphics is used, this is more complex as objects are first converted to physical device co-
- 20 ordinates (PDC) and then scan converted.

Texture or image mapping may be characterised by the dimensionality involved:

- 1-D texture domain → 3-D (WDC) → 2-D (PDC)
- 2-D texture domain → 3-D (WDC) → 2-D (PDC)
- 25 3-D texture domain → 3-D (WDC) → 2-D (PDC)

In embodiments of the present invention, as described above, image or texture maps are defined in two dimensions (u, v) and are mapped onto a three dimensional object. For each point in 3-D space (X_1, Y, Z) there must

be found a corresponding point in the 2D map space and the point on the 3D space must be coloured with the value from the texture map. This is easiest to achieve with parametrically defined surfaces which in general requires two parameters to define. Thus, a surface may be defined by
 5 equations of the form:

$$x = X(u, v); y = Y(u, v); z = Z(u, v).$$

To render the images an equirectangular image is mapped onto the mesh. In the case of a spherical mesh, this may be done as follows:

An angle q is defined as the angle from the x axis ($0 \leq q \leq 2\pi$) where
 10 $\pi = \pi$; and an angle f is defined as the angle from the z axis ($0.0 \leq f \leq \pi$).

The equations for a sphere are:

$$X = R \sin(f) \cdot \cos(q) = R \sin(pv) \cdot \cos(2pu) \quad (1)$$

where $f/p = v$ ($0.0 \leq v \leq 1.0$)

$$15 \quad Y = R \sin(f) \cdot \sin(q) = R \sin(2pu) \quad (2)$$

where $q/2\pi = u$ ($0.0 \leq u \leq 1.0$)

$$Z = R \cos(f) = R \cos(pv) \quad (3)$$

$$\text{From (3) } v = f/p = \arccos(z/r)/p$$

$$\text{From (1) } u = [\arccos(X/R \sin(pv))]/2p$$

$$20 \quad \text{Note: } q = \arccos(x/X) = \arccos(x/R \sin(pv))$$

Thus, if the point on the surface X, Y, Z is known, a point in the u, v texture space can be computed.

Of course, the same method of calculation may be used for any other shape that can be represented parametrically.

It will be appreciated that the process described is repeated for each image in a stream of images. The camera used to acquire images may be a video camera acquiring images in XHD or other format and the image frame will be processed one by one.

- 5 The manner for empirically determining the shape of the texture mapping mesh for a given lens will now be described with respect to Figures 10 a - c. The first step is to view an image of a known shape using the particular lens of the camera being used to capture the series of images. The chosen shape could, for example, be a cube or other angular shape
- 10 having a surface pattern of regular shapes. A good example is a half cube with small squares 1 cm x 1 cm patterned on the surface. This is a good choice of shape as it reveals any distortions of the capture lens. A single image of this shape is then captured with a digital camera using the chosen fish eye lens. The picture is then mapped onto a mesh with the
- 15 parameters varied to correct the shape by eye until the edges of the image are not distorted. Having determined the appropriate mesh empirically, this mesh can now be applied to the entire sequence of images captured with that lens.

- The mesh of Figure 10 is an example of a mesh that may be determined
- 20 empirically using this method. It will be appreciated from a comparison of Figures 7 and 10 that the mesh of Figure 10 is not a true hemisphere but that the edge of the hemisphere is curved outwards to compensate for distortion at the edge of the fish eye lens. It will therefore be apparent that the actual shape of the mesh of Figure 10 will depend on the
- 25 characteristics of the wide angle lens that is used.

It will be understood from the preceding example that the hemispherical, or near hemispherical mesh, can only generate a 180 degree view and that a 360 degree view requires more than one mesh. For example, two identical hemispherical meshes may be used, or two near hemispherical

meshes with each shaped to a respective one of a pair of 180 degree lenses

Video Player

The process of playing out the video will now be described.

- 5 The process of playing a sequence of video frames as provided by the pre-rendering process, is computationally simple and may be undertaken by a standard graphics card on a PC. To play each frame of a sequence of images pre-rendered according to the pre-rendering arrangement, a virtual camera is defined and arranged in software to view the pre-rendered
- 10 scene on the image mesh. User definable parameters, such as pan, tilt and zoom, are received at an input and applied to the virtual camera, so that an appropriate selection of the pixels from the image mesh is made and represented on a 2D screen. The selection of the pixels does not require computationally intensive processes such as interpolation because
- 15 the pre-rendering process has ensured that the pixel arrangement, as transformed onto the mesh, is such that a simple selection of a portion of the pixels in any direction (pan or tilt) or of any size (zoom) is already appropriate for display on a 2D display. The selection merely involves selecting the appropriate part of the image mapped onto the mesh. This
- 20 process can be repeated for each frame of an image thereby creating an video player.

Thus, the 3D image player takes data mapped on the image mesh and transforms this to a 2D image. This step is computationally simple as it only involves taking the pixels mapped on the mesh, and the pan, tilt and

25 zoom parameters input by the user, to select the pixels to be presented in a 2D image.

The virtual camera parameters can be modified in real time. For example, the focal length can be very short, providing a wide angle view. The

sphere or other mesh shape, the texture and the virtual camera are combined in software which the end user controls by adjusting pan, tilt and zoom function of the virtual camera.

5 A graphic user interface 50 generates and send the pan, tilt and zoom parameters to the virtual camera.

An additional possibility in the video player is that each real camera 10 and virtual camera 40 have a common nodal point. If the fish eye cameras and the optional PTZ camera have the same nodal point, the texture relative to the PTZ camera may then be texture mapped onto a spline mesh, inside
10 the 3D world, exactly between the sphere mesh and the virtual camera. If the digital zoom is still in the zoom range, a window can display only the view from the sphere mesh. If the digital zoom is greater than the zoom range, then the window could display the view provided by the PTZ camera instead.

15 Thus, in a practical embodiment of the invention, a user acquires 360° video images using a camera and a twin fish eye lens and some other arrangement. The texture mapping of the image data onto the 3D surface can be performed in real time using a convention PC graphics card and the user therefore has real time pan, tilt and zoom control of the images
20 displayed to him.

Many modifications to the embodiments described are possible and will occur to those skilled in the art without departing from the invention which is defined by the following claims.

Claims

1. A method of processing a stream of images for display, comprising:

acquiring the image stream from an image source;

mapping each image of the image stream onto a three dimensional
5 mesh to provide a three dimensional representation of each image;

selecting a portion of the three dimensional representation of each
image under the control of a degree of freedom control or zoom control
input by a user; and

outputting the selected portion of a 2 dimensional stream of images.
- 10 2. A method according to claim 1, wherein the selection of a portion of
each three dimensional representation comprises defining a virtual camera
at a focus of the three dimensional mesh, the camera having a pre-defined
aspect ratio, and moving the camera to select a desired portion of the
three dimensional representation in accordance with the input at a degree
15 of freedom or zoom control.
3. A method according to claim 1, wherein the mapping of the image
data is texture mapping.
4. A method according to claim 1, or 2, wherein the images of the
image stream are processed to provide equirectangular images prior to
20 mapping.
5. A method according to any of claims 1 to 3, wherein the image
source is a camera having a wide angle lens and providing a stream of
video images for mapping onto the mesh.
6. A method according to claim 5, wherein the mesh shape is
25 determined by the shape of the camera lens.

