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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 July 2017 (13.07.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/120247 A1

(51) International Patent Classification:
H04N 13/04 (2006.01)

(21) International Application Number:
PCT/US2017/012203

(22) International Filing Date:
4 January 2017 (04.01.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/274,897 5 January 2016 (05.01.2016) US

(71) Applicant: **REALD SPARK, LLC** [US/US]; Suite 200, 100 N. Crescent Drive, Beverly Hills, California 90210 (US).

(72) Inventors: **SOMMERLADE, Eric**; Suite 200, 100 North Crescent Drive, Beverly Hills, California 90210 (US). **ROBINSON, Michael G.**; Suite 200, 100 North Crescent Drive, Beverly Hills, California 90210 (US).

(74) Agents: **LOWRY, Penny L.** et al.; 5700 Flatiron Parkway, Boulder, Colorado 80301 (US).

(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, 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).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

[Continued on next page]

(54) Title: GAZE CORRECTION OF MULTI-VIEW IMAGES

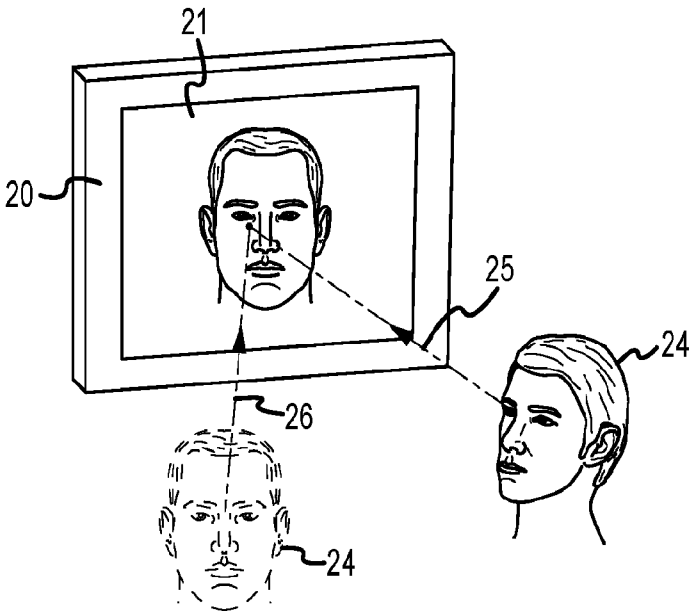


FIG.2

(57) Abstract: Gaze is corrected by adjusting multi-view images of a head. Image patches containing the left and right eyes of the head are identified and a feature vector is derived from plural local image descriptors of the image patch in at least one image of the multi-view images. A displacement vector field representing a transformation of an image patch is derived, using the derived feature vector to look up reference data comprising reference displacement vector fields associated with possible values of the feature vector produced by machine learning. The multi-view images are adjusted by transforming the image patches containing the left and right eyes of the head in accordance with the derived displacement vector field.

WO 2017/120247 A1

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

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

Gaze correction of multi-view images

TECHNICAL FIELD

[01] This application relates to the image processing of multi-view images of head, for example a stereoscopic pair of images of a head, having regard to the perceived gaze of the eyes of the head.

BACKGROUND

[02] In many systems, stereoscopic pair of images, or more generally multi-view images, of a head may be captured in one device and displayed on a different device for viewing by an observer. One non-limiting example is a system for performing teleconferencing between two telecommunications devices. In that case, each device may capture a stereoscopic pair of images, or more generally multi-view images, of a head of the observer of that device and transmit it to the other device over a telecommunications network for display and viewing by the observer of the other device.

[03] When a stereoscopic pair of images, or more generally multi-view images, of a head is captured and displayed, the gaze of the head in the displayed stereoscopic pair of images, or more generally multi-view images, may not be directed at the observer. This may be caused for example by the gaze of the head not being directed at the camera system used to capture the stereoscopic pair of images, for example because the user whose head is imaged is observing a display in the same device as the camera system and the camera system is offset above (or below) that display. In that case, the gaze in the displayed images will be perceived to be downwards (or upwards). The human visual system has evolved high sensitivity to gaze during social interaction, using cues gained from the relative position of the iris and white sclera of other observers. As such errors in the perceived gaze are disconcerting. For example in a system for performing teleconferencing, errors in the perceived gaze can create unnatural interactions between the users.

BRIEF SUMMARY

[04] The present disclosure is concerned with an image processing technique for adjusting the stereoscopic pair of images, or more generally multi-view images, of a head to correct the perceived gaze.

[05] According to a first aspect of the present disclosure, there is provided a method of adjusting multi-view images of a head to correct gaze, the method comprising: in each image of the multi-view images, identifying image patches containing the left and right eyes of the

head, respectively; in respect of the image patches containing the left eyes of the head in each image of the multi-view images, and also in respect of the image patches containing the right eyes of the head in each image of the multi-view images, performing the steps of: deriving a feature vector from plural local image descriptors of the image patch in at least one image of the multi-view images; and deriving a displacement vector field representing a transformation of an image patch, using the derived feature vector to look up reference data comprising reference displacement vector fields associated with possible values of the feature vector; and adjusting each image of the multi-view images by transforming the image patches containing the left and right eyes of the head in accordance with the derived displacement vector field.

[06] In this method, image patches containing the left and right eyes of the head are identified and transformed. To derive a displacement vector field that represents the transformation, a feature vector is derived from plural local image descriptors of the image patch in at least one image of the multi-view images and used to look up reference data comprising reference displacement vector fields associated with possible values of the feature vector. The form of the feature vector may be derived in advance from the reference data using machine learning. This method allows the gaze to be corrected, thereby reducing the disconcerting effect of incorrect gaze when the multi-view images are subsequently displayed.

[07] Various approaches to deriving and using displacement vector fields are possible as follows.

[08] In a first approach, displacement vector fields may be derived in respect of the image patches in each image of the multi-view images independently. This allows for correction of gaze, but there is a risk that the displacement vector fields in respect of each image may be inconsistent with each other, with the result that conflicting transformations are performed which can distort the stereoscopic effect and/or reduce the quality of the image.

[09] However, the following alternative approaches overcome this problem.

[10] A second possible approach is as follows. In the second approach, the plural local image descriptors used in the method are plural local image descriptors in both images of the multi-view images. In this case, the reference data comprises reference displacement vector fields for each image of the multi-view images, which reference displacement vector fields are associated with possible values of the feature vector. This allows a displacement vector field to be derived from the reference data for each image of the multi-view images. As such,

the derived displacement vector fields for each image of the multi-view images are inherently consistent.

[11] A potential downside of this second approach is that it may require the reference data to be derived from stereoscopic or more generally multi-view imagery, which may be inconvenient to derive. However, the following approaches allow the reference data to be derived from monoscopic imagery.

[12] A third possible approach is as follows. In the third approach, the plural local image descriptors are plural local image descriptors in one image of the multi-view images, and the displacement vector fields are derived as follows. A displacement vector field representing a transformation of the image patch in said one image of the multi-view images is derived, using the derived feature vector to look up reference data comprising reference displacement vector fields associated with possible values of the feature vector. Then, a displacement vector field representing a transformation of the image patch in the other multi-view image or images is derived by transforming the derived displacement vector field representing a transformation of the image patch in said one image of the multi-view images in accordance with an estimate of the optical flow between the image patches in the one image and the other multi-view image or images.

[13] Thus, in the third approach, the displacement vector fields derived in respect of each image are consistent, because only one displacement vector field is derived from the reference data, and the other displacement vector field is derived therefrom using a transformation in accordance with an estimate of the optical flow between the image patches in the images of the multi-view images.

[14] A fourth possible approach is as follows. In the fourth approach, the plural local image descriptors are plural local image descriptors in both images of the multi-view images, and the displacement vector fields are derived as follows. An initial displacement vector field representing a notional transformation of a notional image patch in a notional image having a notional camera location relative to the camera locations of the images of the multi-view images, using the derived feature vector to look up reference data comprising reference displacement vector fields associated with possible values of the feature vector. Then, displacement vector fields representing a transformation of the image patches in each image of the multi-view images are derived by transforming the initial displacement vector field in accordance with an estimate of the optical flows between the notional image patches in the notional images and the image patches in the images of the multi-view images.

[15] Thus, in the fourth approach, the displacement vector fields derived in respect of each image are consistent, because only one displacement vector field is derived from the reference data, this representing a notional transformation of a notional image patch in a notional image having a notional camera location relative to the camera locations of the images of the multi-view images. The respective displacement vector fields used to transform the two images of the multi-view images are derived therefrom using a transformation in accordance with an estimate of the optical flow between the notional image patches in the notional images and the images of the multi-view images.

[16] A fifth possible approach is as follows. In the fifth approach, displacement vector fields in respect of the image patches in each image of the multi-view images are derived, but then a merged displacement vector field is derived therefrom and used to transform the image patches containing both the left and right eyes of the head. In this case, the displacement vector fields for each image are consistent because they are the same.

[17] The merging may be performed in any suitable manner. For example, the merging may be a simple average or may be an average that is weighted by a confidence value associated with each derived displacement vector field. Such a confidence value may be derived during the machine learning.

[18] According to a second aspect of the present disclosure, there is provided an apparatus configured to perform a similar method to the first aspect of the disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[19] Non-limitative embodiments are illustrated by way of example in the accompanying figures, in which like reference numbers indicate similar parts, and in which:

[20] **FIGURE 1** is a schematic perspective view of a device that captures a stereoscopic pair of images;

[21] **FIGURE 2** is a schematic perspective view of a device that displays the stereoscopic pair of images;

[22] **FIGURE 3** is a flow chart of a method of adjusting a stereoscopic pair of images;

[23] **FIGURE 4** is a diagram illustrating the processing of the stereoscopic pair of images in the method of FIGURE 3;

[24] **FIGURE 5** is a flow chart of a step of extracting an image patch;

[25] **FIGURE 6** and **FIGURE 7** are flow charts of steps of deriving displacement vector fields according to two alternative approaches;

[26] **FIGURE 8** and **FIGURE 9** are flow charts of two alternatives for a step of

adjusting an image;

[27] **FIGURES 10** is a flow chart of a transformation step within the step of adjusting an image in the methods shown in FIGURE 8 and FIGURE 9; and

[28] **FIGURE 11** is a diagram of a telecommunications system in which the method may be implemented.

