



(51) International Patent Classification:

H04N 13/25 (2018.01) *G06T 5/50* (2006.01)
G06T 7/30 (2017.01)

(21) International Application Number:

PCT/IB2019/057391

(22) International Filing Date:

02 September 2019 (02.09.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant: **TRIEYE LTD.** [IL/IL]; 94 Yigal Alon St.,
6789155 Tel Aviv (IL).

(72) Inventors: **SHKLYAR, Roman**; 28/4 Sderot Yerushalayim,
5835902 Holon (IL). **LEVY, Uriel**; 13 Hakfar St.,
5552507 Kiryat Ono (IL). **DOBRINSKY, Roni**; 79 Yeshua
Ben Nun St., 6249712 Tel Aviv (IL). **KAPACH, Omer**;
9/5 Ha-Rav Shmuel Braukh St., 9738008 Jerusalem (IL).
BAKAL, Avraham; 144 Jabotinsky St., 6299111 Tel Aviv
(IL).

(74) Agent: **NATHAN & ASSOCIATES PATENT AGENTS LTD**; P.O. Box 10178, 6110101 Tel-Aviv (IL).

(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).

Published:

- with international search report (Art. 21(3))
- in black and white; the international application as filed
contained color or greyscale and is available for download
from PATENTSCOPE

(54) Title: MULTI-VIEW IMAGE FUSION BY IMAGE SPACE EQUALIZATION AND STEREO-BASED RECTIFICATION FROM TWO DIFFERENT CAMERAS

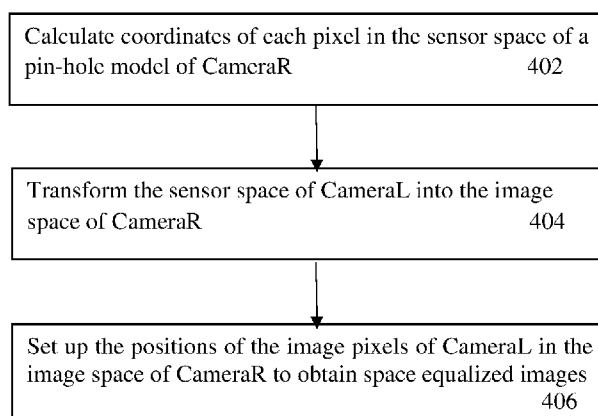


FIG. 4

(57) Abstract: Methods to solve the problem of performing fusion of images acquired with two cameras with different type sensors, for example a visible (VIS) digital camera and an short wave infrared (SWIR) camera, include performing image space equalization on images acquired with the different type sensors before performing rectification and registration of such images in a fusion process.

MULTI-VIEW IMAGE FUSION BY IMAGE SPACE EQUALIZATION AND STEREO-BASED RECTIFICATION FROM TWO DIFFERENT CAMERAS

FIELD

5

Embodiments disclosed herein relate in general to cameras and in particular to image processing in dual-cameras with different type sensors.

BACKGROUND

10

The term “fusion” means in general an approach to extraction of information acquired in several domains. In work with images (e.g. in photography), the goal of “image fusion” (IF) is to integrate complementary multi-sensor, multitemporal and/or multi-view information into one new image containing information with a quality that cannot be achieved otherwise. The term “quality”, its meaning and its measurement depend on the particular application.

15

Existing fusion methods (wavelets transforms, statistical approaches (e. g. principal component analysis or “PCA”), Multi Scale Decomposition (“MSD) and others) need to be performed on a pair of registered images from the same image space. Several known approaches deal with the problem of having inputs for fusion from distinct sensors (e.g. image sensor of two (or “dual”) cameras), based on cost computation methods, such as histogram of oriented gradients (HOG), connected component labeling (CCL), image statistical methods, etc. These approaches work well if the two cameras are identical, having the same optics and capturing the same image. However, when different type sensors are used, they provide different image properties, significantly reducing the practicality of these approaches. Therefore, in fusion, based on inputs from different type sensors, one may observe a degraded image. For example, a ghost image may appear on top of the real image.

20

25

The ghost image problem has known solutions. However, known solutions to this problem require heavy computational power, sometimes even requiring a graphic processing unit (GPU).

30

SUMMARY

Embodiments disclosed herein provide approaches (methods) to solve the problem of performing fusion of images acquired with two cameras with different type sensors, for

example a visible (VIS) digital camera and an short wave infrared (SWIR) camera, with smaller complexity and/or computation power than required by known fusion techniques. Such approaches also solve the problem of ghost images. An exemplary approach includes performing Image Space Equalization (ISE) on images acquired with the different type sensors before performing rectification and registration of such images in a fusion process.

