



(51) International Patent Classification:

Not classified

(21) International Application Number:

PCT/FI2018/050095

(22) International Filing Date:

12 February 2018 (12.02.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1702680.8 20 February 2017 (20.02.2017) GB

(71) Applicant: **NOKIA TECHNOLOGIES OY** [FI/FI];
Karaportti 3, 02610 Espoo (FI).

(72) Inventors: **YOU, Yu**; Aakkulantie 16 A 1, 36220 Kangasala (FI). **WANG, Tinghuai**; Insinöörinkatu 59 B 63, 33720 Tampere (FI). **FAN, Lixin**; Suolijokankatu 2, 33720 Tampere (FI).

(74) Agent: **NOKIA TECHNOLOGIES OY** et al.; Ari Aarnio, IPR Department, Karakaari 7, 02610 Espoo (FI).

(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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: METHODS AND APPARATUSES FOR DETERMINING POSITIONS OF MULTI-DIRECTIONAL IMAGE CAPTURE APPARATUSES

(57) Abstract: This specification describes a method comprising processing a plurality of first images to generate a plurality of stereo-pairs of panoramic images, wherein each first image is captured by a respective camera of a respective one of a plurality of multi-directional image capture apparatuses, and wherein each stereo-pair of panoramic images is generated from first images captured by cameras of a respective one of the plurality of multi-directional image capture apparatuses, performing image re-projection on each panoramic image of the plurality of stereo-pairs of panoramic images, thereby to generate a plurality of re-projected second images which are each associated with a respective virtual camera, processing the plurality of second images to generate respective positions of the virtual cameras associated with the second images, and based on the generated positions of the virtual cameras, determining a position of each of the plurality of multi-directional image capture apparatuses.



Methods and Apparatuses for Determining Positions of Multi-Directional Image Capture Apparatuses

Technical Field

- 5 The present specification relates to methods and apparatuses for determining positions of multi-directional image capture apparatuses.

Background

- 10 Camera pose registration is an important technique used to determine positions and orientations of image capture apparatuses such as cameras. The recent advent of commercial multi-directional image capture apparatuses, such as 360° camera systems, brings new challenges with regard to the performance of camera pose registration in a reliable, accurate and efficient manner.

Summary

- 15 According to a first aspect, this specification describes a method comprising processing a plurality of first images to generate a plurality of stereo-pairs of panoramic images, wherein each first image is captured by a respective camera of a respective one of a plurality of multi-directional image capture apparatuses, and wherein each stereo-pair of
20 panoramic images is generated from first images captured by cameras of a respective one of the plurality of multi-directional image capture apparatuses, performing image re-projection on each panoramic image of the plurality of stereo-pairs of panoramic images, thereby to generate a plurality of re-projected second images which are each associated with a respective virtual camera, processing the plurality of second images to generate
25 respective positions of the virtual cameras associated with the second images, and based on the generated positions of the virtual cameras, determining a position of each of the plurality of multi-directional image capture apparatuses.

The first images may be fisheye images.

- 30 Processing the plurality of first images to generate the plurality of stereo-pairs of panoramic images may comprise de-warping the first images and stitching the de-warped images to generate the panoramic images.

- 35 The second images may be rectilinear images.

- 2 -

The processing of the plurality of second images to generate respective positions of the virtual cameras may comprise processing the second images using a structure from motion algorithm to generate the positions of the virtual cameras.

- 5 The panoramic images of each stereo-pair may be offset from each other by a baseline distance.

The baseline distance to be used may be a predetermined fixed distance.

- 10 The baseline distance to be used may be determined by: minimising a cost function which indicates an error associated with use of each of a plurality of baseline distances, and determining that the baseline distance associated with the lowest error is to be used.

- 15 The processing of the plurality of second images to generate respective positions of the virtual cameras may comprise processing the second images using a structure from motion algorithm to generate the positions of the virtual cameras and the cost function may be a weighted average of: re-projection error from the structure from motion algorithm and variance of calculated baseline distances between stereo-pairs of virtual cameras.

- 20 The method may further comprise determining a pixel to real world distance conversion factor based on the determined positions of the virtual cameras and the baseline distance used.

- 25 The processing of the plurality of second images may generate respective orientations of the virtual cameras, and the first aspect may further comprise: based on the generated orientations of the virtual cameras, determining an orientation of each of the plurality of multi-directional image capture apparatuses.

- 30 According to a second aspect, this specification describes apparatus configured to perform any method described with reference to the first aspect.

- According to a third aspect, this specification describes computer-readable instructions which, when executed by computing apparatus, cause the computing apparatus to perform
35 any method described with reference to the first aspect.

- 3 -

According to a fourth aspect, this specification describes apparatus comprising at least one processor, and at least one memory including computer program code, which when executed by the at least one processor, causes the apparatus to: process a plurality of first images to generate a plurality of stereo-pairs of panoramic images, wherein each first
5 image is captured by a respective camera of a respective one of a plurality of multi-directional image capture apparatuses, and wherein each stereo-pair of panoramic images is generated from first images captured by cameras of a respective one of the plurality of multi-directional image capture apparatuses, perform image re-projection on each panoramic image of the plurality of stereo-pairs of panoramic images, thereby to generate
10 a plurality of re-projected second images which are each associated with a respective virtual camera, process the plurality of second images to generate respective positions of the virtual cameras associated with the second images, and based on the generated positions of the virtual cameras, determine a position of each of the plurality of multi-directional image capture apparatuses.

15 The first images may be fisheye images.

Processing the plurality of first images to generate the plurality of stereo-pairs of panoramic images may comprise de-warping the first images and stitching the de-warped
20 images to generate the panoramic images.

The second images may be rectilinear images.

The processing of the plurality of second images to generate respective positions of the
25 virtual cameras may comprise processing the second images using a structure from motion algorithm to generate the positions of the virtual cameras.

The panoramic images of each stereo-pair may be offset from each other by a baseline distance.

30 The baseline distance to be used may be a predetermined fixed distance.

The baseline distance to be used may be determined by: minimising a cost function which indicates an error associated with use of each of a plurality of baseline distances, and
35 determining that the baseline distance associated with the lowest error is to be used.

- 4 -

The processing of the plurality of second images to generate respective positions of the virtual cameras may comprise processing the second images using a structure from motion algorithm to generate the positions of the virtual cameras and the cost function may be a weighted average of: re-projection error from the structure from motion
5 algorithm and variance of calculated baseline distances between stereo-pairs of virtual cameras.

The computer program code, when executed by the at least one processor, may cause the apparatus to determine a pixel to real world distance conversion factor based on the
10 determined positions of the virtual cameras and the baseline distance used.

The processing of the plurality of second images may generate respective orientations of the virtual cameras, and the computer program code, when executed by the at least one processor, may cause the apparatus to: determine an orientation of each of the plurality of
15 multi-directional image capture apparatuses based on the generated orientations of the virtual cameras.

According to a fifth aspect, this specification describes computer-readable medium having computer-readable code stored thereon, the computer readable code, when executed by at
20 least one processor, causes performance of: processing a plurality of first images to generate a plurality of stereo-pairs of panoramic images, wherein each first image is captured by a respective camera of a respective one of a plurality of multi-directional image capture apparatuses, and wherein each stereo-pair of panoramic images is
25 generated from first images captured by cameras of a respective one of the plurality of multi-directional image capture apparatuses, performing image re-projection on each panoramic image of the plurality of stereo-pairs of panoramic images, thereby to generate a plurality of re-projected second images which are each associated with a respective virtual camera, processing the plurality of second images to generate respective positions of the virtual cameras associated with the second images, and based on the generated
30 positions of the virtual cameras, determining a position of each of the plurality of multi-directional image capture apparatuses.

The computer-readable code stored on the medium of the fifth aspect may further cause performance of any of the operations described with reference to the method of the first
35 aspect.