7. A method according to claim 6, wherein the mesh is hemispherical.
8. A method according to claim 7, wherein the mesh shape is adjusted to compensate for distortion in images produced by the lens.
9. A method according to any of claims 5 to 8, wherein the camera
5 has a plurality of lenses to acquire substantially 360° images.
10. A method according to any of claims 5 to 9, wherein the camera includes a camera with pan, tilt and zoom controls.
11. A method according to any of claims 1 to 4, wherein the images are computer generated images and the mesh is a sphere.
- 10 12. A method according to any preceding claim, wherein the pan, tilt and zoom controls are input by a user via a user interface and control a portion of the images mapped onto the mesh that is selected for conversion back to a two dimensional image and display to the user.
13. A method according to any of claims 5 to 10, wherein the camera
15 provides the video images in XHD format, comprising pre-processing the XHD format images in a video codec prior to mapping onto the mesh.
14. A method according to any preceding claim, comprising defining, prior to processing the stream of images, a three dimensional world including the mesh, and including at least one of texture, light and
20 movement.
15. A method according to claim 2, wherein the parameters of the virtual camera are set to provide 2D equirectangular images for display.
16. A method according to any preceding claim, wherein the frame rate of the image stream is the same as the frame rate of the output stream of
25 selected image portions.

17. A method of playing out video images processed by a method according to any preceding claim, comprising inputting degree of freedom parameters and a zoom parameter to define a portion of the 3D representation of the images to be displayed, displaying the selected
5 portion of the images as two dimensional images, and varying the portion of the images displayed by varying at least one of the degree of freedom parameters and the zoom parameter to select a different image portion.

18. Apparatus for processing a stream of images for display, comprising:

10 means for acquiring the image stream from an image source;

means for mapping each image of the image stream onto a three dimensional mesh to provide a three dimensional representation of each image;

15 means for selecting from the three dimension representation of each image, under the control of user defined controls comprising a degree of freedom control and/or zoom control, a portion of the three dimensional representation of each image; and

means for presenting the selected portion as a two dimensional image.

20 19. Apparatus according to claim 18, wherein the selecting means is a software defining a virtual camera at a focus of the three dimensional mesh, the camera having a pre-defined aspect ratio, and being movable in response to user degree of freedom and/or zoom controls to select a desired portion of the three dimensional representations of the images.

25 20. Apparatus according to claim 18 or 19, wherein the mapping means performs textured mapping of the image data.

21. Apparatus according to claim 18, 19 or 20 comprising a processor to processing the images to provide equirectangular images to the mapping means.
22. Apparatus according to any of claims 18 to 20 wherein the means
5 for acquiring images comprises a video camera having a wide angle lens and providing a stream of images for mapping on to the mesh.
23. Apparatus according to claim 22, wherein the wide angle lens comprises at least one fish eye lens.
24. Apparatus according to claim 22, wherein the at least one fish eye
10 lens provides a substantially 360° field of view.
25. Apparatus according to claim 22, 23 or 24, wherein the image acquisition means comprises a camera having at least one of pan, tilt and zoom controls.
26. Apparatus according to any of claims 18 to 25, comprising a mesh
15 store for storing the mesh, the mesh being selected in accordance with characteristics of the input images.
27. Apparatus according to claim 26, wherein the input images are computer generated and the mesh is spherical.
28. Apparatus according to claim 26, wherein the images are acquired
20 from a camera and lens assembly and the mesh is chosen in accordance with characteristics of the lens.
29. Apparatus according to claim 28, wherein the lens is a fish eye lens and the mesh is hemispherical.
30. Apparatus according to claim 29, wherein the shape of the,
25 hemispherical mesh is adjusted to compensate for distortion in the fish eye lens.

31. Apparatus according to any of claims 18 to 30, comprising a user interface for provision of degree of freedom and zoom controls to the selecting means to enable a desired portion of the three dimensional mapped image to be selected.
- 5 32. A video player for playing the two dimensional images selected from mapped three dimensional images by the method of any of claims 1 to 16, comprising an interface for input of degree of freedom and zoom parameters to define a portion of the 3D representation of the images to be displayed, and means for outputting a stream of selected image
- 10 portions for display, wherein the interface includes means for varying at least one of the degree of freedom and zoom to select a different image portion for output.