DETAILED DESCRIPTION

[29] **FIGURE 1** and **FIGURE 2** illustrate how incorrect gaze is perceived when a stereoscopic pair of images of a head is captured by the device 10 shown in FIGURE 1 which will be referred to as the source device 10 and displayed on a different device 20 shown in FIGURE 2 which will be referred to as the destination device 20.

[30] The capture device 10 includes a display 11 and a camera system 12 comprises two cameras 13 used to capture the stereoscopic pair of images of the head of a source observer 14. The source observer 14 views the display 11, along line 15. The cameras 13 of the camera system 12 are offset from the display 11, in this case being above the display 11. Thus, the cameras 13 effectively look down on the source observer 14 along line 16.

[31] The display device 20 includes a display 21 which is a stereoscopic display of any known type, for example an autostereoscopic display of any known type. The display 21 displays the stereoscopic pair of images is captured by the capture device 10. A destination observer 24 views the display 21. If the destination observer 24 is located in a normal viewing position perpendicular to the center of the display 21, as shown by the hard outline of the destination observer 24, then the gaze of the source observer 14 is perceived by the destination observer 24 to be downwards, rather than looking at the destination observer 24, because the cameras 13 of the source device 10 look down on the source observer 14.

[32] Although the cameras 13 are above the display 11 in this example, the cameras 13 could in general could be in any location adjacent the display 11, and the gaze of the source observer 14 perceived by the destination observer 24 would be correspondingly incorrect.

[33] If the destination observer 24 is located in an offset viewing position, as shown by the dotted outline of the destination observer 24 so that the destination observer 24 views the display 21 along line 26, then the offset of the destination observer 24 creates an additional error in the gaze of the source observer 14 perceived by the destination observer 24. A similar additional error in the perceived gaze of the source observer 14 occurs if the destination observer 24 is located in the normal viewing position along line 25, but the stereoscopic pair of images is displayed on the display 25 in a position offset from the center of the display 25.

[34] A stereoscopic pair of images is an example of multi-view images where there are two images. Although FIGURE 1 illustrates an example where the camera system 12 includes two cameras 13 that capture of a stereoscopic pair of images, alternatively the camera system may include more than two cameras 13 that capture more than two multi-view images, in which case similar issues of incorrect perceived gaze exist on display.

[35] **FIGURE 3** illustrates a method of adjusting multi-view images to correct such errors in the perceived gaze. For simplicity, this method will be described with respect to the adjustment of multi-view images comprising a stereoscopic pair of images. The method may be generalized to multi-view images comprising more than two images, simply by performing the similar processing on a larger number of images.

[36] The method may be performed in an image processor 30. The image processor 30 may be implemented by a processor executing a suitable computer program or by dedicated hardware or by some combination of software and hardware. Where a computer program is used, the computer program may comprise instructions in any suitable language and may be stored on a computer readable storage medium, which may be of any type, for example: a recording medium which is insertable into a drive of the computing system and which may store information magnetically, optically or opto-magnetically; a fixed recording medium of the computer system such as a hard drive; or a computer memory.

[37] The image processor 30 may be provided in the source device 10, the destination device 10 or in any other device, for example a server on a telecommunications network, which may be suitable in the case that the source device 10 and the destination device 10 communicate over such a telecommunications network.

[38] The stereoscopic pair of images 31 are captured by the camera system 12. Although the camera systems 12 is illustrated in FIGURE 1 as including two cameras 13, this is not limitative and more generally the camera system 13 may have the following properties.

[39] The camera system comprises a set of cameras 13, with at least two cameras 13. The cameras are typically spaced apart by a distance less than the average human intrapupilar distance. In the alternative that the method is applied to more than two multi-view images, then there are more than two cameras 13, that is one camera 13 image.

[40] The cameras 13 are spatially related to each other and the display 11. The spatial relationship between the cameras 13 themselves and between the cameras 13 and the display 11 is known in advance. Known methods for finding the spatial relationship may be applied, for example a calibration method using a reference image, or specification *a priori*.

[41] The cameras 13 face in the same direction as the display 11. Thus, when the source observer 14 is viewing the display 11, then the cameras 13 face the source observer 14 and the captured . stereoscopic pair of images are images of the head of the source observer 14. The cameras in the camera system can have different fields of view.

[42] The camera system 12 may include cameras 13 having different sensing modalities, including visible light and infrared.

[43] The main output of the camera system 13 is a stereoscopic pair of images 31 which are typically video images output at a video rate. The output of the camera system 13 may also include data representing the spatial relationship between the cameras 13 and the display 11, the nature of the sensing modalities and internal parameters of the cameras 13 (for example focal length, optical axis) which may be used for angular localization.

[44] The method performed on the stereoscopic pair of images 31 is as follows. To illustrate the method, reference is also made to **FIGURE 4** which shows an example of the stereoscopic pair of images 31 at various stages of the method.

[45] In step S1, the stereoscopic pair of images 31 are analyzed to detect the location of the head and in particular the eyes of the source observer 14 within the stereoscopic pair of images 31. This is performed by detecting presence of a head, tracking the head, and localizing the eyes of the head. Step S1 may be performed using a variety of techniques that are known in the art.

[46] One possible technique for detecting the presence of the head is to use Haar feature cascades, for example as disclosed in Viola and Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", CVPR 2001, pp 1-9 (incorporated herein by reference).

[47] One possible technique for tracking the head is to use the approach of Active Appearance Models to provide the position of the head of the subject, as well as the location of the eyes, for example as disclosed in Cootes et al., "Active shape models - their training and application", Computer Vision and Image Understanding, 61(1):38-59, Jan. 1995 and in Cootes et al. "Active appearance models", IEEE Trans. Pattern Analysis and Machine Intelligence, 23(6):681-685, 2001 (incorporated herein by reference).

[48] In step S1, typically, a set of individual points ("landmarks") are set to regions of the face, typically the eyes, for example corners of the eye, upper and lower lid locations, etc, thereby localizing the eyes.

[49] In step S2, image patches containing the left and right eyes of the head, respectively

are identified in each image 31 of the stereoscopic pair. FIGURE 4 shows the identified image patches 32 of the right eye in each image 31 (the image patches for the left eye being omitted in FIGURE 4 for clarity).

[50] Step S2 may be performed as shown in **FIGURE 5**, as follows.

[51] In step S2-1, image patches 32 containing the left and right eyes of the head are identified in each image 31 of the stereoscopic pair. This is done by identifying an image patch 39in each image 31 located around the identified points ("landmarks") corresponding to features of an eye, as shown for example in FIGURE 4.

[52] In step S2-2, the image patches 32 identified in step S2-1 are transformed into a normalized coordinate system, being the same normalized coordinate system as used in the machine learning process which is described further below. The transformation is chosen to align the points ("landmarks") of the eye within the image patch that were identified in step S1, with predetermined locations in the normalized coordinate system. The transformation may include translation, rotation and scaling, to appropriate extents to achieve that alignment. The output of step S2-2 is identified image patches 33 of the right eye in each image in the normalized coordinate system as shown for example in FIGURE 4.

[53] The following steps are performed separately (a) in respect of the image patches containing the left eyes of the head in each image 31 of the stereoscopic pair, and (b) in respect of the image patches containing the right eyes of the head in each image 31 of the stereoscopic pair. For brevity, the following description will refer merely to image patches and eyes without specifying the left or right eye, but noting the same steps are performed for both left and right eyes.

[54] In step S3, a feature vector 34 is derived from plural local image descriptors of an image patch 33 in at least one image 31 of the stereoscopic pair. Depending on the approach and as described further below, this may be an image patch in a single image 31 of the stereoscopic pair or may be both images 31 of the stereoscopic pair. Thus, the local image descriptors are local image descriptors derived in the normalized coordinate system.

[55] The feature vectors 34 are representations of the image patches 33 that are suitable for use in looking up reference data 35 comprising reference displacement vector fields that represent transformations of the image patch and are associated with possible values of the feature vector.

[56] The reference data 35 is obtained and analyzed in advance using a machine learning technique which derives the form of the feature vectors 34 and associates the reference

displacement vector fields with the possible values of the feature vector. Accordingly, the machine learning technique will now be described before reverting to the method of FIGURE 3.

[57] The training input to the machine learning technique is two sets of images, which may be stereoscopic pairs of images or monoscopic images, as discussed further below. Each set comprises images of the head of the same group of individuals but captured from cameras in different locations relative to the gaze so that the perceived gaze differs as between them.

[58] The first set are input images, being images of each individual with an incorrect gaze where the error is known *a priori*. In particular, the images in the first set may be captured by at least one cameras in a known camera location where the gaze of the individual which is in a different known direction. For example in the case of the source device of FIGURE 1, the camera location may be the location of a camera 13 and while the gaze of the imaged individual is towards the center of the display 11.

[59] The second set are output images, being images of each individual with correct gaze for a predetermined observer location relative to a display location in which the image is to be displayed. In the simplest case, the observer location is a normal viewing position perpendicular to the center of the display location, for example as shown by the hard outline of the destination observer 24 in the case of the destination device 20 of FIGURE 2.

[60] For each image in the two sets, the image is analyzed to detect the location of the head and in particular the eyes using the same technique as used in step S1 described above, and then image patches containing the left and right eyes of the head, respectively, are identified using the same technique as used in step S2 described above. The following steps are then performed separately (a) in respect of the image patches containing the left eyes of the head in each image, and (b) in respect of the image patches containing the right eyes of the head in each image. For brevity, the following description will refer merely to image patches and eyes without specifying the left or right eye, but noting the same steps are performed for both left and right eyes.

