



(51) International Patent Classification:

G01S 7/41 (2006.01) G01S 17/10 (2020.01)

G01S 13/89 (2006.01) G01S 17/89 (2020.01)

G01S 15/89 (2006.01) G06N 3/02 (2006.01)

(21) International Application Number:

PCT/EP2020/063682

(22) International Filing Date:

15 May 2020 (15.05.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1908547.1 14 June 2019 (14.06.2019) GB

(71) Applicant: **THE UNIVERSITY COURT OF THE UNIVERSITY OF GLASGOW** [GB/GB]; Gilbert Scott Building, University Avenue, Glasgow Strathclyde G12 8QQ (GB).

(72) Inventors: **FACCIO, Daniele**; c/o The University Court of the University of Glasgow, Gilbert Scott Building, University Avenue, Glasgow Strathclyde G12 8QQ (GB). **TURPIN, Alejandro**; c/o The University Court of the University of Glasgow, Gilbert Scott Building, University Avenue, Glasgow Strathclyde G12 8QQ (GB). **MUSARRA,**

Gabriella; c/o The University Court of the University of Glasgow, Gilbert Scott Building, University Avenue, Glasgow Strathclyde G12 8QQ (GB).

(74) Agent: **MEWBURN ELLIS LLP**; Aurora Building, Counterslip, Bristol Bristol BS1 6BX (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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,

(54) Title: METHOD AND APPARATUS FOR THREE-DIMENSIONAL IMAGING

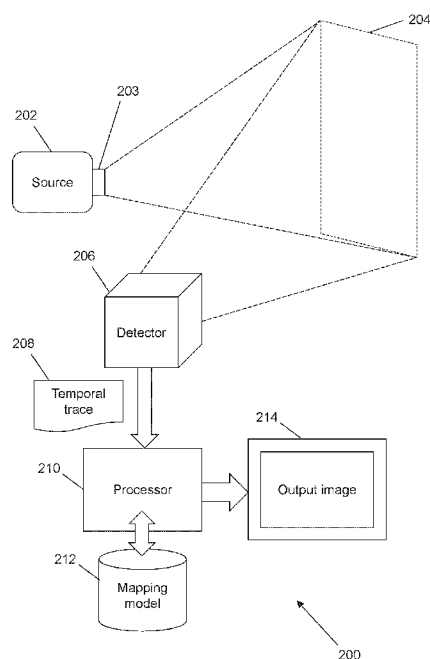


FIG. 2

(57) Abstract: An imaging apparatus configured to obtain a temporally resolvable signal from an entire scene to be imaged. Instead of scanning the scene to build a three-dimensional image, the imaging apparatus maps the temporally resolvable signal from the entire scene to three-dimensional image data using a mapping model, which may comprise a machine learning algorithm. The temporally resolvable signal can be a reflected signal emitted from the scene in response to a measurement signal pulse emitted by the imaging apparatus. The machine learning algorithm may be trained using a training set comprising temporally resolvable signals associated with corresponding three-dimensional images. The measurement signal pulse may be an acoustic signal (e.g. ultrasound), or may comprise visible light, or radiation in the super high frequency (SHF) or extremely high frequency (EHF) ranges.

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

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

METHOD AND APPARATUS FOR THREE-DIMENSIONAL IMAGING**FIELD OF THE INVENTION**

5 The invention relates to techniques for obtaining images
in which objects can be resolved in three dimensions within a
field of view. In particular, the invention relates to a
technique in which temporal resolution of a reflected signal
can be converted or mapped to spatial resolution in an output
10 image.

BACKGROUND TO THE INVENTION

15 Various techniques for obtaining three-dimensional images
are known. Some techniques actively interact with an object
or scene to be imaging in order to measure depth information.
Examples include computed tomography scanning, laser-based
rangefinders, structured light scanning, etc.

20 In such systems, the need to obtain multiple depth
measurements for any given image places a practical limit on
the rate at which any output 3D image can be refreshed. By
way of example, LIDAR (Laser Imaging Detection and Ranging) is
a range finding technology used in multiple areas such as
surveying, robotics, object recognition etc. The operational
25 principle of LIDAR is the following: pulsed laser light is
used to illuminate a scene in a controlled way (either
scanning the scene in 2D or by using multiple laser sources)
and the depth information is recovered by measuring the time
it takes for the light pulse to be reflected by every point on
30 the field of view. As a consequence, LIDAR systems are
generally bulky devices with relatively low frame rate and
scanning parts. This can make them unpractical for use in
scenarios where the detector needs to be compact or in which
the output images need to be provided rapidly.

SUMMARY OF THE INVENTION

35 At its most general, the invention provides an imaging
apparatus that operates by obtaining a temporally resolvable
40 signal from a scene to be imaged, and mapping that temporally

resolvable signal to three-dimensional image data using a suitable algorithm. The temporally resolvable signal can be a reflected signal emitted from the scene in response to a measurement signal pulse emitted by the imaging apparatus.

5 The algorithm may be a machine learning algorithm trained using a training set comprising temporally resolvable signals associated with corresponding three-dimensional images.

Herein, reference to three-dimensional images may mean digital images in which each pixel or pixel group is associated with depth information, i.e. information indicative of a distance to the object depicted by the pixel or pixel group. The output from the algorithm may be a three-dimensional image data, e.g. a 2D matrix of depth information, which may be displayed alone as a 3D image, or combined with a 2D image (e.g. captured by a separate image sensor) to provide enhanced spatial resolution.

The invention proposes a radically different approach compared with traditional 3D scanners, because it does not require scanning parts, nor arrays of sensors or structured illumination. In use, the apparatus operates to expose the whole scene to the measurement signal pulse. In an example where visible light is used to obtain the temporally resolvable signal, the measurement signal pulse may resemble a single flash across a field of view of a camera. Other types of radiation may also be used for the measurement signal pulse, e.g. centimetre wave or millimetre wave electromagnetic radiation (i.e. radiation in the super high frequency (SHF) or extremely high frequency (EHF) having a frequency in a range 3-300 GHz). Alternatively or additionally, the measurement signal pulse may be an acoustic signal, e.g. an ultrasound pulse similar to those used in car parking sensors.

According to a first aspect of the invention, there is provided a three-dimensional imaging apparatus comprising: a source configured to emit a measurement signal pulse towards a scene to be imaged; a detector configured to detect a reflected response to the measurement signal pulse from the scene, wherein the reflected response comprises a temporally resolvable signal; a processor configured to: receive the temporally resolvable signal; input the temporally resolvable signal into a computer-executable mapping model, wherein the mapping model is configured to map the temporally resolvable

signal to three-dimensional image data that represents the scene; and output the three-dimensional image data. The apparatus thus operates to convert a temporally resolvable signal, which may resemble a simple one-dimensional time series data structure, into 3D image data, e.g. a 2D array of depth values.

The apparatus of the invention may provide certain advantages compared with conventional 3D imaging systems, and in particular known LIDAR systems. Firstly, the exposure of the scene to a single pulse (e.g. using flash illumination) means that the energy delivered per area (intensity) is much smaller than for the case of laser scanning LIDAR. Secondly, the apparatus does not require scanning parts or arrays of sensors (e.g. cameras) for gathering the data, which means that the apparatus can be more compact and faster than other approaches.