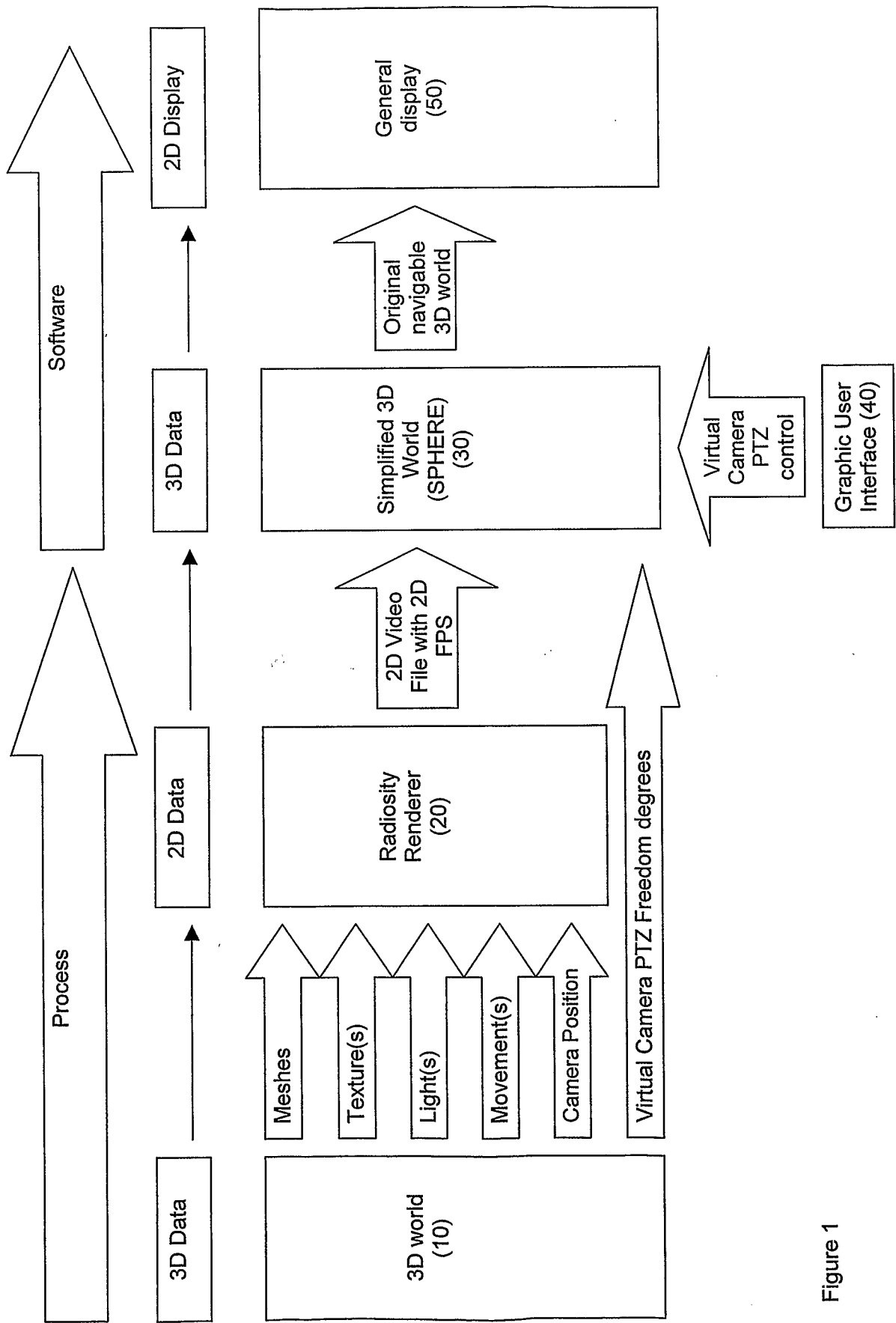


Figure 1

Figure 2

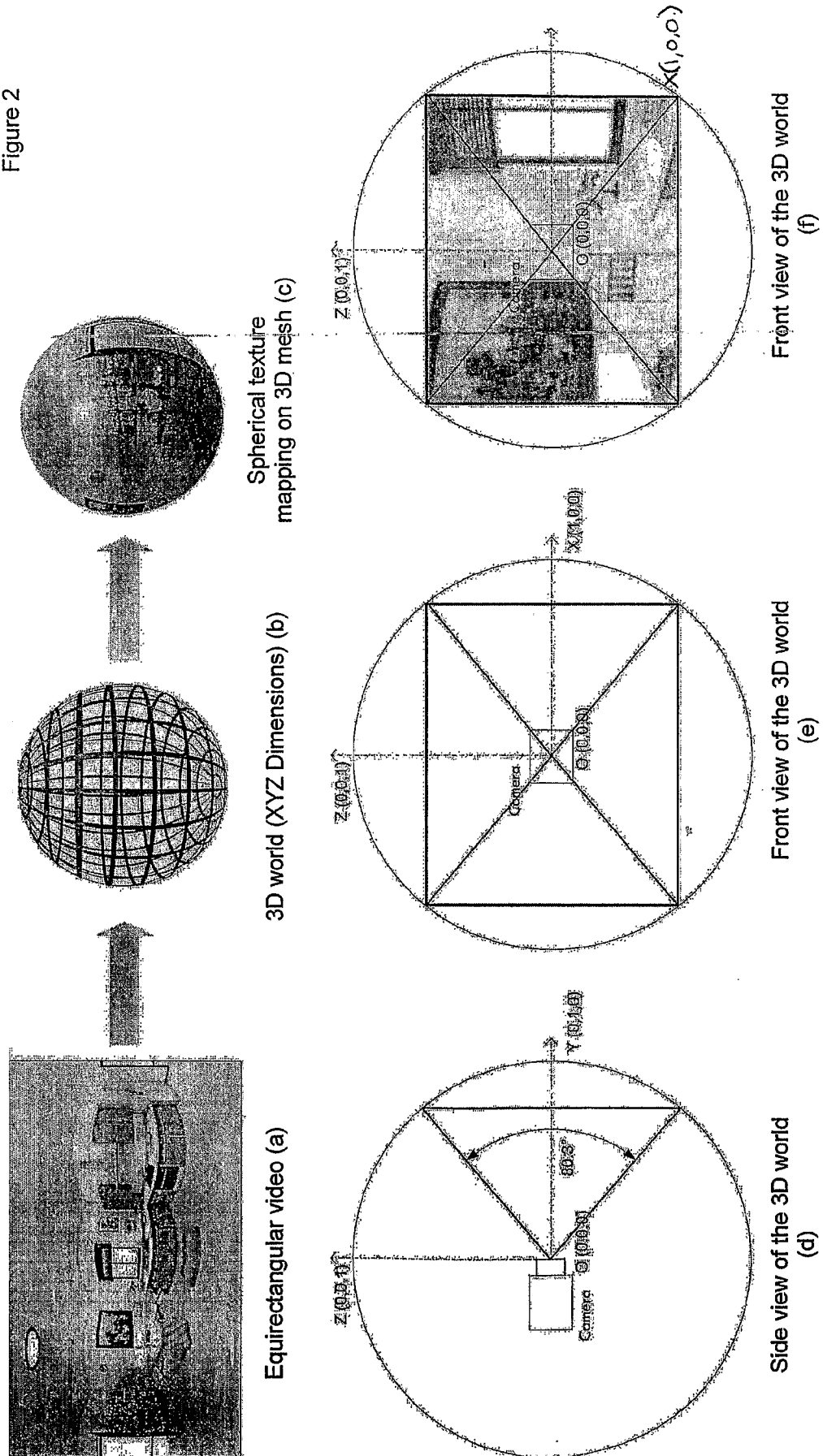
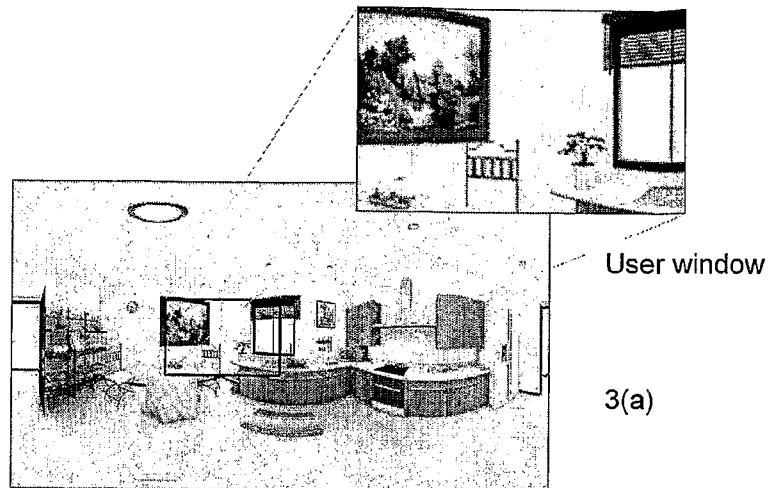
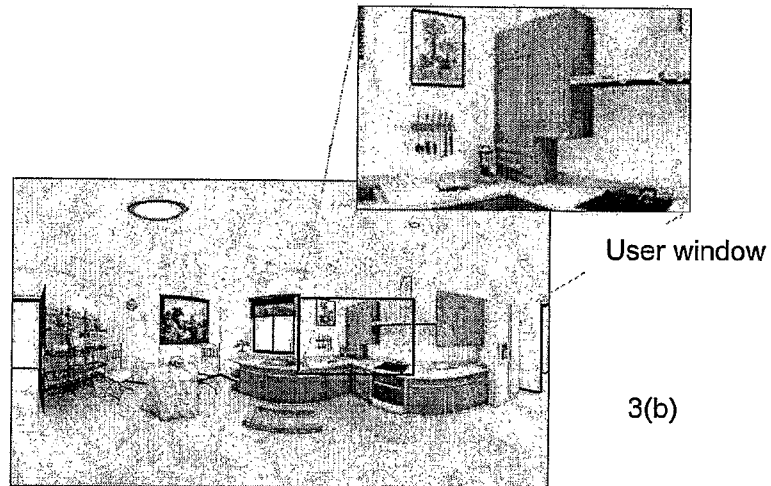