[61] Each image patch is transformed into the same normalized coordinate system as used in step S2 described above. As described above, the transformation is chosen to align points ("landmarks") of the eye with predetermined locations in the normalized coordinate system. The transformation may include translation, rotation and scaling, to appropriate extents to achieve that alignment.

[62] Thus, the image patches input and output images of each individual are aligned in the

normalized coordinate system.

[63] From an input and output image of each individual, there is derived a displacement vector field that represents the transformation of the image patch in the input image required to obtain the image patch of the output image, for example as follows. Defining positions in the image patches by (x,y), the displacement vector field **F** is given by

$$\mathbf{F}=\{u(x, y), v(x, y)\}$$

where u and v define the horizontal and vertical components of the vector at each position (x,y).

[64] The displacement vector field **F** is chosen so that the image patch of the output image O(x,y) is derived from the image patch of the input image I(x,y) as

$$O(x, y) = I(x+u(x, y), y+v(x,y))$$

[65] For image data from more than one camera, the system delivers a displacement vector field for the input image from each camera.

[66] The displacement vector field **F** for an input and output image of an individual may be derived using a process in which a trial feature vector $\mathbf{F}'=\{u',v'\}$ is modified to minimize error, optionally in an iterative process, for example in accordance with:

$$\sum\left|O(x,y) - I\left(x + u'(x,y), y + v'(x,y)\right)\right| = min!$$

[67] By way of non-limitative example, the displacement vector field **F** may be derived as disclosed in Kononenko et al., "Learning To Look Up: Realtime Monocular Gaze Correction Using Machine Learning", Computer Vision and Pattern Recognition, 2015, pp. 4667-4675 (incorporated herein by reference), wherein the displacement vector field **F** is referred to as a "flow field".

[68] A machine learning technique is used to obtain a map from the displacement vector field **F** of each individual to respective feature vectors derived from plural local image descriptors of the image patch of the input image.

[69] The local descriptors capture relevant information of a local part of the image patch of the input image and the set of descriptors usually form a continuous vectorial output.

[70] The local image descriptors input into the machine learning process are of types expected to provide discrimination between different individuals, although the specific local image descriptors are selected and optimized by the machine learning process itself. In general, the local image descriptors may be of any suitable type, some non-limitative examples which may be applied in any combination being as follows.

[71] The local image descriptors may include values of individual pixels or a linear

combination thereof. Such a linear combination may be, for example, a difference between the pixels at two points, a kernel derived within a mask at an arbitrary location, or a difference between two kernels at different locations.

[72] The local image descriptors may include distances of a pixel location from the position of an eye point ("landmark").

[73] The local image descriptors may include SIFT features (Scale-invariant feature transform features), for example as disclosed in Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision 60 (2), pp 91–110 (incorporated herein by reference).

[74] The local image descriptors may include HOG features (Histogram of Oriented Gradients features) , for example as disclosed in Dalal et al. "Histograms of Oriented Gradients for Human Detection", Computer Vision and Pattern Recognition, 2005, pp. 886-893 (incorporated herein by reference).

[75] The derivation of the feature vector from plural local image descriptors depends on the type of machine learning applied.

[76] In a first type of machine learning technique, the feature vector may comprise features that are values derived from the local image descriptors in a discrete space, being binary values or values discretized into more than two possible values. In this case, the machine learning technique associates a reference displacement vector field **F** derived from the training input with each possible value of the feature vector in the discrete space, so the reference data 35 is essentially a look-up table. This allows a reference displacement vector field **F** to be simply selected from the reference data 35 on the basis of the feature vector 34 derived in step S3, as described below.

[77] In the case that the feature vector comprises features that are binary values derived from the local image descriptors, the feature vector has a binary representation. Such binary values may be derived in various ways from the values of descriptors, for example by comparing the value of a descriptor with a threshold, comparing the value of two descriptors, or by comparing the distance of a pixel location from the position of an eye point ("landmark").

[78] Alternatively, the feature vector may comprise features that are discretized values of the local image descriptors. In this case, more than two discrete values of each feature are possible.

[79] Any suitable machine learning technique may be applied, for example using a

decision tree, a decision forest, a decision fern or an ensemble or combination thereof.

[80] By way of example, a suitable machine learning technique using a feature vector comprising features that are binary values derived by comparing a set of individual pixels or a linear combination thereof against a threshold, is disclosed in Ozuysal et al. "Fast Keypoint Recognition in Ten Lines of Code", Computer Vision and Pattern Recognition, 2007, pp. 1-8 (incorporated herein by reference).

[81] By way of further example, a suitable machine learning technique using a distance of a pixel location with the position of an eye landmark is disclosed in Kononenko et al., "Learning To Look Up: Realtime Monocular Gaze Correction Using Machine Learning", Computer Vision and Pattern Recognition, 2015, pp. 4667-4675 (incorporated herein by reference).

[82] By way of further example, a suitable machine learning technique using a random decision forest is disclosed in Ho, "Random Decision Forests", Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995, pp. 278-282 (incorporated herein by reference).

[83] In a second type of machine learning technique, the feature vector may comprise features that are discrete values of the local image descriptors in a continuous space. In this case, the machine learning technique associates a reference displacement vector field **F** derived from the training input with possible discrete values of the feature vector in the continuous space. This allows a displacement vector field **F** to be derived from the reference data 35 by interpolation from the reference displacement vector fields based on the relationship between the feature vector 34 derived in step S3 and the values of the feature vector associated with the reference displacement vector fields.

[84] Any suitable machine learning technique may be applied, for example using support vector regression.

[85] By way of example, a suitable machine learning technique using support vector regression is disclosed in Drucker et al. "Support Vector Regression Machines", Advances in Neural Information Processing Systems 9, NIPS 1996, 155–161, (incorporated herein by reference). The output of the technique is a continuously varying set of interpolation directions that form part of the reference data 35 and are used in the interpolation.

[86] The machine learning technique, regardless of its type, inherently also derives the form of the feature vectors 34 that is used to derive the reference displacement vector fields **F**. This is the form of the feature vectors 34 that is derived in step S3.

[87] Optionally, the output of the machine learning technique may be augmented to provide confidence values associated with derivation of a displacement vector field from the reference data 35.

[88] In the case that the feature vector comprises features that are values in a discrete space, a confidence value is derived for each reference displacement vector field.

[89] One example of deriving a confidence value is to keep, for each resulting index (value of the feature vector) in the resulting look-up table, a distribution of the corresponding part of the input image in the training data. In this case, the confidence value may be the amount of training data that ended up with the same index, divided by the total number of training data exemplars.

[90] Another example of deriving a confidence value is to fit a Gaussian to the distribution of input images in the training data in each bin indexed, and to use the trace of the covariance matrix around the mean value as the confidence value.

[91] In the case that the feature vector comprises features that are discrete values of the local image descriptors in a continuous space, the confidence values may be derived according to the machine learning method used. For example, when using support vector regression, the confidence values may be the inverse of the maximum distance to the support vectors.

[92] Where used, the confidence values are stored as part of the reference data.

[93] The description now reverts to the method of FIGURE 3.

[94] In step S4, at least one displacement vector field 37 representing a transformation of an image patch is derived by using the feature vector 34 derived in step S3 is to look up the reference data 35. Due to the derivation of the displacement vector field 37 from the reference data 35, the transformation represented thereby corrects the gaze that will be perceived when the stereoscopic pair of images 31 are displayed.

[95] In the case that the feature vector 34 comprises features that are values in a discrete space and the reference displacement vector fields of the reference data 35 comprise a reference displacement vector field associated with each possible value of the feature vector in the discrete space, then the displacement vector field for the image patch is derived by selecting the reference displacement field associated with the actual value of the derived feature vector 34.

[96] In the case that the feature vector 34 comprises features that are discrete values of the local image descriptors in a continuous space, then then the displacement vector field for the

image patch is derived by interpolating a displacement vector field from the reference displacement vector fields based on the relationship between the actual value of the derived feature vector 34 and the values of the feature vectors associated with the reference displacement vector fields. In the case that the machine learning technique was support vector regression, This may be done using the interpolation directions that form part of the reference data 35.

[97] Some different approaches to the derivation of the displacement vector field 37 in step S4 will now be described.

[98] In a first approach, in step S4, a displacement vector field 37 is derived in respect of the image patches in each image 31 of the stereoscopic pair independently. This first approach may be applied when the reference data 35 was derived from monoscopic images. This approach provides for correction of gaze, but there is a risk that the displacement vector fields 37 in respect of each image may be inconsistent with each other, with the result that conflicting transformations are subsequently performed which can distort the stereoscopic effect and/or reduce the quality of the image.

[99] Other approaches which overcome this problem are as follows.

[100] In a second possible approach, the plural local image descriptors used in deriving the feature vector 34 in step S3 are plural local image descriptors in both images of the stereoscopic pair. In this case, the reference data 35 similarly comprises pairs of reference displacement vector fields for each image 31 of the stereoscopic image pair, it being the pairs of reference displacement vector fields that are associated with possible values of the feature vector 34.

[101] This second approach allows a pair of displacement vector fields 35 to be derived from the reference data 35, that is one displacement vector field for each image 31 of the stereoscopic pair. As such, the derived displacement vector fields for each image 31 of the stereoscopic pair are inherently consistent since they are derived together from the consistent pairs of reference displacement vector fields in the reference data 35.

[102] The downside of this second approach is that it requires the reference data 35 to be derived from training input to the machine learning technique that is stereoscopic pairs of images. This does not create any technical difficulty, but may create some practical inconvenience as monoscopic imagery is more commonly available. Accordingly, the following approaches may be applied when the reference data 35 is derived from training input to the machine learning technique that is monoscopic images.

