

(51) International Patent Classification:
G01N 21/62 (2006.01)(21) International Application Number:
PCT/US2014/050513(22) International Filing Date:
11 August 2014 (11.08.2014)

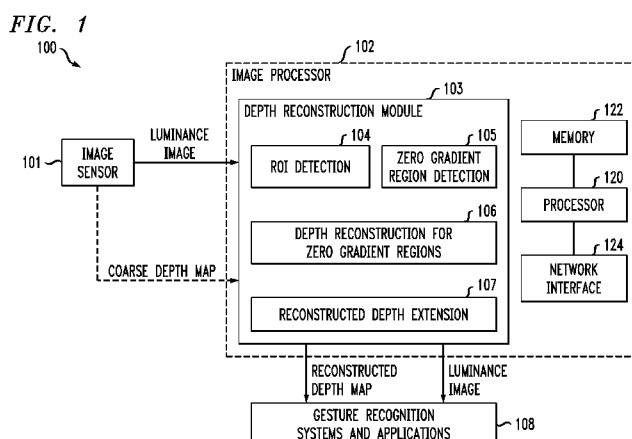
(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2014104445 7 February 2014 (07.02.2014) RU(71) Applicant: LSI CORPORATION [US/US]; 1320 Ridder
Park Drive, San Jose, CA 95131 (US).(72) Inventors: MAZURENKO, Ivan, L.; 36A Molo-
dyezhanaya Street, Apt. 51, Khimki, Moscow 141407
(RU). RADOVANOVIC, Nikola; 4388 El Camino Real,
Unit 69, Los Altos, CA 94022 (US). PARKHOMENKO,
Denis, V.; 5-1 2nd Schelkovsky Street, Apt. 40, Mytyschy,
Moscow, 141007 (RU). KHOLODENKO, Alexander, B.;
35-3 Obrucheva Street, Apt. 16, Moscow 117246 (RU).
PARFENOV, Denis, V.; 52-1 Chertanovskaya Street, Apt.
39, Moscow 117534 (RU).(74) Agent: RYAN, Joseph, B.; Ryan, Mason & Lewis, LLP,
48 South Service Road, Suite 100, Melville, NY 11747
(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, 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, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, 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, 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))

(54) Title: DEPTH IMAGE GENERATION UTILIZING DEPTH INFORMATION RECONSTRUCTED FROM AN AMPLITUDE
IMAGE

(57) Abstract: An image processing system comprises an image processor having image processing circuitry and an associated memory. The image processor is configured to identify a region of interest in an amplitude image, to detect one or more relatively low gradient regions in the region of interest, to reconstruct depth information for said one or more relatively low gradient regions, to extend the reconstructed depth information beyond said one or more relatively low gradient regions to additional pixels of the region of interest, and to generate a depth image utilizing at least portions of the reconstructed depth information and the extended reconstructed depth information. The image processor in some embodiments is adapted for coupling to an active lighting image sensor, such as an infrared sensor that does not provide depth information corresponding to the amplitude image, or an SL or ToF sensor that provides depth information corresponding to the amplitude image.

**DEPTH IMAGE GENERATION UTILIZING DEPTH INFORMATION
RECONSTRUCTED FROM AN AMPLITUDE IMAGE**

Field

5 The field relates generally to image processing, and more particularly to techniques for generating depth images.

Background

10 Depth images are commonly utilized in a wide variety of machine vision applications including, for example, gesture recognition systems and robotic control systems. A depth image may be generated directly using a depth imager such as a structured light (SL) camera or a time of flight (ToF) camera. Such cameras can be configured to provide both depth information and amplitude information, in the form of respective depth and amplitude images. However, the depth information provided by these and other depth imagers is often incomplete, noisy, distorted
15 or of insufficient resolution for a particular application. Other types of imagers, such as infrared imagers, typically provide only amplitude images. Accordingly, a need exists for improved techniques for generating depth images, both in the case of depth imagers such as SL or ToF cameras as well as in infrared imagers and other imagers that do not ordinarily provide depth information.

20

Summary

 In one embodiment, an image processing system comprises an image processor having image processing circuitry and an associated memory. The image processor is configured to identify a region of interest in an amplitude image, to detect one or more relatively low gradient
25 regions in the region of interest, to reconstruct depth information for said one or more relatively low gradient regions, to extend the reconstructed depth information beyond said one or more relatively low gradient regions to additional pixels of the region of interest, and to generate a depth image utilizing at least portions of the reconstructed depth information and the extended reconstructed depth information.

30 By way of example only, the image processor may be implemented in a depth imager such as an SL or ToF camera. It is also possible to utilize the image processor to implement a depth imager using an image sensor that does not ordinarily provide depth information, such as an active lighting infrared image sensor. The image processor can be implemented in a wide variety of other types of processing devices.

Illustrative embodiments of the invention include but are not limited to methods, apparatus, systems, processing devices, integrated circuits, and computer-readable storage media having computer program code embodied therein.

5 **Brief Description of the Drawings**

FIG. 1 is a block diagram of an image processing system that includes an image processor comprising a depth reconstruction module in an illustrative embodiment.

FIG. 2 is a block diagram of an image processing system in which an image processor comprising a depth reconstruction module is implemented within a depth imager in another
10 illustrative embodiment.

FIG. 3 is a flow diagram of an illustrative embodiment of a depth reconstruction process implemented in the image processors of FIGS. 1 and 2.

Detailed Description

15 Embodiments of the invention will be illustrated herein in conjunction with exemplary image processing systems that include image processors configured to generate depth maps or other types of depth images suitable for use in gesture recognition and other applications. It should be understood, however, that embodiments of the invention are more generally applicable to any image processing system or associated device or technique that involves generating a depth
20 image using depth information at least a portion of which is reconstructed from at least one amplitude image.

FIG. 1 shows an image processing system 100 in an illustrative embodiment of the invention. The image processing system 100 comprises an image sensor 101 coupled to an image processor 102. The image processor 102 comprises a depth reconstruction module 103 that is
25 itself comprised of multiple modules. More particularly, the depth reconstruction module 103 includes exemplary modules 104, 105, 106 and 107 for region of interest (ROI) detection, zero gradient region detection, depth reconstruction for zero gradient regions, and reconstructed depth extension, respectively.

The ROI detection module 104 is configured to identify an ROI in a luminance image
30 received from the image sensor 101, which may comprise an active lighting image sensor configured to provide a luminance image. The luminance image is typically in the form of a rectangular matrix of picture elements or “pixels” having respective positive integer or floating values, although other image formats could be used.