The mapping model may comprise a machine learning algorithm, e.g. an artificial neural network or the like. The machine learning algorithm may be trained using a training set comprises entries that associate temporally resolvable signal with known (ground truth) 3D image data. When configured in this way, the apparatus may be capable of further improvement over time, e.g. by upgrading the mapping model based on an expanded training data set. In one example, the apparatus may include or be used in a context where there is also a 3D scanner. The apparatus may thus generate more data whilst in use, so the algorithm can be updated in situ to improve its ability to recognise features.

The temporally resolvable signal may be a data structure that associates variation in a parameter of the reflected response with time. The parameter used may depend on the nature of the reflected response. In one example, the temporally resolvable signal may be a temporal histogram of intensity of the reflected response over time. The intensity of a signal may be indicative of an amount of energy received per unit time, for example.

The detector may comprise a single-point detector. In other words, the reflected response may be detected at a single detection point. In this example, the temporally resolvable signal is a one-dimensional entity, because no spatial information is obtained by the detector. In another

example, the detector may comprise a plurality of single-point detectors, in which each of the single-point detectors obtains a respective (one-dimensional) temporally resolvable signal. The plurality of single-point detectors may be spatially separated from one another. In this example, the outputs from the array of single-point detectors can be used to distinguish between scenes that produce a similar temporally resolvable signal for a single point detector.

The detector may be any device capable of sensing the reflected response to generated a temporally resolvable signal. The type of detector may depend on the source, as discussed below.

The source may be configured to generate optical radiation, i.e. electromagnetic radiation in the visible part of the spectrum. In one example, the source may comprise a laser. It may be desirable to use a laser so that the measurement signal pulse has a well defined wavelength, which may assist detection of the reflected response.

The measurement signal pulse may thus be a light pulse. The source may include an optical system (e.g. lens or lens array) configured to flash illuminate the scene with the light pulse. Where the source generates optical radiation, the detector may be an optical detector. For example, the detector may comprise any of a single photon avalanche diode (SPAD) detector, a photodiode, an avalanche photodiode (APD) detector, or a silicon photomultiplier detector. The temporally resolvable signal may comprise a temporal histogram of photon count over time.

Thus, the apparatus may operate with a single laser system that emits light pulses, while the returned light may be focused and collected with a single pixel SPAD detector. The SPAD detector records arrival times of return photons from the whole scene. If we assume that the SPAD is placed at a position $r_0 = (x_0, y_0, z_0)$, the arrival time of an object placed at position $r = (x, y, z)$ is $t \propto |r - r_0|/c$ (where c is the speed of light). The SPAD registers all the photon events during a defined time window and the processing electronics generate a temporal histogram showing the number of photons that arrived at each time within the window. The mapping model operates to recover the 3D information of the scene from the temporal

histogram. The mapping model may provide a 2D image with intensity-coded depth of field for each new measured histogram (in a single-shot), without requiring scanning.

5 The technique disclosed herein can in principle be applied to any source of radiation/wave energy whose temporal trace can be measured. The source may thus use other types of electromagnetic radiation (e.g. similar to those used in short-range radar) or acoustic radiation.

10 The source may thus comprise an electromagnetic radiation generator. The measurement signal pulse may be an electromagnetic radiation pulse in the SHF or EHF band, e.g. having a frequency in a range from 1 GHz to 300 GHz.

15 Alternatively, the source may comprise an acoustic transducer, whereby the measurement signal pulse is an acoustic pulse. For example, the measurement signal pulse may be an ultrasonic pulse, and the detector may comprise an ultrasonic sensor.

20 In these examples, the same device may be used to emit the measurement signal pulse and detect the reflected response. The apparatus may thus comprise a signal transceiver operable as both the source and the detector. This assists in making the apparatus compact.

25 The apparatus may comprise a display screen. The processor may be further configured to display the three-dimensional image data on the display screen.

Alternatively or additionally, the apparatus may comprise a communication module. The processor may be further configured to transmit, using the communication module, the three-dimensional image data to a remote device.

30 In another aspect, there is provided a method of generating three-dimensional image data, the method comprising: emitting a measurement signal pulse towards a scene to be imaged; detecting a reflected response to the measurement signal pulse from the scene, wherein the reflected response comprises a temporally resolvable signal; receiving, by a processor, the temporally resolvable signal; inputting, by the processor, the temporally resolvable signal into a computer-executable mapping model; mapping, by a mapping model executed by the processor, the temporally resolvable signal to
40 three-dimensional image data that represents the scene; and outputting the three-dimensional image data. This aspect may

have any of the features discussed above with referred to the three-dimensional imaging apparatus.

In a further aspect, there is provided method of training a mapping model for generating three-dimensional image data, the method comprising: (a) emitting a measurement signal pulse towards a scene to be imaged; (b) detecting a reflected response to the measurement signal pulse from the scene, wherein the reflected response comprises a temporally resolvable signal; (c) measuring, by a depth sensor, three-dimensional image data for the scene; (d) associating the temporally resolvable signal with the three-dimensional image data to create an entry in a training data set; (e) repeating steps (a) to (d) to generate a training data set comprising a plurality of entries; and (f) training an artificial neural network using the training data set to generate a mapping model configured to map an input temporally resolvable signal to three-dimensional image data. The depth sensor may be any suitable 3D scanning or imaging apparatus.

The method may include adding simulated data to the training set. For example, the method may further comprise: generating a virtual three-dimensional scene; determining a simulated temporally resolvable signal at a detection point by simulating a reflected response from the virtual three-dimensional scene at the detection point to a measurement signal; calculating three-dimensional image data for the virtual three-dimensional scene at the detection point; and associating the simulated temporally resolvable signal with the calculated three-dimensional image data to create an entry in the training data set.

In a further aspect, there is provided a computer-readable medium having computer-readable instructions stored thereon, which, when executed by a processor, cause the processor to: receive a temporally resolvable signal indicative of a reflected response to a measurement signal pulse from a scene; map the temporally resolvable signal to three-dimensional image data that represents the scene; and output the three-dimensional image data. This aspect may have any of the features of the mapping model discussed above.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention are discussed in detail below, with reference to the accompanying drawings, in which:

Fig. 1 is a schematic illustration of an image generation process that underlies the present invention;

Fig. 2 is a schematic diagram of a three-dimensional imaging apparatus that is an embodiment of the invention;

Fig. 3 is a schematic diagram of an imaging apparatus for gathering training data for an mapping algorithm suitable for use with the invention; and

Fig. 4 illustrate experimental results obtaining by applying the principles of the invention.