In exemplary embodiments, there are disclosed methods for fusing images from two different cameras, a first camera and a second camera, comprising: performing a transformation that translates the image space of the first camera onto the image space of the second camera to obtain a stereo rig in the form of a pair of space equalized images; performing rectification on the pair of space equalized images to obtain rectified first and second images; performing registration on the rectified first and second images to obtain registered first and second images; and fusing the registered first and second images into a fused image.

In an embodiment, the first camera has a first image sensor with a first resolution, the second camera has a second image sensor with a second resolution and the first resolution is higher than the second resolution.

In various embodiments, the first camera differs from the second camera in at least one parameter selected from focal length, sensor resolution, distortion, gain and spectral range.

In various embodiments, the performing a transformation that translates the image space of the first camera onto the image space of the second camera includes calculating coordinates of each pixel in the sensor space of a pin-hole model of the second camera, transforming a sensor space of the first camera into an image space of the second camera, and setting up the positions of image pixels of the first camera in the image space of the second camera to obtain the space equalized images.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting examples of embodiments disclosed herein are described below with reference to figures attached hereto that are listed following this paragraph. Identical structures, elements or parts that appear in more than one figure may be labeled with a same numeral in all the figures in which they appear. The drawings and descriptions are meant to illuminate and clarify embodiments disclosed herein, and should not be considered limiting in any way.

FIG. 1 shows an image of a scene captured by a visible (VIS) camera with a variable focal length of 4.5-13.2 mm and having a 1080x1920 pixel image sensor with 4.08 nm pixel size (width and height);

FIG. 2 shows an image of the same scene captured by a short wave IR (SWIR) camera with a fixed focal length of 35 mm and having a 512x640 pixel image sensor with 15 nm pixel size;

FIG. 3 shows in (a) an image obtained from the resizing of dimensions 1080x1920 of the VIS image of FIG. 1 to 512x640 and in (b) SWIR image of FIG. 2;

FIG. 4 shows in a flow chart main steps in an ISE process disclosed herein;

FIG. 5 shows the images of FIG. 3 after image space equalization;

FIG. 6 shows a depth map obtained from rectification of the stereo rig of equalized VIS and SWIR images;

FIG. 7 shows in a schematic description the projection of a common point M on both cameras used to acquire the images in FIG. 1 and FIG. 2 in a known process;

FIG. 8 shows a schematic drawing of the image rectification process in a known process.

DETAILED DESCRIPTION

If the two cameras are identical, the rectification process is well known. However, assume a more general case in which the cameras are different. Specifically, assume two distinct cameras with different focal lengths, different sensor resolutions, different principal point displacement, different distortion and possibly different gain and spectral range. As used herein, the term “different cameras” reflects the fact that the cameras providing images to be fused into a fused image differ from each other in at least one parameter chosen from focal lengths, sensor resolutions, distortion, gain and spectral range. In some embodiments, the parameter by which different cameras differ may also be principal point position.

For example, assume two cameras, a “left camera” (or “first camera”) CameraL and a “right camera” (or “second camera”) CameraR, where one camera (e.g. CameraL) operates in the VIS spectral regime while the other (e.g. CameraR) works in the SWIR spectral regime. CameraL has the following intrinsic parameters: focal length F_1 , principal point $c_L(p_L, q_L)$, radial distortion coefficients $[K_1^L, K_2^L, K_3^L]$ and tangential distortion $[T_1^L, T_2^L]$. CameraR has the following intrinsic parameters: focal length F_2 , principal point $c_R(p_R, q_R)$, radial distortion coefficients $[K_1^R, K_2^R, K_3^R]$ and tangential distortion $[T_1^R, T_2^R]$.

In an example, FIG. 1 shows an image **100** of a scene captured by a VIS camera with a variable focal length of 4.5-13.2 mm and having a 1080x1920 pixel image sensor with 4.08 nm

pixel size. FIG. 2 shows an image **100'** of the same scene captured by a SWIR camera with a fixed focal length of 35 mm and having a 512x640 pixel image sensor with 15 nm pixel size. Both images show a person with a head **102**. The task is to build a transformation that translates the image space of the first camera onto the image space of the second camera to obtain a stereo rig of as would be obtained with two cameras with same focal length, sensor properties and distortion coefficients.

