



(51) International Patent Classification:

Not classified

(21) International Application Number:

PCT/US2019/045477

(22) International Filing Date:

07 August 2019 (07.08.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/717,037 10 August 2018 (10.08.2018) US

(71) Applicant: **UNIVERSITY OF CONNECTICUT** [US/US]; 400 Farmington Avenue MC6400, Farmington, CT 06032 (US).

(72) Inventors: **JAVIDI, Bahram**; 13 Quail Run Road, Storrs Manfield, CT 06268 (US). **MARKMAN, Adam**; 46 Rolling Ridge Road, Orange, CT 06477 (US).

(74) Agent: **NABULSI, Basam E.** et al.; McCarter & English, LLP, 201 Broad Street, Stamford, CT 06901 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

(54) Title: METHODS AND SYSTEMS FOR OBJECT RECOGNITION IN LOW ILLUMINATION CONDITIONS

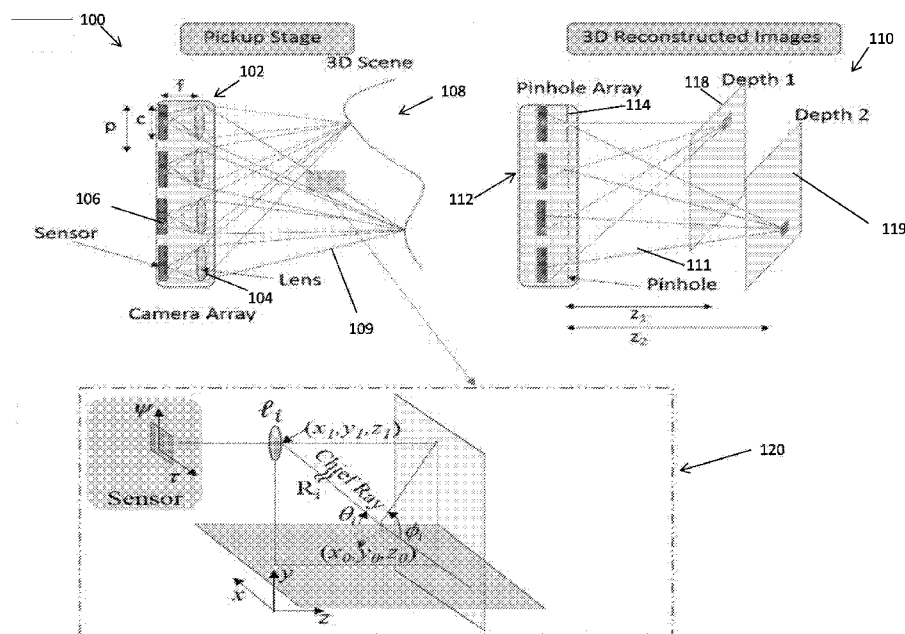


FIG. 1

(57) **Abstract:** Described herein is an object recognition system in low illumination conditions. A 3D InIm system can be trained in the low illumination levels to classify 3D objects obtained under low illumination conditions. Regions of interest obtained from 3D reconstructed images are obtained by de-noising the 3D reconstructed image using total-variation regularization using an augmented Lagrange approach followed by face detection. The regions of interest are then inputted into a trained CNN. The CNN can be trained using 3D InIm reconstructed under low illumination after TV-denoising. The elemental images were obtained under various low illumination conditions having different SNRs. The CNN can effectively recognize the 3D reconstructed faces after TV- denoising.



MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

METHODS AND SYSTEMS FOR OBJECT RECOGNITION IN LOW ILLUMINATION CONDITIONS

STATEMENT OF GOVERNMENT INTEREST

5 This invention was made with government support under grant number W909MY-12-D-0008 awarded by the Department of Defense Army Research Office. The government has certain rights in the invention.

FIELD OF THE DISCLOSURE

10 The present disclosure contained herein is in the general field of imaging methods and systems for 3D object recognition.

BACKGROUND OF THE DISCLOSURE

15 Imaging a scene in low illumination conditions using conventional image sensors that operate in the visible spectrum is difficult as the captured images become read-noise dominant. Thus, signal-to-noise ratio (SNR) suffers resulting in poor scene visualization in addition to making object recognition a difficult task.

20 The electron-multiplying CCD cameras (EM-CCD) and scientific CMOS (sCMOS) cameras, used to for imaging in low-light conditions, are expensive and bulky. In particular, the EM-CCD needs to be cooled to around -55°C prior to operation. Night vision operates by amplifying the number of photons in the scene. Accordingly, if too few photons are available, an active near infrared source is required to illuminate the scene. Infrared cameras have a lower resolution compared with visible range cameras and may require bulkier and more expensive optics.

25 These and other needs are effectively addressed by the systems and methods of the present disclosure.

SUMMARY OF THE DISCLOSURE

30 Embodiments of the present disclosure include systems and methods for detecting and classifying 3D objects in a 3D scene in a range of illumination conditions including in low illumination conditions using three-dimensional (3D) integral imaging (InIm). Low illumination conditions are defined as having low light flux in the scene resulting in degraded captured images compromised by detector noise which may be very difficult to visualize and/or recognize. Low cost passive image sensors that operate in visible spectrum can be

used to capture elemental images of 3D scene in low illumination conditions. The image sensors may also operate beyond the visible spectrum. The elemental images are read-noise dominant. Additionally, 3D InIm is naturally optimal in the maximum likelihood sense for read-noise dominant image as this follows a Gaussian distribution. Upon 3D InIm

5 reconstruction of the 3D scene, signal-to-noise ratio (SNR) increases resulting in improved image visualization. The 3D scene can be de-noised with total-variation regularization using an augmented Lagrange approach (TV-denoising). Regions of interest can be identified and input into a trained convolutional neural network (CNN) for facial/object recognition. The regions/objects may partially occluded from view. The CNN can be pre-trained specifically
10 for object recognition in low-light conditions by training the CNN using 3D reconstructed images from elemental images obtained under low-light conditions. The claimed system can perform object detection and classification in low light levels using the CNN trained in the dark. As an example, an object can include physical objects such as human body parts. In this regard, the claimed system can perform face detection and recognition in low illumination
15 conditions.

In one embodiment, an object recognition system, method, and computer readable medium can detect objects in low illumination using 3D integral imaging. The system can include a computing system including a database and hosting a deep learning engine which implements a deep learning model and an object detection application. The system further
20 includes one or more image capturing devices in communication with the computing system. The one or more image capturing devices are configured to capture plurality of elemental images of a 3D scene including one or more objects and in low illumination levels, and transmit the plurality of elemental images to the computing system. The 3D scene and/or object(s) in the 3D scene may be partially occluded from view. The computing system is
25 configured to receive the plurality of elemental images of the 3D scene, reconstruct the 3D scene using the plurality of elemental images of the 3D scene, and de-noise the reconstructed 3D scene. Of note, the illumination conditions associated with the elemental images received by the computing system are unknown, i.e., illumination conditions are not transmitted to the computing system for use in downstream processing operations by the computing system.
30 The computing system is configured to input the 3D scene in the object detection application to identify one or more regions of interest in the 3D scene. The computing system is further configured to retrieve a set of training data from the database, and train the deep learning engine using the training data on at least one instance. The computing system can input the

identified regions of interest in the deep learning engine. The computing system, via the deep learning engine implementing the deep learning model can identify the one or more objects in the one or more regions of interest.

The object detection application can be configured to detect regions of interest including objects. The objects can include physical objects such as human body parts. As an example, the object detection application can detect regions of interest with human faces.

The deep learning model can be a CNN. The training data includes a set of elemental images including elemental images captured in various illumination levels below the threshold amount and having various levels of SNR. The training data further includes augmented (i.e., additional) data obtained by taking an original set of elemental images and perturbing them by adding Gaussian noise, and then translating the perturbed data. The computing system is configured to de-noise the reconstructed 3D scene using total-variation regularization using an augmented Lagrange approach. The computing system is configured to implement a virtual pinhole array. The computing system is configured to reconstruct the 3D image by back-propagating captured light rays from the plurality of elemental images through the virtual pinhole to a specified depth plane at a specified distance. The one or more image capturing devices is one or more of a lenslet array, an array of cameras, and a moving camera. The plurality of elemental images of the 3D scene include elemental images capturing the 3D scene from different perspectives. The one or more objects may include at least one human face. The plurality of elemental images are one or more of a still image and a moving image.

Additional advantageous features, functions and benefits of the present disclosure will be apparent from the description which follows, particularly when read in conjunction with the accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

Features and aspects of embodiments are described below with reference to the accompanying drawings, in which elements are not necessarily depicted to scale.