DETAILED DESCRIPTION; FURTHER OPTIONS AND PREFERENCES

Overview

The principle of operation of the imaging apparatus described herein arises from the recognition that it is possible to map a temporally resolvable signal that is indicative of a reflected response from a scene into three-dimensional image data for that scene. This approach is based on a realisation that every object in the scene has a different reflectivity (depending on its size and material composition) and different arrival time for the measurement signal pulse. This means that the time and magnitude of the reflection back from each point in the scene varies such that the temporally resolvable signal can be expressed as a histogram peak with certain amplitude and peak position. The amplitude (height) of the peak is related to the object reflectivity and distance from the laser/detector. The peak position (along the time axis) is related to the object distance from the laser/detector.

The temporally resolvable signal is provided by a detector that is responsive to the reflected response. The type of detector is selected depending on the nature of the reflected response. For example, the measurement signal pulse may use a laser, whereby the scene is flood illuminated with a laser pulse and the reflected light is focused and collected by an optical detector. In one example, the optical detector may be a single pixel SPAD (single photon avalanche diode)

detector. The SPAD detector can record arrival times of return photons from the whole scene. A time-correlated single-photon counter may be used with the SPAD detector to generate the temporally resolvable signal, which in this example may be a temporal histogram of received photon counts over time.

In other examples, the measurement signal pulse may be electromagnetic radiation or an acoustic signal, and the detector may be an antenna, transceiver or transducer arranged to receive the reflected response and record a variation in intensity of the response with time.

Fig. 1 is a schematic illustration of an image generation process 100 that underlies the present invention. The process begins with obtaining the temporally resolvable signal, which in this example is a histogram 102 indicating a count of received photon over time. The temporally resolvable signal is provided as an input to an algorithm 104 that maps the input to three-dimensional image data that can be displayed as a three-dimensional image 106 in any suitable manner. In the three-dimensional image 106 different colours or shades of grey are used to indicate the depth information.

The algorithm may take any suitable form, as discussed in more detail below. In one example, the process uses a machine learning algorithm that is pre-trained to reconstruct a 3D image of scene. For example, the machine learning algorithm may be multilayer perceptron (MLP), which is a class of feed-forward artificial neural network (ANN) where the all the elements of the input layer are connected to all the elements of the output layer through a fully-connected layer. In this example, the input layer corresponds to the temporally resolvable signal (e.g. histogram or intensity trace), while the output layer is a 2D image with intensity-encoded depth (i.e. a 2D digital image in which the intensity of each pixel or pixel group is indicative of depth).

The training set for the machine learning algorithm may be assembled in any suitable way. The training set comprises 3D image information associated with corresponding temporally resolvable signals. This information can be obtained through real measurement of 3D image data for scenes that are exposed to the measurement signal pulse, e.g. at the same time as measuring a temporally resolvable signal. The 3D image data may be measured by any suitable device for obtaining spatially

resolved depth measurements. For example, a time-of-flight camera may be used to record a series of 3D scenes together with their corresponding temporally resolvable signals. In another example, a conventional LIDAR system can be used to obtain depth information from a scene, which can be associated with a corresponding temporally resolvable signal.

The success of the proposed approach is surprising, since it would not normally be possible to obtain 3D information (spatially resolved depth information) from a 1D input (variation of photon count or reflected signal intensity with time). Objects having the same reflectivity (thus peak amplitude on the temporally resolvable signal) located at positions within the scene from which the reflected signal has the same arrival time at the detector would be indistinguishable in the temporally resolvable signal. However, the inventors have realised that, in reality, the risk of features in a real world scene possessing the kind of symmetry that would yield indistinguishable temporally resolvable signal is very low. This risk diminishes as the temporal resolution of the temporally resolvable signal increases, because the process increasingly capable of distinguishing between subtle differences in the signals from slightly different objects. Moreover, since the machine learning algorithm can be trained on real-life scenarios, which means that it naturally evolves to match temporally-resolvable signals to correspond real life scenes. This can be distinguished from (and is clearly advantageous over) a scheme in which an attempt was made to reconstruct one or more hypothetical 3D scenes from the temporally resolvable information.

3D imaging apparatus

Fig. 2 is a schematic diagram of a three-dimensional imaging apparatus 200 that is an embodiment of the invention. The apparatus 200 comprises a source 202 configured to deliver a measurement signal pulse onto a scene 204 containing objects for which relative depth information is wanted. The source 202 is selected depending on the nature of the measurement signal pulse. As discussed above, the measurement signal pulse may comprise any type of energy that can be reflected by

objects in the scene 204 such that a reflected response from the scene can be captured by a detector 206. The detector 206 is configured such that the scene 204 is located in its detection zone.

5 The measurement signal pulse is short burst of energy having an amplitude selected to ensure that the reflected response can be distinguished from any background signal that may be recorded by the detector 206. The detector 206 may be arranged to filter out the background signal using any
10 conventional technique. The duration of the pulse may be equal to or less than 1 ns, preferably less than 100 ps, e.g. 20 ps or the like. Having a very short pulse duration reduces the overlap between reflected responses from different points in the scene. When paired with a detector capable of high
15 temporal resolution, this improves the ability of apparatus to distinguish between similar scenes.

 The source may emit a series of pulses, e.g. in a periodic manner. The separate between the pulses is selected to avoid interference or overlap between received reflected
20 responses. Each pulse may be processed to provide 3D image information. The apparatus may thus be capable of outputting a stream of 3D image information. As the mapping from the reflected response (temporally resolvable signal) can be processed rapidly, the apparatus effectively outputs a real-
25 time stream of 3D image information. Tracking algorithms may also be implemented to improve image reconstruction based on previously acquired images.

 The detector 206 is configured to measure change in intensity of the whole reflected response with time. The
30 detector is therefore arranged to capture a reflected response from the whole scene 204, but need not attempt to spatially resolve that response. The output from the detector 206 is a temporally resolvable signal (also referred to herein as a temporal trace 208) that encodes the variation of intensity
35 with time. In this context, the term "intensity" is intended to be indicative of the energy in the received response, which may have different parameters or units depending on the type of measurement signal pulse used. For example, for visible light, the "intensity" may be measured as the rate of received
40 photons. For acoustic radiation or electromagnetic radiation,

the "intensity" may be amplitude of the detected reflected response.

In one example, the source includes a light source such as a laser or LED configured to output a pulse of radiation, e.g. in the visible or infrared parts of the spectrum. To provide a desirable level of temporal resolution in the reflected response, the pulse preferably has a duration of around 1 ns. An optical system 203 may be mounted at the output of the laser to disperse the output beam so that it illuminates the whole scene 204. In this example, the detector 206 comprises an optical sensor configured to detect a reflected response from the scene 204 that corresponds to each measurement signal pulse. As explained above, the optical sensor may be any suitable device for producing an output signal that is indicative of the change in intensity of a received response corresponding to the measurement signal pulse. In one example, the detector comprises a lens for collecting reflected light from the scene 204 and directing it on to a single photon avalanche diode (SPAD) sensor. The output from the sensor may be coupled to a time-correlated single photon counting (TSCPC) module to output a temporal trace that is indicative of change in received photon count with time.