FIG. 3 shows a scene capture by a stereo rig of VIS CameraL and SWIR CameraR. FIG. 3 shows in (a) image **100** (with a head **102'**) obtained from the resizing of dimensions 1080x1920 of the VIS image of FIG. 1 to dimensions 512x640 of SWIR image, and in (b) the SWIR image of FIG. 2. However, if we only resize the dimensions of the image of CameraL in FIG. 3(a) to the dimensions of the image of CameraR in FIG. 3(b), as done in a known typical fusion process, true fusion will not be obtained because of wrong representation of the resized objects from the CameraL in the image space of CameraR. As seen, there is a significant distortion in the left (CameraL) figure. For example, head **102'** in (a) has an elliptic form with much more eccentricity than head **102** in (b). In addition, one may easily observe ghost images, FIG 3 (c). Thus, the task is to make the left image in FIG. 3(a) look similar to the original image in FIG. 2 so that the structure similarity index (SSIM) will be as close as possible to 100%. To achieve this goal, it is proposed herein to perform an “image space equalization” (ISE) process before applying fusion, to obtain a ghost-free image after fusion regardless of the chosen fusion method. To emphasize, performing ISE as disclosed herein removes the need for ghost “cleaning” processes now used in known art to reach the same result.

Image Space Equalization

We construct a lookup table that transfers the positions of all the sensor pixels from the image sensor of CameraL to the sensor pixels of the image sensor of CameraR. This represents a transform from a sensor with bigger (higher) resolution into a sensor with smaller (lower) resolution. First we define a scaling factor for the dimensions $d = \left[\frac{H_L}{H_R}, \frac{W_L}{W_R} \right]$ where the H_L, W_L, H_R, W_R the height and width of the left and right sensor respectively. Next, we build the lookup table for CameraR using the following procedure, FIG. 4.

Assume that each image sensor has a width dimension in direction X and a height dimension in direction Y. Step **402** receives as inputs the intrinsic parameters of CameraL and

CameraR and images captured by both cameras. In step **402**, calculate coordinates x, y of each pixel in the sensor space of a pin-hole model of the CameraR as follows:

$$x = \frac{x_R - p_R}{F_R}, \quad y = \frac{y_R - q_R}{F_R}$$

5

where p_R, q_R are the pixel positions (in mm) of the principal point in horizontal and the vertical direction and F_R is the pixel relative focal length.

In step **404** and using coordinates x, y of each pixel in the sensor of CameraL, we transform the sensor space of the CameraL into the image space of the CameraR. That is, the scene imaged by CameraR is as if CameraR is mounted on the position of CameraL. This is performed by defining a transformation $(x_R, y_R) \rightarrow (u, v)$ in the following way:

10

$$u(x_R) = p_R + \frac{x \cdot F_L}{d(1)} = p_R + \frac{(x_R - p_R) \cdot F_L}{F_R \cdot d(1)}$$

$$v(y_R) = q_R + \frac{y \cdot F_L}{d(2)} = q_R + \frac{(y_R - q_R) \cdot F_L}{F_R \cdot d(2)}$$

15

where $u(x_R), v(x_R)$ are the new pixel positions in cameraL and $d(1) = \frac{H_L}{H_R}, d(2) = \frac{W_L}{W_R}$. The result is a look-up table of pixel positions in the image (transformed from the sensor of CameraL to the image obtained by CameraR).

In step **406**, using the information in the look-up table, set up the positions of the image pixels of CameraL in the image space of CameraR as follows:

20

$$Image_{Left}^{New}(i, j) = Image_{Left}^{Old}(u(i), v(j)) \text{ for all } i, j \text{ pixel positions of CameraR.}$$

where $Image_{Left}^{New}$ is the image from the image space of the CameraL represented in the image space of the CameraR without distortion correction, see FIG. 5. The output of step **406** is relocation of all pixels of CameraL according to the lookup table.

25

FIG. 5 shows the images of FIG. 3 after image space equalization.

The ISE can be performed on any pair of cameras. The fusion by ISE can be applied on any pair of still images or videos, captured by the cameras. The ISE process above allows to perform stereo rectification, so the image registration will be very easily calculated, since the rectified images need to be only translated along stereo base direction. The ISE process above

30

may be applied in all known fusion methods, and the only error is the rectification error that can be easily handled. In the case of the video, auto-rectification needs to be applied because of possible new misalignments, caused by cameras moving

Stereo rectification can now be performed on the pair of space equalized images from the same image space that lead to ability to calculate disparity and depth maps. The stereo rectification is an evaluation of the new camera extrinsic positions such that the optical axes of the cameras will be parallel. The term “extrinsic position” refers to the position of a first camera in the coordinate space of a second camera, i.e. position after rotation and translation of the first camera-relative to the second camera. The rectification can be achieved by computing a pair of transformation matrices that map the epipoles to a point at infinity. Known methods for image rectification that may be used are described for example in “Quasi-Euclidean Epipolar Rectification”, Image Processing On Line, 1 (2011), pp. 187–199, or in “Computing Rectifying Homographies for Stereo Vision”, by Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399, USA (1999). These methods assume that the cameras have the same intrinsic parameters, and ISE removes that constraint. FIG. 6 shows a depth map obtained from rectification of the stereo rig of equalized VIS and SWIR images.