- 5 -

According to a sixth aspect, this specification describes apparatus comprising means for processing a plurality of first images to generate a plurality of stereo-pairs of panoramic images, wherein each first image is captured by a respective camera of a respective one of a plurality of multi-directional image capture apparatuses, and wherein each stereo-pair of
5 panoramic images is generated from first images captured by cameras of a respective one of the plurality of multi-directional image capture apparatuses, means for performing image re-projection on each panoramic image of the plurality of stereo-pairs of panoramic images, thereby to generate a plurality of re-projected second images which are each associated with a respective virtual camera, means for processing the plurality of second
10 images to generate respective positions of the virtual cameras associated with the second images, and means for determining a position of each of the plurality of multi-directional image capture apparatuses based on the generated positions of the virtual cameras.

The apparatus of the sixth aspect may further comprise means for causing performance of
15 any of the operations described with reference to the method of the first aspect.

Brief Description of the Drawings

For a more complete understanding of the methods, apparatuses and computer-readable instructions described herein, reference is now made to the following description taken in
20 connection with the accompanying drawings, in which:

Figure 1 illustrates an example of multiple multi-directional image capture apparatuses in an environment;

Figures 2A and 2B illustrate examples of ways in which images captured by a multi-
25 directional image capture apparatus are processed;

Figures 3A and 3B illustrate the determination of the position and orientation of a multi-directional image capture apparatus relative to a reference coordinate system;

Figure 4 is a flowchart illustrating examples of various operations which may be performed by an image processing apparatus based on a plurality of images captured by a
30 plurality of multi-directional image capture apparatuses;

Figure 5 is a schematic diagram of an example configuration of an image processing apparatus configured to perform various operations including those described with reference to Figure 4;

Figure 6 illustrates an example of a computer-readable storage medium with computer
35 readable instructions stored thereon.

Detailed Description

- 6 -

In the description and drawings, like reference numerals may refer to like elements throughout.

Figure 1 illustrates a plurality of multi-directional image capture apparatuses 10 located within an environment. The multi-directional image capture apparatuses 10 may, in general, be any apparatus capable of capturing images of a scene 13 from multiple different perspectives simultaneously. For example, multi-directional image capture apparatus 10 may be a 360° camera system (also known as an omnidirectional camera system or a spherical camera system). However, it will be appreciated that multi-directional image capture apparatus 10 does not necessarily have to have full angular coverage of its surroundings and may only cover a smaller field of view.

The term “image” used herein may refer generally to visual content. This may be visual content captured by, or derived from visual content captured by, multi-directional image capture apparatus 10. For example, an image may be a photograph or a single frame of a video.

As illustrated in Figure 1, each multi-directional image capture apparatus 10 may comprise a plurality of cameras 11. The term “camera” used herein may refer to a sub-part of a multi-directional image capture apparatus 10 which performs the capturing of images. As illustrated, each of the plurality of cameras 11 of multi-directional image capture apparatus 10 may be facing a different direction to each of the other cameras 11 of the multi-directional image capture apparatus 10. As such, each camera 11 of a multi-directional image capture apparatus 10 may have a different field of view, thus allowing the multi-directional image capture apparatus 10 to capture images of a scene 13 from different perspectives simultaneously.

Similarly, as illustrated in Figure 1, each multi-directional image capture apparatus 10 may be at a different location to each of the other multi-directional image capture apparatuses 10. Thus, each of the plurality of multi-directional image capture apparatuses 10 may capture images of the environment (via their cameras 11) from different perspectives simultaneously.

In the example scenario illustrated in Figure 1, a plurality of multi-directional image capture apparatuses 10 are arranged to capture images of a particular scene 13 within the environment. In such circumstances, it may be desirable to perform camera pose registration in order to determine the position and orientation of each of the multi-

- 7 -

directional image capture apparatuses 10. In particular, it may be desirable to determine these positions and orientations relative to a particular reference coordinate system. This allows the overall arrangement of the multi-directional image capture apparatuses 10 relative to each other to be determined, which may be useful for a number of functions.

5 For example, such information may be used for any of the following: performing 3D reconstruction of the captured environment, performing 3D registration of the multi-directional image capture apparatuses 10 with respect to other sensors such as LiDAR (Light Detection and Ranging) or infrared (IR) depth sensors, audio positioning of audio sources, playback of object-based audio with respect to multi-directional image capture
10 apparatus 10 location, and presenting multi-directional image capture apparatuses positions as 'hotspots' to which a viewer can switch during virtual reality (VR) viewing.

One way of determining the positions of multi-directional image capture apparatuses 10 is to use Global Positioning System (GPS) localization. However, GPS only provides position
15 information and does not provide orientation information. In addition, position information obtained by GPS may not be very accurate and may be susceptible to changes in the quality of the satellite connection. One way of determining orientation information is to obtain the orientation information from magnetometers and accelerometers installed in the multi-directional image capture apparatuses 10. However, such instruments may be
20 susceptible to local disturbance (e.g. magnetometers may be disturbed by a local magnetic field), so the accuracy of orientation information obtained in this way is not necessarily very high.

Another way of performing camera pose registration is to use a computer vision method.
25 For example, position and orientation information can be obtained by performing structure from motion (SfM) analysis on images captured by a multi-directional image capture apparatus 10. Broadly speaking, SfM works by determining point correspondences between images (also known as feature matching) and calculating location and orientation based on the determined point correspondences.

30 However, when multi-directional image capture apparatuses 10 are used to capture a scene which lacks distinct features/textures (e.g. a corridor), determination of point correspondences between captured images may be unreliable due to the lack of distinct features/textures in the limited field of view of the images. In addition, since multi-
35 directional image capture apparatuses 10 typically capture fish-eye images, it may not be possible to address this by capturing fish-eye images with increased field of view, as this

- 8 -

will lead to increased distortion of the images which may negatively impact point correspondence determination.

A computer vision method for performing camera pose registration which may address some or all of the challenges mentioned above will now be described.

Figure 2A illustrates one of the plurality of multi-directional image capture apparatuses 10 of Figure 1. Each of the cameras 11 of the multi-directional image capture apparatus 10 may capture a respective first image 21. Each first image 21 may be an image of a scene within the field of view 20 of its respective camera 11. In some examples, the lens of the camera 11 may be a fish-eye lens and so the first image 21 may be a fish-eye image (in which the camera field of view is enlarged). However, the method described herein may be applicable for use with lenses and resulting images of other types. More specifically, the camera pose registration method described herein may also be applicable to images captured by a camera with a hyperbolic mirror in which the camera optical centre coincides with the focus of the hyperbola, and images captured by a camera with a parabolic mirror and an orthographic lens in which all reflected rays are parallel to the mirror axis and the orthographic lens is used to provide a focused image.

The first images 21 may be processed to generate a stereo-pair of panoramic images 22. Each panoramic image 22 of the stereo-pair may correspond to a different view of a scene captured by the first images 21 from which the stereo-pair is generated. For example, one panoramic image 22 of the stereo-pair may represent a left-eye panoramic image and the other one of the stereo-pair may represent a right-eye panoramic image. As such, the stereo-pair of panoramic images 22 may be offset from each other by a baseline distance B. By generating panoramic images 22 as an initial step, the effective field of view may be increased, which may allow the methods described herein to better deal with scenes which lack distinct textures (e.g. corridors).

The generated panoramas may be referred to as spherical (or part-spherical) panoramas in the sense that they may include image data from a sphere (or part of a sphere) around the multi-directional image capture apparatus 10.