Exemplary embodiments of the present disclosure are further described with reference to the appended figures. It is to be noted that the various features, steps and combinations of features/steps described below and illustrated in the figures can be arranged and organized differently to result in embodiments which are still within the scope of the present disclosure. To assist those of ordinary skill in the art in making and using the disclosed assemblies,

systems and methods, reference is made to the appended figures, wherein:

Fig. 1 is a block diagram of the 3D scene capture and reconstruction in accordance with an exemplary embodiment;

Fig. 2 illustrates 3D reconstructed images for training a CNN in accordance with an exemplary embodiment;

Fig. 3 is a block diagram illustrating an object recognition system according to an exemplary embodiment;

Fig. 4 illustrates sample elemental images with corresponding SNR and photon/pixel in accordance with an exemplary embodiment;

Fig. 5 illustrates three-dimensional (3D) reconstructed images in accordance with an exemplary embodiment;

Fig. 6 illustrates graphs of SNR as illumination decreases and SNR as photons/pixels increase in accordance with an exemplary embodiment;

Fig. 7 illustrates a block diagram of an example computing device for implementing exemplary embodiments of the present disclosure;

Fig. 8 illustrates an exemplary process of the object recognition system in accordance with an exemplary embodiment; and

Fig. 9 illustrates an exemplary process of the object recognition system in accordance with an exemplary embodiment.

DETAILED DESCRIPTION OF THE DISCLOSURE

The exemplary embodiments disclosed herein are illustrative of methods and related systems. Described herein is an object recognition system that is operable in a range of illumination conditions including in low illumination conditions using 3D integral imaging. Low illumination conditions are defined as having low light flux in the scene resulting in degraded captured images compromised by detector noise which may be very difficult to visualize and/or recognize. The system includes a computing system including a database and hosting a deep learning engine, and an object detection application. The system further includes one or more image capturing devices in communication with the computing system. The one or more image capturing devices are configured to capture a plurality of elemental images of a 3D scene including one or more objects and in an illumination level lower than a

threshold amount, and transmit the plurality of elemental images to the computing system. The 3D scene and/or the object(s) in the 3D scene may be partially occluded from view. The computing system is configured to receive the plurality of elemental images of the 3D scene, retrieve a set of training data from the database, train the deep learning engine using the
5 training data on at least one instance, reconstruct the 3D scene using the plurality of or more elemental images of the 3D scene, de-noise the reconstructed 3D scene, input the 3D scene in the object detection application to identify one or more regions of interest in the 3D scene and identify the one or more objects in the 3D scene by inputting the one or more regions of interest of the 3D scene in the deep learning engine.

10 The object detection application can be configured to detect regions of interest including objects. The objects can include physical objects such as human body parts. As an example, the object detection application can detect regions of interest with human faces. As noted above, the region(s) of interest and/or object(s) may be partially occluded from view.

The claimed system can use integral imaging for object recognition in a 3D scene
15 which can perform efficiently in degraded environments because of the large number of apertures used to capture scene information which can mitigate the effects of degradations and obscurations of the scene. In this regard, object recognition under low light conditions is possible using the 3D reconstructed scene. Regions of interest of the 3D reconstructed scene can be found and inputted into a convolutional neural networks (CNN) (implemented by the
20 deep learning engine). This CNN can be pre-trained specifically for low-light facial recognition by training the CNN using 3D reconstructed images after de-noising from elemental images obtained under low-light conditions having different SNRs. As the CNN has detected and/or seen objects obtained under various illumination conditions, it is possible to quickly perform object recognition.

25 The claimed system can use 3D integral imaging to classify objects in a scene under low illumination conditions using passive imaging sensors that operate in the visible spectrum. The disclosed passive imaging sensors may operate in other spectra across the electromagnetic spectrum, i.e., beyond the visible spectra. For example, in addition to operating in the visible spectrum, the disclosed passive imaging sensors may operate in
30 spectra that include infrared, ultraviolet, x-ray wavelengths, or some combination thereof. The disclosed passive imaging sensors may also be effective for hyperspectral imaging. The 2D elemental images are read-noise dominant. Upon 3D InIm reconstruction, signal-to-noise ratio increases resulting in improved image visualization. The 3D scene can be de-noised

with total-variation regularization using an augmented Lagrange approach (TV-denoising). Regions of interest can be obtained to detect faces in the 3D scene using the Viola-Jones object detection framework which are then inputted into a trained CNN for facial/object recognition. The CNN can be trained specifically for low-light facial/object recognition by training the CNN using 3D reconstructed images after TV-denoising from elemental images obtained under low-light conditions having different SNRs.

Fig. 1 is a block diagram of the 3D scene 108 capture and reconstruction in accordance with an exemplary embodiment. In one embodiment, 3D integral imaging (InIm) can be used to capture a 3D scene 108. It can be appreciated that the 3D scene 108 can include humans and other physical objects. The 3D scene 108 can be a still or moving scene. Integral imaging uses many apertures such as an $M \times N$ array of image capturing devices 102 to capture the 3D scene 108 and create 3D reconstructions of the 3D scene where M and N denote the number of cameras in the x - and y - directions, respectively. Capturing the 3D scene 108 can be implemented in a 3D InIm pickup stage 100. The 3D InIm pickup stage 100 can include image capturing devices 100 such as a lenslet array, an array of cameras, or a moving camera to capture multiple 2D images of different perspectives of the 3D scene 108, known or elemental images. Each lens or lenslet 104 can contain a unique projective transformation that maps the 3D scene 108 onto the elemental images and is a function of the lens or lenslet 104 position and the focal length. For example, the lens or lenslet 104 can capture the 3D scene 108 at different angles and positions as shown by the projected rays 109. As a result, an array of inverted real images is formed on an image sensor 106 of the image capturing devices 102. The 3D InIm pickup stage 100 captures both intensity and angular information during the integral imaging.

Once the elemental images have been acquired, the 3D scene 108 can be reconstructed. During the 3D InIm reconstruction phase 110, the 3D scene 108 can be reconstructed by back-propagating (or back-projecting) the captured light rays 111 through a virtual pinhole 114 of a virtual pinhole array 112 to a particular depth plane a distance z away. As an example, the 3D InIm reconstruction phase 110 can include a depth plane 118 at depth 1 and a depth plane 119 at depth 2.

The 3D InIm reconstruction phase 110 can use a computer synthesized virtual pinhole array 112 for inverse mapping of each elemental image into the object space. The virtual pinhole array 112 can include virtual pinholes 114. Each of the elemental images are computationally overlapped afterward. The intensity distribution can be reconstructed at

arbitrary planes inside the 3D object volume (of the 3D objects in the 3D scene 108). The 3D scene 108 includes depth information that can be recovered by post-processing the elemental images. Each elemental image is back-projected through its own viewpoint on the particular depth plane (e.g., depth plane 118 or depth plane 119). The chief ray, R_i 120 from the object surface in 3D space (x, y, z) at location (x_0, y_0, z_0) with azimuth angle θ_i and zenith angle ϕ_i can be imaged by the i -th lens located at (x_l, y_l, z_l) and arriving at the sensor plane at (τ, ψ) .

In one embodiment, using the elemental images, the 3D InIm reconstruction phase 110 can be performed optically or computationally. Computational 3D InIm reconstruction is implemented as follows:

10

$$I(x, y; z) = \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} E^{k,b} \left(x - k \frac{L_x \times p_x}{c_x \times M}, y - b \frac{L_y \times p_y}{c_y \times M} \right), \quad (1)$$

where (x, y) is pixel index, z is reconstruction distance, $O(x, y)$ is the overlapping number on (x, y) , K and B are the total number of elemental images obtained in each column and row, respectively; $E^{k,b}$ is the elemental image in the k th column and b th row, L_x and L_y are the total number of pixels in each column and row, respectively, for each $E_{k,b}$, M is the magnification factor and equals z/g , g is the focal length, p_x and p_y is the pitch between image sensors, c_x and c_y are the size of the image sensor 106.

A captured image can be defined as $E(x, y) = I(x, y)r(x, y)$ where $I(x, y) > 0$ is the illumination factor and $r(x, y)$ is the reflection coefficient between 0 and 1. As the illumination for the 3D scene 108 decreases, the reflection coefficient diminishes. Moreover, read-noise becomes greater for the 3D scene 108 than the 3D scene 108 signal hindering adequate scene visualization. Accordingly, the captured image becomes read-noise dominant. Read-noise results from on-chip sensor noise, is additive, and can be modeled as a zero-mean Gaussian distribution. Using Equation 1, the 3D InIm reconstruction of the 3D scene with read-noise is:

$$\begin{aligned} I(x, y; z) &= \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} (E^{k,b}(x', y') + \varepsilon_r^{k,b}(x', y')), \\ &= \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} E^{k,b}(x', y') + \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} \varepsilon_r^{k,b}(x', y'), \end{aligned} \quad (2)$$

where $\epsilon_r^{k,b}(x', y')$ is zero mean additive Gaussian noise (i.e. read noise) for the elemental image in the k th column and b th row at location (x', y') , $x' = x - k(L_x \times p_x)/(c_x \times M)$ and $y' = y - k(L_y \times p_y)/(c_y \times M)$.

5 Taking the variance of Equation 2, the variance of the noise component for a fixed z , assuming that noise is wide sense stationary, is:

$$\text{var}\left(\frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} \epsilon_r^{k,b}(x', y')\right) = \frac{1}{O(x, y)} \sigma^2, \quad (3)$$

where $\text{var}(\cdot)$ is variance and σ^2 is the variance of the read-noise.