Although the rectification after ISE may be done in well-known ways, it is described next in detail for completeness. The rectification process begins by defining the epipolar geometry between a pair of images (marked next I_L and I_R , or as in FIG. 8, I and I'). Let us define the pair of the pin-hole cameras C_L (left camera) and C_R (right camera) in a 3-dimensional (3D) space. Let us define M as a common point, visible from both cameras. We define m_L and m_R as the projections of point M (m_L and m_R are the coordinates in units of pixels) called correspondent points into, respectively, left image I_L and right image I_R . The geometry of this definition is shown in FIG. 7. The problem is that m_L and m_L can be anywhere on an epipolar line, FIG. 8.

As seen in FIG. 8, the epipolar line is the intersection of the plane, defined by point P and optical centers of the left and right cameras with left and right image spaces accordingly. All epipolar lines that belong to different matching points intersect in one point (“epipole”). For each common point P there are two epipolar lines (dashed) as shown. A change in the position of the common point yields new epipolar lines. All distinct epipolar lines before rectification intersect in an epipole (not shown). After rectification, the epipolar lines become parallel and epipoles lie at infinity.

The problem above is expressed by a known epipolar constraint defined as $(m_L)^T F m_R = 0$ for all pairs of image correspondences. Matrix F is a “fundamental” matrix, a 3×3 matrix with rank 2 that maps pixels from image I_L to image I_R such that each corresponding point from one camera is mapped onto the epipolar line on the second camera. The fundamental matrix can be estimated directly from intrinsic and extrinsic camera parameters. Intrinsic parameters include for example focal length, principal point, distortion, pixel size in nanometers, sensor physical dimensions in millimeters and, optionally, a skew factor. Extrinsic parameters include for example a rotation matrix and translation vectors of one (e.g. the first) camera relative to another (e.g. the second) camera. There are several known techniques to calculate the fundamental matrix.

The new orientation of the left and right camera with respect to world coordinates is explained with reference to FIG. 8. In the figure, H_1 is a projective transformation that rotates the first image to be perpendicular to the baseline connecting O and O' . H_2 is the projective transformation that takes the rotated image and twists it so that the horizontal axis aligns with the baseline. Finally $H = H_2 \circ H_1$. H is defined as the rectification transformation of the (left) image I .

Following the rectification process, stereo matching is performed, benefitting from the ISE. Stereo matching is used to recover 3D depth information from at least two images captured from different point of views. One of the major problems here is a correspondent problem - find the matching between the different projections of the same point in real space. If a point in one image is given, its corresponding point must lie on the epipolar line in the other image. If the two cameras are placed side by side on the same baseline and have the same intrinsic parameters, then the obtained images are known as a rectified pair of stereo images. On these images, matching points must lie on a same horizontal line. So the solution of the problem is to calculate a pair of warping transforms for a pair of stereo images that convert the epipolar geometry to the ideal state which has been described above (i.e. the state where all epipolar lines are parallel and the epipoles lie at infinity).

In conclusion, known methods dealing with what problem of ghost images in images fused after acquisition with two cameras with different type sensors involve iterative non-linear optimizations that lead to heavy computations. The ISE process disclosed herein allows to omit the heavy calculation routines needed in such known methods. By defining a lookup table and including an ISE process, the complexity is only $O(mn)$, where $m \times n$ is the dimension of CameraR, instead of at least $O(kmn)$ of any other iterative non-linear method, where k is

number of iterations, which is commonly inversely proportional to a desired accuracy. Therefore, we return the solution of the problem of ghost images in images fused after acquisition with two cameras with different type sensors back into solution by all known methods for the equal cameras with low computing cost, as mentioned above.

5 Unless otherwise stated, the use of the expression “and/or” between the last two members of a list of options for selection indicates that a selection of one or more of the listed options is appropriate and may be made.

It should be understood that where the claims or specification refer to "a" or "an" element, such reference is not to be construed as there being only one of that element.