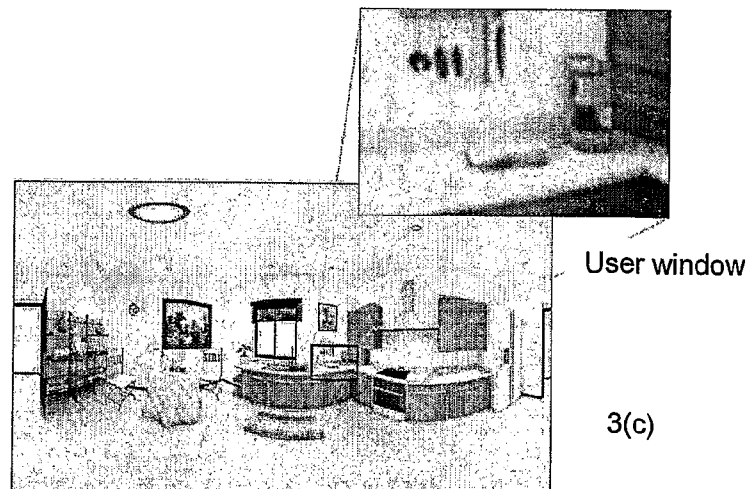
Figure 3



Pan and tilt player function



Pan and tilt player function



Zoom player function

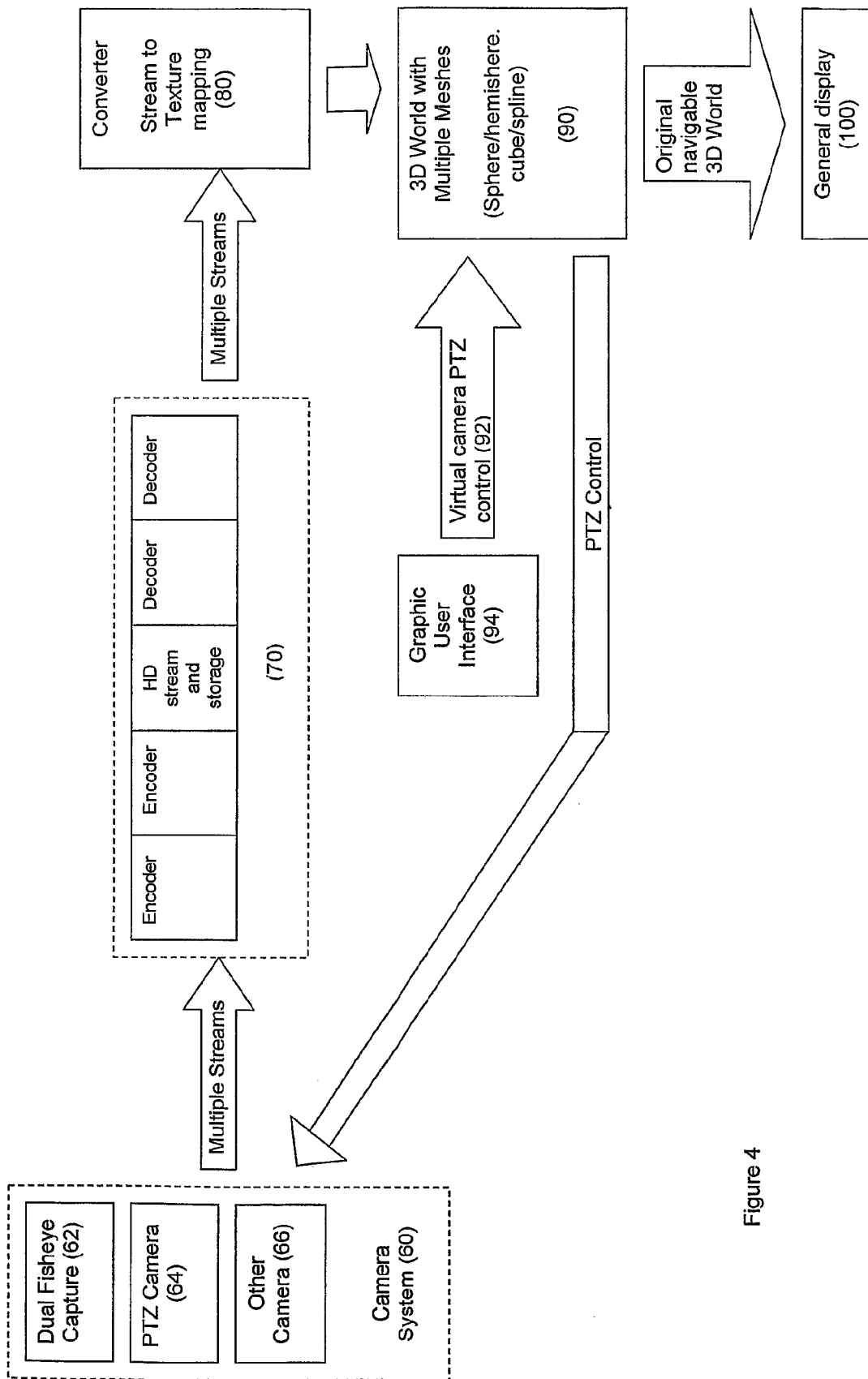