In other embodiments, other types of intensity images or more generally amplitude images may be used. The term “amplitude image” as used herein is intended to be broadly construed so as to encompass a luminance image, intensity image or other type of image providing amplitude information. As noted above, such amplitude information for a given amplitude image is typically arranged in the form of a rectangular array of pixels. The depth reconstruction module 103 is configured to reconstruct depth information from such a luminance image or other amplitude image, and in this embodiment more particularly from a luminance image provided by the image sensor 101, possibly in combination with a corresponding coarse depth map or other depth image, as will be described in more detail below.

By way of example, the image sensor 101 may comprise an active lighting image sensor such as an SL or ToF image sensor that produces both amplitude and depth information, or an active lighting infrared image sensor that produces only amplitude information. A wide variety of other types of image sensors providing different types of image output at fixed or variable frame rates can also be used.

The zero gradient region detection module 105 is configured to detect one or more regions within the ROI that have gradients sufficiently close to zero gradients. Such regions are more generally referred to herein as “relatively low gradient regions” of the ROI in that these regions have gradients that are closer to zero gradients than other regions of the ROI. Although the present embodiment identifies regions that have substantially zero gradients, other embodiments can use other types of relatively low gradient regions. For example, relatively low gradient regions can be identified as regions having respective gradients that are at or below a specified gradient threshold, where the threshold is a non-zero threshold. Accordingly, it is to be appreciated that references herein to zero gradients are exemplary only, and other embodiments can be implemented using other types of relatively low gradients.

The depth reconstruction for zero gradient regions module 106 is configured to reconstruct depth information for the zero gradient regions detected by module 105, but not for other portions of the ROI, such as those portions that have relatively high gradients. Thus, in the present embodiment, the depth reconstruction module 103 reconstructs depth information for only the zero gradient regions of the ROI.

The reconstructed depth extension module 107 is configured to extend the reconstructed depth information generated by module 106 beyond the zero gradient regions to additional pixels of the ROI.

The output of the depth reconstruction module 103 in the present embodiment comprises a depth image illustratively in the form of a reconstructed depth map generated utilizing at least portions of the reconstructed depth information generated by module 106 and the extended reconstructed depth information generated by module 107. The original luminance image is also
5 output by the module 103 in this embodiment. The reconstructed depth map and the luminance image are provided as inputs to gesture recognition systems and applications 108, which may be implemented on one or more other processing devices coupled to or otherwise associated with the image processor 102. In other embodiments, at least portions of the gesture recognition systems and applications 108 can be implemented on the image processor 102, rather than on an
10 associated separate processing device.

The gesture recognition systems and applications 108 are illustratively configured to recognize particular gestures utilizing reconstructed depth maps and corresponding luminance images supplied by the image processor 102 and to take appropriate actions based on the recognized gestures. For example, a given gesture recognition system can be configured to
15 recognize a gesture from a specified gesture vocabulary and to generate a corresponding gesture pattern identifier (ID) and possibly additional related parameters for delivery to one or more of the applications. A given such application can translate that information into a particular command or set of commands to be executed by that application. The gesture recognition system may comprise, for example, separate subsystems or recognition modules for static pose
20 recognition, dynamic gesture recognition and cursor gesture recognition.

As indicated previously, the depth reconstruction module 103 in some implementations of the FIG. 1 embodiment also utilizes an input depth map provided by the image sensor 101, via the dashed arrow in the figure, as a supplement to the luminance image in order to facilitate detection of the ROI in module 104. For example, such an input depth map can be provided in
25 embodiments that include an image sensor that is configured to utilize SL or ToF imaging techniques. This input depth map is an example of what is more generally referred to herein as a “coarse depth map” or still more generally as a “coarse depth image” because it typically has a substantially lower resolution than the reconstructed depth map generated at the output of the depth reconstruction module 103.

30 In an implementation in which the coarse depth map is utilized, the ROI is detected in module 104 using both the luminance image and the coarse depth map. Moreover, the output reconstructed depth map in such an embodiment may be generated in module 103 by combining the coarse depth map with a reconstructed depth map that is generated utilizing at least portions

of the reconstructed depth information generated by module 106 and the extended reconstructed depth information generated by module 107.

It should be noted that the term “depth image” as broadly utilized herein may in some embodiments encompass an associated amplitude image. Thus, a given depth image may
5 comprise depth information as well as corresponding amplitude information. For example, the amplitude information may be in the form of a grayscale image or other type of intensity image that is generated by the same image sensor 101 that generates the depth information. An amplitude image of this type may be considered part of the depth image itself, or may be implemented as a separate image that corresponds to or is otherwise associated with the depth
10 image. Other types and arrangements of depth images comprising depth information and having associated amplitude information may be generated in other embodiments.

Accordingly, references herein to a given depth image should be understood to encompass, for example, an image that comprises depth information only, as well as an image that comprises a combination of depth and amplitude information. The depth and amplitude
15 images mentioned previously in the context of the description of depth reconstruction module 103 therefore need not comprise separate images, but could instead comprise respective depth and amplitude portions of a single image. An “amplitude image” as that term is broadly used herein comprises amplitude information and possibly other types of information, and a “depth image” as that term is broadly used herein comprises depth information and possibly other types
20 of information.

It should be understood that the particular functional modules 103, 104, 105, 106 and 107 utilized in image processor 102 in the FIG. 1 embodiment are exemplary only, and other embodiments of the invention can be configured using other arrangements of additional or alternative modules or other components.

25 The image processor 102 in the present embodiment is assumed to be implemented using at least one processing device and comprises a processor 120 coupled to a memory 122. The processor 120 executes software code stored in the memory 122 in order to control the performance of image processing operations.

The image processor 102 also comprises a network interface 124 that supports
30 communication over one or more networks. The network interface 124 may comprise one or more conventional transceivers. Accordingly, the image processor 102 is assumed to be configured to communicate with a computer or other processing device of the image processing system 100 over a network or other type of communication medium. Depth images generated by

the image processor 102 can be provided to other processing devices for further processing in conjunction with implementation of functionality such as gesture recognition. Such depth images can additionally or alternatively be displayed, transmitted or stored using a wide variety of conventional techniques.

5 In other embodiments, the image processor 102 need not be configured for communication with other devices over a network, and in such embodiments the network interface 124 may be eliminated.

The processor 120 may comprise, for example, a microprocessor, an application-specific integrated circuit (ASIC), a field-programmable gate array (FPGA), a central processing unit
10 (CPU), an arithmetic logic unit (ALU), a digital signal processor (DSP), or other similar processing device component, as well as other types and arrangements of image processing circuitry, in any combination. A “processor” as the term is generally used herein may therefore comprise portions or combinations of a microprocessor, ASIC, FPGA, CPU, ALU, DSP or other image processing circuitry.