If the first images 21 are fish eye images, processing the first images to generate the panoramic images may comprise de-warping the first images 21 and then stitching the de-warped images. De-warping the first images 21 may comprise re-projecting each of the first images to convert the first images 21 from a fish eye projection to a spherical

- 9 -

projection. Fish eye to spherical re-projections are generally known in the art and will not be described here in detail. Stitching the de-warped images may, in general, be performed using any suitable image stitching technique. Many image stitching techniques are known in the art and will not be described here in detail. Generally, image stitching involves
5 connecting portions of images together based on point correspondences between images (which may involve feature matching).

Following the generation of the stereo-pair of panoramic images 22, the stereo pair may be processed to generate one or more second images 23. More specifically, image re-
10 re-projection may be performed on each of the panoramic images 22 to generate one or more re-projected second images 23. For example, if the panoramic image 22 is not rectilinear (e.g. if it is curvilinear), it may be re-projected to generate one or more second images 23 which are rectilinear images. As illustrated in Figure 2A, a corresponding set of second images 23 may be generated for each panoramic image 22 of the stereo pair. The type of
15 re-projection may be dependent on the algorithm used to analyse the second images 23. For instance, as is explained below, structure from motion algorithms, which are typically used to analyse rectilinear images, may be used, in which case the re-projection may be selected so as to generate rectilinear images. However, it will be appreciated that, in general, the re-projection may generate any type of second image 23, as long as the image
20 type is compatible with the algorithm used to analyse the re-projected images 23.

Each re-projected second image 23 may be associated with a respective virtual camera. A virtual camera is an imaginary camera which does not physically exist, but which corresponds to a camera which would have captured the re-projected second image 23
25 with which it is associated. A virtual camera may be defined by virtual camera parameters which represent the configuration of the virtual camera required in order to have captured to the second image 23. As such, for the purposes of the methods and operations described herein, a virtual camera can be treated as a real physical camera. For example, each virtual camera has, among other virtual camera parameters, a position and
30 orientation which can be determined.

As illustrated by Figure 2B, the processing of each panoramic image 22 may be performed by resampling the panoramic image 22 based on a horizontal array of overlapping sub-
portions 22-1 of the panoramic image 22. The sub-portions 22-1 may be chosen to be
35 evenly spaced so that adjacent sub-portions 22-1 are separated by the same distance (as illustrated by Figure 2B). As such, the viewing directions of adjacent sub-portions 22-1 may differ by the same angular distance. A corresponding re-projected second image 23

- 10 -

may be generated for each sub-portion 22-1. This may be performed by casting rays following the pinhole camera model (which represents a first order approximation of the mapping from the spherical (3D) panorama to the 2D second images) based on a given field of view (e.g. 120 degrees) of each sub-portion 22-1 from a single viewpoint to the panoramic image 23. As such, each re-projected second image 23 may correspond to a
5 respective virtual pinhole camera. The virtual pinhole cameras associated with second images 23 generated from one panoramic image 22 may all have the same position, but different orientations (as illustrated by Figure 3A).

10

Each second image 23 generated from one of the stereo-pair of panoramic images 22 may form a stereo pair with a second image 23 from the other one of the stereo-pair of panoramic images 22. As such, each stereo-pair of second images 23 may correspond to a stereo-pair of virtual cameras. Each stereo-pair of virtual cameras may be offset from each
15 other by the baseline distance as described above.

It will be appreciated that, in general, any number of second images 23 may be generated. Generally speaking, generating more second images 23 may lead to less distortion in each of the second images 23, but may also increase computational complexity. The precise
20 number of second images 23 may be chosen based on the scene/environment being captured by the multi-directional image capture apparatus 10.

The methods described with reference to Figures 2A and 2B may be performed for each of a plurality of multi-directional image capture apparatuses 10 which are capturing the
25 same general environment, e.g. the plurality of multi-directional images capture apparatuses 10 as illustrated in Figure 1. In this way, all of the first images 21 captured by a plurality of multi-directional image capture apparatuses 10 of a particular scene may be processed as described above.

30 It will be appreciated that the first images 21 may correspond to images of a scene at a particular moment in time. For example, if the multi-directional image capture apparatuses 10 are capturing video images, a first image 21 may correspond to a single video frame of a single camera 11, and all of the first images 21 may be video frames that are captured at the same moment in time.

35

Figures 3A and 3B illustrate the process of determining the position and orientation of a multi-directional image capture apparatus 10. In Figures 3A and 3B, each arrow 31, 32

- 11 -

represents the position and orientation of a particular element in a reference coordinate system 30. The base of the arrow represents the position and the direction of the arrow represents the orientation. More specifically, each arrow 31 in Figure 3A represents the position and orientation of a virtual camera associated with a respective second image 23,
5 and the arrow 32 in Figure 3B represents the position and orientation of the multi-directional image capture apparatus 10.

After generating the second images 23, the second images 23 may be processed to generate respective positions of the virtual cameras associated with the second images 23.

10 The output of the processing for one multi-directional image capture apparatus 10 is illustrated by Figure 3A. The processing may include generating the positions of a set of virtual cameras for each panoramic image 22 of the stereo-pair of panoramic images. As illustrated by Figure 3A, one set of arrows 33A may correspond to virtual cameras of one of the stereo-pair of panoramic images 22, and the other set of arrows 33B may
15 correspond to virtual cameras of the other one of the stereo-pair of panoramic images. The generated positions may be relative to the reference coordinate system 30. The processing of the second images may also generate respective orientations of the virtual cameras relative to the reference coordinate system 30. As mentioned above and illustrated by Figure 3A, all of the virtual cameras of each set of virtual cameras, which correspond to
20 the same panoramic image 22, may have the same position but different orientations.

It will be appreciated that, in order to perform the processing for a plurality of multi-directional image capture apparatuses 10, it may be necessary for the multi-directional image capture apparatuses 10 to have at least partially overlapping fields of view with each
25 other (for example, in order to allow point correspondence determination as described below).

The above described processing may be performed by using a structure from motion (SfM) algorithm to determine the position and orientation of each of the virtual cameras. The
30 SfM algorithm may operate by determining point correspondences between various ones of the second images 23 and determining the positions and orientations of the virtual cameras based on the determined point correspondences. For example, the determined point correspondences may impose certain geometric constraints on the positions and orientations of the virtual cameras, which can be used to solve a set of quadratic equations
35 to determine the positions and orientations of the virtual cameras relative to the reference coordinate system 30. More specifically, in some examples, the SfM process may involve any one of or any combination of the following operations: extracting images features,

- 12 -

matching image features, estimating camera position, reconstructing 3D points, and performing bundle adjustment.

Once the positions of the virtual cameras have been determined, the position of the multi-directional image capture apparatus 10 relative to the reference coordinate system 30 may be determined based on the determined positions of the virtual cameras. Similarly, once the orientations of the virtual cameras have been determined, the orientation the multi-directional image capture apparatus 10 relative to the reference coordinate system 30 may be determined based on the determined orientations of the virtual cameras. The position of the multi-directional image capture apparatus 10 may be determined by averaging the positions of the two sets 33A, 33B of virtual cameras illustrated by Figure 3A. For example, as illustrated, all of the virtual cameras of one set 33A may have the same position as each other and all of the virtual cameras of the other set 33B may also have the same position as each other. As such, the position of the multi-directional image capture apparatus 10 may be determined to be the average of the two respective positions of the two sets 33A, 33B of virtual cameras.

Similarly, the orientation of the multi-directional image capture apparatus 10 may be determined by averaging the orientation of the virtual cameras. In more detail, the orientation of the multi-directional image capture apparatus 10 may be determined in the following way.

The orientation of each virtual camera may be represented by rotation matrix R_l . The orientation of the multi-directional image capture apparatus 10 may be represented by rotation matrix R_{dev} . The orientation of each virtual camera relative to the multi-directional image capture apparatus 10 may be known, and may be represented by rotation matrix R_{ldev} . Thus, the rotation matrices R_l of the virtual cameras may be used to obtain a rotation matrix for multi-directional image capture apparatus 10 the according to:

$$R_{dev} = R_l R_{ldev}^{-1}$$

Put another way, the rotation matrix of a multi-direction image capture apparatus (R_{dev}) can be determined by multiplying the rotation matrix of a virtual camera (R_l) onto the inverse of the matrix representing the orientation of the virtual camera relative to the orientation of the multi-directional image capture apparatus (R_{ldev}^{-1}).