[103] In a third possible approach, a feature vector 34 is derived from plural local image descriptors that are plural local image descriptors derived from one image of the stereoscopic pair. In that case, the displacement vector fields 37 are derived as shown in **FIGURE 6** as follows.

[104] In step S4-1, a first displacement vector field 37 representing a transformation of the image patch in said one image 31 of the stereoscopic pair (which may be either image 31) is derived. This is done using the derived feature vector 34 to look up the reference data 35.

[105] In step S4-2, a displacement vector field 37 representing a transformation of the image patch in the other image 31 of the stereoscopic pair is derived. This is done by transforming the displacement vector field derived in step S4-1 in accordance with an estimate of the optical flow between the image patches in the images 31 of the stereoscopic pair.

[106] The optical flow represents the effect of the different camera locations as between the images 31 of the stereoscopic pair. Such an optical flow is known in itself and may be estimated using known techniques for example as disclosed in Zach et al., "A Duality Based Approach for Realtime TV-L1 Optical Flow", Pattern Recognition (Proc. DAGM), 2007, pp. 214-223 (incorporated herein by reference).

[107] By way of example, if the first displacement vector field 37 derived in step S4-1 is for the left image L_o, L_i (where the subscripts o and i represent the output and input images respectively), and the optical flow to the right image R_o is represented by a displacement vector field **G** given by

$$\mathbf{G}=\{s(x, y), t(x, y)\}$$

then the second displacement vector field 37 may be derived in accordance with

$$R_o(x,y) = L_o(x + s(x,y), y + t(x,y)) = L_i(x + s + u(x + s, y + t, y + t + v(x + s, y + t)$$

[108] Thus, in the third approach, the displacement vector fields 37 derived in respect of each image 31 of the stereoscopic pair are consistent, because only one displacement vector field is derived from the reference data 35, and the other displacement vector field is derived therefrom using a transformation which maintains consistency because it is derived in accordance with an estimate of the optical flow between the image patches in the images 31 of the stereoscopic pair.

[109] In a fourth possible approach, a feature vector 34 is derived from plural local image descriptors that are plural local image descriptors derived from both images of the

stereoscopic pair. In that case, the displacement vector fields 37 are derived as shown in **FIGURE 7** as follows.

[110] In step S4-3, an initial displacement vector field representing a notional transformation of a notional image patch in a notional image having a notional camera location in a predetermined location relative to the camera locations of the images 31, in this example between the camera locations of the images 31. This may be thought of as a Cyclopean eye. This is done using the derived feature vector 34 to look up the reference data 35 which comprises reference displacement vector fields associated with possible values of the feature vector. This means that the reference data 35 is correspondingly structured, but may still be derived from training input that comprises monoscopic images.

[111] In step S4-4, a displacement vector fields 37 representing transformations of the image patches in each image 31 of the stereoscopic pair are derived. This is done by transforming the initial displacement vector field derived in step S4-3 in accordance with an estimate of the optical flow between the notional image patches in the notional images and the image patches in the images 31 of the stereoscopic pair.

[112] The optical flow represents the effect of the different camera locations as between the notional image and the images 31 of the stereoscopic pair. Such an optical flow is known in itself and may be estimated using known techniques for example as disclosed in Zach et al., "A Duality Based Approach for Realtime TV-L1 Optical Flow", Pattern Recognition (Proc. DAGM), 2007, pp. 214-223 (as cited above and incorporated herein by reference).

[113] By way of example, if the optical flow from the left image L to the right image R is represented by a displacement vector field **G** given by

$$\mathbf{G}=\{s(x, y), t(x, y)\}$$

then the transformation deriving the notional image C is given by

$$C(x,y) = R\left(x - \frac{s(x,y)}{2}, y - \frac{t(s,y)}{2}\right) = L(x + \frac{s(x,y)}{2}, y + \frac{t(x,y)}{2})$$

[114] Thus, in this example, the initial displacement vector field F derived in step S4-3 for this notional image C is transformed in step S4-4 to derive the flow fields F_{rc} and F_{lc} for the right and left images in accordance with

$$F_{rc}(x,y) = \left\{u(x,y) + \frac{s(x,y)}{2}, v(x,y) + \frac{t(x,y)}{2}\right\}$$

$$F_{lc}(x,y) = \left\{u(x,y) - \frac{s(x,y)}{2}, v(x,y) - \frac{t(x,y)}{2}\right\}$$

[115] Thus, in the fourth approach, the displacement vector fields 37 derived in respect of each image 31 of the stereoscopic pair are consistent, because only one displacement vector field is derived from the reference data 35, this representing a notional transformation of a notional image patch in a notional image, and the displacement vector fields for the left and right images are derived therefrom using a transformation which maintains consistency because it is derived in accordance with an estimate of the optical flow between the image patches in the notional image and in the images 31 of the stereoscopic pair.

[116] In step S5, each image 31 of the stereoscopic pair is adjusted by transforming the image patches containing the left and right eyes of the head in accordance with the derived displacement vector fields 37. This produces an adjusted stereoscopic pair of images 38 as shown in FIGURE 4, in which the gaze has been corrected. In particular, the adjustment may be performed using two alternative methods, as follows.

[117] A first method for performing step S5 is shown in **FIGURE 8** and performed as follows.

[118] In step S5-1, the image patch is transformed in the normalised coordinate system in accordance with the corresponding displacement vector field 37 in respect of the same image, thereby correcting the gaze. As described above, for a displacement vector field **F** the transformation of the image patch of the input image I(x,y) provides the output image O(x,y) in accordance with

$$O(x, y) = I(x+u(x, y), y+v(x,y))$$

[119] In step S5-2, the transformed image patch output from step S5-1 is transformed out of the normalised coordinate system, back into the original coordinate system of the corresponding image 31. This is done using the inverse transformation from that applied in step S2-2.

[120] In step S5-3, the transformed image patch output from step S5-2 is superimposed on the corresponding image 31. This may be done with a full replacement within an eye region corresponding to the eye itself, and a smoothed transition between the transformed image patch and the original image 31 over a boundary region around the eye region. The width of the boundary region may be of fixed size or a percentage of the size of the image patch in the original image 31.

[121] A second method for performing step S5 is shown in **FIGURE 9** and performed as follows.

[122] In this second, alternative method, the transformation back into the coordinate system of the corresponding image 31 occurs before the transformation of the image patch in accordance with the transformed displacement vector field **F**.

[123] In step S5-4, the displacement vector field **F** is transformed out of the normalised coordinate system, back into the original coordinate system of the corresponding image 31. This is done using the inverse transformation from that applied in step S2-2.

[124] In step S5-5, the image patch 32 in the coordinate system of the image 31 is transformed in accordance with the displacement vector field **F** that has been transformed into the same coordinate system in step S5-4. As described above, for a displacement vector field **F** the transformation of the image patch of the input image I(x,y) provides the output image O(x,y) in accordance with

$$O(x, y) = I(x+u(x, y), y+v(x,y))$$

but this is now performed in the coordinate system of the original image 31.

[125] Step S5-6 is the same as S5-3. Thus, in step S5-6, the transformed image patch output from step S5-5 is superimposed on the corresponding image 31. This may be done with a full replacement within an eye region corresponding to the eye itself, and a smoothed transition between the transformed image patch and the original image 31 over a boundary region around the eye region. The width of the boundary region may be of fixed size or a percentage of the size of the image patch in the original image 31.

[126] The displacement vector fields 37 used in step S5 will now be discussed.

[127] One option is that the displacement vector fields 37 derived in step S4 in respect of the left and right images are used directly in step S5. That is, the image patch in respect of each image 31 of the stereoscopic patch is transformed in accordance with the displacement vector field 37 in respect of that image 31. This is appropriate if the displacement vector fields 37 are sufficiently accurate, for example because they have been derived from reference data 35 that is itself derived from stereoscopic imagery in accordance with the second approach described above.

[128] An alternative option in accordance with a fifth approach is that a merged displacement vector field 39 is derived and used. This may be applied in combination with any of the first to fourth approaches discussed above. In this case, step S5 additionally includes step S5-a as shown in **FIGURE 10** which is performed before step S5-1 in the first method of FIGURE 8 or before step S5-4 in the second method of FIGURE 9. In step S5-a, a

merged displacement vector field 39 is derived from the displacement vector fields 37 derived in step S4 in respect of the image patches in each image 31 of the stereoscopic pair.

[129] The rest of step S5 is then performed using the merged displacement vector field 39 in respect of each image 31. That is, in the first method of FIGURE 8, the image patch 33 in respect of each image 31 of the stereoscopic pair is transformed in step S5-1 of in accordance with the merged displacement vector field 39. Similarly, in the second method of FIGURE 9, in step S5-4 the merged displacement vector field 39 is transformed and in step S5-5 the image patch 33 in respect of each image 31 of the stereoscopic pair is transformed in accordance with that merged displacement vector field 39.

[130] In this case, the displacement vector fields for each image are consistent because they are the same.

[131] The merging in step S5-1a may be performed in any suitable manner.

[132] In one example, the merging in step S5-1a may be a simple average of the displacement vector fields 37 derived in step S4

[133] In another example, the merging in step S5-1a may be an average that is weighted by a confidence value associated with each derived displacement vector field 37. In this case, confidence values form part of the reference data 35, in the manner described above, and in step S4 the confidence values are derived from the reference data 35, together with the derived displacement vector field 37.

[134] By way of example, denoting the derived displacement vector field 37 as F_i, the merged displacement vector field 39 as F_{avg}, and the confidence values as a_i, then the merged displacement vector field 39 may be derived as

$$F_{avg} = \frac{\sum a_i F_i}{\sum a_i}$$

[135] In the example described above, gaze is corrected for a destination observer 24 in an observer location that is a normal viewing position perpendicular to the center of the display location, for example as shown by the hard outline of the destination observer 24 in the case of the destination device 20 of FIGURE 2. This is sufficient in many situations. However, there will now be described an optional modification which allows gaze is corrected for a destination observer 24 in a different observer location, for example as shown by the dotted outline of the destination observer 24 in the case of the destination device 20 of FIGURE 2.