As the number of overlapping images increases, the variance or noise power, of the read noise decreases. 3D InIm reconstruction is naturally optimal in the maximum-likelihood
10 sense for read-noise limited images as the distribution is approximately Gaussian. Without photon counting devices to measure the flux density, the SNR of the image is estimated as:

$$SNR = \sqrt{(\langle g_o^2 \rangle - \langle N^2 \rangle) / \langle N^2 \rangle}, \quad (4)$$

where g_o is the average power of the object region in the elemental image, $\langle N^2 \rangle$ is average noise power defined as $\langle N^2 \rangle = (\Phi_o + \Phi_b)Q_e t + Dt + n_r^2$, where Φ_o and Φ_b are the photon flux of the object
15 and background (photons/pixel/second), D is dark current (electrons /pixel/second), Q_e is the quantum efficiency (electrons/photons), t is exposure time (seconds), n_r is read noise (RMS electrons/pixel/second), and $\langle \rangle$ denotes mean ensemble, respectively.

The number of photons per pixel ($N_{photons}$) can be estimated as:

$$\Phi_o t = N_{photons} \approx SNR \times n_r / Q_e, \quad (5)$$

20 where $N_{photons}$ is the estimated number of photons.

FIG. 2 is a block diagram illustrating the object recognition system 250 according to an exemplary embodiment. The object recognition system 250 can include one or more databases 205, one or more servers 210, one or more computing systems 200, and one or more image capturing devices 102. The computing system 200 can be in communication with
25 the databases 205, the server(s) 210, and the image capturing devices 102, via a communications network 215. In one embodiment, the image capturing devices 102 can be directly connected to the computing system 200. The computing system 200 can implement

at least one instance of a control engine 220. The routing engine can be an executable application configured to implement the object recognition system 250. The control engine 220 can also execute a deep learning engine 222. The deep learning engine 222 can implement deep learning algorithms (i.e., a deep learning model) to implement the object
5 recognition system 250. The computing system 200 can further execute a virtual pinhole array 112, and an object detection application 230. The object detection application can be configured to detect regions of interest including objects in a 3D scene. The objects can include physical objects such as human body parts. As an example, the object detection application can detect regions of interest with human faces. The region(s) of interest and/or
10 the object(s) may be partially occluded from view.

In an example embodiment, one or more portions of the communications network 215 can be an ad hoc network, an intranet, an extranet, a virtual private network (VPN), a local area network (LAN), a wireless LAN (WLAN), a wide area network (WAN), a wireless wide area network (WWAN), a metropolitan area network (MAN), a portion of the Internet, a
15 portion of the Public Switched Telephone Network (PSTN), a cellular telephone network, a wireless network, a WiFi network, a WiMax network, any other type of network, or a combination of two or more such networks.

The computing system 200 includes one or more processors configured to communicate with the databases 205 and image capturing devices 102 via the network 215.
20 The server 210 hosts one or more applications configured to interact with one or more components of the object recognition system 250. The databases 205 may store information/data, as described herein. For example, the databases 205 can include a training data database 225, a 3D scenes database 225, and a regions of interest (ROI) database 240. The training data database 225 can store information associated with training data for training
25 deep learning algorithms such as CNN. The 3D scene database 225 can store data associated with 3D scenes. The ROI database 240 can store data associated with regions of interest in a 3D scene. The databases 205 and server 210 can be located at one or more geographically distributed locations from each other or from the computing system 200. Alternatively, the databases 205 can be included within server 210 or computing system 200.

30 In one embodiment, the one or more image capturing devices 102 can capture a 3D scene by capturing a plurality of elemental images capturing different perspectives of the 3D

scene, using 3D InIm. As a non-limiting example, a 3D scene can be reconstructed using at least two elemental images. The 3D scene can be captured in an illumination level lower than a threshold amount. The low illumination levels can have low light flux. In one embodiment, the elemental images can be 2D. The one or more image capturing devices 102 can be one or
5 more of a lenslet array, an array of cameras, and a moving camera. The 3D scene can include one or more objects. The objects can include human faces. The one or more image capturing devices 102 can transmit the captured elemental images to the computing system 300.

The computing system 300 can execute the control engine 220 in response to receiving the elemental images of the 3D scene. The control engine 220 can implement the
10 virtual pinhole array 112 to reconstruct the 3D scene using the elemental images, using 3D InIm. As described above, the 3D scene can be reconstructed by back-propagating (or back-projecting) the captured light rays of the elemental images through a virtual pinhole of a virtual pinhole array 112 to a particular depth plane a specified distance away. In one embodiment, the control engine 220 can store the reconstructed 3D scene in the 3D scene
15 database 235. It can be appreciated, the 3D scene can be reconstructed optically and computationally.

The reconstructed 3D scene can have an increased SNR resulting in improved image visualization. The control engine 220 can de-noise the reconstructed 3D scene using total-variation regularization using an augmented Lagrange approach (TV-denoising). As an
20 example, the reconstructed 3D scene can have high Gaussian noise. Gaussian noise can be caused by camera sensor due to the low illumination conditions. The augmented Lagrange approach can be used for image restoration and to solve space-time TV minimization problems of motion blur and noise. By using the total-variation regularization using an augmented Lagrange approach on the reconstructed 3D scenes, the control engine 220 can
25 reduce and/or eliminate the blur and noise in the reconstructed 3D scene.

Once the reconstructed 3D scene has been de-noised, the control engine 220 can input the image in the object detection application 230 to identify regions of interest. The object detection application 230 can implement a Viola-Jones face detector. The Viola-Jones face detector uses a face detection style algorithm which combines motion and image intensity
30 information to detect regions of interest such as human faces in an image (or moving image). The Viola-Jones face detector uses rectangle filters to detect regions of interest such as

human faces in the image (or moving image). In response to the object detection application 230 detecting the regions of interest, the control engine 220 can input the regions of interest in the deep learning engine 222 to identify and classify the objects such as human faces. In one embodiment, the control engine 220 can store the regions of interest in the ROI database 240. As described above, it can be appreciated that the object detection application can be configured to detect regions of interest including objects. The objects can be any particular physical objects including human faces.

The deep learning engine 222 can implement a trained CNN. CNN is a type of deep learning architecture which can be used for analyzing imagery. CNN can be trained on low illumination data for object/face recognition from the regions of interest. Feature extraction when implementing CNN is not needed at the expense of computational complexity. The CNN can be trained using training data. For example, the control engine 220 can retrieve training data from the training data database 225. The training data can be training images were 3D InIm reconstructed images of faces/objects after TV-denoising obtained under different illumination conditions. Training in photon-starved environments can improve the classifier's (i.e., CNN's) ability to discern between different faces/objects in low illumination conditions.

As non-limiting example, training data can be generated by the image capturing devices 102 capturing elemental images of a 3D scene at varying illumination levels. The image capturing devices 102 can transmit the images to the computing system. The control engine 220 can reconstruct the 3D scene using the virtual pinhole array (as described above). The control engine 220 can de-noise the reconstructed 3D scene using TV-denoising. The reconstructed and de-noised 3D scene can be embodied as training data and can be used to train the CNN. The control engine 220 can store the training data in the training data database 225.

As a non-limiting example, normalized correlation can used to demonstrate the difficulty in discriminating faces under low illumination conditions. With reference to Fig. 3, image 300, 302, and 304 are images of people in low illumination conditions. Image 300 shows a 3D reconstruction reference image at $z = 4.5$ m after TV-denoising obtained using elemental images under an SNR of 10.41 dB. The image 300 can be correlated with 3D reconstructed images after TV-denoising whose elemental images were obtained under an

SNR of -12.75 dB, shown in image 302 (true class object) and 304 (false class object), with correlation values of 0.58 and 0.48, respectively. Accordingly, it is difficult to discriminate objects under low-light conditions without training the classifier information about what object appears in low-light. Image 300 can correlate well with both image 302 and image 304 even though the person in image 304 is a different than the person in image 300 and image 302.

With reference to Fig. 2, in response to inputting the regions of interest into the deep learning engine 222, the deep learning engine 222 can identify and classify faces/objects from the regions of interest. As a non-limiting example, the deep learning engine 222 can execute face recognition from the regions of interest. It can be appreciated the deep learning engine 222 can implement other deep learning or machine learning architectures.

As a non-limiting example, the image capturing devices 102 can be embodied as Allied Vision Mako-192 camera with 86.4 mm x 44 mm x 29 mm camera dimensions, in the object recognition system 250. Additionally, the sensor (image sensor 106 of the image capturing devices 102 as shown in FIG. 1) can be embodied as an e2v EV 76C570 and a CMOS sensor type. The F/# is F/1.8 can be a focal length of 50 mm, pixel size of 4.5 μ m x 4.5 μ m, sensor size of 7.2 mm (H) x 5.4 mm (V), and image size of 1600 (H) x 1200 (V). The camera read-noise can be 20.47 electrons/pixel/sec and the quantum efficiency at 525 nm can be 0.44 electron/photons. A gain of 0 can be used. The 3D InIm setup in this non-limiting example, includes 72 elemental images using 3 x 24 array with 10 mm (H) x 80 mm (V) a pitch and exposure time of 0.015 s.