10 Citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present disclosure.

While this disclosure has been described in terms of certain embodiments and generally associated methods, alterations and permutations of the embodiments and methods will be apparent to those skilled in the art. The disclosure is to be understood as not limited by the
15 specific embodiments described herein, but only by the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A method for fusing images from two different cameras, a first camera and a second camera, comprising:
 - a) performing a transformation that translates the image space of the first camera onto the image space of the second camera to obtain a stereo rig in the form of a pair of space equalized images;
 - b) performing rectification on the pair of space equalized images to obtain rectified first and second images;
 - c) performing registration on the rectified first and second images to obtain registered first and second images; and
 - d) fusing the registered first and second images into a fused image.
2. The method of claim 1, wherein the first camera has a first image sensor with a first resolution, wherein the second camera has a second image sensor with a second resolution and wherein the first resolution is higher than the second resolution.
3. The method of claim 1, wherein the first camera differs from the second camera in at least one parameter selected from focal length, sensor resolution, distortion, gain and spectral range.
4. The method of claim 2, wherein the first camera differs from the second camera in at least one parameter selected from focal length, sensor resolution, distortion, gain and spectral range.
5. The method of claim 1, wherein the performing a transformation that translates the image space of the first camera onto the image space of the second camera includes calculating coordinates of each pixel in the sensor space of a pin-hole model of the second camera, transforming a sensor space of the first camera into an image space of the second camera, and setting up the positions of image pixels of the first camera in the image space of the second camera to obtain the space equalized images.
6. The method of claim 2, wherein the performing a transformation that translates the image space of the first camera onto the image space of the second camera includes calculating

coordinates of each pixel in the sensor space of a pin-hole model of the second camera, transforming a sensor space of the first camera into an image space of the second camera, and setting up the positions of image pixels of the first camera in the image space of the second camera to obtain the space equalized images.

7. The method of claim 3, wherein the performing a transformation that translates the image space of the first camera onto the image space of the second camera includes calculating coordinates of each pixel in the sensor space of a pin-hole model of the second camera, transforming a sensor space of the first camera into an image space of the second camera, and setting up the positions of image pixels of the first camera in the image space of the second camera to obtain the space equalized images.

8. The method of claim 4, wherein the performing a transformation that translates the image space of the first camera onto the image space of the second camera includes calculating coordinates of each pixel in the sensor space of a pin-hole model of the second camera, transforming a sensor space of the first camera into an image space of the second camera, and setting up the positions of image pixels of the first camera in the image space of the second camera to obtain the space equalized images.