[136] In this case, the method further comprises using location data 40 representing the

observer location relative to a display location of the stereoscopic pair of images 31. This location data 40 may be derived in the destination device 20, for example as described below. In that case, if the method is not performed in the destination device 20, then the location data 40 is transmitted to the device in which the method is performed.

[137] The relative observer location may take into account the location of the observer with respect to the display 21. This may be determined using a camera system in the destination device 20 and an appropriate head tracking module to detect the location of the destination observer 24.

[138] The relative observer location may assume that the image is displayed centrally on the display 21. Alternatively, relative observer location may take into account both the location of the observer with respect to the display 21 and the location of the image displayed on the display 21. In this case, the location of the image displayed on the display 21 may be derived from the display geometry (for example, the position and area of a display window and the size of the display 21).

[139] To account for different observer locations, the reference data 34 comprises plural sets of reference displacement vector fields, each set being associated with different observer locations. This is achieved by the training input to the machine learning technique comprising plural second sets of output images, each second set being images of each individual with correct gaze for a respective, predetermined observer location relative to a display location in which the image is to be displayed. Thus, in step S4, the displacement vector fields 37 are derived by looking up the set of reference displacement vector fields associated with the observer location represented by the location data.

[140] As described above, the method may be implemented in an image processor 30 provided in various different devices. By way of non-limitative example, there will now be described an particular implementation in a telecommunications system which is shown in **FIGURE 11** and arranged as follows.

[141] In this implementation, the source device 10 and the destination device 10 communicate over such a telecommunications network 50. For communication over the telecommunications network 50, the source device 10 includes a telecommunications interface 17 and the destination device 20 includes a telecommunications interface 27.

[142] In this implementation, the image processor 30 is provided in the source device 10and is provided with the stereoscopic pair of images directly from the camera system 12. The telecommunications interface 17 is arranged to transmit the adjusted stereoscopic pair of

images 38 over the telecommunications network 50 to the destination device 20 for display thereon.

[143] The destination device 20 includes an image display module 28 that controls the display 26. The adjusted stereoscopic pair of images 38 are received in the destination device 20 by the telecommunications interface 27 and supplied to the image display module 28 which causes them to be displayed on the display 26.

[144] The following elements of the destination device 20 are optionally included in the case that the method corrects gaze for a destination observer 24 in an observer location other than a normal viewing position perpendicular to the center of the display location. In this case, the destination device 20 includes a camera system 23 and an observer location module 29. The camera system 23 captures an image of the destination observer 24. The observer location module 29 derives the location data 40. The observer location module 29 includes a head tracking module that uses the output of the camera system 23 to detect the location of the destination observer 24. The observer location module 29. Where the relative observer location also takes into account the location of the image displayed on the display 21, the observer location module 29 obtains the location of the image displayed on the display 21 from the image display module 28. The telecommunications interface 17 is arranged to transmit the location data 40 over the telecommunications network 50 to the source device 10 for use thereby.

[145] Although the above description refers to a method applied to images supplied from a source device 10 to a destination device 20, the method may equally be applied to images supplied from in the opposite direction from the destination device 20 to the source device 10, in which case the destination device 20 effectively becomes the "source device" and the source device 10 effectively becomes the "destination device". Where images are supplied bi-directionally, the labels "source" and "destination" may be applied to both devices, depending on the direction of communication being considered.

[146] While various embodiments in accordance with the principles disclosed herein have been described above, it should be understood that they have been presented by way of example only, and not limitation. Thus, the breadth and scope of this disclosure should not be limited by any of the above-described exemplary embodiments, but should be defined only in accordance with any claims and their equivalents issuing from this disclosure. Furthermore, the above advantages and features are provided in described embodiments, but

shall not limit the application of such issued claims to processes and structures accomplishing any or all of the above advantages.

[147] Additionally, the section headings herein are provided for consistency with the suggestions under 37 CFR 1.77 or otherwise to provide organizational cues. These headings shall not limit or characterize the embodiment(s) set out in any claims that may issue from this disclosure. Specifically and by way of example, although the headings refer to a "Technical Field," the claims should not be limited by the language chosen under this heading to describe the so-called field. Further, a description of a technology in the "Background" is not to be construed as an admission that certain technology is prior art to any embodiment(s) in this disclosure. Neither is the "Summary" to be considered as a characterization of the embodiment(s) set forth in issued claims. Furthermore, any reference in this disclosure to "invention" in the singular should not be used to argue that there is only a single point of novelty in this disclosure. Multiple embodiments may be set forth according to the limitations of the multiple claims issuing from this disclosure, and such claims accordingly define the embodiment(s), and their equivalents, that are protected thereby. In all instances, the scope of such claims shall be considered on their own merits in light of this disclosure, but should not be constrained by the headings set forth herein.

Claims:

1. A method of adjusting multi-view images of a head to correct gaze, the method comprising

in each image, identifying image patches containing the left and right eyes of the head, respectively;

in respect of the image patches containing the left eyes of the head in each image of the multi-view images, and also in respect of the image patches containing the right eyes of the head in each image of the multi-view images, performing the steps of:

deriving a feature vector from plural local image descriptors of the image patch in at least one image of the multi-view images; and

deriving a displacement vector field representing a transformation of an image patch, using the derived feature vector to look up reference data comprising reference displacement vector fields associated with possible values of the feature vector; and

adjusting each image of the multi-view images by transforming the image patches containing the left and right eyes of the head in accordance with the derived displacement vector field.

2. A method according to claim 1, wherein

said plural local image descriptors are plural local image descriptors in each images of the multi-view images, and

said reference data comprising pairs of reference displacement vector fields for each image of the multi-view images, which pairs of reference displacement vector fields are associated with possible values of the feature vector.

3. A method according to claim 1, wherein

said plural local image descriptors are plural local image descriptors in one image of the multi-view images,

said step of deriving displacement vector fields comprises:

deriving a displacement vector field representing a transformation of the image patch in said one image of the multi-view images, using the derived feature vector to look up reference data comprising reference displacement vector fields associated with possible values of the feature vector; and

deriving a displacement vector field representing a transformation of the image patch in the other multi-view image or images by transforming the derived displacement vector field representing a transformation of the image patch in said one image of the multi-view images in accordance with an estimate of the optical flow between the image patches in said one image of the multi-view images and the other multi-view image or images.

4. A method according to claim 1, wherein

said plural local image descriptors are plural local image descriptors in both images of the stereoscopic pair,

said step of deriving displacement vector fields comprises:

deriving an initial displacement vector field representing a notional transformation of a notional image patch in a notional image having a notional camera location relative to the camera locations of the multi-view images, using the derived feature vector to look up reference data comprising reference displacement vector fields associated with possible values of the feature vector; and

deriving displacement vector fields representing a transformation of the image patches in each image of the multi-view images by transforming the initial displacement vector field in accordance with an estimate of the optical flows between the notional image patches in the notional images and the image patches in the images of the multi-view images.

5. A method according to claim 4, wherein the multi-view images are a stereoscopic pair of images and the notional camera location is between the camera locations of the images of the stereoscopic pair.

6. A method according to any one of the preceding claims, wherein

the step of deriving a displacement vector field comprises deriving displacement vector fields in respect of the image patches in each image of the multi-view images, and

the step of transforming the image patches containing the left and right eyes of the head is performed in accordance with the displacement vector fields derived in respect of the image patches in each image of the multi-view images.

7. A method according to any one of claims 1 to 5, wherein

the step of deriving a displacement vector field comprises deriving displacement

vector fields in respect of the image patches in each image of the multi-view images, and further deriving a merged displacement vector field from the displacement vector fields derived in respect of the image patches in each image of the multi-view images, and

the step of transforming the image patches containing the left and right eyes of the head is performed in accordance with the merged displacement vector field.

8. A method according to claim 7, wherein

the reference displacement vector fields are further associated with confidence values,

the step of deriving displacement vector fields in respect of the image patches in each image of the multi-view images further comprises deriving a confidence value associated with each derived displacement vector field, and

the merged displacement vector field is an average of the displacement vector fields derived in respect of the image patches in each image of the multi-view images weighted by the derived confidence values.

9. A method according to any one of the preceding claims, wherein

the method uses location data representing an observer location relative to a display location of the multi-view images,

said reference data comprises plural sets of reference displacement vector fields associated with possible values of the feature vector, which sets are associated with different observer locations, and

said step of deriving displacement vector fields representing a transformation of an image patch is performed using the derived feature vector to look up the set of reference displacement vector fields associated with the observer location represented by the location data.

10. A method according to any one of the preceding claims, wherein the local image descriptors are local image descriptors derived in a normalized coordinate system, and the reference and derived displacement vector fields are displacement vector fields in the same normalized coordinate system.

11. A method according to any one of the preceding claims, wherein

the feature vector comprises features that are values derived from the local image descriptors in a discrete space,

the reference displacement vector fields comprise a reference displacement vector field associated with each possible value of the feature vector in the discrete space, and

the step of deriving a displacement vector field for the image patch comprises selecting the reference displacement field associated with the actual value of the derived feature vector.

12. A method according to any one of claims 1 to 10, wherein

the feature vector comprises features that are discrete values of the local image descriptors in a continuous space, and

the step of deriving a displacement vector field for the image patch comprises interpolating a displacement vector field from the reference displacement vector fields based on the relationship between the actual value of the derived feature vector and the values of the feature vector associated with the reference displacement vector fields.

13. A method according to any one of the preceding claims, wherein the multi-view images are a stereoscopic pair of images.

14. A computer program capable of execution by a processor and arranged on execution to perform a cause the processor to perform a method according to any one of the preceding claims.