In this non-limiting example, the conditions for capturing the 3D include low illumination conditions and include a 3D InIm setup with 6 subjects (i.e., objects/faces) located a distance 4.5 m away from the camera array (i.e., image capturing devices 102). Of note, the disclosed systems and methods are operable regardless of the distance of the object(s) from the image capturing device(s). Thus, the object(s) in a scene may be located an arbitrary distance within the field of view (FOV) of the imaging capturing device(s) according to the disclosed systems and methods, i.e., the disclosed systems/methods are not dependent on the object(s) being a predetermined distance from the imaging capturing device(s) for the disclosed system/method to be efficacious.

Each subject can be captured under different illumination conditions resulting in different SNR levels. The illumination conditions can be altered by adjusting the intensity of a light source. With reference to FIGS. 4 and 5, image 400 in FIG. 4 depicts the elemental image with an SNR of 10.41 dB (i.e. good illumination) and image 500 in FIG. 5 shows the
 5 3D reconstructed image (of the image 400 in FIG. 4) at $z = 4.5$ m with an SNR of 12.39 dB.

Prior to 3D reconstruction, the elemental images can be registered and aligned due to the conditions (e.g., unbalanced camera array). Fifty bias frames can be taken and averaged for each camera in the camera array and subtracted from the elemental images. The SNR can be computed by assuming $\langle g_o^2 \rangle$ is the object (i.e., the person's face) while $\langle N^2 \rangle$ is an area of
 10 scene that is completely dark. The elemental images acquired using the 3D InIm under low illumination conditions are shown in FIG. 4 images 402-410. In FIG 4 image 402, the SNR can be 1.20 dB with approximately 48.75 photons/pixel on the object. The person captured is still visible. In FIG. 4 image 404, the SNR decreases to -9.13 dB with 19.26 photons/pixel. The average power of object is lower than the noise power for the images shown in
 15 FIG. 2 (d)-(f). As a result, SNR cannot be computed as $\langle N^2 \rangle > \langle g_o^2 \rangle$ resulting in an imaginary number in Equation 4.

With reference to FIG. 5, 3D reconstructed images corresponding to the elemental images shown in FIG. 4 images 400-410, at $z = 4.5$ m are shown. In FIG. 5 images 502-510, the SNR increases to 8.93 dB, 0.96 dB, -5.28 dB, -9.67 dB, and -12.38 dB, respectively.
 20 Moreover, the corresponding photons/pixel for FIG. 5 images 502-510 can be 130.02, 51.95 photons/pixel, 25.34 photons/pixel, 15.27 photons/pixel, and 11.18 photons/pixel, respectively.

With reference to FIG. 6 depicts a first graph 600 of SNR as a function of decreasing illumination and a second graph 602 displaying SNR (in dB) as a function of the number of
 25 photons/pixel. In the first graph 600, the SNR of the 3D reconstructed images is higher than that of the 2D elemental images. The SNR may not be computed for elemental images with SNRs below -21.35 dB as the noise can be greater than the signal. With reference to the second graph 602, the 3D reconstructed images have a higher number of photons/pixel relative to their corresponding 2D elemental image.

Continuing with the non-limiting example, a CNN implemented by the deep learning engine 222, can be trained to perform facial recognition using data from the 3D InIm reconstructed data. A data set (i.e., training data set retrieved from the training data database 225) can include 6 different subjects, and 17 different illumination conditions acquired using the 3D InIm. The images can be then computationally reconstructed over distances of 4 m to 5 meters with a step size of 50 mm where the true object distance can be 4.5 meter. The dataset can be split into testing and training whereas 4 randomly chosen illumination conditions having SNRs of approximately -1.41 dB, -8.322 dB, -8.971 dB, and -12.75 dB may not be used to train the classifier and the other 13 illumination conditions were used for training. The training images can be grayscale images of size 256 x 256 pixels and can be perturbed by adding additive Gaussian noise with mean 0 and standard deviation of 0.01, 0.05 and 0.1, rotated -1, -0.5, 0.5 and 1 degrees, and translated -5, -3, 3, and 5 pixels in both the x- and y- directions generating a total of 29,232 images. The data can be then de-noised using total-variation regularization using an augmented Lagrange approach with a regularization parameter of 20000 (as shown in FIG. 3 302). The CNN can include a convolution layer [13 x 13, 20 filters], rectified linear unit layer (ReLU), 2 x 2 max pooling, convolution layer [11 x 11, 20 filters], ReLU, 2 x 2 max pooling, fully connected layer, and a SoftMax layer [6 outputs]. For training, stochastic gradient descent can be used with a learning rate of .0001 and a maximum of 10 epoch used along with the cross-entropy metric to evaluate model performance. In total, the model can take approximately 4 hours to train on a high performance computer utilizing a GPU Tesla K40m running CUDA 8.0 and implemented using MATLAB.

For classification, regions of interest of the 3D reconstructed image including of information from 72 elemental images can be considered to reduce the size of the input image. The 3D reconstructed image can be de-noised using total-variation regularization using an augmented Lagrange approach (i.e., TV-denoising) with a regularization parameter of 20000. The object detection application 230 implementing the Viola-Jones face detector can be used to find regions of interest. The regions of interest can be input into the pre-trained CNN classifier. This process can be repeated over all z . In the event, the same face appears in the same region and detected over multiple depths, the face with the highest mean value can be assumed to be the correct reconstruction depth as. The correct reconstruction depth is assumed to be signal plus noise whereas incorrect depths can be considered noise.

Table 1. summarizes the results of classifying the regions of interest using CNN.

		Predicted					
		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Actual	Class 1	4	0	0	0	0	0
	Class 2	0	4	0	0	0	0
	Class 3	0	0	4	0	0	0
	Class 4	0	0	0	4	0	0
	Class 5	0	0	0	0	4	0
	Class 6	0	0	0	0	0	4

Table 1. Face detection results using CNN under low illumination conditions using 3D reconstructed images

FIG. 7 is a block diagram of an exemplary computing device suitable for use in an embodiment. The computing device 700 may be, but is not limited to, a smartphone, laptop, tablet, desktop computer, server or network appliance. The computing device 700 can be embodied as part of the computing system. The computing device 700 includes one or more non-transitory computer-readable media for storing one or more computer-executable instructions or software for implementing exemplary embodiments. The non-transitory computer-readable media may include, but are not limited to, one or more types of hardware memory, non-transitory tangible media (for example, one or more magnetic storage disks, one or more optical disks, one or more flash drives, one or more solid state disks), and the like. For example, memory 706 included in the computing device 700 may store computer-readable and computer-executable instructions or software (e.g., applications 730 such as the control engine 220, deep learning engine 222, virtual pinhole array 112, and object detection application 230) for implementing exemplary operations of the computing device 700. The computing device 700 also includes configurable and/or programmable processor 702 and associated core(s) 704, and optionally, one or more additional configurable and/or programmable processor(s) 702' and associated core(s) 704' (for example, in the case of computer systems having multiple processors/cores), for executing computer-readable and computer-executable instructions or software stored in the memory 706 and other programs for implementing exemplary embodiments of the present disclosure. Processor 702 and processor(s) 702' may each be a single core processor or multiple core (704 and 704') processor. Either or both of processor 702 and processor(s) 702' may be configured to execute one or more of the instructions described in connection with computing device 700.

Virtualization may be employed in the computing device 700 so that infrastructure and resources in the computing device 700 may be shared dynamically. A virtual machine 712 may be provided to handle a process running on multiple processors so that the process appears to be using only one computing resource rather than multiple computing resources.

5 Multiple virtual machines may also be used with one processor.

Memory 706 may include a computer system memory or random access memory, such as DRAM, SRAM, EDO RAM, and the like. Memory 706 may include other types of memory as well, or combinations thereof.

10 A user may interact with the computing device 700 through a visual display device 714, such as a computer monitor, which may display one or more graphical user interfaces 716, multi touch interface 720, a pointing device 718, an image capturing device 734 and a scanner 732.

15 The computing device 700 may also include one or more computer storage devices 726, such as a hard-drive, CD-ROM, or other computer readable media, for storing data and computer-readable instructions and/or software that implement exemplary embodiments of the present disclosure (e.g., applications). For example, exemplary storage device 726 can include one or more databases 728 for storing information regarding training data, 3D scenes, and objects. The databases 728 may be updated manually or automatically at any suitable time to add, delete, and/or update one or more data items in the databases.

20 The computing device 700 can include a network interface 708 configured to interface via one or more network devices 724 with one or more networks, for example, Local Area Network (LAN), Wide Area Network (WAN) or the Internet through a variety of connections including, but not limited to, standard telephone lines, LAN or WAN links (for example, 802.11, T1, T3, 56kb, X.25), broadband connections (for example, ISDN, Frame Relay, 25 ATM), wireless connections, controller area network (CAN), or some combination of any or all of the above. In exemplary embodiments, the computing system can include one or more antennas 722 to facilitate wireless communication (e.g., via the network interface) between the computing device 700 and a network and/or between the computing device 700 and other computing devices. The network interface 708 may include a built-in network adapter, 30 network interface card, PCMCIA network card, card bus network adapter, wireless network adapter, USB network adapter, modem or any other device suitable for interfacing the

computing device 700 to any type of network capable of communication and performing the operations described herein.