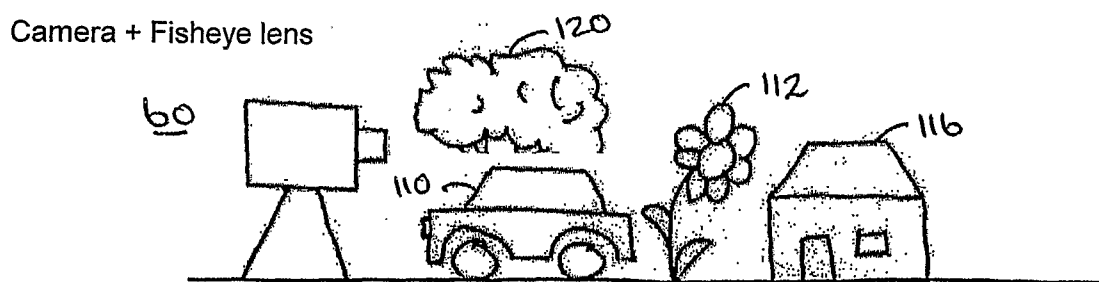
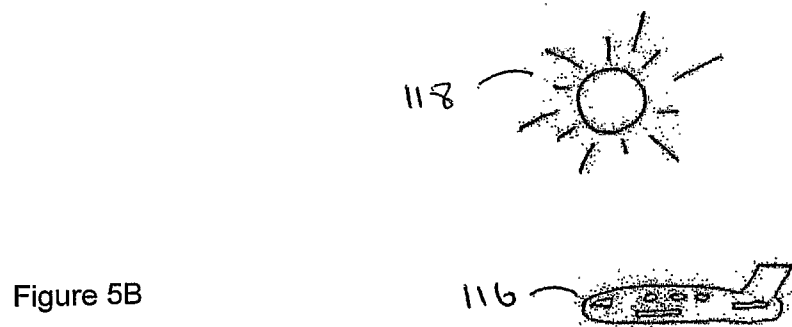
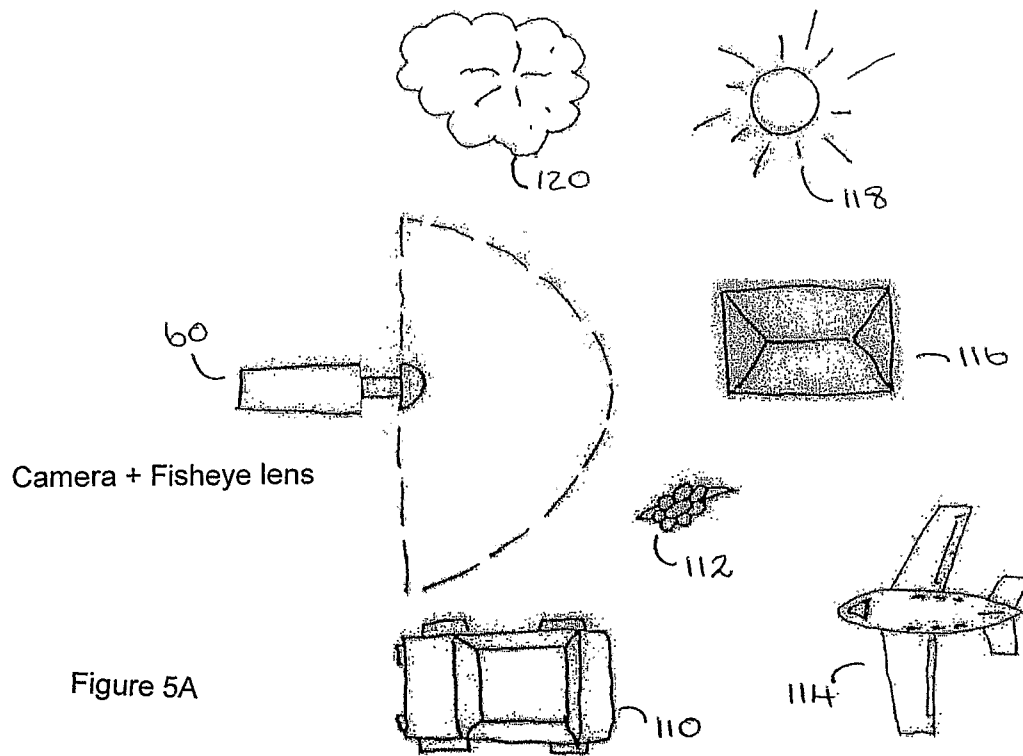
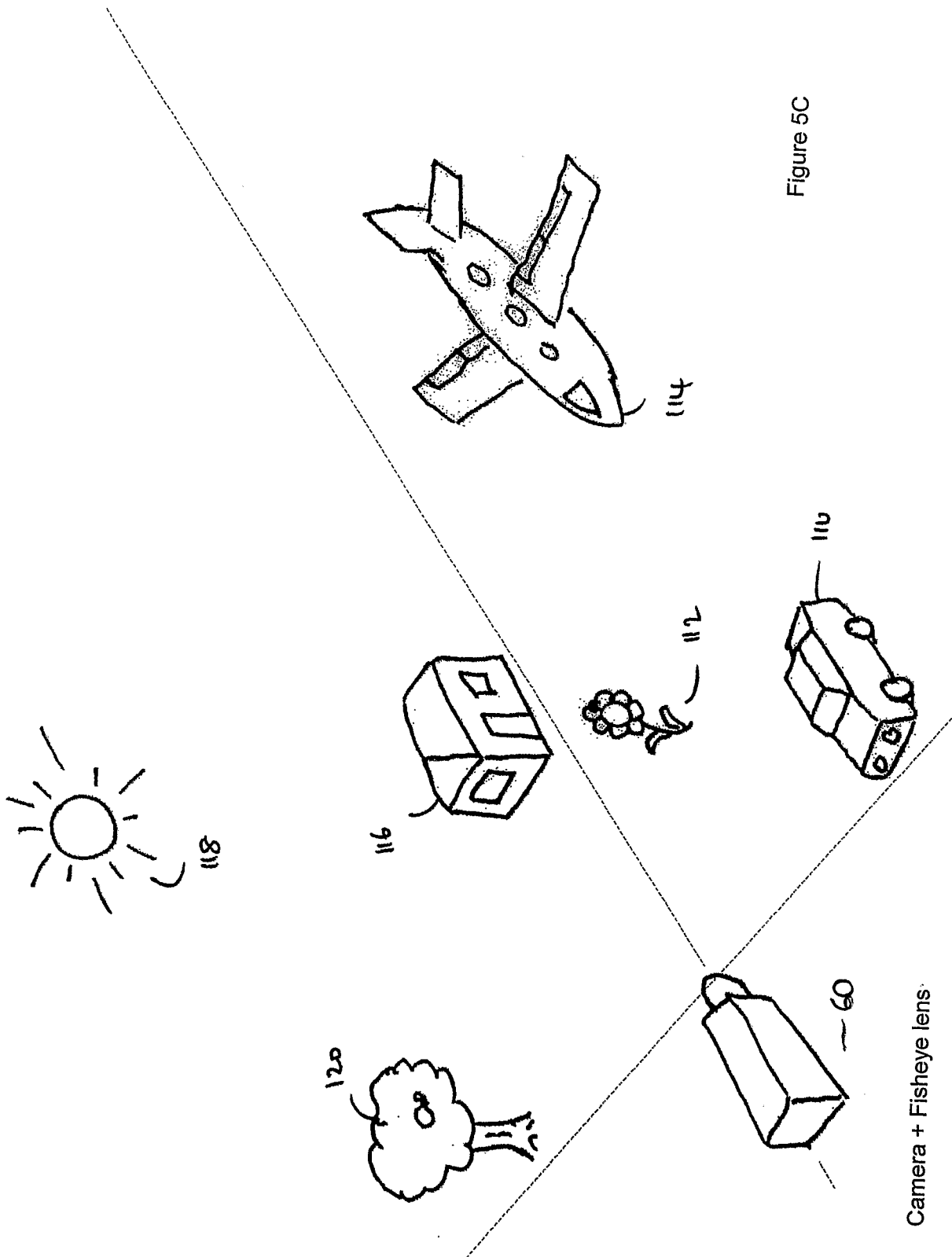
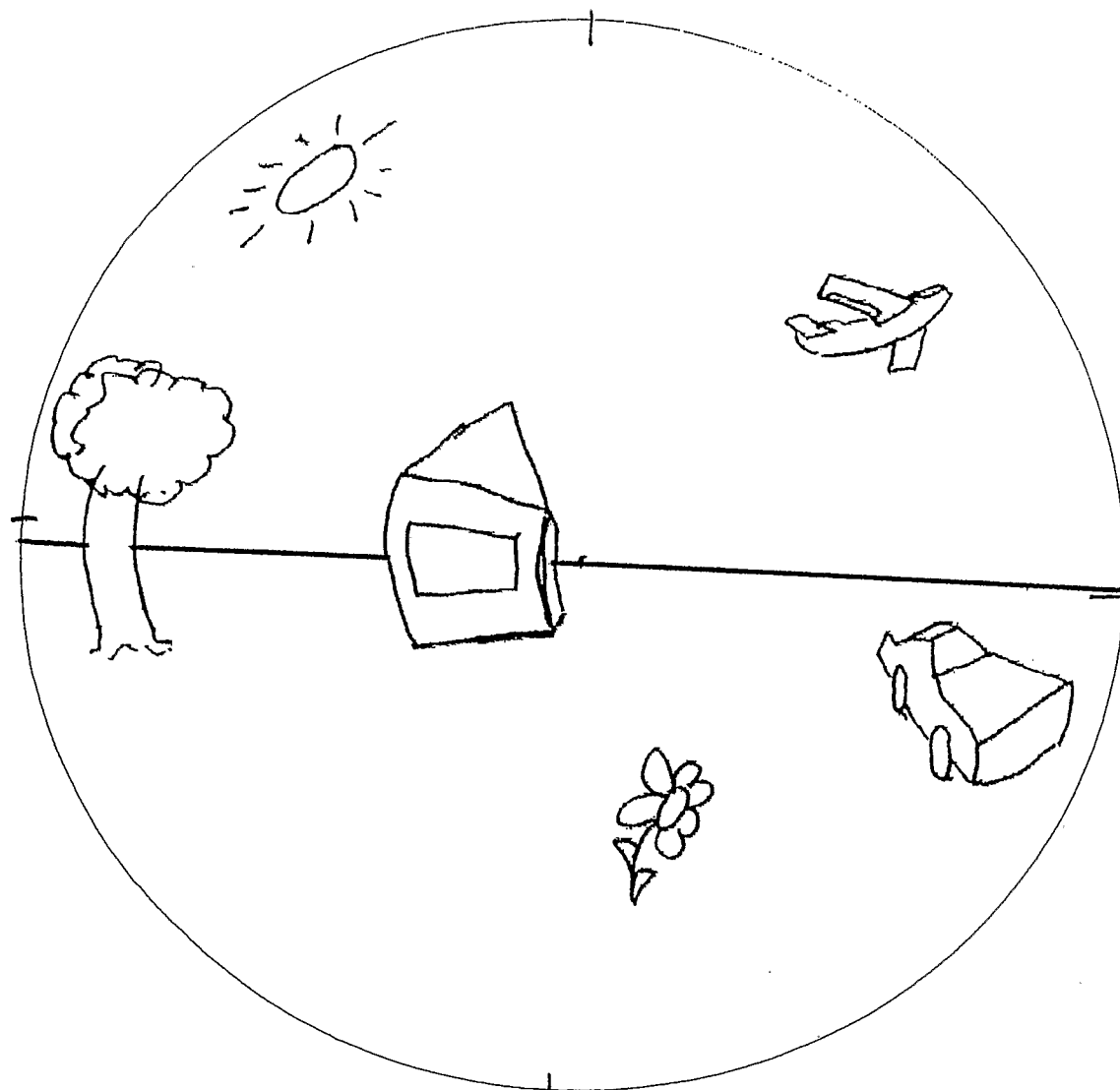