15 As noted above, the memory 122 stores software code for execution by the processor 120 in implementing portions of the functionality of image processor 102, such as one or more of the modules 104, 105, 106 and 107 of the depth reconstruction module 103. A given such memory that stores software code for execution by a corresponding processor is an example of what is more generally referred to herein as a computer-readable storage medium having computer
20 program code embodied therein, and may comprise, for example, electronic memory such as random access memory (RAM) or read-only memory (ROM), magnetic memory, optical memory, or other types of storage devices in any combination.

Articles of manufacture comprising such computer-readable storage media are considered embodiments of the invention. The term “article of manufacture” as used herein should be
25 understood to exclude transitory, propagating signals.

It should also be appreciated that embodiments of the invention may be implemented in the form of integrated circuits. In a given such integrated circuit implementation, identical die are typically formed in a repeated pattern on a surface of a semiconductor wafer. Each die includes an image processor or other image processing circuitry as described herein, and may
30 include other structures or circuits. The individual die are cut or diced from the wafer, then packaged as an integrated circuit. One skilled in the art would know how to dice wafers and package die to produce integrated circuits. Integrated circuits so manufactured are considered embodiments of the invention.

The particular configuration of image processing system 100 as shown in FIG. 1 is exemplary only, and the system 100 in other embodiments may include other elements in addition to or in place of those specifically shown, including one or more elements of a type commonly found in a conventional implementation of such a system.

5 For example, in some embodiments, the image processing system 100 is implemented as a video gaming system or other type of gesture-based system that processes image streams in order to recognize user gestures. The disclosed techniques can be similarly adapted for use in a wide variety of other systems requiring a gesture-based human-machine interface, and can also be applied to other applications, such as machine vision systems in robotics and other industrial
10 applications that utilize gesture recognition.

It should be noted that embodiments of the invention are not limited to use in recognition of hand gestures, but can be applied to other types of gestures as well. The term “gesture” as used herein is therefore intended to be broadly construed. Moreover, depth images generated by an image processor in the manner disclosed herein are also suitable for use in a wide variety of
15 applications other than gesture recognition.

The image processor 102 in some embodiments may be implemented on a common processing device with a computer, mobile phone or other device that processes images. By way of example, a computer or mobile phone may be configured to incorporate the image sensor 101 and the image processor 102, as well as at least portions of the gesture recognition systems and
20 applications 108.

It is also to be appreciated that the image processor 102 may itself comprise multiple distinct processing devices, such that different portions of the depth reconstruction module 103 are implemented using two or more processing devices.

Accordingly, the particular arrangement of components shown in image processor 102 in
25 the FIG. 1 embodiment can be varied in other embodiments. For example, an otherwise conventional image processing integrated circuit or other type of image processing circuitry suitably modified to perform processing operations as disclosed herein may be used to implement at least a portion of one or more of the modules 103, 104, 105, 106 and 107 of image processor 102. One possible example of image processing circuitry that may be used in one or more
30 embodiments of the invention is an otherwise conventional graphics processor suitably reconfigured to perform functionality associated with one or more of the modules 103, 104, 105, 106 and 107.

Also, the particular number of modules can be varied in other embodiments. For example, in other embodiments two or more of these modules may be combined into a lesser number of modules, or the disclosed depth reconstruction and depth image generation functionality may be distributed across a greater number of modules.

5 The term “image processor” as used herein is intended to be broadly construed so as to encompass these and other arrangements.

 The image sensor 101 and image processor 102 may be implemented within a depth imager configured to generate both depth and amplitude images, although other implementations may be used in other embodiments.

10 An illustrative embodiment in which the image processor 102 is implemented within an exemplary depth imager is shown in FIG. 2. In this embodiment, an information processing system 200 comprises a depth imager 201 coupled to previously-described gesture recognition systems and applications 108. The depth imager 201 incorporates the image processor 102 comprising depth reconstruction module 103 having component modules 104, 105, 106 and 107,
15 also as previously described.

 The depth imager 201 in the FIG. 2 embodiment comprises a light emitting diode (LED) emitter 202 that generates modulated light for imaging a scene. The LED emitter 202 may comprise an array of LEDs or a single LED. The modulated light illustratively comprises, for example, infrared light, although other types of light sources may be used. The corresponding
20 reflected light is detected by a semiconductor photonic sensor 204 that produces a raw luminance image. The raw luminance image is applied to a luminance demodulator 205 along with the modulated signal from LED emitter 202 or the corresponding modulator phase information. The luminance demodulator 205 processes these inputs to generate the demodulated luminance image, and possibly an associated coarse depth map, which are applied as respective inputs to the
25 depth reconstruction module 103 of image processor 102.

 By way of example, in an active lighting infrared image sensor implementation of the FIG. 2 embodiment, the luminance demodulator 205 comprises a shutter synchronized with the modulation of the LED emitter 202, such that the shutter is open for a short period of time corresponding to the LED emitter being in its “on” state. Such an arrangement generally does
30 not provide an associated coarse depth map, and therefore the reconstructed depth map generated by depth reconstruction module 103 is generated entirely by reconstruction of depth information from the luminance image.

As another example, in an implementation utilizing ToF depth sensing, the luminance demodulator is more particularly configured as a ToF demodulator which reconstructs both the amplitude and the phase of the reflected light and converts the phase to per-pixel coarse depth information which collectively provides the coarse depth map. In this case, the reconstructed
5 depth map is generated by the depth reconstruction module 103 using not only the luminance image but also the coarse depth map.

Other types of image sensing techniques can be used, providing at least a luminance image or other type of amplitude image and possibly an associated depth map or other type of depth image, for further processing by the depth reconstruction module 103.

10 It is therefore apparent that embodiments of the invention allow a depth image to be generated without requiring the use of a depth image sensor.

The operation of the depth reconstruction module 103 of image processor 102 in the image processing system embodiments of FIGS. 1 and 2 will now be described in greater detail with reference to FIG. 3. This figure illustrates an exemplary process 300 that is implemented in the
15 image processor 102 using depth reconstruction module 103. The process 300 includes steps 302 through 310 as shown, with steps 302, 304, 306 and 308 being performed by respective modules 104, 105, 106 and 107 of the depth reconstruction module 103, and step 310 being performed by one or more other components of the image processor 102. As indicated previously, portions of the process may be implemented at least in part utilizing software executing on image processing
20 circuitry of the image processor 102.

The process 300 is applied to a given luminance image and possibly also utilizes a corresponding coarse depth map if available. Such exemplary amplitude and depth images may be subject to various preprocessing operations such as filtering and noise reduction prior to application of the steps of process 300.

25 In step 302, an ROI is detected in the luminance image. Detection of the ROI in step 302 may also make use of a coarse depth map, if available. This step in the present embodiment more particularly involves defining an ROI mask for a region in the luminance image that corresponds to one or more hands of a user in an imaged scene, also referred to as a hand region of the luminance image. The output of the ROI detection step 302 in the present embodiment includes
30 a binary ROI mask for the hand region in the input luminance image. It can be in the form of an image having the same size as the input luminance image, or a sub-image containing only those pixels that are part of the ROI.