The computing device 700 may run any operating system 710, such as versions of the Microsoft® Windows® operating systems, different releases of the Unix and Linux operating systems, versions of the MacOS® for Macintosh computers, embedded operating systems, real-time operating systems, open source operating systems, proprietary operating systems, or any other operating system capable of running on the computing device 700 and performing the operations described herein. In exemplary embodiments, the operating system 710 may be run in native mode or emulated mode. In an exemplary embodiment, the operating system 710 may be run on one or more cloud machine instances.

FIG. 8 is a flowchart illustrating an exemplary process of the object detection system in accordance with an exemplary embodiment. In operation 800, one or more image capturing devices (e.g., image capturing devices 102 as shown in FIGS. 1-2) can capture a plurality of elemental images of a 3D scene (e.g., 3D scene 108 as shown in FIG. 1) including one or more objects. The 3D scene can be captured in an illumination level than a threshold amount. In operation 802, the one or more image capturing devices can transmit the elemental images to a computing system. The computing system can include a database host a deep learning engine (e.g., deep learning engine 222 as shown in FIG. 2) and an object detection application 230 (e.g., object detection application 230 as shown in FIG. 2). In operation 804, the computing system can receive the plurality of elemental images of the 3D scene. The illumination conditions associated with the elemental images received by the computing system are unknown, i.e., illumination conditions are not transmitted to the computing system for use in downstream processing operations by the computing system. In operation 806, the computing system can reconstruct the 3D scene using the plurality of elemental images of the 3D scene. In operation 808, the computing system can de-noise the reconstructed 3D scene. In operation 810, the computing system can input the 3D scene in an object detection application to identify one or more regions of interest in the 3D scene. In operation 812, the computing system can retrieve a set of training data from the database (e.g., training data database 225 as shown in FIG. 2). In operation 814, the computing system can train the deep learning engine using the training data on at least one instance. Accordingly, once the deep learning engine is trained, it generally need not be retrained for future implementations of the disclosed system/method. In operation 816, the computing

system can input the one or more identified regions of interest in the deep learning engine. In operation 818, the computing system, via the deep learning engine, can identify the one or more objects the one or more regions of interest.

FIG. 9 is a flowchart illustrating an exemplary process of the object detection system in accordance with an exemplary embodiment. The object detection system can be implementing by executing process 900 and process 902. In process 900 and in operation 904, one or more image capturing devices (e.g., image capturing devices 102 as shown in FIG. 1 and 2) can capture elemental images of a 3D scene (e.g., 3D scene 108 as shown in FIG. 1) in various perspectives and transmit the elemental images to a computing system (e.g., computing system 200 as shown in FIG. 2). The computing system can host virtual pinhole array (e.g., virtual pinhole array 112), a deep learning engine (e.g., deep learning engine 222 as shown in FIG. 2), and an object detection application (e.g., object detection application 230 as shown in FIG. 2). In operation 906, the computing system can reconstruct the 3D scene using the elemental images by back-propagating captured light rays from the plurality of elemental images through virtual pinholes of a virtual pinhole array to a specified depth plane at a specified distance. In operation 908, the computing system can de-noise the reconstructed 3D scene. In operation 910, the computing system can identify regions of interest in the reconstructed and de-noised 3D scene using the object detection application.

With reference to process 902, prior to operation 918 of process 900, the computing system can implement process 902 of training a CNN being implemented by the deep learning engine. In operation 912, image capturing devices can capture elemental images of a 3D scene at various illumination levels. In operation 914, the computing system can reconstruct the 3D scene using the elemental images captured in various illumination levels. In operation 916, the reconstructed 3D scene is de-noised and can be embodied as training data. In operation 918, the CNN is trained using the reconstructed and de-noised 3D scene.

Turning back to process 900, in operation 920, the identified regions of interest can be input in the trained CNN being implemented by the deep learning engine. In operation 922, the deep learning engine can identify and classify objects and/or human faces in the regions of interest.

All documents cited herein and the following listed document, all referenced publications cited therein, and the descriptions and information contained in these documents

are expressly incorporated herein in their entirety to the same extent as if each document or cited publication was individually and expressly incorporated herein:

5 A. Markman and B. Javidi unpublished document entitled "Learning in the dark: 3D Integral Imaging object recognition in very low illumination conditions using Convolutional Neural Networks," [8 pages].

10 In describing exemplary embodiments, specific terminology is used for the sake of clarity. For purposes of description, each specific term is intended to at least include all technical and functional equivalents that operate in a similar manner to accomplish a similar purpose. Additionally, in some instances where a particular exemplary embodiment includes
15 a multiple system elements, device components or method steps, those elements, components or steps may be replaced with a single element, component or step. Likewise, a single element, component or step may be replaced with multiple elements, components or steps that serve the same purpose. Moreover, while exemplary embodiments have been shown and described with references to particular embodiments thereof, those of ordinary skill in the art
20 will understand that various substitutions and alterations in form and detail may be made therein without departing from the scope of the present disclosure. Further still, other aspects, functions and advantages are also within the scope of the present disclosure.

20 Exemplary flowcharts are provided herein for illustrative purposes and are non-limiting examples of methods. One of ordinary skill in the art will recognize that exemplary methods may include more or fewer steps than those illustrated in the exemplary flowcharts, and that the steps in the exemplary flowcharts may be performed in a different order than the order shown in the illustrative flowcharts.

References

1. N. Levin, Q. Zhang, "A global analysis of factors controlling VIIRS nighttime light levels from densely populated areas," *Remote Sens. Res.* 190, 366-382 (2017).
2. B. Phillips, D. Gruber, G. Vasan, C. Roman, V. Pieribone, and J. Sparks,
5 "Observations of in situ deep-sea marine bioluminescence with a high-speed, high-resolution sCMOS camera," *DEEP-SEA RES PT 1* 111, 102-109 (2016).
3. D. Boening, T. Groemer, and J. Klingauf, "Applicability of an EM-CCD for spatially resolved TIR-ICS," *Opt. Express* 18(13), 13516-13528 (2010).
4. Z. Petrášek and K. Suhling, "Photon arrival timing with sub-camera exposure time
10 resolution in wide-field time-resolved photon counting imaging," *Opt. Express* 18(24), 24888-24901 (2010).
5. G. Lippmann, "Épreuves réversibles donnant la sensation du relief," *J. Phys. Theory Appl.* 7(1), 821-825 (1908).
6. J. Jang and B. Javidi, "Three-dimensional synthetic aperture integral imaging," *Opt.*
15 *Lett.* 27(13), 1144-1146 (2002).
7. A. Llavador, E. Sánchez-Ortiga, G. Saavedra, B. Javidi, and M. Martínez-Corral, "Free-depths reconstruction with synthetic impulse response in integral imaging," *Opt. Exp.* 23(23), 30127-30135 (2015).
8. F. Okano, H. Hoshino, J. Arai, and I. Yuyama, "Real-time pickup method for a 3D
20 image based on integral photography," *Appl. Opt.* 36(7), 1598-1603 (1997).
9. H. Hoshino, F. Okano, H. Isono, and I. Yuyama, "Analysis of resolution limitation of integral photography," *J. Opt. Soc. Am. A* 15(8), 2059-2065 (1998).
10. M. Yamaguchi and R. Higashida, "3D touchable holographic light-field display," *Appl. Opt.* 55(3), A178-A183 (2016).
- 25 11. B. Tavakoli, B. Javidi, and E. Watson, "Three dimensional visualization by photon counting computational integral imaging," *Opt. Express.* 16(7), 4426-4436 (2008).
12. A. Stern, D. Doron and B. Javidi, "Experiments with three-dimensional integral imaging under low light levels," *IEEE Photon. J.* 4(4), 1188-1195 (2012).
13. A. Markman, X. Shen, and B. Javidi, "Three-dimensional object visualization and
30 detection in low light illumination using integral imaging," *Opt. Lett.* 42(16), 3068-3017 (2017).

14. S. Chan, R. Ramsin, K. Gibson, P. Gill, and T. Nguyen, "An augmented Lagrangian method for total variation video restoration," *IEEE Trans. Image Process.* 20(11), 3097-3111 (2011).
- 5 15. P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *CVPR* 63(2), 153-161 (2005).
16. A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *NIPS*, 1097-1105 (2012).
17. S. Lawrence, C. Giles, A. Chung Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.* 8(1), 98-113
10 (1997).
18. R. Gonzalez and R. Woods, *Digital Image Processing* (Pearson, 2008).
19. F. Sadjadi and A. Mahalanobis, "Automatic Target Recognition," *Proc. SPIE* 10648 (2018).