In another example, the source includes an ultrasonic transducer configured to emit an ultrasonic pulse across the scene 204. The ultrasonic pulse preferably has a pulse duration selected in conjunction with the detector sensitivity to obtain a temporally resolvable signal that provides a temporal resolution corresponding to a desirable level of spatial resolution, e.g. to permit objects having a size in the range 10-30 cm to be resolvable, for example. In this example, the detector comprises an ultrasonic sensor configured to measure acoustic energy reflected from the scene 204. The output of the detector in this example may be a temporal trace 208 indicative of variation in an amplitude of the detected acoustic signal over time.

In a further example, the source may comprise an electromagnetic radiation generator configured to generate and emit a pulse of electromagnetic radiation. As mentioned above, the electromagnetic radiation may be in the SHF or EHF bands of the spectrum. The electromagnetic radiation

generator may include an antenna configured to direct the emitted pulse towards the scene. In this example, the detector may comprise an antenna configured to detect radiation at the frequency of the measurement signal pulse.

5 The antenna may be a directional antenna configured to receive signals primarily from the scene 204. The detector 206 may include a filter configured to remove a background signal from the output of the antenna. The output of the detector in this example may be a temporal trace 208 indicative of variation in
10 detected power of the reflected electromagnetic signal over time.

The apparatus includes a processor 210 configured to receive temporal trace 208. The processor 210 is part of a computing device (not shown) that includes memory and software
15 instructions that, when executed by the processor, enable the apparatus to perform the functions discussed herein. The processor 210 has access to a mapping model 212, which is an example of the algorithm discussed below, that can translate or transform the temporal trace 208 into 3D image information.
20 The processor 210 inputs the temporal trace 208 to the mapping model 212 to obtain the 3D image information. The processor 210 can use the obtained 3D image information to generate and cause display of output image 214.

As discussed above, the output image 214 may be a
25 representation of the 3D image information itself, e.g. as a 2D image in which each pixel or pixel group has an intensity (e.g. colour) that encodes depth information. Alternatively, the output image 214 may be a combination of the 3D image information from the mapping model 212 with a 2D image of the
30 scene 214 captured by a separate camera (not shown).

The source 202, detector 204 and processor 210 may be provided as a unit, e.g. within a common housing. The unit may include a display for showing the output image. Alternatively or additionally, the unit may include a
35 communication module capable of transmitting the 3D image information (or the output image) to another device in order for it to be displayed. The communication module may be arranged to communicate wirelessly over a suitable network.

40 *Gathering data for training*

Fig. 3 is a schematic diagram of an imaging apparatus 300 for gathering training data for an algorithm (mapping model) suitable for use with the invention.

The apparatus 300 comprises a source 202 configured to deliver a measurement signal pulse onto a scene 304 containing objects for which relative depth information is wanted. The source 302 is selected depending on the nature of the measurement signal pulse. As discussed above, the measurement signal pulse may comprise any type of energy that can be reflected by objects in the scene 304 such that a reflected response from the scene can be captured by a detector 306. The detector 306 is configured such that the scene 304 is located in its detection zone. The source 302 and detector 306 may be configured to output a temporal trace 308 in the same way as discussed above with respect to the apparatus 200 in Fig. 2. In the imaging apparatus 300 for gathering training data, the temporal trace formed part of a training data set 320.

The apparatus 300 further comprises a depth sensor 310 that is configured to measure 3D image information from the scene 304. In particular, the depth sensor 310 is configured to detect a distance to each spatially resolvable object in the scene 304. The depth sensor 310 may be any suitable 3D imaging device, e.g. 3D scanner, time-of-flight camera, LIDAR system, etc. The output of the depth sensor 310 is 3D image data 312 that also forms part of the training data set 320. The depth sensor 310 and detector 306 are synchronised so that each received temporal trace 308 can be associated with corresponding 3D image data 312 from the depth sensor 310. The 3D image data 312 thus represents "ground truth" data for training the algorithm.

The training data set 320 is provided to a model generator 322, which may be a suitable computing device configured to train a machine learning algorithm to map an input temporal trace to 3D image data. The algorithm is discussed in more detail below.

The training data set 320 may consist entirely or actual measurement obtained by the depth sensor 310 and detector 306 for a variety of different scenes 304.

However, in a development of the apparatus 300, the training data set 320 may also include simulated data. The

simulated data may comprise a simulated 3D scenes 314 that can be generated by any suitable computer modelling software. As the scenes are simulated, the depth information from any given point of view can be calculated in a straightforward manner.

5 The apparatus 300 may then include a reflection model 316, which receives each simulated 3D scene as an input. The reflection model 316 is configured to simulate a reflected response of the 3D scene to an incident measurement single pulse (e.g. an incident spherical wave from a given point
10 source) measured at a certain detection point. The simulation may be based on assumptions in relation to the reflectivity of various objects in the scene. This assumptions may be made based on known information about the reflectivity of real world objects in order to make the simulated reflected
15 response as realistic as possible. The output of the reflection model 316 is thus a simulated temporal trace 318 corresponding to the combination of the given point source and detection point location. The simulated temporal trace 318 can be associated with calculated depth information for the
20 detection point to provide simulated training data for the training data set 320.

Having simulated training data may permit the training data set 320 to be rapidly expanded because it provides an efficient way of obtaining temporal traces from multiple
25 different detection points for multiple different scenes.

In the discussion of the apparatus 200, 300 above, the detector 206, 306 provides a single detection point for the temporal trace. However, in order to further mitigate the risk of certain scenes being difficult to distinguish because
30 they result in similar temporal traces, it is possible for the apparatus to include multiple detectors. For example, higher confidence can be obtained by using two spatially-separated detectors. In this case, for all points of the same spherical cap, the temporal trace will be different, and it is therefore
35 possible to determine unambiguously which object-position corresponds to peaks in the temporal trace. It may be noted, that even under this configuration there exists a set of surfaces for which the reflected response at both detectors is the same, but this set is expected to be smaller and consist
40 of much more complicated shape than the rotational symmetry limitation of a single detector. Ultimately, an array of

time-resolving detectors (such as a SPAD camera, for example) would substantially negate this risk. However, this is not believed to be necessary in practical operating conditions/scenarios. That is, the invention is advantageous because it does not require multiple detectors (or some kind of scanning detection) in order to obtain the 3D information.

In another arrangement, it is possible to use a plurality of different sources, each of which flash-illuminate the scene at different locations, in conjunction with a single detector for measurement respective temporal traces. The plurality of sources preferably have the same modality so that the same detection mechanism is used. For example, the plurality of sources may be light sources whose output light has differing wavelengths. The measurement signal pulses from the plurality of sources are emitted in a non-overlapping sequence, so that the reflected response from objects in the scene have different arrival times at the detector.

Algorithm (Mapping model)

The underlying idea of the invention is that every object in the scene has a different reflectivity (depending on its size and material composition) and different arrival time for the reflected response, e.g. photons emitted by the laser and reflected back from the object to the detector. Thus, each object in the scene manifests as a peak in the temporal trace with certain amplitude and temporal position. The amplitude (height) of the peak is related to the object reflectivity and distance from the source/detector. The peak position (along the time axis) is related to the object distance from the source/detector.