For further description below, it is assumed that the binary ROI mask is an image having the same size as the input luminance image. Thus, by way of example, if the input luminance image comprises an $H \times W$ matrix of pixels, the binary ROI mask generated in step 302 also comprises an $H \times W$ matrix of pixels, with the pixels within the ROI having a certain binary value, illustratively a logic 1 value, and pixels outside the ROI having the complementary binary value, illustratively a logic 0 value.

Amplitude values and possibly also depth values are associated with respective pixels of the ROI defined by the binary ROI mask. These ROI pixels are assumed to be part of one or more input images, such as the input luminance image.

A variety of different techniques can be used to detect the ROI in step 302. For example, it is possible to use techniques such as those disclosed in Russian Patent Application No. 2013135506, filed July 29, 2013 and entitled "Image Processor Configured for Efficient Estimation and Elimination of Background Information in Images," which is commonly assigned herewith and incorporated by reference herein.

As another example, the binary ROI mask can be determined using threshold logic applied to pixel values of the input luminance image. More particularly, one can select only those pixels with luminance values greater than some predefined threshold. For active lighting imagers such as SL or ToF imagers or active lighting infrared imagers, the closer an object is to the imager, the higher the luminance values of the corresponding image pixels, not taking into account reflecting materials. Accordingly, selecting only those pixels with relatively high luminance values for the ROI allows one to preserve close objects from an imaged scene and to eliminate far objects from the imaged scene.

It should be noted that for ToF imagers, pixels with lower luminance values tend to have higher error in their corresponding depth values, and so removing pixels with low luminance values from the ROI additionally protects one from using incorrect depth information.

In embodiments in which the coarse depth map is available in addition to the luminance image, the ROI can be detected at least in part by selecting only those pixels with depth values falling between predefined minimum and maximum threshold depths D_{min} and D_{max} . These thresholds are set to appropriate distances between which the hand region is expected to be located within the image. For example, the thresholds may be set as $D_{min}=0$, $D_{max}=0.5$ meters (m), although other values can be used.

In conjunction with detection of the ROI, opening or closing morphological operations utilizing erosion and dilation operators can be applied to remove dots and holes as well as other

spatial noise in the image. Other exemplary noise reduction techniques that may be utilized in conjunction with detection of the ROI are described in PCT International Application PCT/US13/56937, filed on August 28, 2013 and entitled "Image Processor With Edge-Preserving Noise Suppression Functionality," which is commonly assigned herewith and incorporated by reference herein.

One possible implementation of a threshold-based ROI determination technique using both amplitude and depth thresholds is as follows:

1. Set $ROI_{ij} = 0$ for each i and j .
2. For each depth pixel d_{ij} set $ROI_{ij} = 1$ if $d_{ij} \geq d_{min}$ and $d_{ij} \leq d_{max}$.
3. For each amplitude pixel a_{ij} set $ROI_{ij} = 1$ if $a_{ij} \geq a_{min}$.

4. Coherently apply an opening morphological operation comprising erosion followed by dilation to both ROI and its complement to remove dots and holes comprising connected regions of ones and zeros having area less than a minimum threshold area A_{min} .

It is also possible in some embodiments to detect a palm boundary and to remove from the ROI any pixels below the palm boundary, leaving essentially only the palm and fingers in a modified hand image. Such a step advantageously eliminates, for example, any portions of the arm from the wrist to the elbow, as these portions can be highly variable due to the presence of items such as sleeves, wristwatches and bracelets, and in any event are typically not useful for hand gesture recognition.

Exemplary techniques suitable for use in implementing the above-noted palm boundary determination in the present embodiment are described in Russian Patent Application No. 2013134325, filed July 22, 2013 and entitled "Gesture Recognition Method and Apparatus Based on Analysis of Multiple Candidate Boundaries," which is commonly assigned herewith and incorporated by reference herein.

Alternative techniques can be used. For example, the palm boundary may be determined by taking into account that the typical length of the human hand is about 20-25 centimeters (cm), and removing from the ROI all pixels located farther than a 25 cm threshold distance from the uppermost fingertip, possibly along a determined main direction of the hand. The uppermost fingertip can be identified simply as the uppermost 1 value in the binary ROI mask.

It should be appreciated, however, that palm boundary detection need not be applied in determining the binary ROI mask in step 302. For example, in embodiments in which the

detected ROI comprises a pair of hands of a given user, the above-described palm boundary detection can be eliminated.

In step 304, one or more regions having a close to zero gradient within the detected ROI are determined. A region having a “close to zero” gradient is an example of what is more generally referred to herein as a “relatively low gradient region” in that it exhibits a substantially lower gradient than other portions of the detected ROI. For example, it may comprise a portion, area or other region of the ROI that has a gradient at or below a specified gradient threshold, or a region of the ROI that has a substantially zero gradient. The latter is also referred to herein as a zero gradient region, although it is to be appreciated that the gradient need not be exactly zero. In these and other relatively low gradient regions, it is assumed for purposes of the process 300 that hand surface is primarily locally perpendicular to the direction from the imager and therefore most incident light is reflected back to the image sensor. As a result, there is a strong dependency in such regions between the luminance value for a given pair of pixel index coordinates and depth to the corresponding point on the hand.

An exemplary technique for detecting regions within the ROI with close to zero gradient in step 304 will now be described. First, it is assumed that the detection of the ROI in step 302 involves smoothing the input luminance image. More particularly, input luminance image A is smoothed using a low-pass filter or other type of filter configured to suppress speckle noise in the input luminance image A. For example, a Gaussian filter with $\sigma=3$ may be used. Let \hat{A} be the smoothed luminance image. It is further assumed that the ROI in step 302 is detected in this smoothed luminance image. The binary ROI mask generated in step 302 is therefore generated using the smoothed luminance image \hat{A} . Alternatively, the input luminance image A may be applied as an input not only to step 302 but also as an input to step 304, with the ROI being detected in the input luminance image A in step 302 and the smoothed luminance image \hat{A} being generated in step 304 separately from the ROI detection.

The exemplary technique for detecting regions with close to zero gradient includes the following steps:

Step 1. For each pixel (i,j) ($i>1, j>1$) within the ROI determined in step 302, estimate luminance gradient $d\hat{A}(i,j)$ as $d\hat{A}(i,j)=(\hat{A}(i,j)-\hat{A}(i,j-1), \hat{A}(i,j)-\hat{A}(i-1,j))$. Other types of gradient estimation may be used.