Claims:

1. An object recognition system operable in a range of illumination conditions including in low illumination conditions using 3D integral imaging, the system comprising:

5 a computing system including a database and hosting a deep learning engine, and an object detection application;

one or more image capturing devices in communication with the computing system, the one or more image capturing devices configured to:

capture a plurality of elemental images of a 3D scene in low illumination conditions having low light flux, the 3D scene including one or more objects;

10 transmit the plurality of elemental images to the computing system;

wherein the computing system is configured to:

retrieve a set of training data from the database;

train the deep learning engine using the retrieved training data on at least one instance;

15 wherein the computing system is further configured to:

receive the plurality of elemental images of the 3D scene;

reconstruct the 3D scene using the plurality of elemental images of the 3D scene;

de-noise the reconstructed 3D scene;

20 input the reconstructed and de-noised 3D scene in the object detection application to identify one or more regions of interest in the reconstructed and de-noised 3D scene;

input the one or more regions of interest in the deep learning engine; and

25 identify, via the deep learning engine that has been trained on the at least one instance, the one or more objects in the one or more regions of interest.

2. The system of claim 1, wherein the deep learning engine implements a Convolutional Neural Network (CNN).

3. The system of claim 2, wherein the training data includes (i) a set of elemental images including elemental images captured in various low illumination levels and having various levels of low signal-to-noise ratio (SNR), and (ii) augmented data obtained by perturbing elemental images with Gaussian noise.

5 4. The system of claim 1, wherein the computing system is configured to denoise the reconstructed 3D scene using a denoising algorithm.

5. The system of claim 4, wherein the denoising algorithm comprises total-variation regularization using an augmented Lagrange approach.

10 6. The system of claim 1, wherein the computing system is configured to implement a virtual pinhole array for reconstructing the 3D scene.

7. The system of claim 6, wherein the computing system is configured to reconstruct the 3D scene by back-propagating captured light rays from the plurality of elemental images through virtual pinholes of the virtual pinhole array to a specified depth plane at a specified distance.

15 8. The system of claim 1, wherein the one or more image capturing devices is one or more of a lenslet array, an array of cameras, or a moving camera.

9. The system of claim 1, wherein the plurality of elemental images of the 3D scene include elemental images capturing the 3D scene from different perspectives.

20 10. The system of claim 1, wherein the one or more objects include at least one human face.

11. The system of claim 1, wherein the one or more identified regions of interest include the one or more objects.

12. The system of claim 1, wherein the image capturing devices operate in multiple bands across the electromagnetic spectrum.

25 13. The system of claim 12, wherein the image capturing devices operate in the visible spectrum and at least one additional spectrum selected from the group consisting of infrared, ultraviolet and x-ray, and combinations thereof.

14. The system of claim 1, wherein the 3D scene or at least one of the one or more objects in the 3D scene is partially occluded from view.

15. The system of claim 1, wherein the one or more objects in the 3D scene may be located any distance from the one or more imaging capturing devices provided the
5 distance is within the field of view of the one or more imaging capturing devices.

16. The system of claim 1, wherein the computing system does not receive the illumination conditions associated with the elemental images of the 3D scene.

17. An object recognition method operable in a range of illumination conditions including in low illumination conditions using 3D integral imaging, the system comprising:
10 capturing, via one or more image capturing devices in communication with a computing system, a plurality of elemental images of a 3D scene in low illumination conditions having low light flux, the 3D scene including one or more objects;
transmitting, via the one or more image capturing devices, the plurality of elemental images to the computing system;
15 receiving, via the computing system including a database and hosting a deep learning engine, and an object detection application, the plurality of elemental images of the 3D scene;
reconstructing, via the computing system, the 3D scene using the plurality of elemental images of the 3D scene;
de-noising, via the computing system, the reconstructed 3D scene;
20 inputting, via the computing system, the reconstructed and de-noised 3D scene in the object detection application to identify one or more regions of interest in the reconstructed and de-noised 3D scene;
retrieving, via the computing system, a set of training data from the database;
training, via the computing system, the deep learning engine using the training data on
25 at least one instance;
inputting, via the computing system, the regions of interest in the deep learning engine;

identifying, via the deep learning engine hosted on the computing system, the one or more objects in the one or more regions of interest.

18. The method of claim 17, wherein the deep learning engine implements a Convolutional Neural Network (CNN).

5 19. The method of claim 18, wherein the training data includes (i) a set of elemental images including elemental images captured in various low illumination levels and having various levels of low signal-to-noise ratio (SNR), and (ii) augmented data obtained by perturbing elemental images with Gaussian noise.

10 20. The method of claim 17, further comprising de-noising, via the computing system, the reconstructed 3D scene using a denoising algorithm.

21. The method of claim 20, wherein the denoising algorithm comprises total-variation regularization using an augmented Lagrange approach.

22. The method of claim 17, further comprising implementing, via the computing system, a virtual pinhole array for reconstructing the 3D scene.

15 23. The method of claim 22, further comprising reconstructing, via the computing system, the 3D scene by back-propagating captured light rays from the plurality of elemental images through virtual pinholes of a virtual pinhole array to a specified depth plane at a specified distance.

20 24. The method of claim 17, wherein the one or more image capturing devices is one or more of a lenslet array, an array of cameras, or a moving camera.

25. The method of claim 17, wherein the plurality of elemental images of the 3D scene include elemental images capturing the 3D scene from different perspectives.

26. The method of claim 17, wherein the one or more objects include at least one human face.

25 27. The method of claim 17, wherein the one or more identified regions of interest include the one or more objects.

28. The method of claim 17, wherein the image capturing devices operate in multiple bands across the electromagnetic spectrum.

29. The method of claim 28, wherein the image capturing devices operate in the visible spectrum and at least one additional spectrum selected from the group consisting of
5 infrared, ultraviolet and x-ray, and combinations thereof.

30. The method of claim 17, wherein the 3D scene or at least one of the one or more objects in the 3D scene is partially occluded from view.

31. The method of claim 17, wherein the one or more objects in the 3D scene may be located any distance from the one or more imaging capturing devices provided the
10 distance is within the field of view of the one or more imaging capturing devices.

32. The method of claim 17, wherein the computing system does not receive the illumination conditions associated with the elemental images of the 3D scene.

33. A non-transitory computer readable memory medium storing instructions, wherein the instructions are executable by a processor to:

15 capture, via one or more image capturing devices in communication with a computing system, a plurality of elemental images of a 3D scene including one or more objects and in an illumination level lower than a threshold amount;

transmit, via the one or more image capturing devices, the plurality of elemental images to the computing system;

20 receive, via the computing system including a database and hosting a deep learning engine, and an object detection application, the plurality of elemental images of the 3D scene;

reconstruct, via the computing system, the 3D scene using the plurality of elemental images of the 3D scene;

de-noise, via the computing system, the reconstructed 3D scene;

25 input, via the computing system, the reconstructed and de-noised 3D scene in the object detection application to identify one or more regions of interest in the reconstructed and de-noised 3D scene;

retrieve, via the computing system, a set of training data from the database;

input, via the computing system, the regions of interest in the deep learning engine;
and

identify, via the deep learning engine hosted on the computing system, the one or more objects in the one or more regions of interest.