- 13 -

For example, if there are twelve virtual cameras (six from each panoramic image 22 of the stereo-pair of panoramic images) corresponding to the multi-directional image capture apparatus 10 (as illustrated in Figure 3A) then twelve rotation matrices are obtained for the orientation of the multi-directional image capture apparatus 10. Each of these rotation

5 matrices may then be converted into corresponding Euler angles to obtain a set of Euler angles for the multi-directional image capture apparatus 10. The set of Euler angles may then be averaged and converted into a final rotation matrix representing the orientation of the multi-directional image capture apparatus 10.

10 The set of Euler angles may then be averaged according to:

$$\theta_l = \arctan \frac{\sum_{i=0}^5 \sin(\theta_i)}{\sum_{i=0}^5 \cos(\theta_i)}$$

Where θ_l represents the averaged Euler angles for a multi-directional image capture

15 apparatus 10 and θ_i represents the set of Euler angles. Put another way, the averaged Euler angles are determined by calculating the sum of the sines of the set of Euler angles divided by the sum of the cosines of the set of Euler angles, and taking the arctangent of the ratio. θ_l may then be converted back into a rotation matrix representing the final determined orientation of multi-directional image capture apparatus 10.

20

It will be appreciated that the above formula is for the specific example in which there are nine virtual cameras – the maximum value of i may vary according to the number of virtual cameras generated. For example, if there are twelve virtual cameras as illustrated in Figure 3A, then i may take values from zero to eleven.

25

In some examples, unit quaternions may be used instead of Euler angles for the abovementioned process. The use of unit quaternions to represent orientation is a known mathematical technique and will not be described in detail here. Briefly, quaternions $q_1, q_2, \dots, q_N$ corresponding to the virtual camera rotation matrices may be determined. Then,

30 the quaternions may be transformed, as necessary, to ensure that they are all on the same side of the 4D hypersphere. Specifically, one representative quaternion q_M is selected and the signs of any quaternions q_l where the product of q_M and q_l is less than zero may be inverted. Then, all quaternions q_l (as 4D vectors) may be summed into an average quaternion q_A , and q_A may be normalised into a unit quaternion q_A' . The unit quaternion

35 q_A' may represent the averaged orientation of the camera and may be converted back to

- 14 -

other orientation representations as desired. Using unit quaternions to represent orientation may be more numerically stable than Euler angles.

It will be appreciated that the generated positions of the virtual cameras (e.g. from the SfM algorithm) may be in units of pixels. Therefore, in order to enable scale conversions between pixels and a real world distance (e.g. metres), a pixel to real world distance conversion factor may be determined. This may be performed by determining the baseline distance B of a stereo-pair of virtual cameras in both pixels and in a real world distance. The baseline distance in pixels may be determined from the determined positions of the virtual cameras in the reference coordinate system 30. The baseline distance in a real world distance (e.g. metres) may be known already from being set initially during the generation of the panoramic images 22. The pixel to real world distance conversion factor may then be simply calculated by taking the ratio of the two distances. This may be further refined by calculating the conversion factor based on each of the stereo-pairs of virtual cameras, determining outliers and inliers (as described in more detail below), and averaging the inliers to obtain a final pixel to real world distance conversion factor. The pixel to real world distance conversion factor may be denoted $S_{\text{pixel2meter}}$ in the present specification.

The inlier and outlier determination may be performed according to:

$$d_i = |S_i - \text{Median}(S)|, \forall S_i \in S$$

$$d_\sigma = \text{Median}(\{d_0, \dots, d_N\})$$

$$\text{inliers} = \frac{d_i}{d_\sigma} < m, \forall i \in N$$

where S is the set of pixel to real world distance ratios of all stereo-pairs of virtual cameras, d_i is a measure of the difference between a pixel to real world distance ratio and the median of all pixel to real world distance ratios, d_σ is the median absolute deviation (MAD), m is a threshold value below which a determined pixel to real world distance ratio is considered an inlier (for example, m may be set to be 2). The MAD may be used as it may be a robust and consistent estimator of inlier errors, which follow a Gaussian distribution.

It will therefore be understood from the above expressions that a pixel to real world distance ratio may be determined to be an inlier if the difference between its value and the median value divided by the median absolute deviation is less than a threshold value. That is to say, for a pixel to real world distance ratio to be considered an inlier, the difference

- 15 -

between its value and the median value must be less than a threshold number of times larger than the median absolute deviation.

Once final positions for a plurality of multi-directional image capture apparatuses 10 has been determined, the relative positions of the plurality of multi-directional image capture apparatuses may be determined according to:

$$\begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} = \frac{c_{dev}^j - c_{dev}^i}{s_{pixel2meter}}$$

In the above equation, $\begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix}$ represents the relative positions of one of the plurality of multi-directional image capture apparatuses (apparatus j) relative to another one of the plurality of multi-directional image capture apparatuses (apparatus i). c_{dev}^j is the position of apparatus j and c_{dev}^i is the position of apparatus i . $s_{pixel2meter}$ is the pixel to real world distance conversion factor.

As will be understood from the above expression, a vector representing the relative position of one of the plurality of multi-directional image capture apparatuses relative to another one of the plurality of multi-directional image capture apparatuses may be determined by taking the difference between their positions. This may be divided by the pixel-to-real world distance conversion factor depending on the scale desired.

As such, the positions of all of the multi-directional image capture apparatuses 10 relative to one another may be determined in the reference coordinate system 30.

The baseline distance B described above may be chosen in two different ways. One way is to set a predetermined fixed baseline distance (e.g. based on the average human interpupillary distance) to be used to generate stereo-pairs of panoramic images. This fixed baseline distance may then be used to generate all of the stereo-pairs of panoramic images.

An alternative way is to treat B as a variable within a range (e.g. a range constrained by the dimensions of the multi-directional image capture apparatus) and to evaluate a cost function for each value of B within the range. For example, this may be performed by minimising a cost function which indicates an error associated with the use of each of a

- 16 -

plurality of baseline distances, and determining that the baseline distance associated with the lowest error is to be used.

The cost function may be defined as the weighted average of the re-projection error from the structure from motion algorithm and the variance of calculated baseline distances between stereo-pairs of virtual cameras. An example of a cost function which may be used is $E(B) = w_0 \times R(B) + w_1 \times V(B)$, where $E(B)$ represents the total cost, $R(B)$ represents the re-projection error returned by the SfM algorithm by aligning the generated second images from the stereo-pairs displaced by value B , $V(B)$ represents the variance of calculated baseline distances, and w_0 and w_1 are constant weighting parameters for $R(B)$ and $V(B)$ respectively.

As such, the above process may involve generating stereo-pairs of panoramic images for each value of B , generating re-projected second images from the stereo-pairs, and inputting the second images for each value of B into a structure from motion algorithm, as described above. It will be appreciated that the re-projection error from the structure from motion algorithm may be representative of a global registration quality and the variance of calculated baseline distances may be representative of the local registration uncertainty.

It will be appreciated that, by evaluating a cost function as described above, the baseline distance with the lowest cost (and therefore lowest error) may be found, and this may be used as the baseline distance used to determine the position/orientation of the multi-directional image capture apparatus 10.

Figure 4 is a flowchart showing examples of operations as described herein.

At operation 4.1, a plurality of first images 21 which are captured by a plurality of multi-directional image capture apparatuses 10 may be received. For example, image data corresponding to the first images 21 may be received at image processing apparatus 50 (see Figure 5).