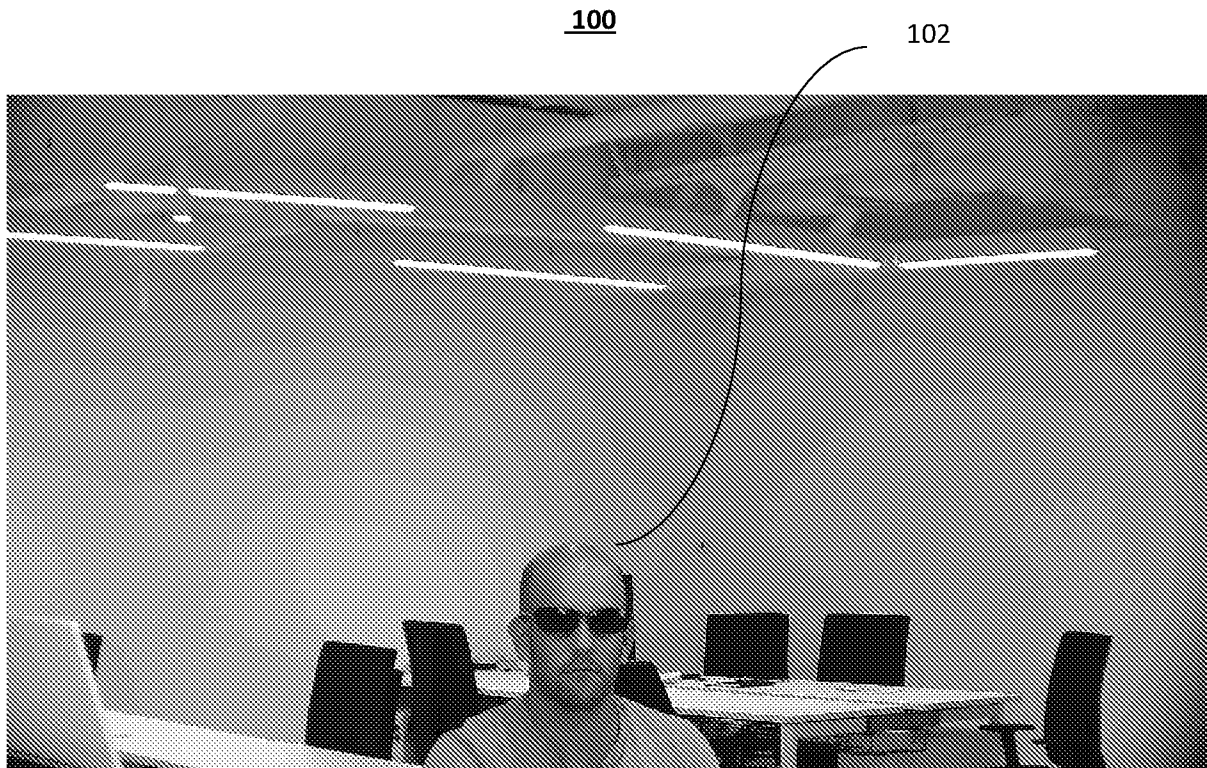


FIG. 1

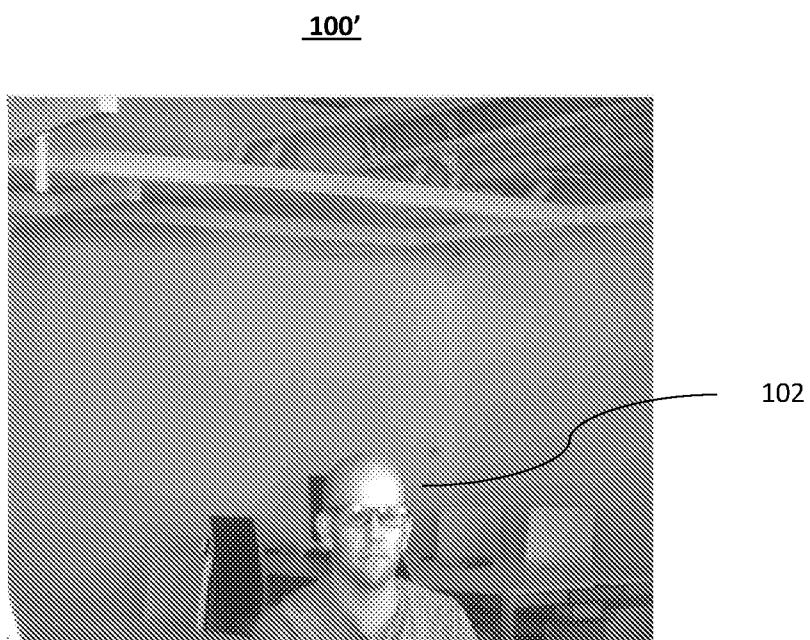


FIG. 2

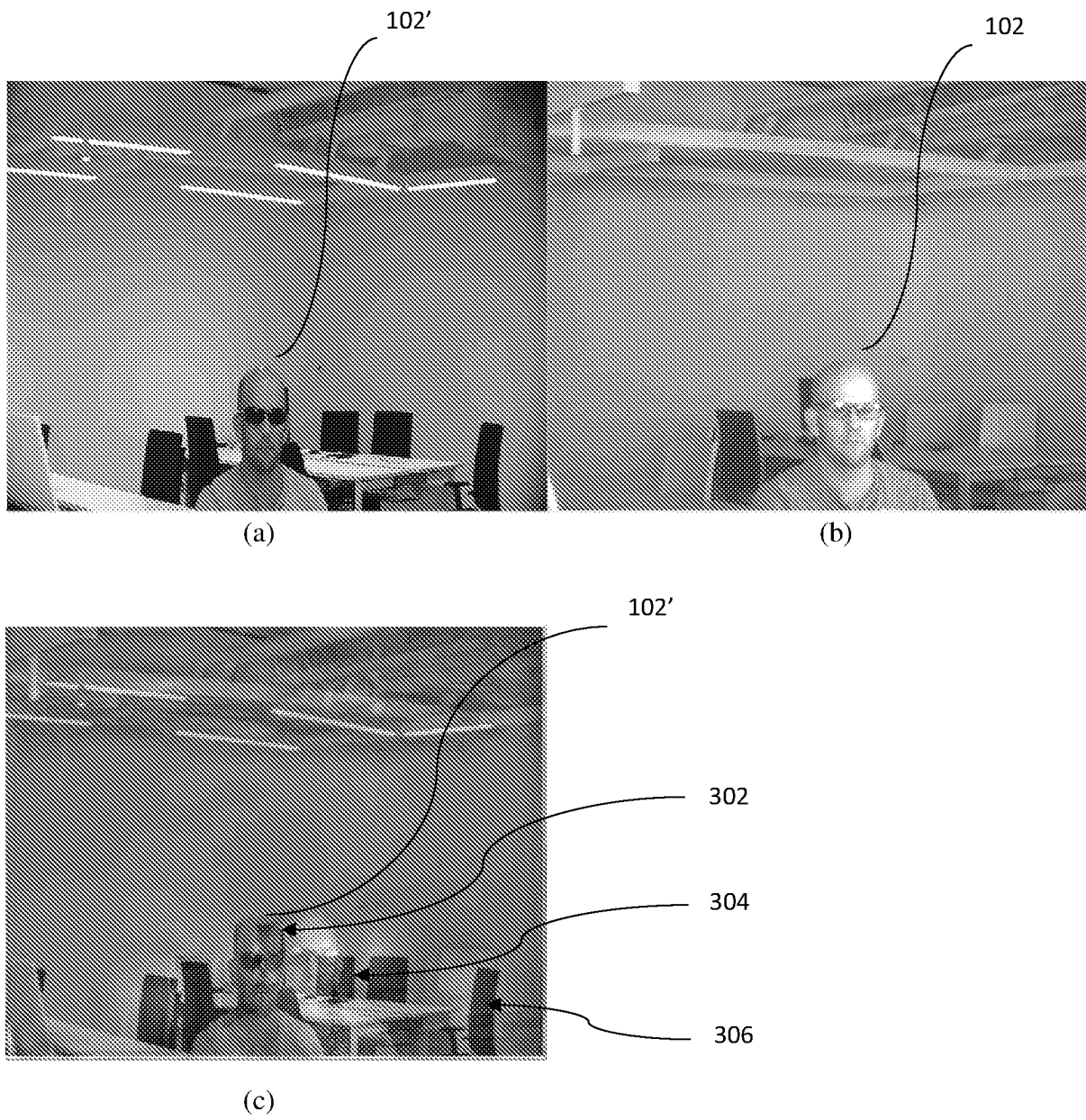


FIG. 3

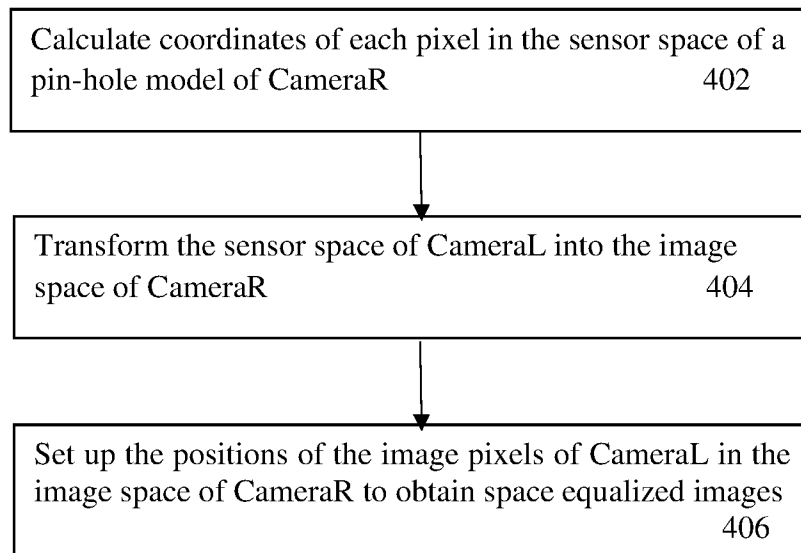


FIG. 4



FIG. 5



FIG. 6

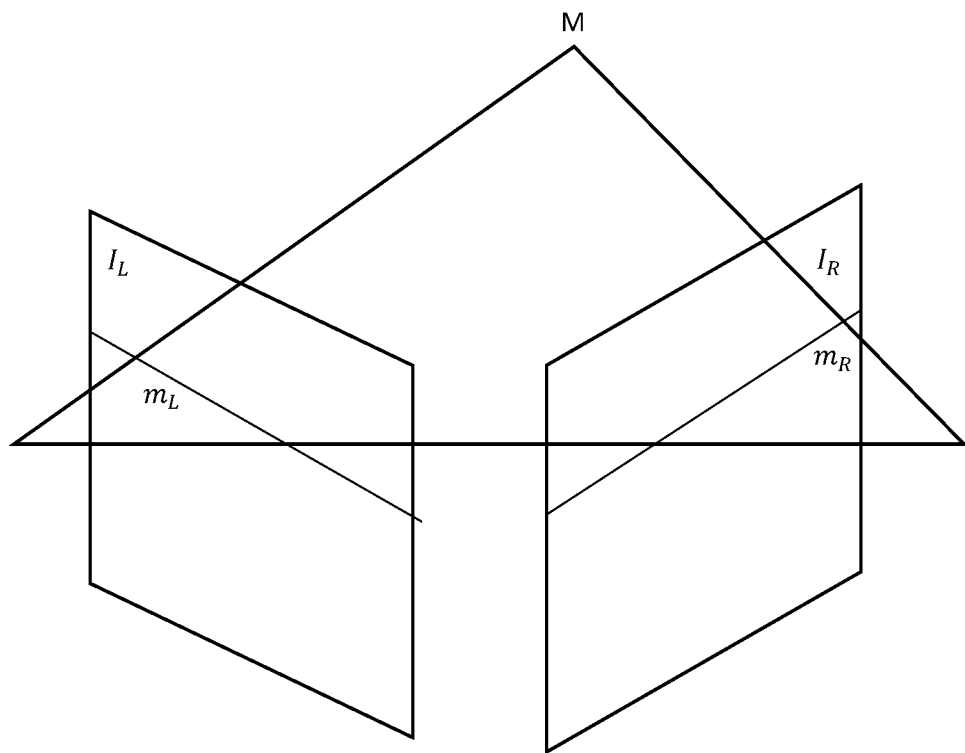
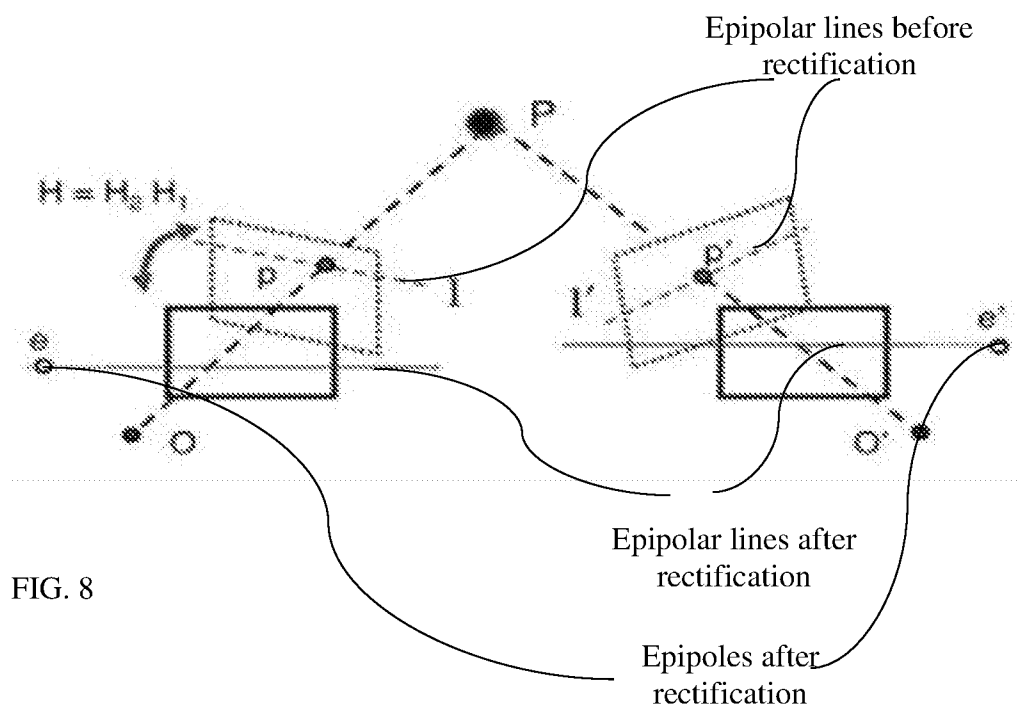


FIG. 7



INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB19/57391

A. CLASSIFICATION OF SUBJECT MATTER

IPC - H04N 13/25; G06T 7/30, 5/50 (2019.01)

CPC - H04N 5/2258, 13/25; G06T 7/30, 5/50

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)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2019/0149721 A1 (COREPHOTONICS LTD.) 16 May 2019; figures 2, 4A, 4B; paragraphs [0013], [0017], [0032], [0048]; claim 17.	1-8
Y	US 2014/0267614 A1 (SEIKO EPSON CORPORATION) 18 September 2014; figures 6, 7, 19; paragraphs [0107]-[0109], [0114], [0115], [0139];	1-8
A	WO 2018/063482 A1 (QUALCOMM INCORPORATED) 05 April 2018; entire document.	1-8

☐ 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

"D" document cited by the applicant in the international application

"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

01 December 2019 (01.12.2019)

Date of mailing of the international search report

30 DEC 2019

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

Telephone No. PCT Helpdesk: 571-272-4300