5

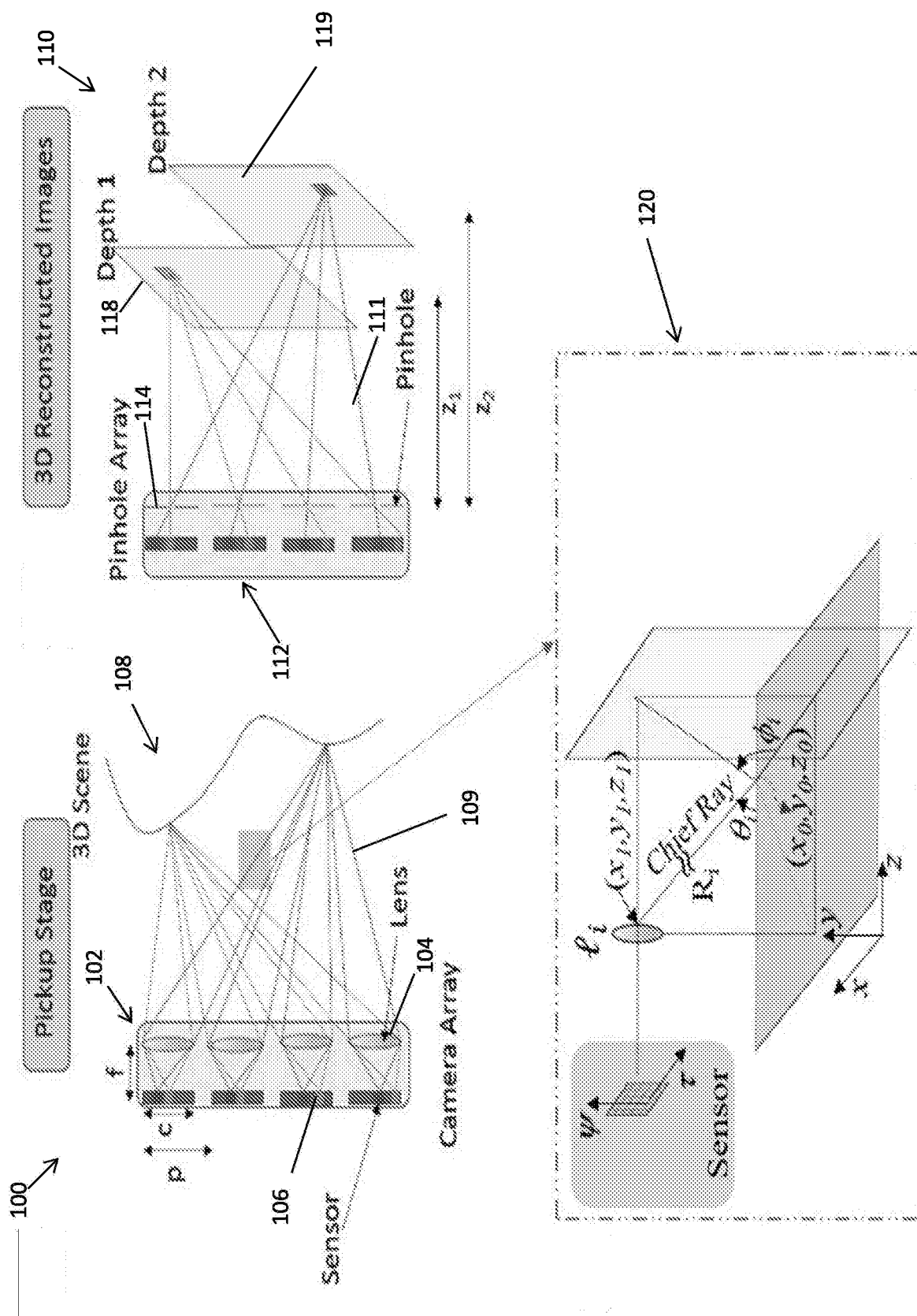


FIG. 1

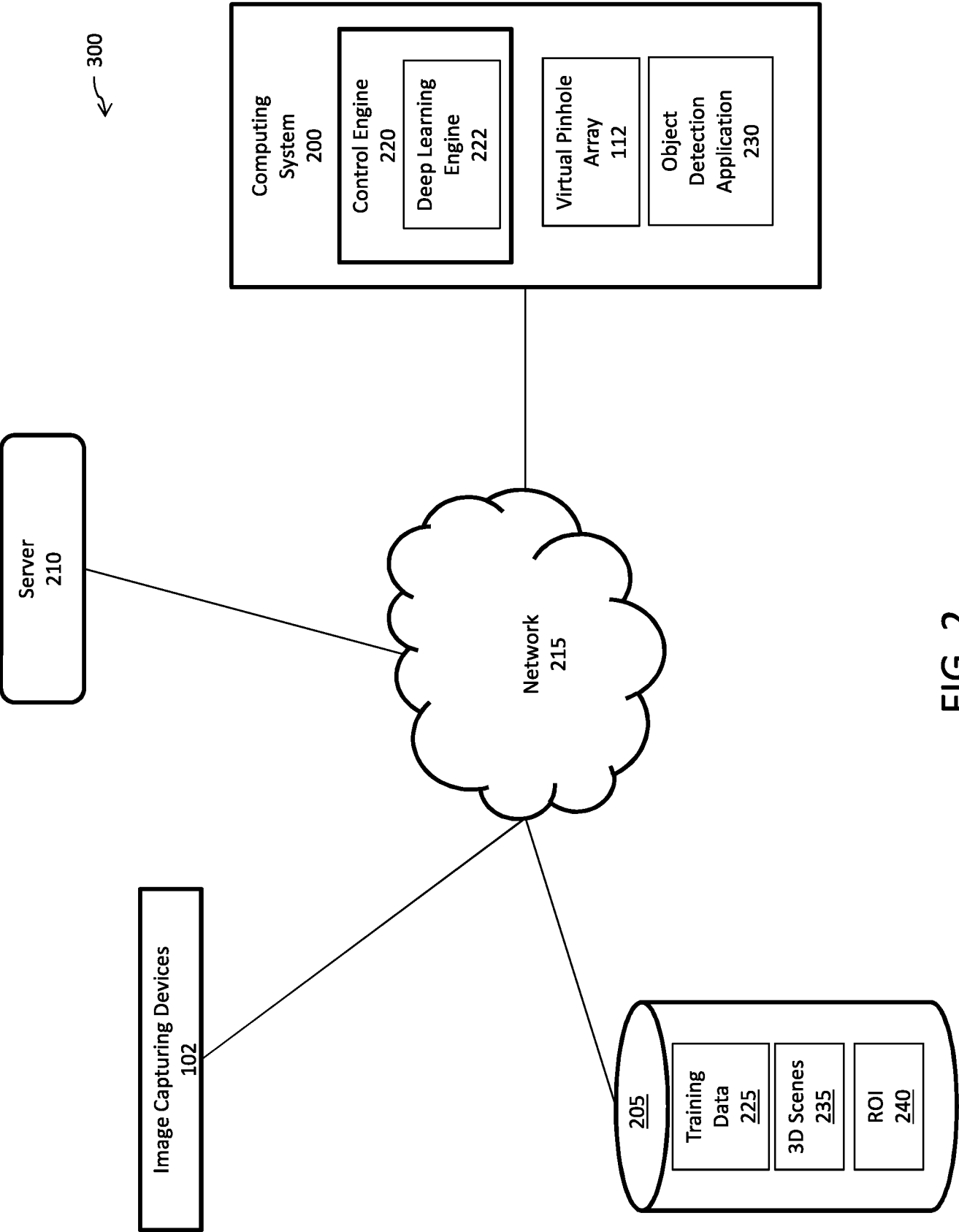


FIG. 2

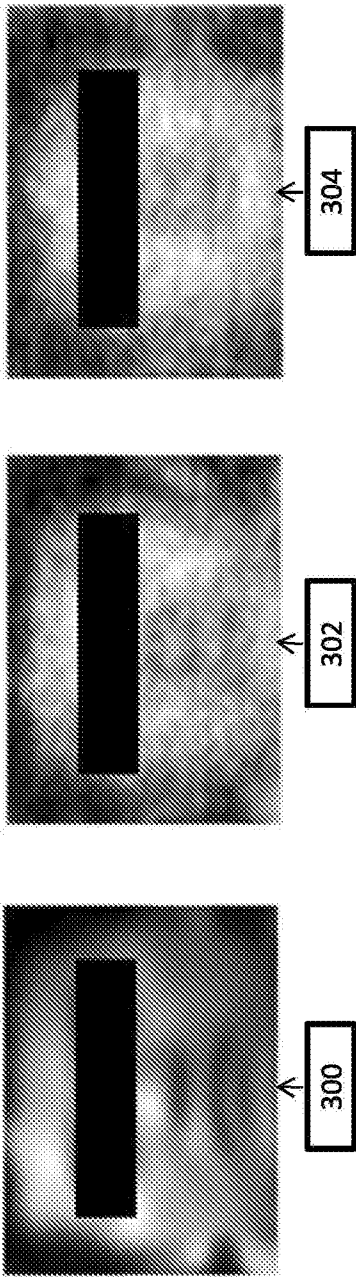


FIG. 3