15. A computer readable storage medium storing a computer program according to claim 14.

16. A device for adjusting multi-view images of a head to correct gaze, the device comprising an image processor arranged to process the multi-view images of a head by:

in each image, identifying image patches containing the left and right eyes of the head, respectively;

in respect of the image patches containing the left eyes of the head in each image of the multi-view images, and also in respect of the image patches containing the right eyes of the head in each image of the multi-view images, performing the steps of:

deriving a feature vector from plural local image descriptors of the image patch in at least one image of the multi-view images; and

deriving a displacement vector field representing a transformation of an image patch, using the derived feature vector to look up reference data comprising reference displacement vector fields associated with possible values of the feature vector; and

adjusting each image of the multi-view images by transforming the image patches containing the left and right eyes of the head in accordance with the derived displacement vector field.

17. A device according to claim 16, further comprising a camera system arranged to capture the multi-view images.

18. A device according to claim 16 or 17, further comprising a telecommunications interface arranged to transmit the adjusted images over a telecommunications network to a destination device for display thereon.

19. A device according to any one of claims 16 to 18, wherein the multi-view images are a stereoscopic pair of images.

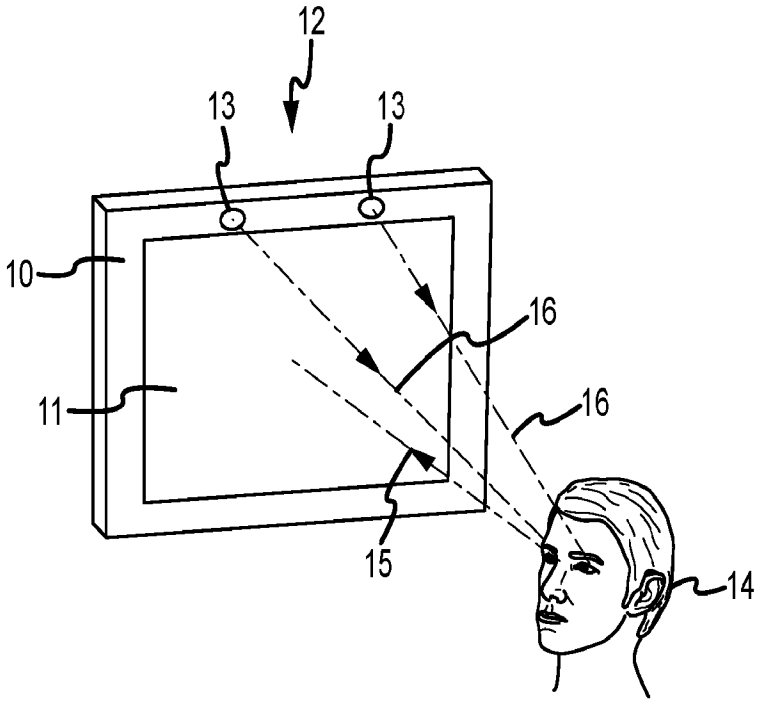


FIG.1

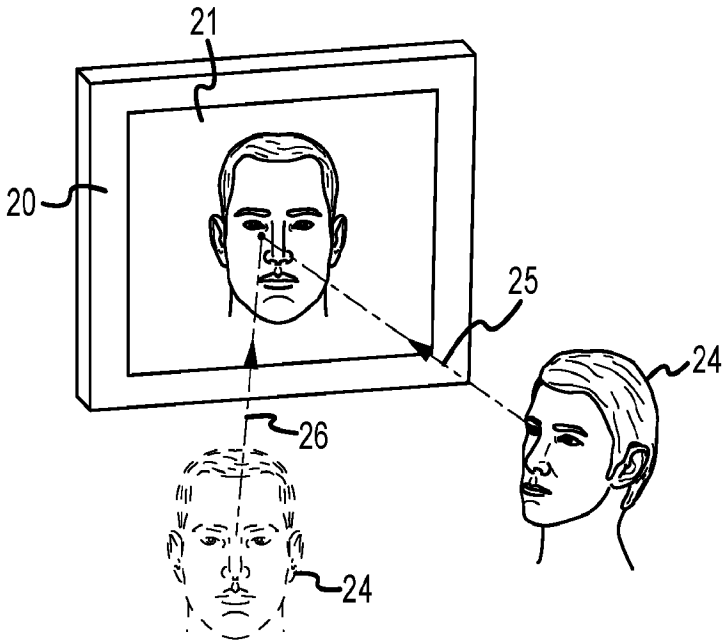


FIG.2

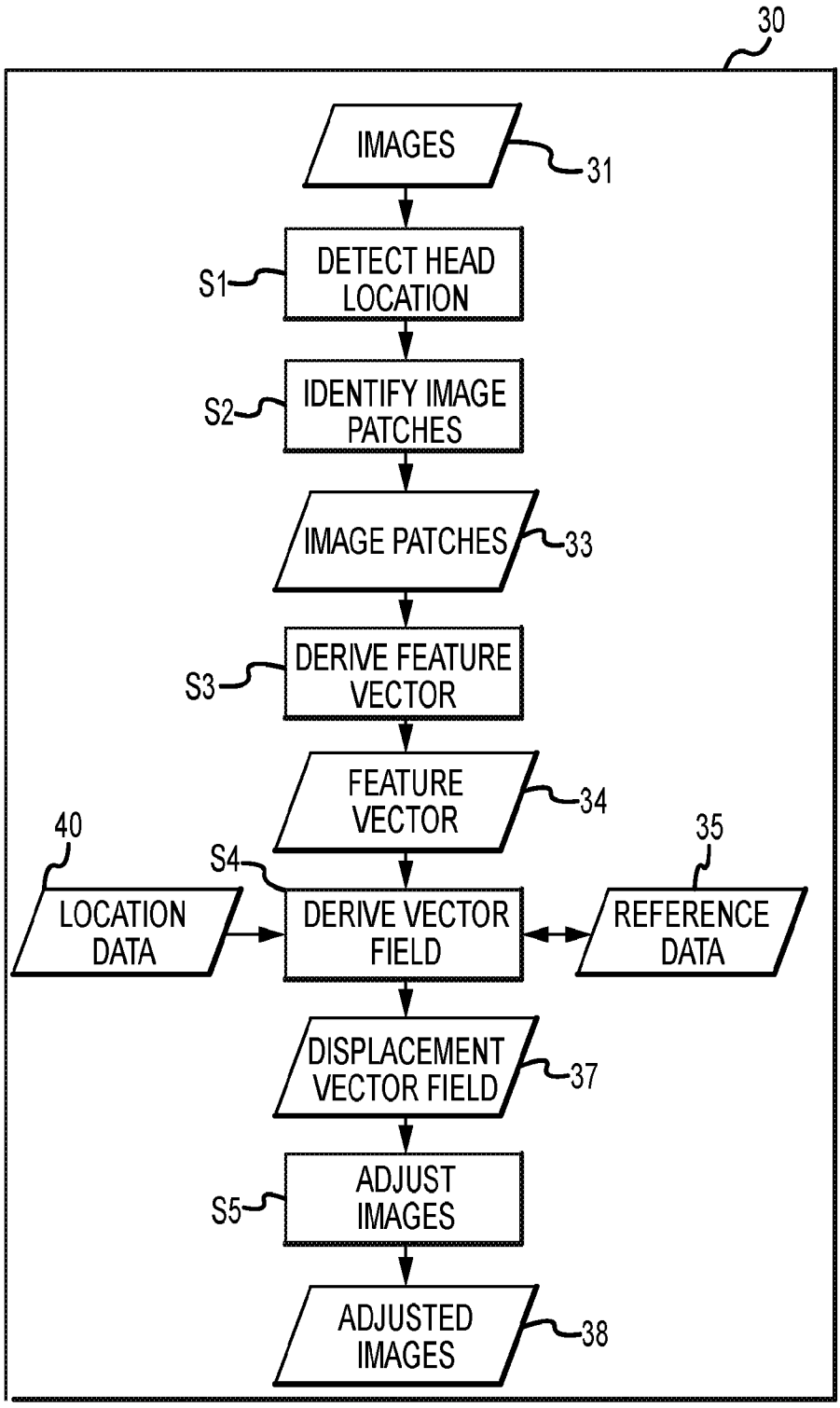

30
IMAGES
31
S1
DETECT HEAD LOCATION
S2
IDENTIFY IMAGE PATCHES
IMAGE PATCHES
33
S3
DERIVE FEATURE VECTOR
FEATURE VECTOR
34
40
35
S4
LOCATION DATA
DERIVE VECTOR FIELD
REFERENCE DATA
DISPLACEMENT VECTOR FIELD
37
S5
ADJUST IMAGES
ADJUSTED IMAGES
38