Although in principle there are an infinite number of objects with different combinations of dimensions/material that could lead to the same reflectivity (thus, same peak amplitude in our temporal traces), in reality the type of objects that appear in a practical scenario are very restricted. For instance, using the proof-of-principle experiment discussed below, it has been shown to be possible to distinguish between scenes composed by multiple people, also with different clothing, and even including completely

different objects, such as pieces of cardboard moving around the people.

Following the discussion above, our approach works with any data-driven algorithm that is able to associate a 3D image of the illuminated scene to the temporal trace measured by the time-resolving detector. In other words, the problem to solve by the algorithm is: what is the mathematical operation $\mathbf{M}$ performed on the measured histogram $\mathbf{H}$ that provides the scene $\mathbf{S}$? i.e. $\mathbf{M}(\mathbf{H}) = \mathbf{S}$.

In this context, the machine learning approach using a multilayer perceptron (MLP) discussed above is just one example. Indeed, it may be possible to implement the invention using a model that is not based on machine learning. This alternative approach (referred to as inverse retrieval) is discussed further below.

In one example, the algorithm comprises an artificial neural network (ANN) consisting of a single layer that fully connects the temporal traces with their corresponding 3D scene through a matrix operation. A mapping model corresponding to such an algorithm can be expressed as $\mathbf{M} \cdot \mathbf{H} = \mathbf{S}$, which is the simplest approach to solving the problem. Given that the temporal traces may contain thousands of resolvable time bins and that the scenes to be retrieved may contain thousands of pixels, the equation $\mathbf{M} \cdot \mathbf{H} = \mathbf{S}$ becomes a system of linear equations, where $\mathbf{M}$ is a matrix with millions of unknown parameters. To solve the equation, we first obtain thousands of temporal trace-3D image data pairs that we feed to our ANN during the training process. This process can be thought of simply as a way of retrieving the elements of $\mathbf{M}$. After the ANN is trained (i.e. after the weights/elements of the matrix $\mathbf{M}$ are obtained), retrieving the 3D scene from a temporal trace can be done at very high speeds by just performing a matrix operation.

The high speed matrix operation approach is advantageous in application in which it is desirable to rapidly update the 3D image information, e.g. to effectively provide a 3D video.

It should be noted that the matrix operation discussed above is only one way of implementing the invention. Other implementations are possible using more complex machine learning algorithms that can represent the mapping $\mathbf{M}(\mathbf{H}) = \mathbf{S}$.

Those skilled in the art are aware of a vast number of different algorithms that are applicable to the approach set out herein, where the aim is to retrieve 3D spatial data from a single 1D temporally resolvable signal (e.g. time histogram).

As mentioned above, as an alternative to implementing the mapping model using a machine learning algorithm, an inverse retrieval methodology can be used. In this approach, a simulation model may be derived that to simulate reflections from a given scene. The simulation model can produce a simulated temporal trace for a (predicted) candidate scene. The inverse retrieval technique comprises an iterative process in which the candidate scene is adjusted to minimise differences between the simulated temporal trace and a measured temporal trace. In other words, a simulated temporal trace may be expressed as $A(S) = H'$, where mathematical operation A acts on a predicted scene S to produce a simulated temporal trace H' . From an initial guess for the predicted scene S , a numerical simulation can be performed using A to produce the simulated temporal trace H' , which can be compared with a measured temporal trace H . This is done for example by calculating the difference magnitude $\|H' - H\|^2$. The true scene can be obtained by minimising the difference magnitude. The iterative adjustment to the predicted scene S can be performed in any conventional manner.

An advantage of the inverse retrieval approach is that usually no prior knowledge of any kind is required of the object. This can be contrasted with the machine learning technique, in which learning is based on a set of examples that inevitably create a preferred output from the retrieval.

However, the inverse retrieval approach may also require a regularisation function that needs to be added to the difference magnitude in order to ensure that it converges to a minimum. The inverse retrieval method is suitable for apparatus that is configured to obtain multiple temporal traces of the same scene from different detection points in order to discriminant between candidate images that share circular symmetry.

Experimental proof of principle

An apparatus having a configuration similar to that shown in Fig. 3 has been tested in in laboratory conditions. The apparatus operates to flash-illuminate a scene using a
5 supercontinuum fiber pulsed laser (20 ps) with 78 MHz pulse repetition rate (NKT SuperK Extreme) with a narrow filter at 532 nm. The detector is a single pixel detector consisting of a fiber-coupled SPAD sensor (Excelitas) that, together with a TCSPC electronics card (Becker&Hickl), yields 350 ps temporal
10 resolution, which corresponds to a spatial resolution of 10 cm. A time-of-flight (ToF) camera is used as a depth sensor to gather 3D images to train a machine learning algorithm. In this example, the ToF camera was a PMD CardBoard Pico Flexx, offering a depth range up to 4 m and an image resolution of
15 224 × 171 pixels. In this experiment, the obtained images were resized to 24 × 24 pixels.

The mapping model in this experiment consists of a MLP algorithm, implemented as a single-fully-connected layer using Keras on TensorFlow. Training the MLP algorithm involved
20 feeding it with 10,000 temporal trace-3D image data pairs, all corresponding to a scene involving people moving in random positions within an area of roughly 22 m².

Fig. 4 shows some examples of experimental results 400 obtained with the apparatus described above.

25 The first row 402 comprises three experimentally recorded temporal histograms that were not used during the training process of the machine learning algorithm.

The second row 404 shows a 3D image of the scene obtained from the mapping model for each of the temporal histograms
30 through a direct matrix multiplication.

For comparison, the third row 406 depicts the corresponding ground-truth 3D images obtained with the ToF camera. As can be seen, the mapping model is able to distinguish the general features of the scene/person and its
35 distance from the sensor. It is interesting to note that the bright area on the person's head (due to high reflection from glasses that 'fools' the ToF camera) was included in the training set for the mapping model, and therefore also appears in the output from the model.

40 The proof-of-principle results demonstrate the possibility of achieving a scanless and compact 3D imaging

apparatus. The apparatus can operate using a single pixel detector (e.g. SPAD detector). The apparatus can produce 3D images in real time, thereby opening the door to new approaches for high speed 3D pattern recognition.

5 The proof-of-principle tests used a time-of-flight camera to obtain depth images. The camera used above was restricted in range (up to 4 metres) and image resolution (224×171 pixels). However, as explained above, the training data set can be assembled using any suitable spatially resolving
10 sensor. This may be an enhanced time-of-flight camera, e.g. with 13 m range at 640×480 pixels, or a commercial LIDAR system.

Applications

15 The apparatus disclosed herein may find application in a variety of fields, such as autonomous-driving vehicles, machine vision, robotics, etc. The apparatus may find use as a replacement for or alternative to LIDAR in fields such as
20 surveying, robotics, or object recognition. However, the ability of the apparatus disclosed herein to return 3D image information in real time makes it particularly suitable in cases where rapid scene recognition is required. The recent escalating demand for driverless cars means that LIDAR is now
25 a key technology in the automobile industry. The apparatus disclosed herein represents an important development within that field, as it can be used to provide a fast, reliable, yet compact LIDAR-like system.