At operation 4.2, the first images 21 may be processed to generate a plurality of stereo-pairs of panoramic images 22.

At operation 4.3, the stereo-pairs of panoramic images 22 may be re-projected to generate re-projected second images 23.

- 17 -

At operation 4.4, the second images 23 from operation 4.3 may be processed to obtain positions and orientations of virtual cameras. For example, the second images 23 may be processed using a structure from motion algorithm.

- 5 At operation 4.5, a pixel-to-real world distance conversion factor may be determined based on the positions of the virtual cameras determined at operation 4.4 and a baseline distance between stereo-pairs of panoramic images 22.

At operation 4.6, positions and orientations of the plurality of multi-directional image
10 capture apparatuses 10 may be determined based on the positions and orientations of the virtual cameras 11 determined at operation 4.4.

At operation 4.7, positions of the plurality of multi-directional image capture apparatuses
10 relative to each other may be determined based on the positions of the plurality of
15 multi-directional image capture apparatuses 10 determined at operation 4.7.

It will be appreciated that, as described herein, the position of a virtual camera may be the position of the centre of a virtual lens of the virtual camera. The position of the multi-directional image capture apparatus 10 may be the centre of the multi-directional image
20 capture apparatus (e.g. if a multi-directional image capture apparatus is spherically shaped, its position may be defined as the geometric centre of the sphere).

Figure 5 is a schematic block diagram of an example configuration of image processing (or more simply, computing) apparatus 50, which may be configured to perform any of or any
25 combination of the operations described herein. The computing apparatus 50 may comprise memory 51, processing circuitry 52, an input 53, and an output 54.

The processing circuitry 52 may be of any suitable composition and may include one or more processors 52A of any suitable type or suitable combination of types. For example,
30 the processing circuitry 52 may be a programmable processor that interprets computer program instructions and processes data. The processing circuitry 52 may include plural programmable processors. Alternatively, the processing circuitry 52 may be, for example, programmable hardware with embedded firmware. The processing circuitry 52 may be termed processing means. The processing circuitry 52 may alternatively or additionally
35 include one or more Application Specific Integrated Circuits (ASICs). In some instances, processing circuitry 52 may be referred to as computing apparatus.

- 18 -

The processing circuitry 52 described with reference to Figure 5 may be coupled to the memory 51 (or one or more storage devices) and may be operable to read/write data to/from the memory. The memory 51 may store thereon computer readable instructions 512A which, when executed by the processing circuitry 52, may cause any one of or any
5 combination of the operations described herein to be performed. The memory 51 may comprise a single memory unit or a plurality of memory units upon which the computer-readable instructions (or code) 512A is stored. For example, the memory 51 may comprise both volatile memory 511 and non-volatile memory 512. For example, the computer
10 readable instructions 512A may be stored in the non-volatile memory 512 and may be executed by the processing circuitry 52 using the volatile memory 511 for temporary storage of data or data and instructions. Examples of volatile memory include RAM, DRAM, and SDRAM etc. Examples of non-volatile memory include ROM, PROM, EEPROM, flash memory, optical storage, magnetic storage, etc. The memories 51 in general may be referred to as non-transitory computer readable memory media.

15 The input 53 may be configured to receive image data representing the first images 21 described herein. The image data may be received, for instance, from the multi-directional image capture apparatuses 10 themselves or may be received from a storage device. The output 54 may be configured to output any of or any combination of the camera pose
20 registration information described herein. As discussed above, the camera pose registration information output by the computing apparatus 50 may be used for various functions as described above with reference to Figure 1.

Figure 6 illustrates an example of a computer-readable medium 60 with computer-
25 readable instructions (code) stored thereon. The computer-readable instructions (code), when executed by a processor, may cause any one of or any combination of the operations described above to be performed.

Embodiments of the present invention may be implemented in software, hardware,
30 application logic or a combination of software, hardware and application logic. The software, application logic and/or hardware may reside on memory, or any computer media. In an example embodiment, the application logic, software or an instruction set is maintained on any one of various conventional computer-readable media. In the context
of this document, a “memory” or “computer-readable medium” may be any media or
35 means that can contain, store, communicate, propagate or transport the instructions for use by or in connection with an instruction execution system, apparatus, or device, such as a computer.

Reference to, where relevant, “computer-readable storage medium”, “computer program product”, “tangibly embodied computer program” etc., or a “processor” or “processing circuitry” etc. should be understood to encompass not only computers having differing
5 architectures such as single/multi-processor architectures and sequencers/parallel architectures, but also specialised circuits such as field programmable gate arrays FPGA, application specific circuits ASIC, signal processing devices and other devices. References to computer program, instructions, code etc. should be understood to express software for a programmable processor firmware such as the programmable content of a hardware
10 device as instructions for a processor or configured or configuration settings for a fixed function device, gate array, programmable logic device, etc.

As used in this application, the term “circuitry” refers to all of the following: (a) hardware-only circuit implementations (such as implementations in only analogue and/or digital
15 circuitry) and (b) to combinations of circuits and software (and/or firmware), such as (as applicable): (i) to a combination of processor(s) or (ii) to portions of processor(s)/software (including digital signal processor(s)), software, and memory(ies) that work together to cause an apparatus, such as a server, to perform various functions) and (c) to circuits, such as a microprocessor(s) or a portion of a microprocessor(s), that require software or
20 firmware for operation, even if the software or firmware is not physically present.

If desired, the different functions discussed herein may be performed in a different order and/or concurrently with each other. Furthermore, if desired, one or more of the above-described functions may be optional or may be combined. Similarly, it will also be
25 appreciated that the flow diagram of Figure 4 is an example only and that various operations depicted therein may be omitted, reordered and/or combined. For example, it will be appreciated that operation S4.5 as illustrated in Figure 4 may be omitted.

Although various aspects of the invention are set out in the independent claims, other
30 aspects of the invention comprise other combinations of features from the described embodiments and/or the dependent claims with the features of the independent claims, and not solely the combinations explicitly set out in the claims.

It is also noted herein that while the above describes various examples, these descriptions
35 should not be viewed in a limiting sense. Rather, there are several variations and modifications which may be made without departing from the scope of the present invention as defined in the appended claims.

Claims

1. A method comprising:
processing a plurality of first images to generate a plurality of stereo-pairs of
5 panoramic images, wherein each first image is captured by a respective camera of a
respective one of a plurality of multi-directional image capture apparatuses, and wherein
each stereo-pair of panoramic images is generated from first images captured by cameras
of a respective one of the plurality of multi-directional image capture apparatuses;
performing image re-projection on each panoramic image of the plurality of stereo-
10 pairs of panoramic images, thereby to generate a plurality of re-projected second images
which are each associated with a respective virtual camera;
processing the plurality of second images to generate respective positions of the
virtual cameras associated with the second images; and
based on the generated positions of the virtual cameras, determining a position of
15 each of the plurality of multi-directional image capture apparatuses.
2. The method of claim 1, wherein the first images are fisheye images.
3. The method of claim 2, wherein processing the plurality of first images to generate
20 the plurality of stereo-pairs of panoramic images comprises:
de-warping the first images; and
stitching the de-warped images to generate the panoramic images.
4. The method of any one of the preceding claims, wherein the second images are
25 rectilinear images.
5. The method of any one of the preceding claims, wherein the processing of the
plurality of second images to generate respective positions of the virtual cameras
comprises processing the second images using a structure from motion algorithm to
30 generate the positions of the virtual cameras.
6. The method of any one of the preceding claims, wherein the panoramic images of
each stereo-pair are offset from each other by a baseline distance.
- 35 7. The method of claim 6, wherein the baseline distance to be used is a
predetermined fixed distance.

- 21 -

8. The method of claim 6, wherein the baseline distance to be used is determined by:
minimising a cost function which indicates an error associated with use of each of a
plurality of baseline distances; and
determining that the baseline distance associated with the lowest error is to be

5 used.