Step 2. Initialize an $H \times W$ zero gradient binary matrix G with zeros. For all pixels (i,j) from the ROI, set $G(i,j)$ to 1 if $\text{abs}(d\hat{A}(i,j)_1) \leq \text{Athresh1}$ and $\text{abs}(d\hat{A}(i,j)_2) \leq \text{Athresh1}$, where abs denotes absolute value and where Athresh1 is a positive small threshold having a value that

significantly depends on an average value of luminance for a typical scene. The value of $A_{thresh1}$ is sufficiently small to ensure selection of only those regions corresponding to surfaces substantially perpendicular to the direction to the image sensor.

Step 3. For each pixel (i,j) from the ROI for which $G(i,j)=0$ but there exists at least one neighboring pixel (i_1,j_1) for which $G(i_1,j_1)=1$, if $abs(\hat{A}(i,j)-\hat{A}(i_1,j_1)) < A_{thresh2}$ then set $G(i,j)=1$, where $i-1 \leq i_1 \leq i+1$ and $j-1 \leq j_1 \leq j+1$. The value of threshold $A_{thresh2}$ is also a small positive threshold having substantially the same order as the value of $A_{thresh1}$. This step is repeated for multiple passes, and more particularly for k passes, where a suitable value for k in some embodiments is $k=3$. Such multiple passes ensure that the initial zero gradient areas are extended to encompass adjacent pixels having sufficiently close luminance values.

In step 306, depth information is reconstructed from the luminance values for the zero gradient regions identified in step 304. This step can be configured to assume that for small surfaces with homogeneous reflective characteristics which are oriented perpendicular to the image sensor direction, luminance is approximately inversely proportional to the square of distance to the corresponding surface, if the position of that surface within the frame and therefore lighting angle for that surface does not change. In this fixed position scenario, the relationship between luminance and depth for pixel (i,j) can be expressed as $A(i,j) \sim K/d^2(i,j)$, where K denotes a coefficient that can be experimentally determined by taking into account the particular type of image sensor used and other implementation-specific factors.

Additionally or alternatively, if the surface position and therefore lighting angle is instead assumed to change, a similar dependency between luminance and depth is observed, but is more particularly modeled in this scenario as $A(i,j) \sim K(r)/d^2(i,j)$ where $r = \sqrt{(i-i_0)^2 + (j-j_0)^2}$ denotes pixel distance from frame center (i_0,j_0) . For the assumed frame dimension $H \times W$, $i_0=(H-1)/2$ and $j_0=(W-1)/2$. It has been observed that dependency between luminance and the inverse square of depth for different values of r is approximately linear. Non-linear dependency for small depth values is attributed to luminance saturation effects.

In determining the coefficient function $K(r)$, the distribution of light from the active lighting imager should be taken into account. For example, position of an LED lighting source relative to the corresponding image sensor is typically an implementation-specific factor that should be taken into account in determining $K(r)$. Also, relative luminance value as a function of radiation angle is often known for various types of LED sources and can be used in determining $K(r)$ in a given embodiment.

As a more particular example, in an embodiment with a single infrared LED source, a luminance value $A(\alpha)$ for a given radiation angle α can be approximated as $A_0 \cdot \cos(\alpha)$ where A_0 is the luminance value at the same distance for $\alpha=0$. It can be shown that for this case the coefficient function can be approximated as $K(r) = c/\sqrt{\sqrt{1+4 \cdot r^2/W^2}}$ where c is a positive constant. In the case of more complex LED sources such as arrays of multiple LEDs, a more complex approximation of $K(r)$ can be used.

In these and other cases, the coefficient function $K(r)$ can be determined at least in part by applying a calibration process to the particular imager configuration. Such a calibration process can be implemented using the following steps:

Step 1. An image of a plane surface with reflecting characteristics close to those of human skin is acquired by the image sensor so that depth information for each pixel is simultaneously measured with the luminance information. Multiple measurements of this kind are made for each pixel (i,j) , $i=0 \dots H-1$, $j=0 \dots W-1$.

Step 2. For each value of $r=0 \dots \sqrt{(W/2)^2 + (H/2)^2}$, depth and luminance information for pixels (i,j) is collected, where as indicated previously $r=\sqrt{(i-i_0)^2 + (j-j_0)^2}$. Using least means squared (LMS) or other regression-like techniques, a particular coefficient of linear dependency between depth value d and $\sqrt{1/A}$ is estimated. Each such coefficient is an element of the coefficient function $K(r)$ for a particular value of r .

Step 3. The coefficient function $K(r)$ is approximated using a polynomial of fixed degree (e.g., 3) or other similar function.

Using the foregoing calibration process, the coefficient function for an exemplary commercially-available PMD Nano image sensor was approximated as $K(r) = 4e-7 \cdot r^3 - 6e-5 \cdot r^2 + 0.00053 \cdot r + 0.3$.

The resulting coefficient function is then utilized in the manner previously described to reconstruct depth information from luminance information for the zero gradient regions.

In step 308, the reconstructed depth information determined for the zero gradient regions in step 306 is extended to other portions of the ROI. More particularly, in this embodiment, the reconstructed depth information is extended to substantially all of the pixels of the ROI that were not part of any of the zero gradient regions. These additional pixels in the context of the present embodiment are those pixels that are not part of the above-described zero gradient mask G .

The extension of the reconstructed depth information in step 308 can be implemented using a number of different techniques. For example, depth values for pixels that are not part of the zero gradient mask G can be defined as a mean or other function of the depth values of the

pixels that are part of the zero gradient mask G . A low-pass filter or other type of filter can then be used to smooth the resulting reconstructed depth map. In one embodiment, a Gaussian filter is used that has a smoothing factor sufficient to make depth transitions between pixels in G and pixels not in G smaller than a designated depth measurement precision. For the above-noted

5 PMD Nano image sensor, a smoothing factor of $\sigma=10$ was utilized. In generating the reconstructed depth map, depth values for pixels within the ROI but not part of the zero gradient mask G are replaced with the corresponding smoothed depth values.

Another exemplary technique for extension of the reconstructed depth information in step 308 is as follows. It is assumed in this technique that curvature at the edges of a human hand

10 follows a function $\delta(\text{dist})$, where dist is distance between a given reconstructed pixel and the closest pixel in the zero gradient mask G . This distance may instead be calculated using an average distance of the given reconstructed pixel from multiple pixels in G . The function itself assumes a spherical surface having a specified radius of curvature, such as a radius of approximately 1 cm. Using this function, depth values for all pixels within the ROI but not in

15 the zero gradient mask G are reconstructed as $d_1 - \delta(\text{dist})$, where d_1 denotes the depth value for the closest pixel (i_1, j_1) from G for the pixel (i, j) within the ROI but not in G , where dist is the distance between (i, j) and (i_1, j_1) recalculated in meters.