CLAIMS

1. A three-dimensional imaging apparatus comprising:
a source configured to emit a measurement signal pulse
5 towards a scene to be imaged;

a detector configured to detect a reflected response to
the measurement signal pulse from the scene, wherein the
reflected response comprises a temporally resolvable signal;

a processor configured to:

10 receive the temporally resolvable signal;

input the temporally resolvable signal into a
computer-executable mapping model, wherein the mapping model
is configured to map the temporally resolvable signal to
three-dimensional image data that represents the scene; and

15 output the three-dimensional image data.

2. The three-dimensional imaging apparatus of claim 1,
wherein the mapping model comprising a machine learning
algorithm.

3. The three-dimensional imaging apparatus of claim 1
or 2, wherein the temporally resolvable signal is a data
structure that associates variation in a parameter of the
reflected response with time.

4. The three-dimensional imaging apparatus of claim 3,
wherein the temporally resolvable signal is a temporal
histogram of intensity of the reflected response over time.

5. The three-dimensional imaging apparatus of any
preceding claim, wherein the detector comprises a single-point
detector.

6. The three-dimensional imaging apparatus of any one
35 of claims 1 to 4, wherein the detector comprises a plurality
of single-point detectors.

7. The three-dimensional imaging apparatus of any
preceding claim, wherein the source is configured to generate
40 optical radiation.

8. The three-dimensional imaging apparatus of claim 7, wherein the source comprises a laser.

5 9. The three-dimensional imaging apparatus of claim 7 or 8, wherein the measurement signal pulse is a light pulse, and wherein the source includes an optical system configured to flash illuminate the scene with the light pulse.

10 10. The three-dimensional imaging apparatus of any one of claims 7 to 9, wherein the detector comprises a single photon avalanche diode detector, and wherein the temporally resolvable signal comprises a temporal histogram of photon count over time.

15 11. The three-dimensional imaging apparatus of any one of claims 1 to 6, wherein the source comprises an electromagnetic radiation generator.

20 12. The three-dimensional imaging apparatus of claim 11, wherein the measurement signal pulse is an electromagnetic radiation pulse in the SHF or EHF band.

25 13. The three-dimensional imaging apparatus of any one of claims 1 to 6, wherein the source comprises an acoustic transducer, whereby the measurement signal pulse is an acoustic pulse.

30 14. The three-dimensional imaging apparatus of claim 13, wherein the measurement signal pulse is an ultrasonic pulse, and wherein the detector comprises an ultrasonic sensor.

35 15. The three-dimensional imaging apparatus of any one of claims 11 to 14 comprising a signal transceiver operable as both the source and the detector.

40 16. The three-dimensional imaging apparatus of any preceding claim further comprising a display screen, wherein the processor is further configured to display the three-dimensional image data on the display screen.

17. The three-dimensional imaging apparatus of any preceding claim further comprising a communication module, wherein the processor is further configured to transmit, using the communication module, the three-dimensional image data to a remote device.

18. A method of generating three-dimensional image data, the method comprising:

emitting a measurement signal pulse towards a scene to be imaged;

detecting a reflected response to the measurement signal pulse from the scene, wherein the reflected response comprises a temporally resolvable signal;

receiving, by a processor, the temporally resolvable signal;

inputting, by the processor, the temporally resolvable signal into a computer-executable mapping model;

mapping, by a mapping model executed by the processor, the temporally resolvable signal to three-dimensional image data that represents the scene; and

outputting the three-dimensional image data.

19. A method of training a mapping model for generating three-dimensional image data, the method comprising:

(a) emitting a measurement signal pulse towards a scene to be imaged;

(b) detecting a reflected response to the measurement signal pulse from the scene, wherein the reflected response comprises a temporally resolvable signal;

(c) measuring, by a depth sensor, three-dimensional image data for the scene;

(d) associating the temporally resolvable signal with the three-dimensional image data to create an entry in a training data set;

(e) repeating steps (a) to (d) to generate a training data set comprising a plurality of entries; and

(f) training an artificial neural network using the training data set to generate a mapping model configured to map an input temporally resolvable signal to three-dimensional image data.

20. The method of claim 19 further comprising:
generating a virtual three-dimensional scene;
determining a simulated temporally resolvable signal at a
detection point by simulating a reflected response from the
virtual three-dimensional scene at the detection point to a
5 measurement signal;
calculating three-dimensional image data for the virtual
three-dimensional scene at the detection point; and
associating the simulated temporally resolvable signal
10 with the calculated three-dimensional image data to create an
entry in the training data set.

21. A computer-readable medium having computer-readable
instructions stored thereon, which, when executed by a
15 processor, cause the processor to:
receive a temporally resolvable signal indicative of a
reflected response to a measurement signal pulse from a scene;
map the temporally resolvable signal to three-dimensional
image data that represents the scene; and
20 output the three-dimensional image data.

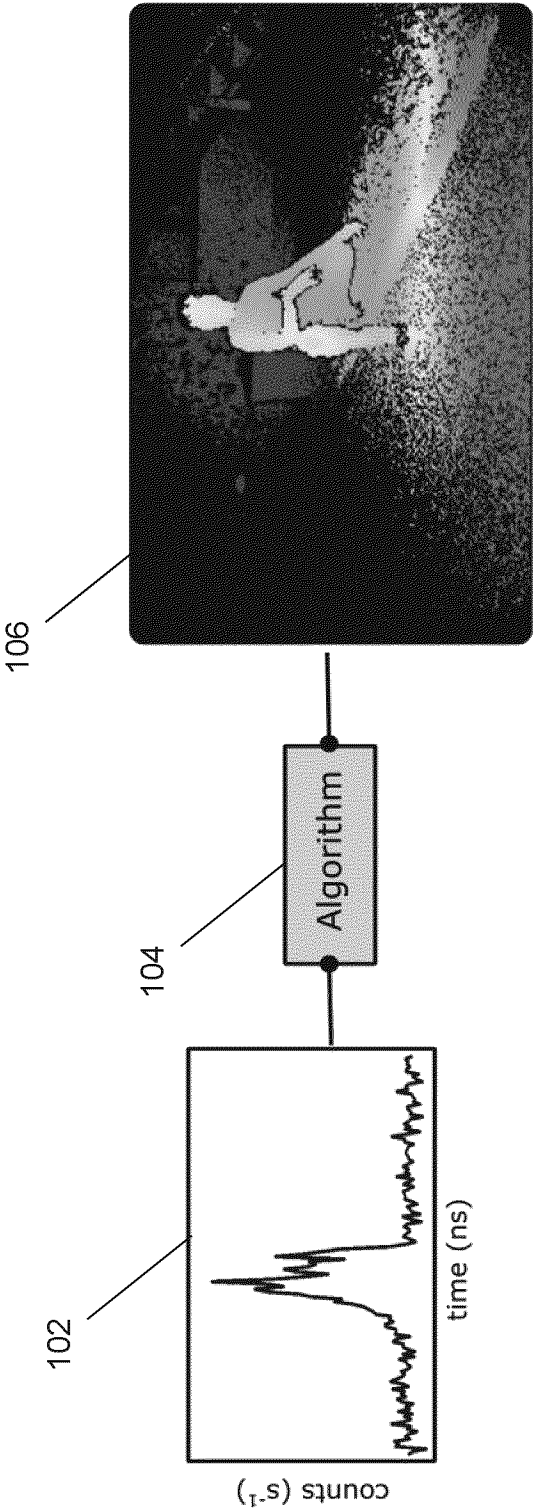
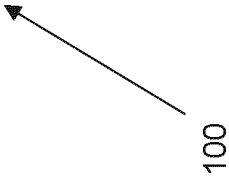