9. The method of claim 8 wherein the processing of the plurality of second images to
generate respective positions of the virtual cameras comprises processing the second
images using a structure from motion algorithm to generate the positions of the virtual
10 cameras and wherein the cost function is a weighted average of:
re-projection error from the structure from motion algorithm; and
variance of calculated baseline distances between stereo-pairs of virtual cameras.

10. The method of any one of claims 6 to 9, further comprising:
15 determining a pixel to real world distance conversion factor based on the
determined positions of the virtual cameras and the baseline distance used.

11. The method of any one of the preceding claims, wherein the processing of the
plurality of second images generates respective orientations of the virtual cameras, and
20 the method further comprises:

based on the generated orientations of the virtual cameras, determining an
orientation of each of the plurality of multi-directional image capture apparatuses.

12. Apparatus configured to perform a method according to any one of claims 1 to 11.

13. Computer-readable instructions which, when executed by computing apparatus,
cause the computing apparatus to perform a method according to any one of claims 1 to 11.

14. Apparatus comprising:
30 at least one processor; and
at least one memory including computer program code, which when executed
by the at least one processor, causes the apparatus to:

process a plurality of first images to generate a plurality of stereo-pairs of
panoramic images, wherein each first image is captured by a respective camera of a
35 respective one of a plurality of multi-directional image capture apparatuses, and wherein
each stereo-pair of panoramic images is generated from first images captured by cameras
of a respective one of the plurality of multi-directional image capture apparatuses;

- 22 -

perform image re-projection on each panoramic image of the plurality of stereo-pairs of panoramic images, thereby to generate a plurality of re-projected second images which are each associated with a respective virtual camera;

process the plurality of second images to generate respective positions of the virtual cameras associated with the second images; and

based on the generated positions of the virtual cameras, determine a position of each of the plurality of multi-directional image capture apparatuses.

15. The apparatus of claim 14, wherein the first images are fisheye images.

16. The apparatus of claim 15, wherein processing the plurality of first images to generate the plurality of stereo-pairs of panoramic images comprises:

de-warping the first images; and

stitching the de-warped images to generate the panoramic images.

17. The apparatus of any one of claims 14 to 16, wherein the second images are rectilinear images.

18. The apparatus of any one of claims 14 to 17, wherein the processing of the plurality of second images to generate respective positions of the virtual cameras comprises processing the second images using a structure from motion algorithm to generate the positions of the virtual cameras.

19. The apparatus of any one of claims 14 to 18, wherein the panoramic images of each stereo-pair are offset from each other by a baseline distance.

20. The apparatus of claim 19, wherein the baseline distance to be used is a predetermined fixed distance.

21. The apparatus of claim 19, wherein the baseline distance to be used is determined by:

minimising a cost function which indicates an error associated with use of each of a plurality of baseline distances; and

determining that the baseline distance associated with the lowest error is to be used.

- 23 -

22. The apparatus of claim 21, wherein the processing of the plurality of second images to generate respective positions of the virtual cameras comprises processing the second images using a structure from motion algorithm to generate the positions of the virtual cameras and wherein the cost function is a weighted average of:

5 re-projection error from the structure from motion algorithm; and
variance of calculated baseline distances between stereo-pairs of virtual cameras.

23. The apparatus of any one of claims 19 to 22, wherein the computer program code, when executed by the at least one processor, causes the apparatus to:

10 determine a pixel to real world distance conversion factor based on the determined positions of the virtual cameras and the baseline distance used.

24. The apparatus of any one of claims 14 to 23, wherein the processing of the plurality of second images generates respective orientations of the virtual cameras, and the

15 computer program code, when executed by the at least one processor, causes the apparatus to:

determine an orientation of each of the plurality of multi-directional image capture apparatuses based on the generated orientations of the virtual cameras.

20 25. A computer-readable medium having computer-readable code stored thereon, the computer readable code, when executed by at least one processor, causes performance of:

processing a plurality of first images to generate a plurality of stereo-pairs of panoramic images, wherein each first image is captured by a respective camera of a respective one of a plurality of multi-directional image capture apparatuses, and wherein
25 each stereo-pair of panoramic images is generated from first images captured by cameras of a respective one of the plurality of multi-directional image capture apparatuses;

performing image re-projection on each panoramic image of the plurality of stereo-pairs of panoramic images, thereby to generate a plurality of re-projected second images which are each associated with a respective virtual camera;

30 processing the plurality of second images to generate respective positions of the virtual cameras associated with the second images; and

based on the generated positions of the virtual cameras, determining a position of each of the plurality of multi-directional image capture apparatuses.

35 26. Apparatus comprising:

means for processing a plurality of first images to generate a plurality of stereo-pairs of panoramic images, wherein each first image is captured by a respective camera of

- 24 -

a respective one of a plurality of multi-directional image capture apparatuses, and wherein each stereo-pair of panoramic images is generated from first images captured by cameras of a respective one of the plurality of multi-directional image capture apparatuses;

5 means for performing image re-projection on each panoramic image of the plurality of stereo-pairs of panoramic images, thereby to generate a plurality of re-projected second images which are each associated with a respective virtual camera;

means for processing the plurality of second images to generate respective positions of the virtual cameras associated with the second images; and

10 means for determining a position of each of the plurality of multi-directional image capture apparatuses based on the generated positions of the virtual cameras.