The above reconstructed depth extension techniques are exemplary only, and other techniques may be used to determine depth values for portions of the ROI outside of the zero

20 gradient regions in other embodiments. All such techniques are intended to be encompassed by general references herein to “extending” or “extension of” reconstructed depth information beyond one or more zero gradient regions.

In step 310, the coarse depth map if available is combined with the reconstructed depth map from step 308 to generate an output reconstructed depth map. This combination is

25 illustratively computed as $\text{Reconstructed_depth} = (\text{actual_depth} * W_1 + \text{depth_from_luminance} * W_2) / (W_1 + W_2)$ where $W_1 = 1/\sigma_1$, $W_2 = 1/\sigma_2$ and σ_1 and σ_2 denote average standard deviations of actual and reconstructed depth jitter (e.g., noise). Other techniques can be used to combine the reconstructed depth map and the coarse depth map. In embodiments in which there is no coarse depth map available, step 310 is eliminated and the output of step 308 is utilized as the

30 reconstructed depth map.

The particular types and arrangements of processing blocks shown in the embodiment of FIG. 3 are exemplary only, and additional or alternative blocks can be used in other embodiments. For example, blocks illustratively shown as being executed serially in the figures can be

performed at least in part in parallel with one or more other blocks or in other pipelined configurations in other embodiments.

The reconstructed depth maps or other reconstructed depth images generated in the illustrative embodiments provide significantly improved gesture recognition performance relative to conventional arrangements. For example, these embodiments provide substantial improvements relative to use of a coarse depth map alone in SL or ToF camera implementations. Moreover, the disclosed techniques provide accurate and efficient generation of depth maps or other depth images in infrared imagers and other types of active lighting imagers that would not otherwise provide depth information.

The depth information in these and other embodiments can be generated at low cost, with low jitter and high precision. Accordingly, problems attributable to incomplete, noisy, distorted or poor resolution depth images provided by some conventional depth imagers are advantageously overcome. Also, capabilities of active lighting imagers are enhanced. Performance in the corresponding gesture recognition systems and applications is accelerated while ensuring a high degree of accuracy in the recognition process.

It should again be emphasized that the embodiments of the invention as described herein are intended to be illustrative only. For example, other embodiments of the invention can be implemented utilizing a wide variety of different types and arrangements of image processing circuitry, modules, processing blocks and associated operations than those utilized in the particular embodiments described herein. In addition, the particular assumptions made herein in the context of describing certain embodiments need not apply in other embodiments. These and numerous other alternative embodiments within the scope of the following claims will be readily apparent to those skilled in the art.

Claims

What is claimed is:

1. A method comprising steps of:

identifying a region of interest in an amplitude image;

5 detecting one or more relatively low gradient regions in the region of interest;

reconstructing depth information for said one or more relatively low gradient regions in the region of interest;

extending the reconstructed depth information beyond said one or more relatively low gradient regions to additional pixels of the region of interest; and

10 generating a depth image utilizing at least portions of the reconstructed depth information and the extended reconstructed depth information;

wherein the steps are implemented in an image processor comprising a processor coupled to a memory.

15 2. The method of claim 1 wherein said one or more relatively low gradient regions comprise regions having respective gradients at or below a specified gradient threshold.

3. The method of claim 1 wherein said one or more relatively low gradient regions comprise regions having respective substantially zero gradients.

20

4. The method of claim 1 wherein identifying a region of interest comprises generating a binary region of interest mask in which pixels within the region of interest all have a first binary value and pixels outside the region of interest all have a second binary value complementary to the first binary value.

25

5. The method of claim 1 wherein reconstructing depth information for said one or more relatively low gradient regions comprises reconstructing depth information for said one or more relatively low gradient regions but not for other portions of the region of interest.

30 6. The method of claim 1 wherein the amplitude image comprises a luminance image generated by an active lighting imager.

7. The method of claim 6 wherein the active lighting imager comprises one of an active lighting infrared image sensor, an SL image sensor and a ToF image sensor.

8. The method of claim 1 wherein reconstructing depth information for said one or more relatively low gradient regions in the region of interest comprises:

determining at least one coefficient that relates amplitude to depth for a particular image sensor configuration; and

utilizing the coefficient to estimate one or more depth values from one or more corresponding amplitude values for respective pixels of said one or more relatively low gradient regions.

9. The method of claim 8 wherein the determined coefficient comprises an element of a coefficient function that relates amplitude to depth for a given pixel as a function of a distance between that pixel and an approximate center of an image frame in the particular image sensor configuration.

10. The method of claim 1 wherein extending the reconstructed depth information beyond said one or more relatively low gradient regions to additional pixels of the region of interest comprises computing one or more depth values for pixels outside of said one or more relatively low gradient regions as a function of depth values determined for pixels within said one or more relatively low gradient regions.

11. The method of claim 10 wherein a depth value for a given pixel outside of said one or more relatively low gradient regions is computed utilizing a mean of a plurality of depth values for respective pixels within said one or more relatively low gradient regions.

12. The method of claim 1 wherein identifying a region of interest comprises identifying the region of interest utilizing the amplitude image and a corresponding coarse depth image.

13. The method of claim 1 wherein generating a depth image utilizing at least portions of the reconstructed depth information and the extended reconstructed depth information further comprises:

generating a reconstructed depth image comprising at least portions of the reconstructed depth information and the extended reconstructed depth information; and
combining a coarse depth image with the reconstructed depth image.

5 14. An article of manufacture comprising a computer-readable storage medium having computer program code embodied therein, wherein the computer program code when executed in the image processor causes the image processor to perform the method of claim 1.

10 15. An apparatus comprising:

an image processor adapted for coupling to an image sensor;

the image processor comprising image processing circuitry and an associated memory;

wherein the image processor is configured:

to identify a region of interest in an amplitude image;

15 to detect one or more relatively low gradient regions in the region of interest;

to reconstruct depth information for said one or more relatively low gradient regions in the region of interest;

to extend the reconstructed depth information beyond said one or more relatively low gradient regions to additional pixels of the region of interest; and

20 to generate a depth image utilizing at least portions of the reconstructed depth information and the extended reconstructed depth information.

16. An integrated circuit comprising the apparatus of claim 15.

25 17. An imager comprising:

an image sensor; and

an image processor coupled to the image sensor;

wherein the image processor is configured:

to identify a region of interest in an amplitude image;

30 to detect one or more relatively low gradient regions in the region of interest;

to reconstruct depth information for said one or more relatively low gradient regions in the region of interest;

to extend the reconstructed depth information beyond said one or more relatively low gradient regions to additional pixels of the region of interest; and

to generate a depth image utilizing at least portions of the reconstructed depth information and the extended reconstructed depth information.