FIG. 1



100

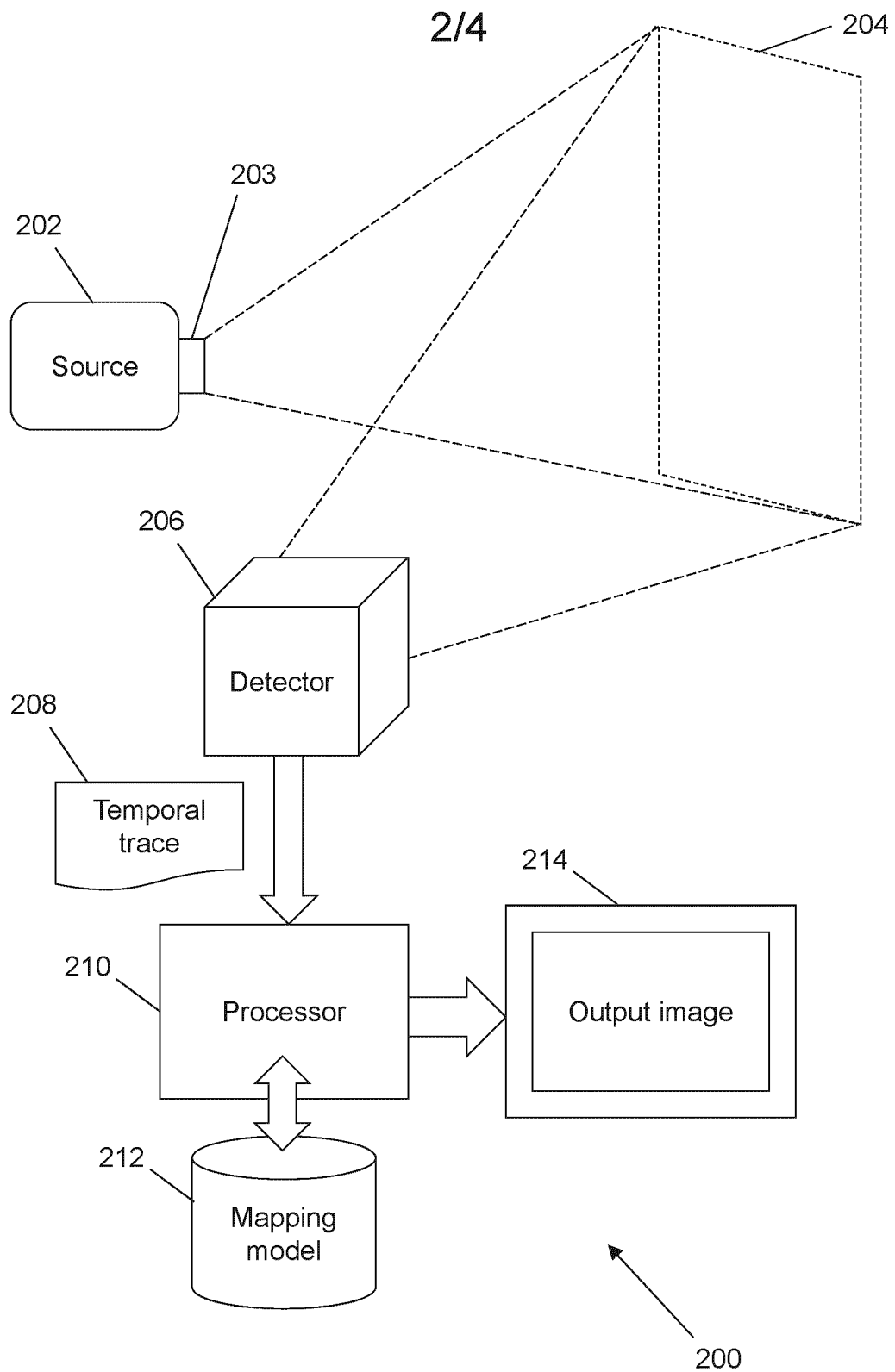


FIG. 2

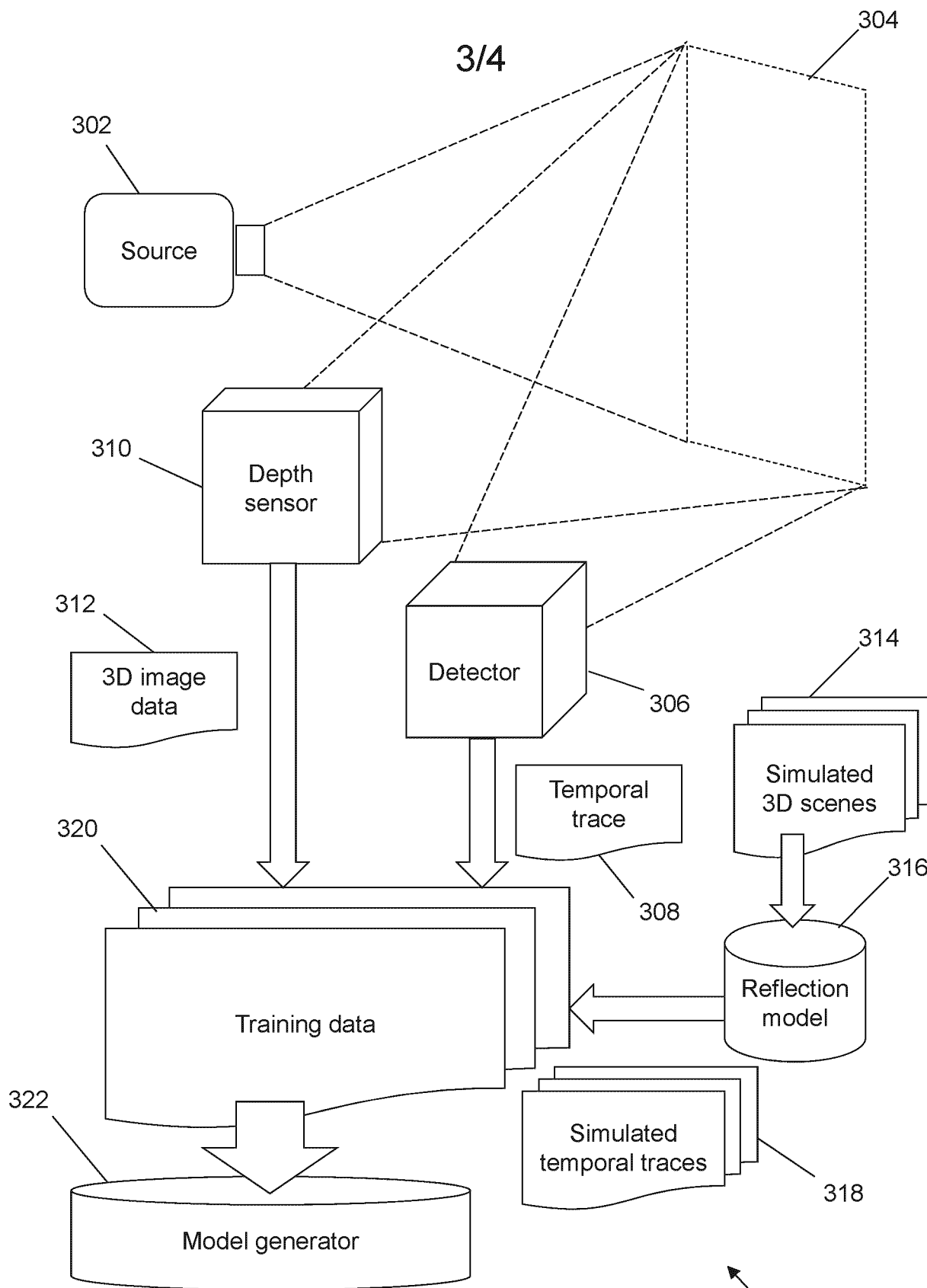


FIG. 3

300

4/4

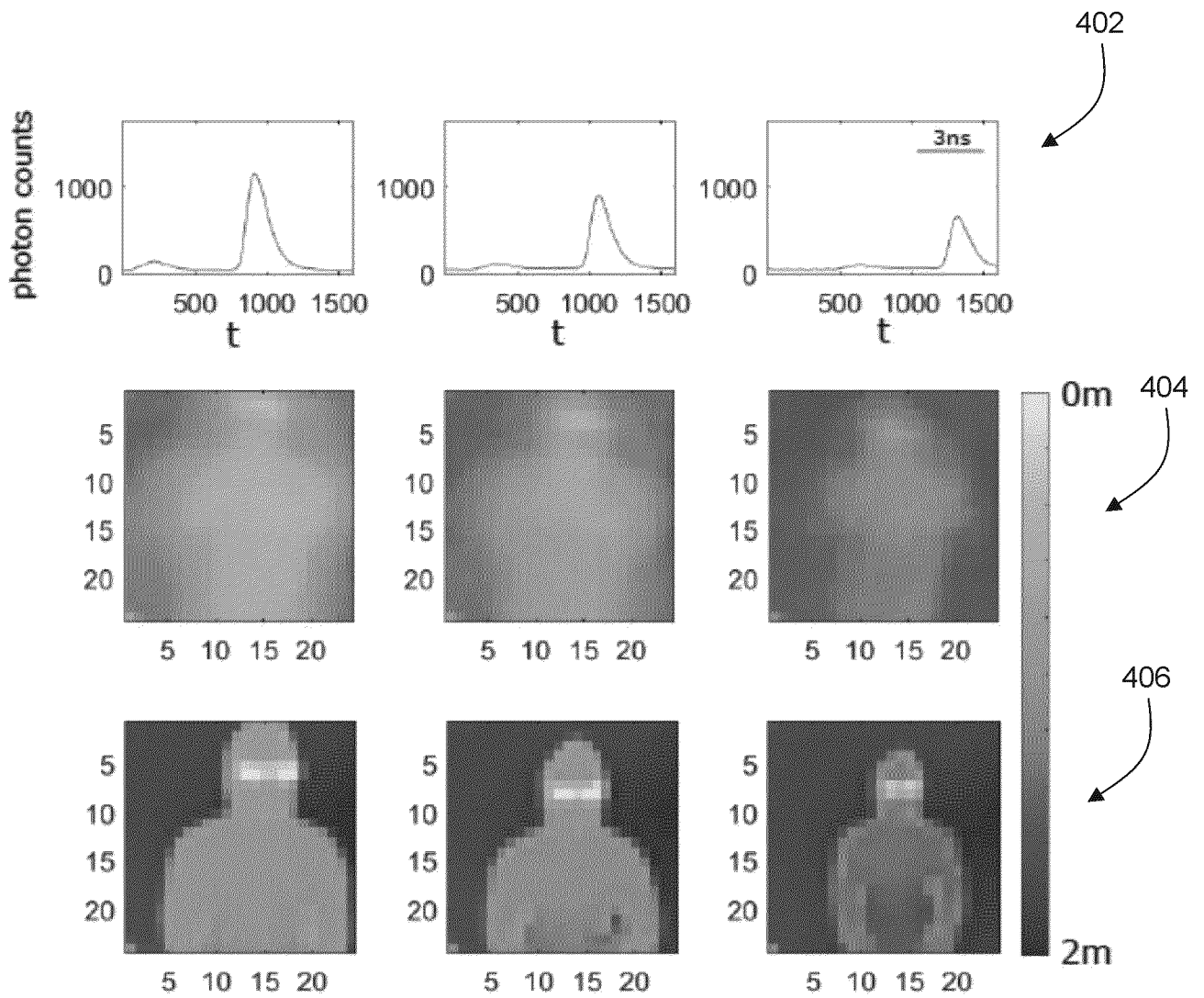


FIG. 4

400

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/063682

A. CLASSIFICATION OF SUBJECT MATTER

INV. G01S7/41 G01S13/89 G01S15/89 G01S17/10 G01S17/89
G06N3/02

ADD.

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)

G01S G06N

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2014/028834 A1 (GOODMAN VERNON R [US]) 30 January 2014 (2014-01-30) paragraphs [0017] - [0029], [0042] - [0043]; figures 1-3 -----	1,5-9, 16-18,21
X	US 2019/004535 A1 (HUANG YU [US] ET AL) 3 January 2019 (2019-01-03) paragraphs [0042] - [0043], [0073] - [0075], [0088]; figures 1-3,5,6 -----	1,2,5-9, 16-21
X	US 10 106 153 B1 (XIAO XINHUA [US] ET AL) 23 October 2018 (2018-10-23) Columns 6, 11-14 and 24; figures 1,2 -----	1,2,5-9, 16-21



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

21 July 2020

Date of mailing of the international search report

18/09/2020

Name and mailing address of the ISA/

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

Authorized officer

TheiBing, Nikolaus

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2020/063682

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.:
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:
3. ☐ Claims Nos.:
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:

see additional sheet

1. ☐ As all required additional 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 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.:

1, 2, 5-9, 16-21

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.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1, 2, 5-9, 16-21

Alternative mapping model

2. claims: 3, 4, 10

Alternative signal

3. claims: 11-15

Alternative type of sensor

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2020/063682

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014028834	A1	30-01-2014	NONE
US 2019004535	A1	03-01-2019	NONE
US 10106153	B1	23-10-2018	US 10106153 B1 23-10-2018
			US 10239521 B1 26-03-2019
			US 2019291720 A1 26-09-2019
			WO 2019182621 A1 26-09-2019