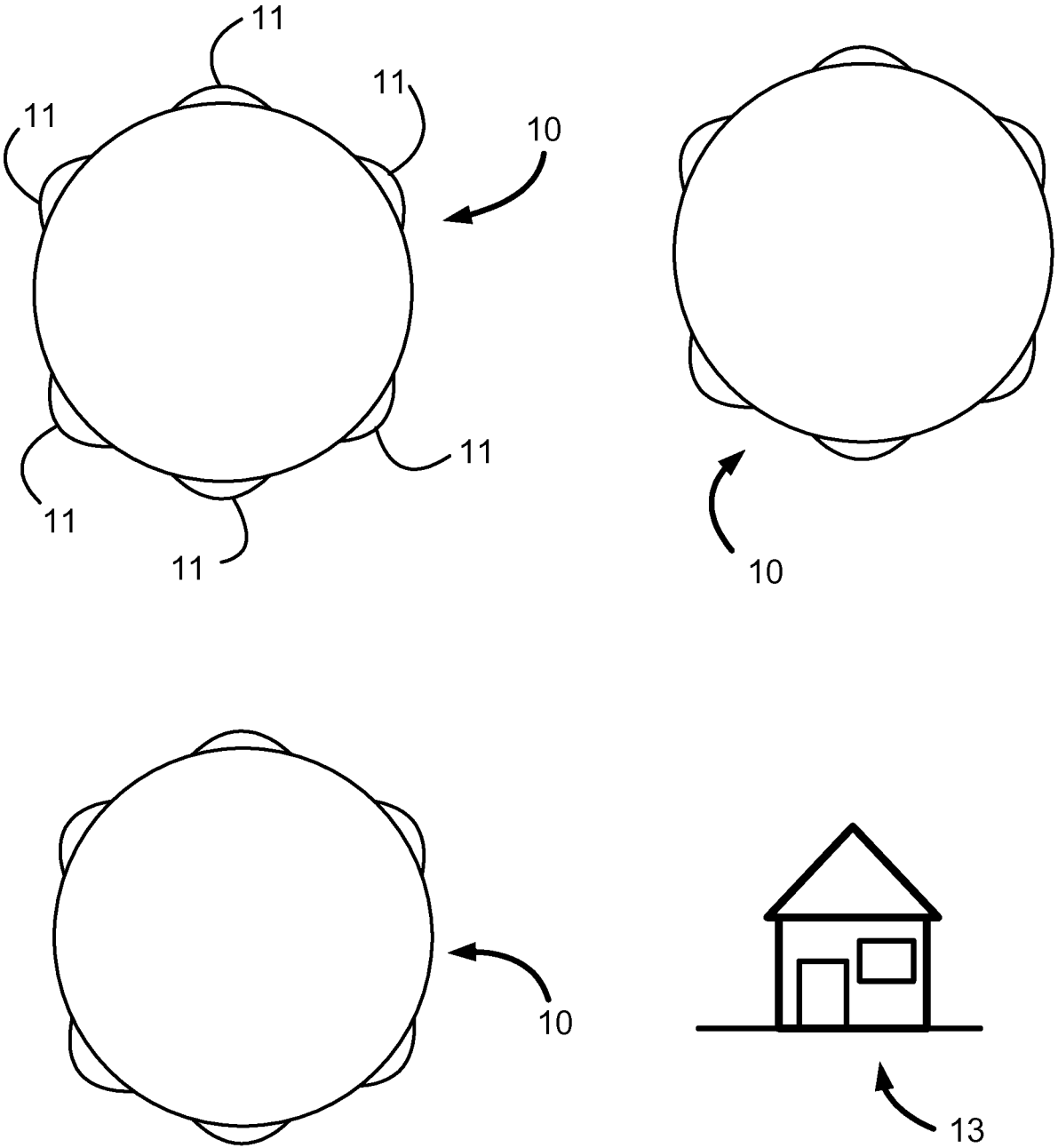


FIGURE 1

2/5

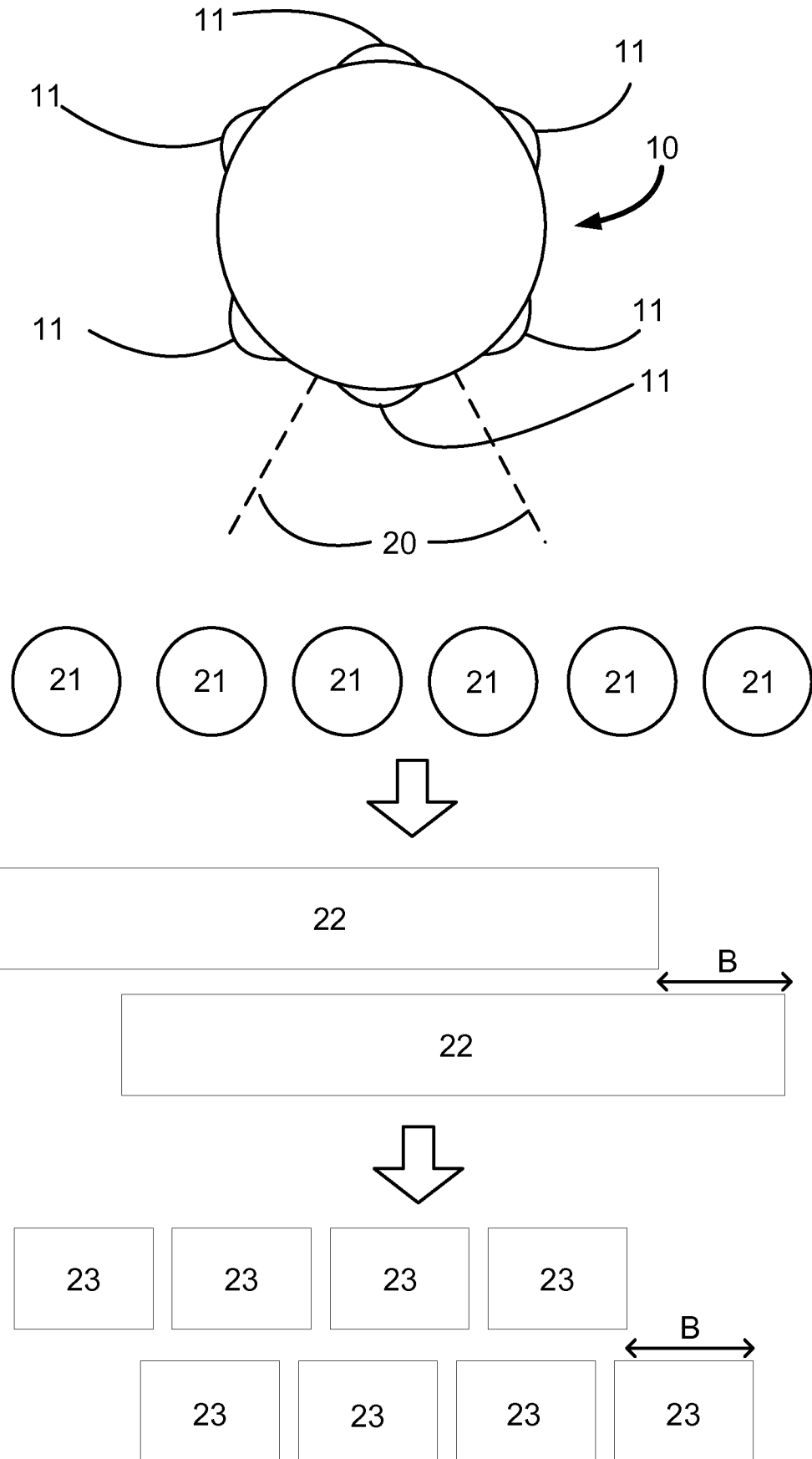


FIGURE 2A

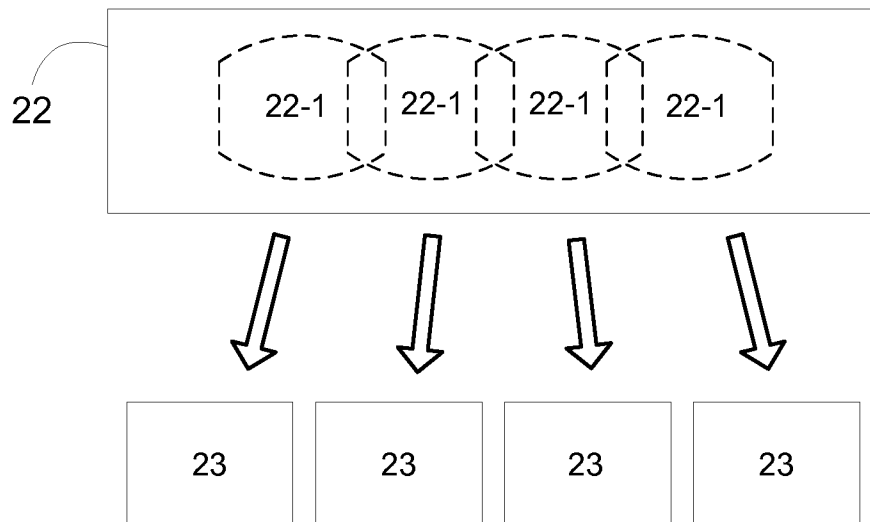


FIGURE 2B

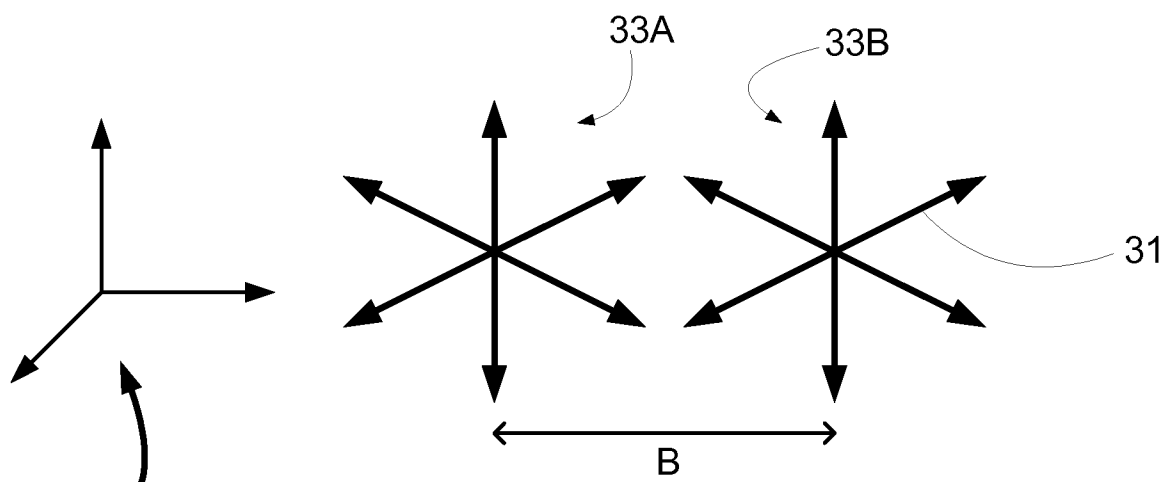


FIGURE 3A