5

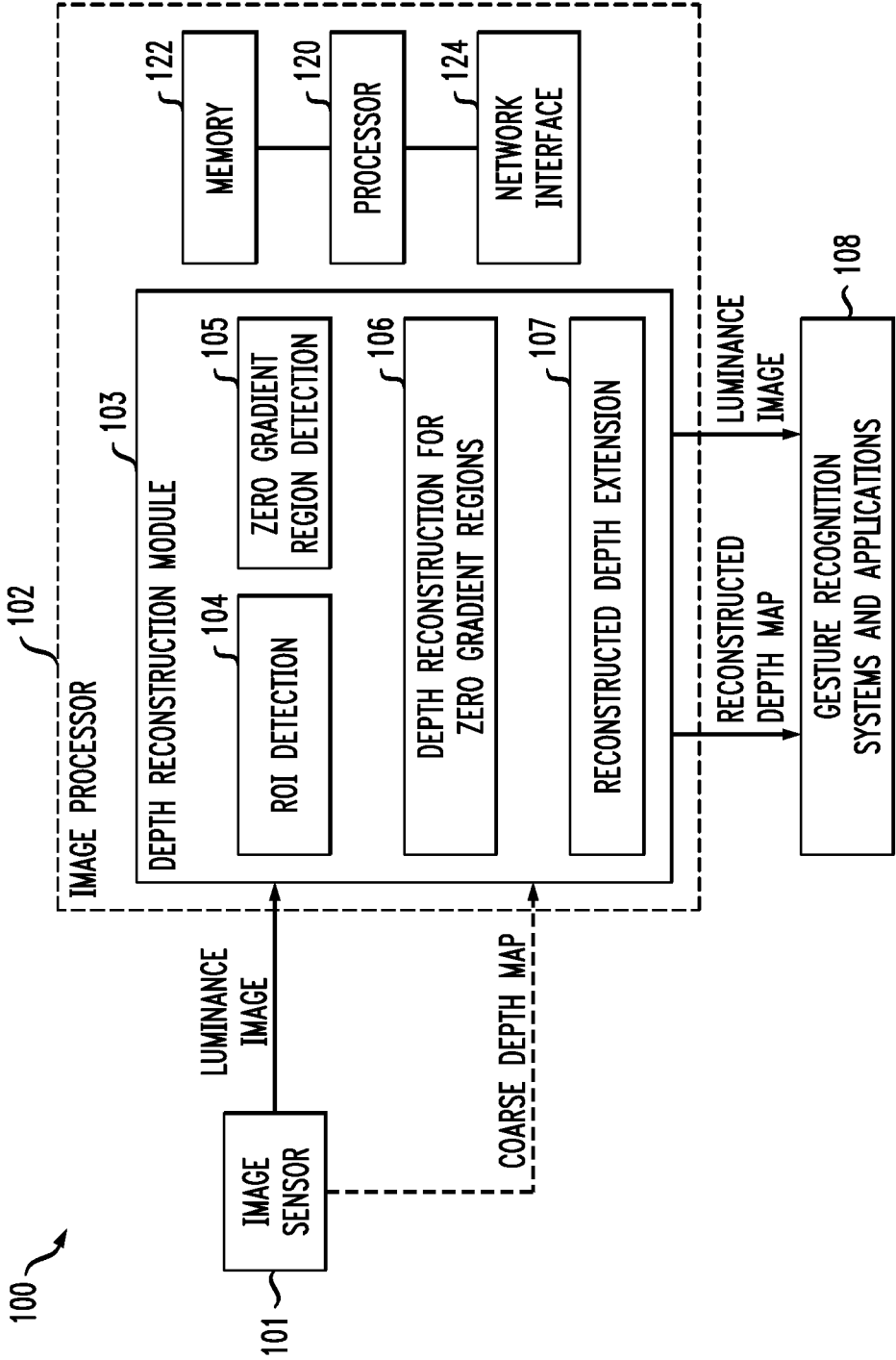
18. The imager of claim 17 wherein the image sensor comprises an active lighting infrared image sensor that does not provide depth information corresponding to the amplitude image.

10

19. The imager of claim 17 wherein the image sensor comprises an SL or ToF image sensor that provides depth information corresponding to the amplitude image.

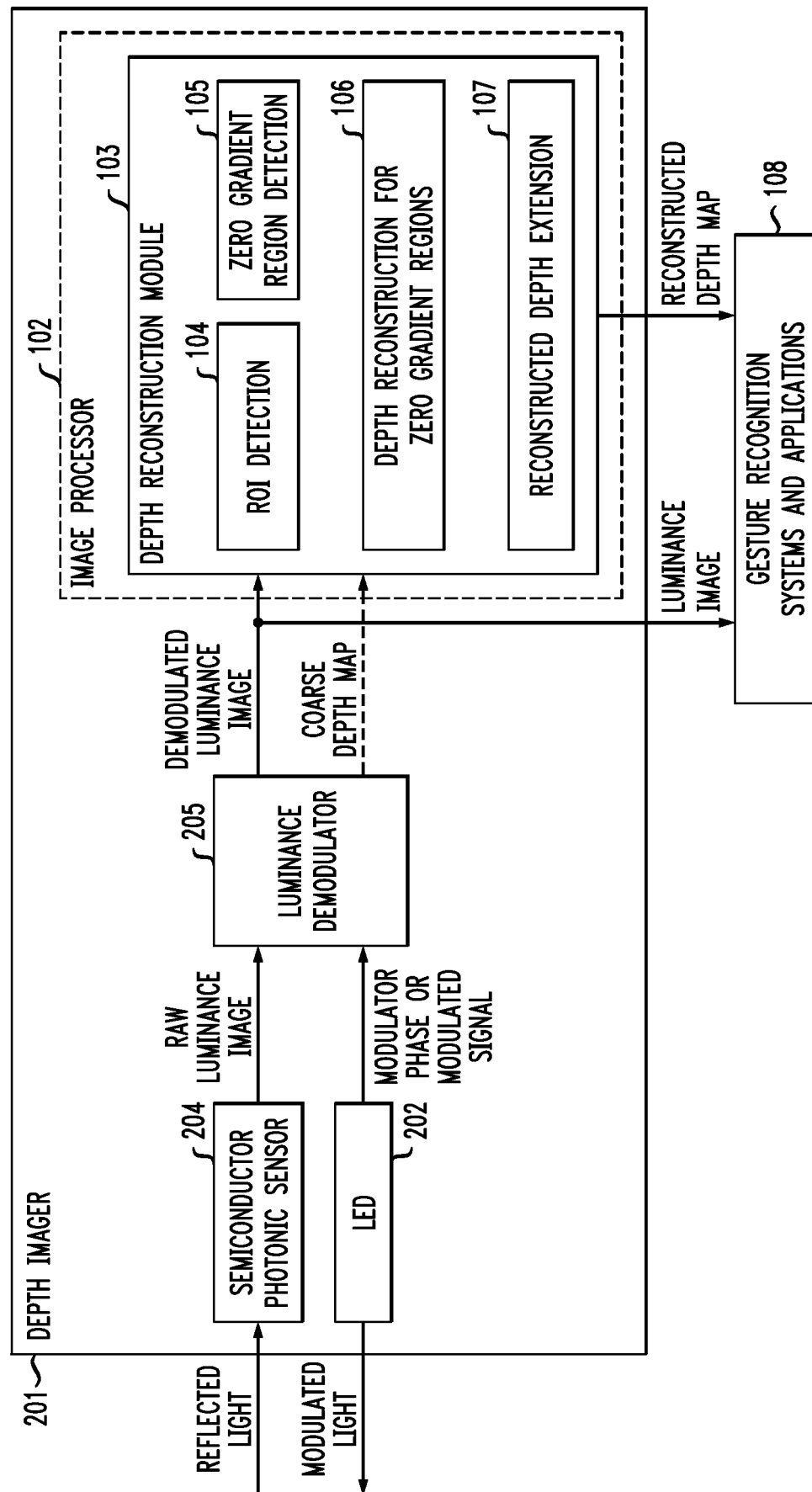
20. An image processing system comprising the imager of claim 17.