Figure 4



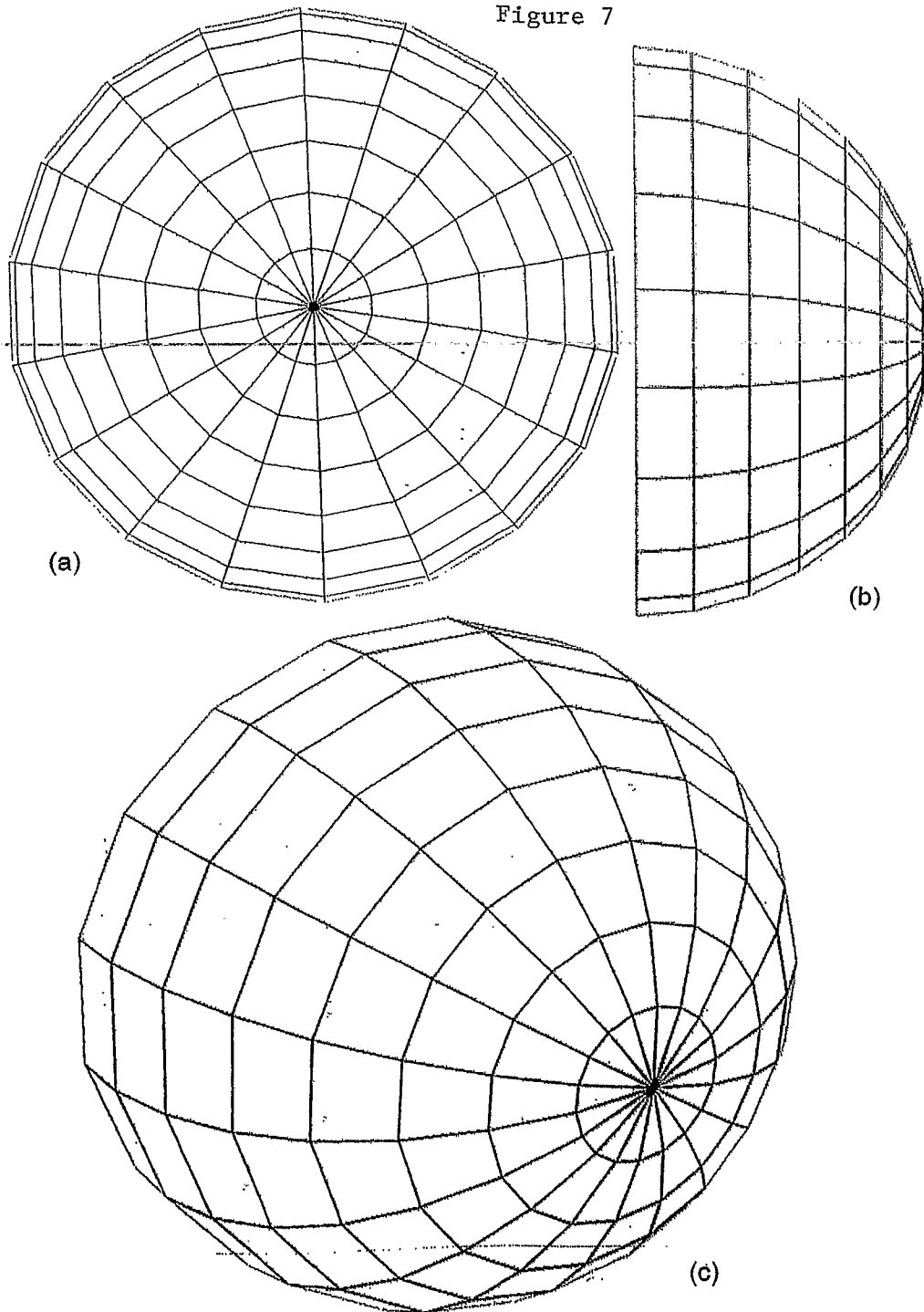




Fisheye view

Figure 6

Figure 7



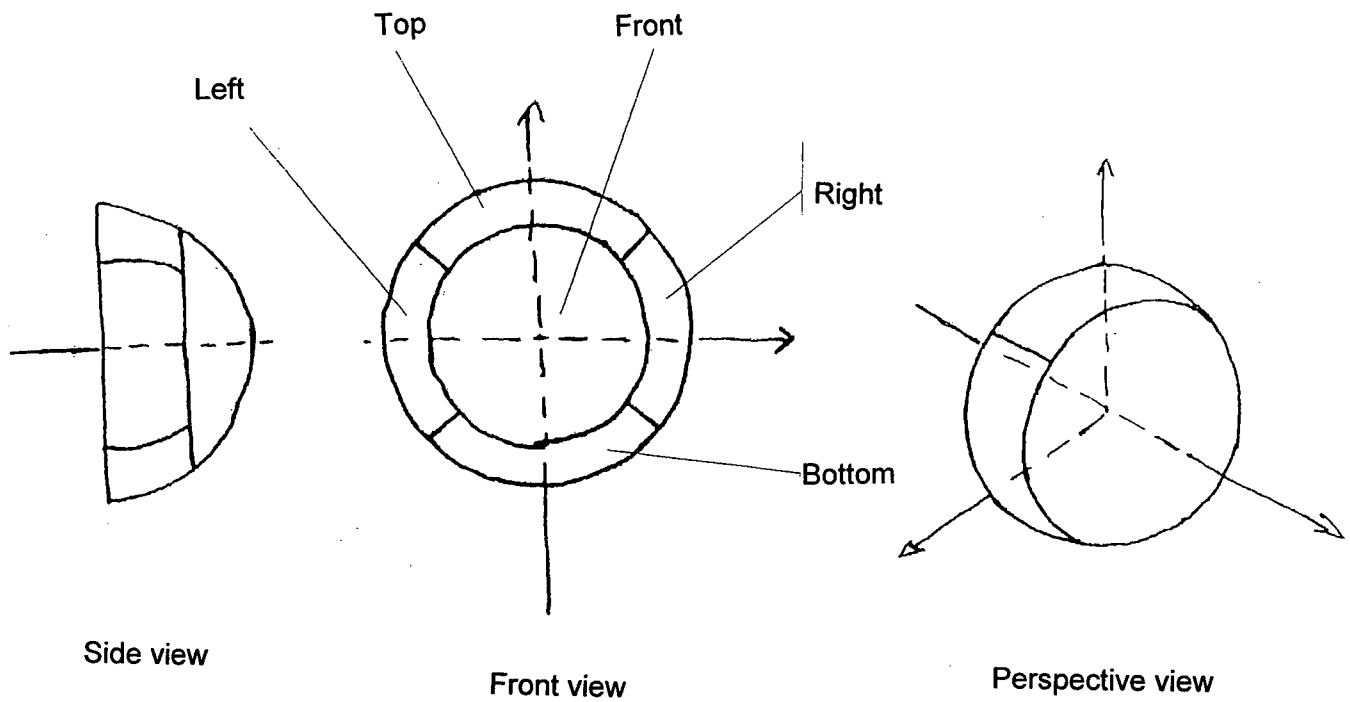


Figure 8a

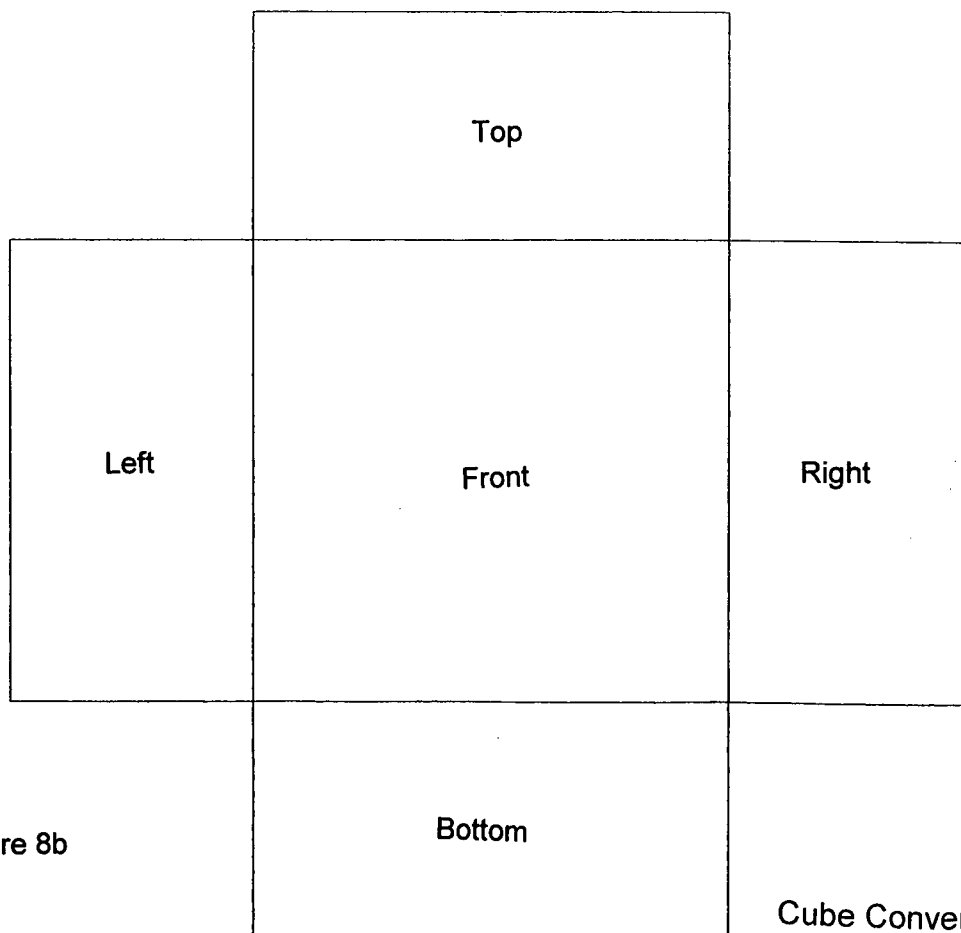


Figure 8b

Cube Conversion

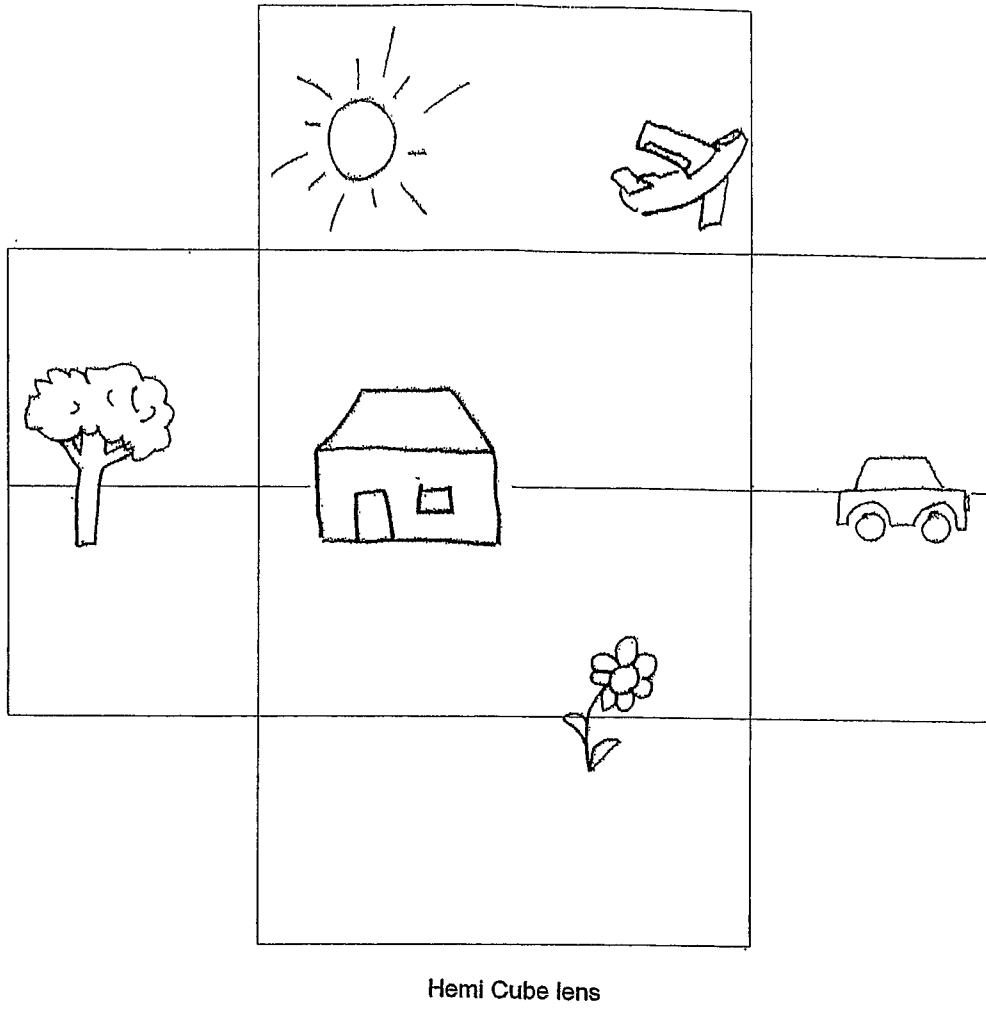


Figure 9

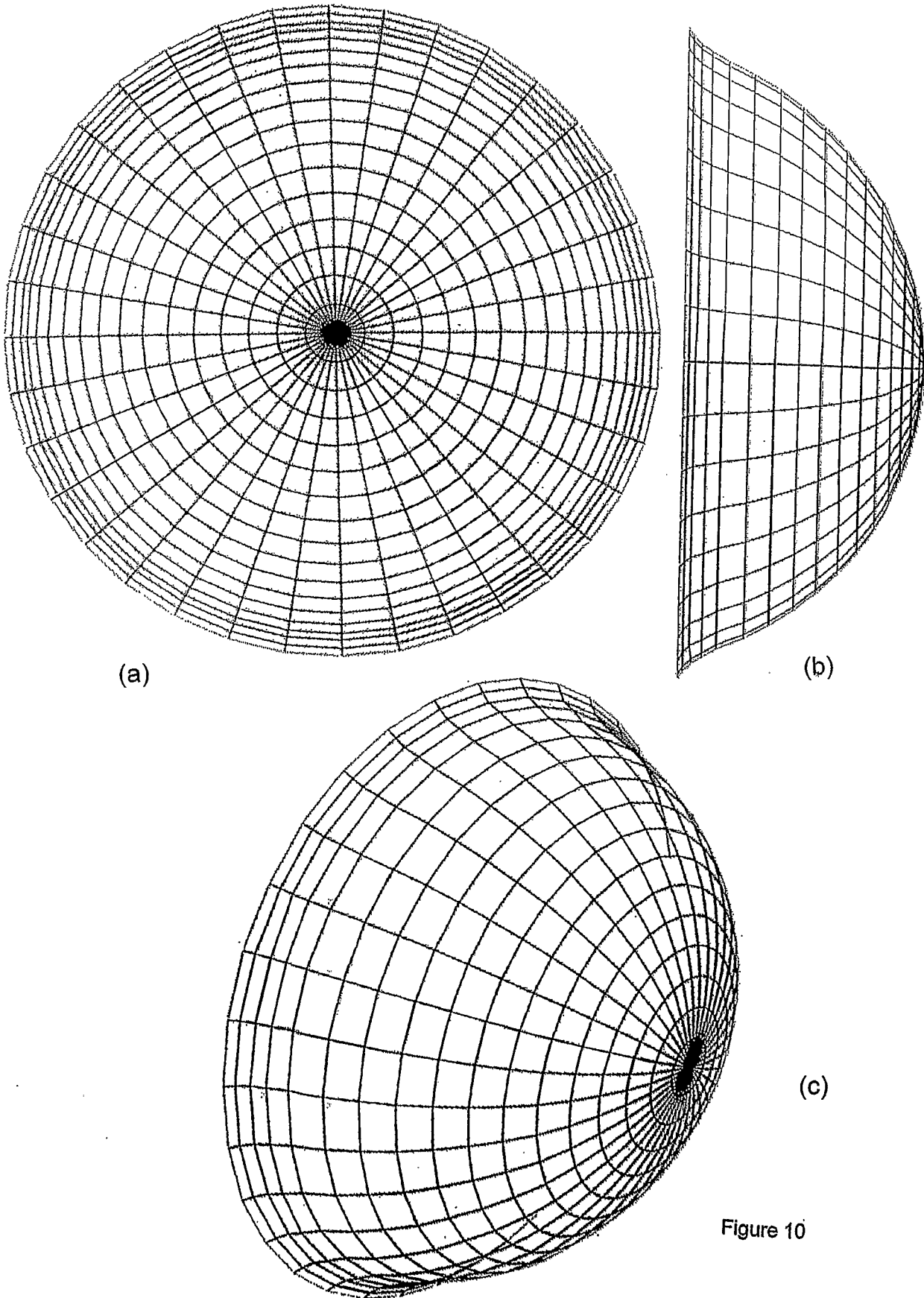


Figure 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2008/001369

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06T3/00 H04N5/262 G06T15/20

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)

G06T H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and where practical, search terms used)

EPO-Internal , WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document with indication, where appropriate of the relevant passages	Relevant to claim No
X	WO 01/95608 A (INTERACTIVE IMAGING SYSTEMS IN [US]; TRAVERS PAUL [US]; THIERRY WESLEY) 13 December 2001 (2001-12-13) the whole document -----	1-32