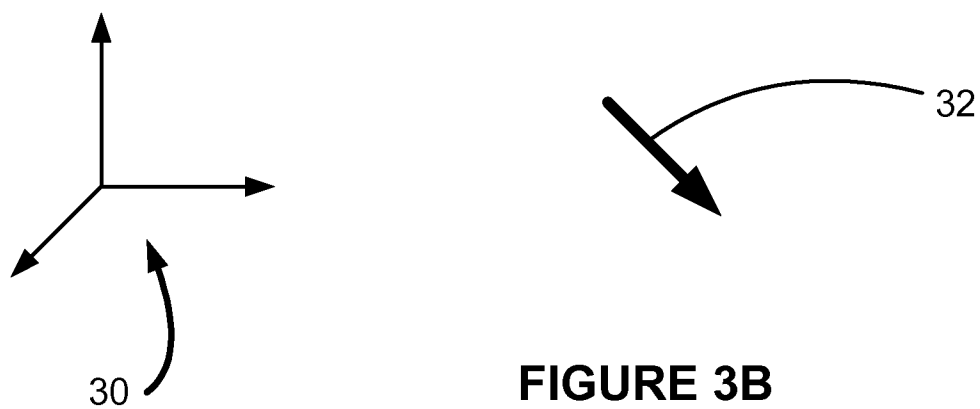
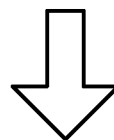
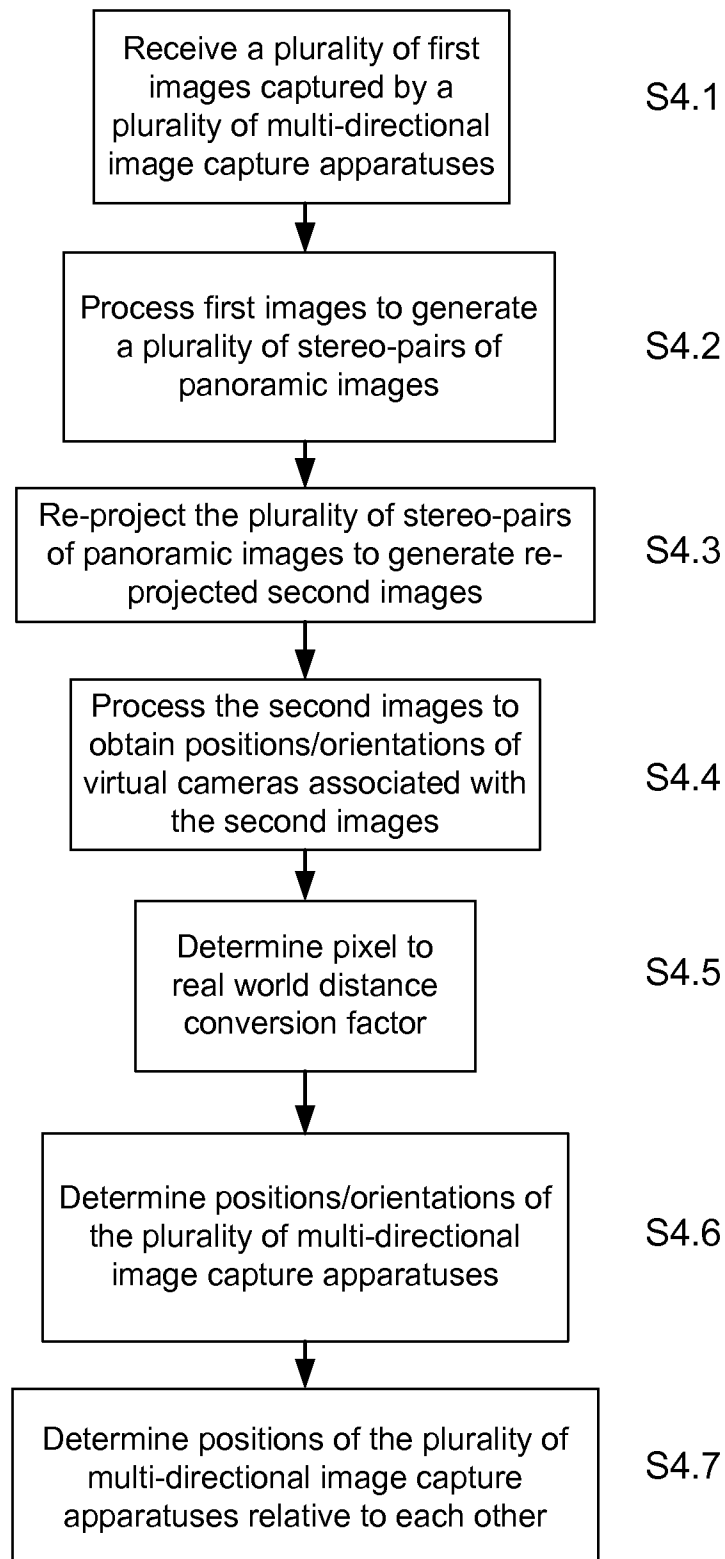
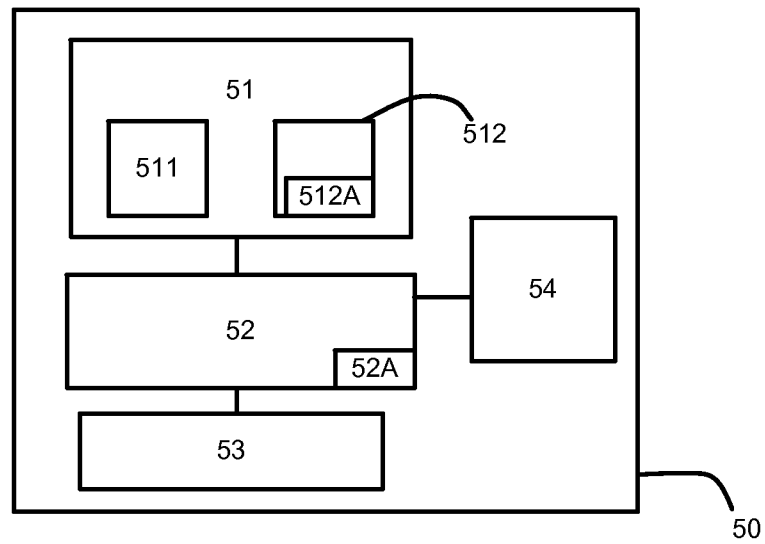
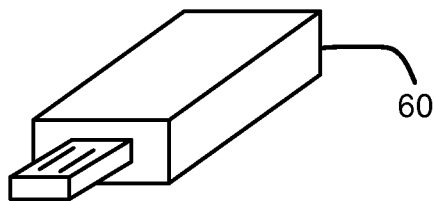


FIGURE 3B

4/5

**FIGURE 4**

**FIGURE 5****FIGURE 6**