FIG. 1
100



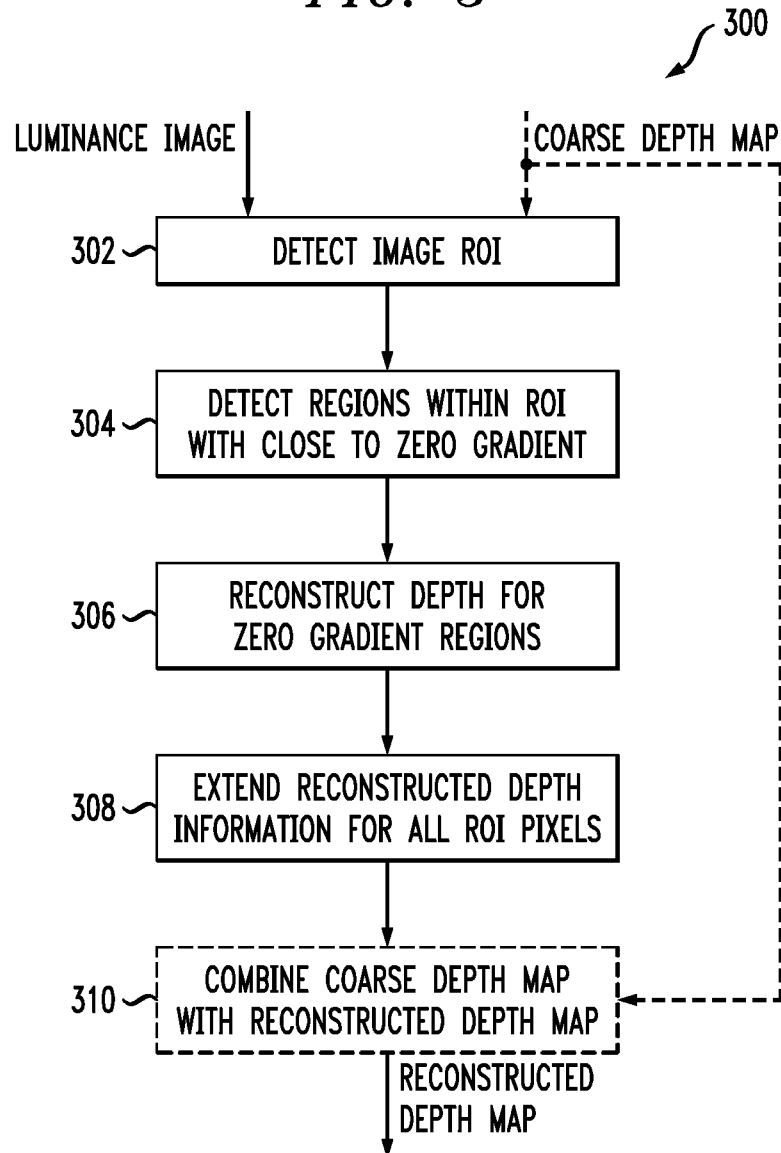
2/3

FIG. 2
200



3/3

FIG. 3



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 14/50513

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G01N 21/62 (2014.01)

CPC - G01S 17/89

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)

IPC: G01N 21/62 (2014.01)

CPC: G01S 17/89

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

IPC: G01N 21/62 (2014.01); CPC: G01N 21/4795

USPC: 378/62, 63 (keyword delimited)

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

Patbase, Google scholar, Google patents Search Terms Used: Depth, image, amplitude, intensity, luminance, region of interest, ROI, zero, 0, minimal, reconstruct, map, bugger, gradient, calculate, determine, compute, make, create, generate, infrared, sensor, TOF, time of flight, SL, distance, center, correlate, relate, coefficient

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y --- A	US 2010/0097389 A1 (LIM et al.) 22 April 2010 (22.04.2010), para [0010], [0021], [0043], [0049], [0079], [0080], [0087], [0151]	1, 6, 7, 12-17, 19, 20 ----- 2-5, 8, 10, 11, 18 ----- 9
Y	US 2004/0184667 A1 (RASKAR et al.) 23 September 2004 (23.09.2004), para [0022], [0097], [0154]	2, 4, 5
Y	US 2004/0076335 A1 (KIM) 22 April 2004 (22.04.2004), para [0014], [0063]	3
Y --- A	US 2013/0201183 A1 (PARK et al.) 08 August 2013 (08.08.2013), para [0010], [0050], [0055], [0061], [0064],	8 --- 9
Y	US 2012/0249801 A1 (NOZAKI) 04 October 2012 (4.10.2012), para [0006]-[0007], [0009], [0015], [0017], [0018]	18
Y	US 2004/0057613 A1 (NOTO et al.) 25 March 2004 (25.03.2004), para [0042]	10, 11
X, P	WO 2014/099048 A2 (PETYUSHKO et al.) 26 June 2014 (26.06.2014), pg. 4, ln 2-11; (pg. 6, ln 8-13; pg. 11, ln 12-21; pg. 16, ln 8-pg. 17, ln 22	1-20
A, P	WO 2014/044569 A1 (MIRBACH et al.) 27 March 2014 (27.03.2014), para [0010], [0035], [0036]	2-5

☐ Further documents are listed in the continuation of Box C.

* 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

29 October 2014 (29.10.2014)

Date of mailing of the international search report

05 DEC 2014

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. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774