X	US 2007/124783 A1 (AHISKA BARTU [GB] ET AL) 31 May 2007 (2007-05-31) abstract figures 1,3-6 paragraph [0013] - paragraph [0014] paragraph [0029] - paragraph [0030] paragraph [0032] - paragraph [0040] paragraph [0049] - paragraph [0053] ----- - / - -	1-32



Further documents are listed in the continuation of Box C



See patent family annex

* Special categories of cited documents

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'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)

'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

'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

'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

'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

'&' document member of the same patent family

Date of the actual completion of the international search

10 October 2008

Date of mailing of the international search report

20/10/2008

Name and mailing address of the ISA/

European Patent Office, P B 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel (+31-70) 340-2040,
Fax (+31-70) 340-3016

Authorized officer

Salvador, Elena

INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2008/001369

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
A	<p>CRIMINISI A ET AL: "Image-based interactive exploration of real-world environments"</p> <p>IEEE COMPUTER GRAPHICS AND APPLICATIONS, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 24, no. 3, 1 May 2004 (2004-05-01), pages 52-63, XP011112360</p> <p>ISSN: 0272-1716</p> <p>the whole document</p> <p>-----</p>	1-32
A	<p>US 2005/007478 A1 (AHISKA YAVUZ [GB])</p> <p>13 January 2005 (2005-01-13)</p> <p>the whole document</p> <p>-----</p>	1-32
A	<p>US 6 243 099 B1 (OXAAL FORD [US])</p> <p>5 June 2001 (2001-06-05)</p> <p>the whole document</p> <p>-----</p>	1-32

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/IB2008/001369

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0195608	A	13-12-2001	AU 7200601 A	17-12-2001
			JP 2003536160 T	02-12-2003
			US 2002190987 A1	19-12-2002
US 2007124783	A1	31-05-2007	NONE	
us 2005007478	A1	13-01-2005	NONE	
us 6243099	B1	05-06-2001	NONE	