FIG.3

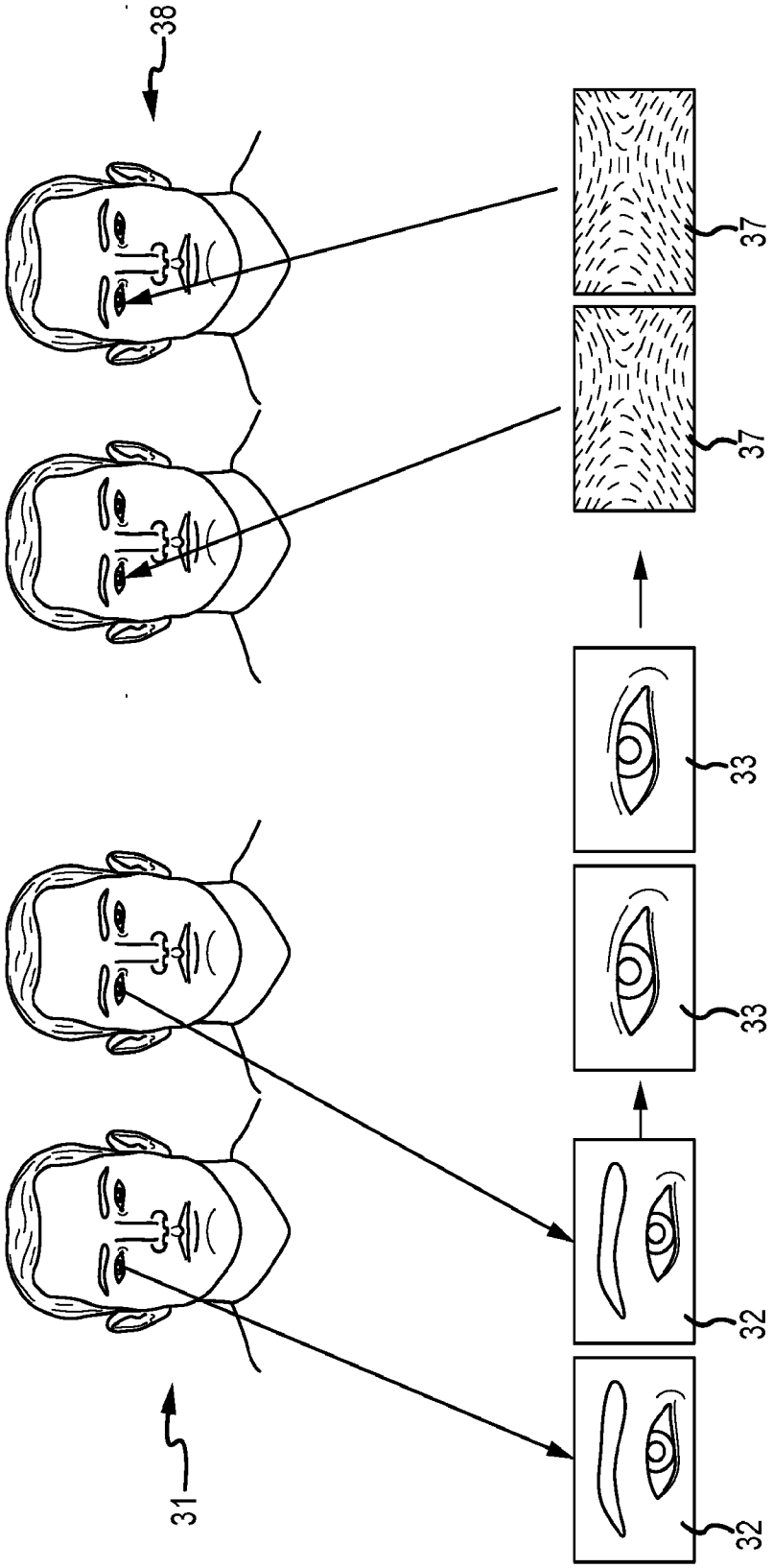


FIG.4

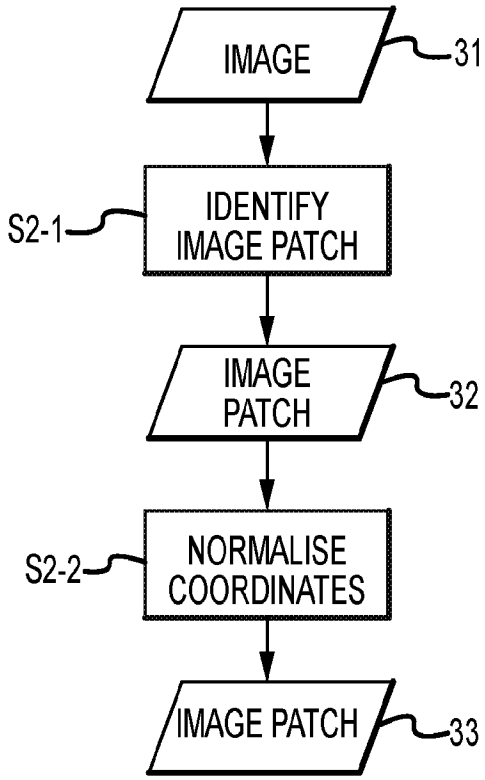

IMAGE
31
S2-1
IDENTIFY IMAGE PATCH
IMAGE PATCH
32
S2-2
NORMALISE COORDINATES
IMAGE PATCH
33

FIG.5

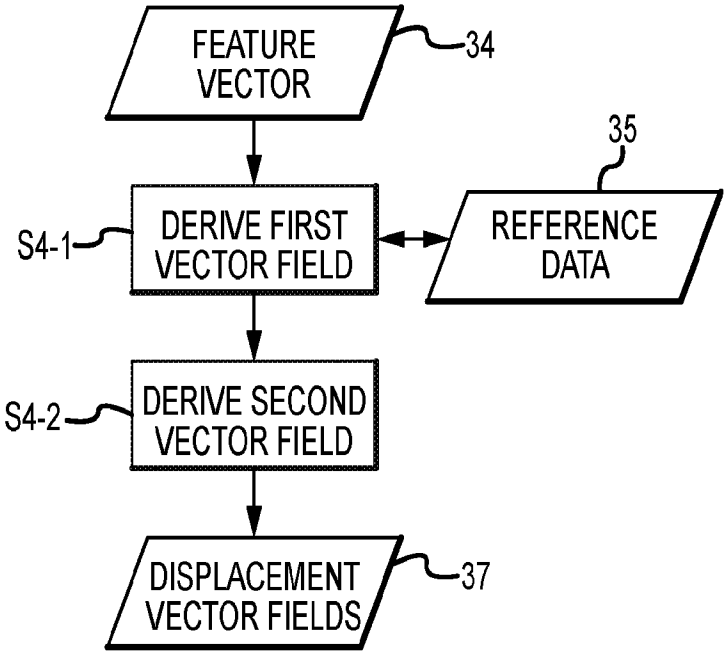

FEATURE VECTOR
34
35
S4-1
DERIVE FIRST VECTOR FIELD
REFERENCE DATA
S4-2
DERIVE SECOND VECTOR FIELD
DISPLACEMENT VECTOR FIELDS
37

FIG.6

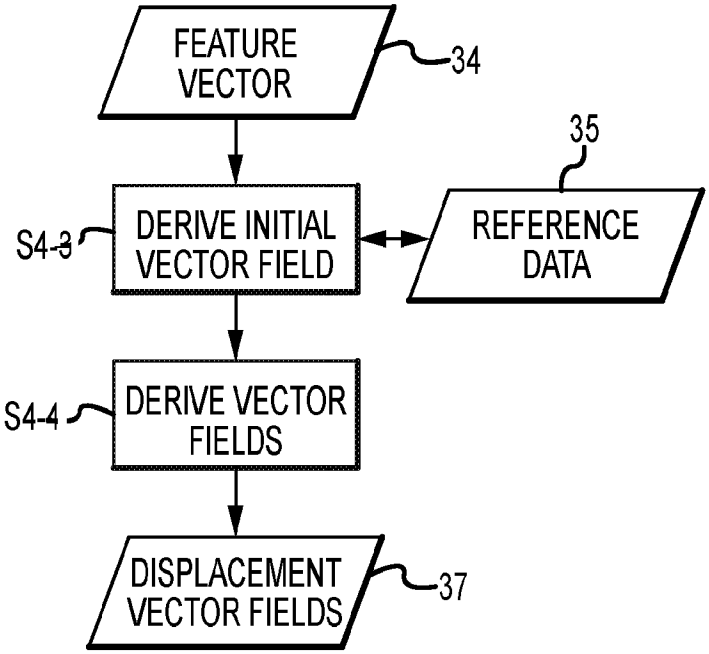

FEATURE VECTOR
34
35
S4-3
DERIVE INITIAL VECTOR FIELD
REFERENCE DATA
S4-4
DERIVE VECTOR FIELDS
DISPLACEMENT VECTOR FIELDS
37

FIG.7

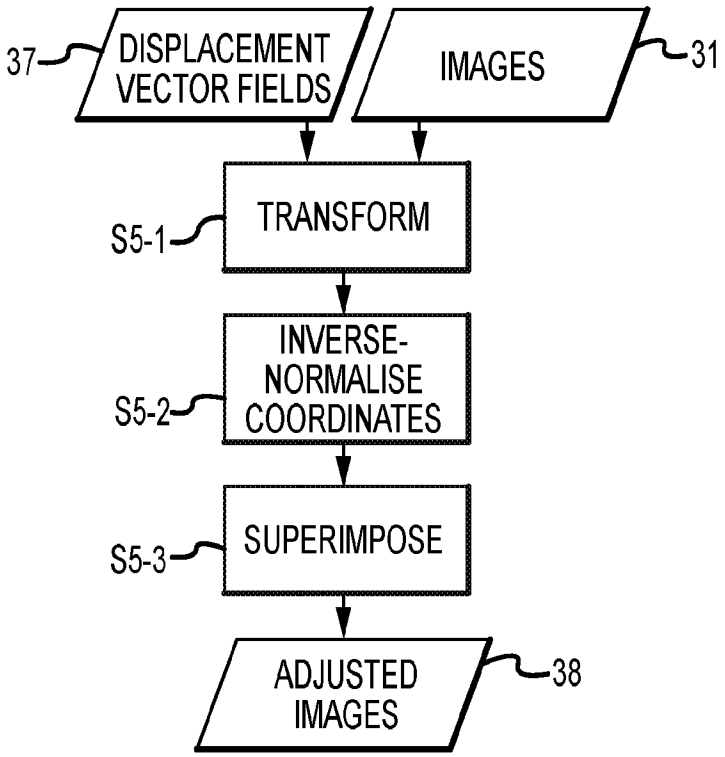

37
DISPLACEMENT VECTOR FIELDS
IMAGES
31
S5-1
TRANSFORM
S5-2
INVERSE-NORMALISE COORDINATES
S5-3
SUPERIMPOSE
ADJUSTED IMAGES
38

FIG.8

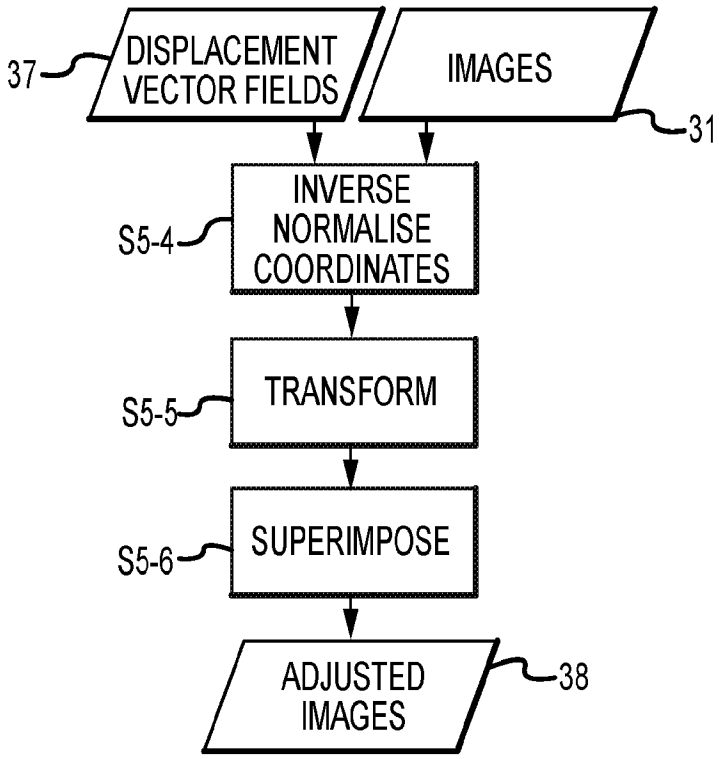

37
DISPLACEMENT VECTOR FIELDS
IMAGES
31
S5-4
INVERSE NORMALISE COORDINATES
S5-5
TRANSFORM
S5-6
SUPERIMPOSE
ADJUSTED IMAGES
38

FIG.9

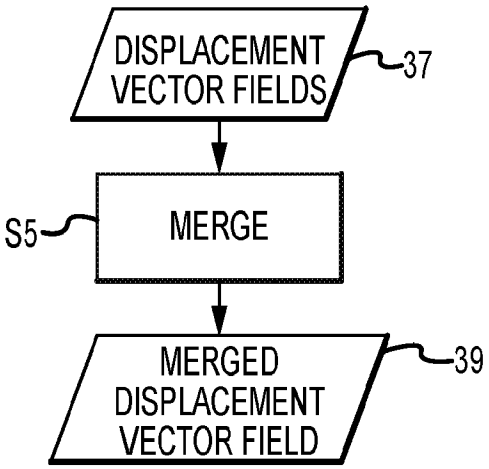

DISPLACEMENT VECTOR FIELDS
37
S5
MERGE
MERGED DISPLACEMENT VECTOR FIELD
39

FIG.10

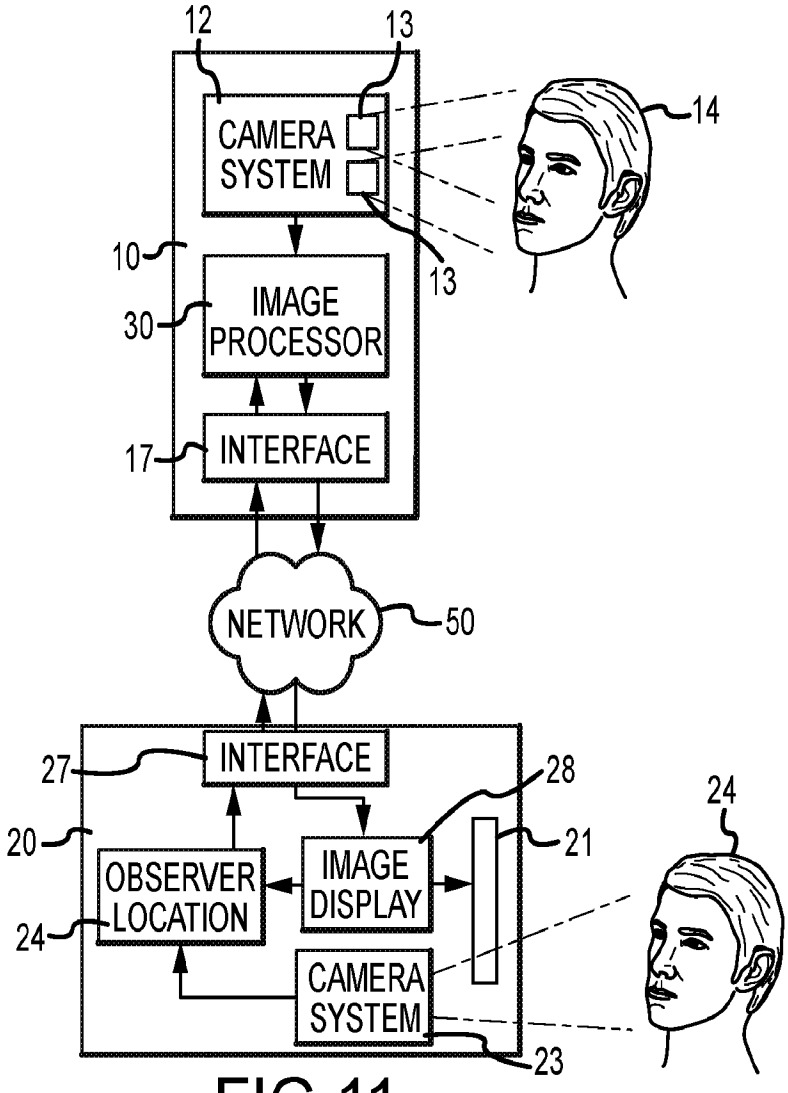

12
13
CAMERA SYSTEM
14
13
10
30
IMAGE PROCESSOR
17
INTERFACE
NETWORK
50
27
INTERFACE
28
20
OBSERVER LOCATION
IMAGE DISPLAY
21
24
24
CAMERA SYSTEM
23

FIG.11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/012203

A. CLASSIFICATION OF SUBJECT MATTER

H04N 13/04(2006.01)i

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)
H04N 13/04; G06K 9/00; G06K 9/60; G06T 7/20; A61B 3/14; H04N 7/14

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: gaze, camera, multi-view, vector, reference

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2015-0339512 A1 (SAMSUNG ELECTRONICS CO., LTD. et al.) 26 November 2015 See paragraphs [0062], [0096]-[0098]; claims 1-11; and figures 1-22.	1-7,16-18
Y		8
Y	US 8942434 B1 (KENNETH M. KARAKOTSIOS et al.) 27 January 2015 See column 4, line 33-column 8, line 38; claims 1-5; and figures 1-11.	8
A	US 2014-0267584 A1 (FRAUNHOFER-GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.) 18 September 2014 See paragraphs [0026]-[0054]; claims 1-11; and figures 1-7.	1-8,16-18
A	US 2012-0219180 A1 (MEHRA SUMAT) 30 August 2012 See paragraphs [0096]-[0107]; claims 1-12; and figures 1A-5B.	1-8,16-18
A	JP 2007-109255 A (EYEMATIC INTERFACES INC.) 26 April 2007 See paragraphs [0012]-[0031]; claim 1; and figures 1-5.	1-8,16-18

☐ 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 application or patent 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	Date of mailing of the international search report
17 April 2017 (17.04.2017)	**18 April 2017 (18.04.2017)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
AHN, Jeong Hwan
Telephone No. +82-42-481-8633

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/012203

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 15
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
Claim 15 is not clear, contrary to the requirement of PCT Article 6, because it refers to a claim which does not comply with PCT Rule 6.4(a).

3. ☒ Claims Nos.: 9-14,19
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required addtional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of any additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2017/012203

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2015-0339512 A1	26/11/2015	KR 10-2014-0010541 A	27/01/2014
		US 2014-0016871 A1	16/01/2014
		US 9111171 B2	18/08/2015
		US 9613286 B2	04/04/2017
US 8942434 B1	27/01/2015	None	
US 2014-0267584 A1	18/09/2014	US 9521362 B2	13/12/2016
		WO 2013-079607 A1	06/06/2013
US 2012-0219180 A1	30/08/2012	US 2012-218398 A1	30/08/2012
		US 8836777 B2	16/09/2014
JP 2007-109255 A	26/04/2007	AU 1999-34904 A1	01/11/1999
		AU 1999-35541 A1	01/11/1999
		AU 1999-36396 A1	01/11/1999
		AU 3639699 A	01/11/1999
		CA 2326816 A1	21/10/1999
		CA 2326816 C	05/04/2005
		CA 2327304 A1	21/10/1999
		EP 1072014 A1	31/01/2001
		EP 1072014 B1	24/11/2004
		EP 1072018 A1	31/01/2001
		EP 1072018 B1	27/08/2003
		EP 1580684 A1	28/09/2005
		EP 1580684 B1	31/12/2008
		JP 2002-511617 A	16/04/2002
		JP 2002-511620 A	16/04/2002
		JP 3970520 B2	05/09/2007
		JP 4177402 B2	05/11/2008
		KR 10-0530812 B1	28/11/2005
		KR 10-0653288 B1	04/12/2006
		US 2001-0033675 A1	25/10/2001
		US 2002-0015512 A1	07/02/2002
		US 2002-0031253 A1	14/03/2002
		US 2002-0067362 A1	06/06/2002
		US 2002-0118195 A1	29/08/2002
		US 2003-0007666 A1	09/01/2003
		US 2003-0031381 A1	13/02/2003
		US 6272231 B1	07/08/2001
		US 6301370 B1	09/10/2001
		US 6580811 B2	17/06/2003
		US 6714661 B2	30/03/2004
		US 6940454 B2	06/09/2005
		US 7050624 B2	23/05/2006
		US 7050655 B2	23/05/2006
		WO 99-53427 A1	21/10/1999
		WO 99-53430 A1	21/10/1999
		WO 99-53443 A1	21/10/1999