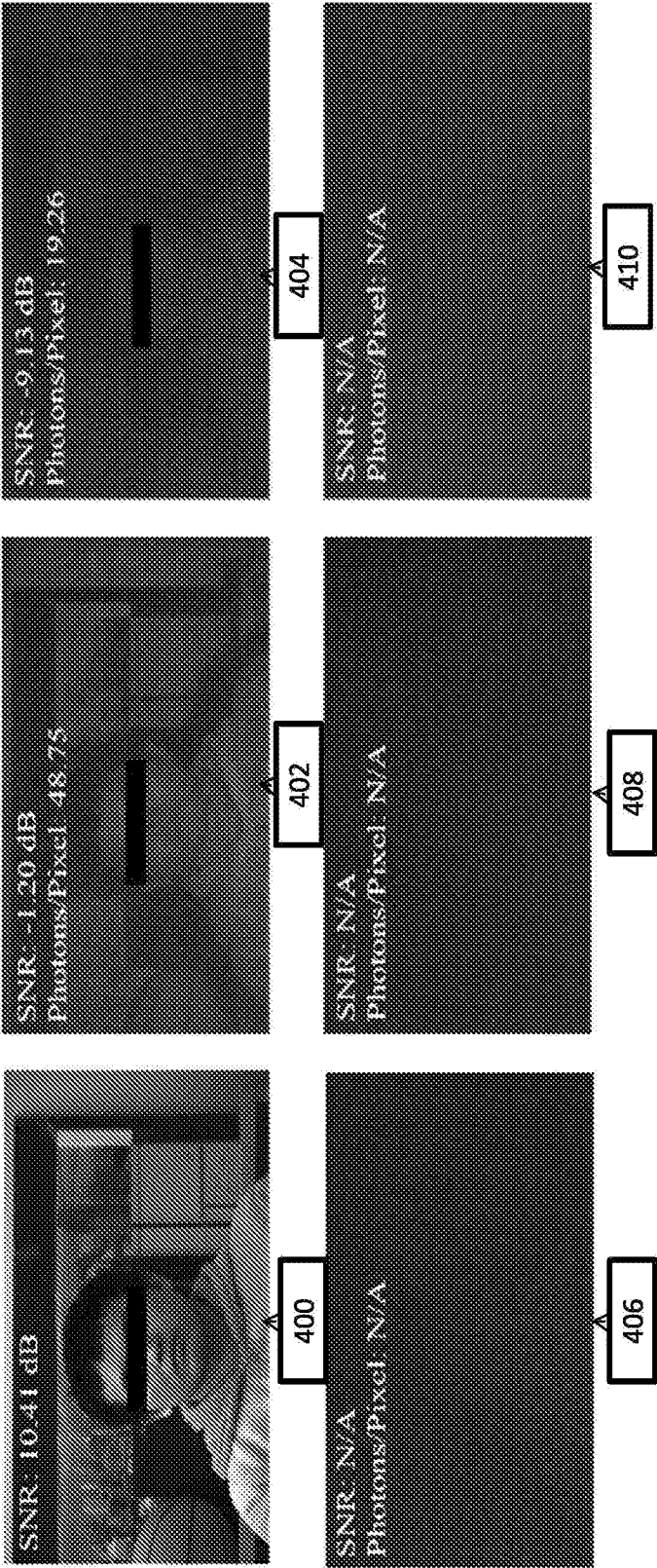


FIG. 4

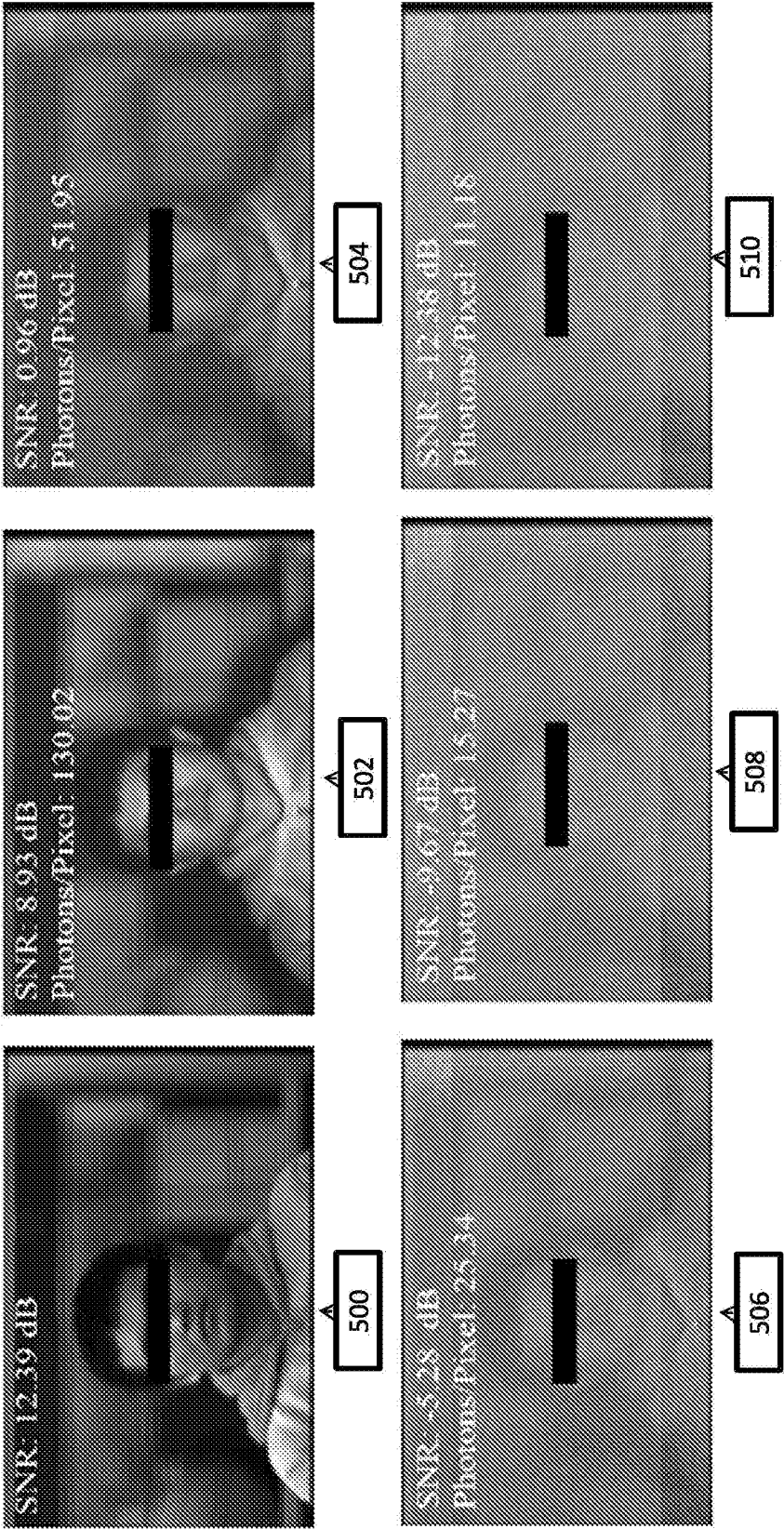


FIG. 5

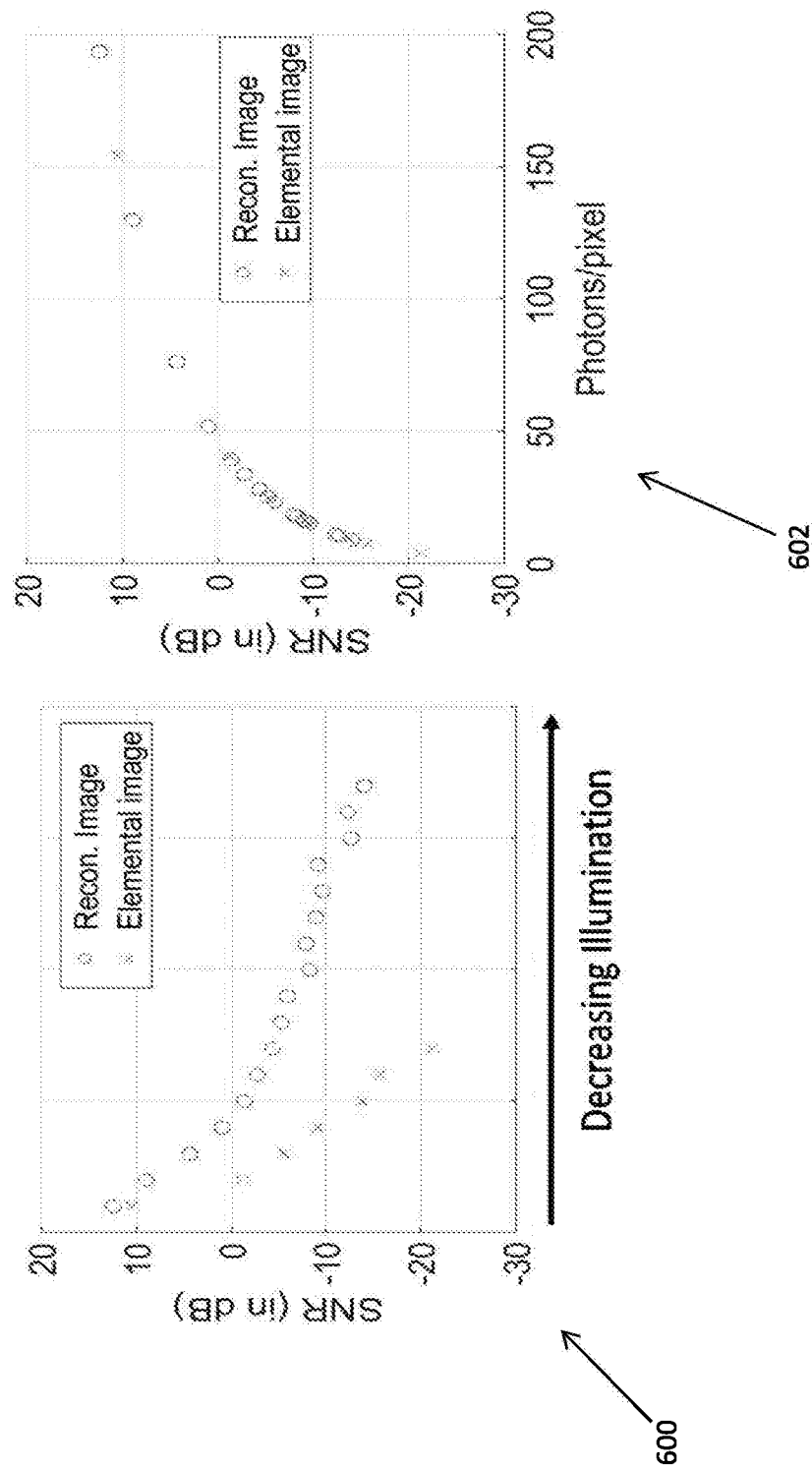


FIG. 6

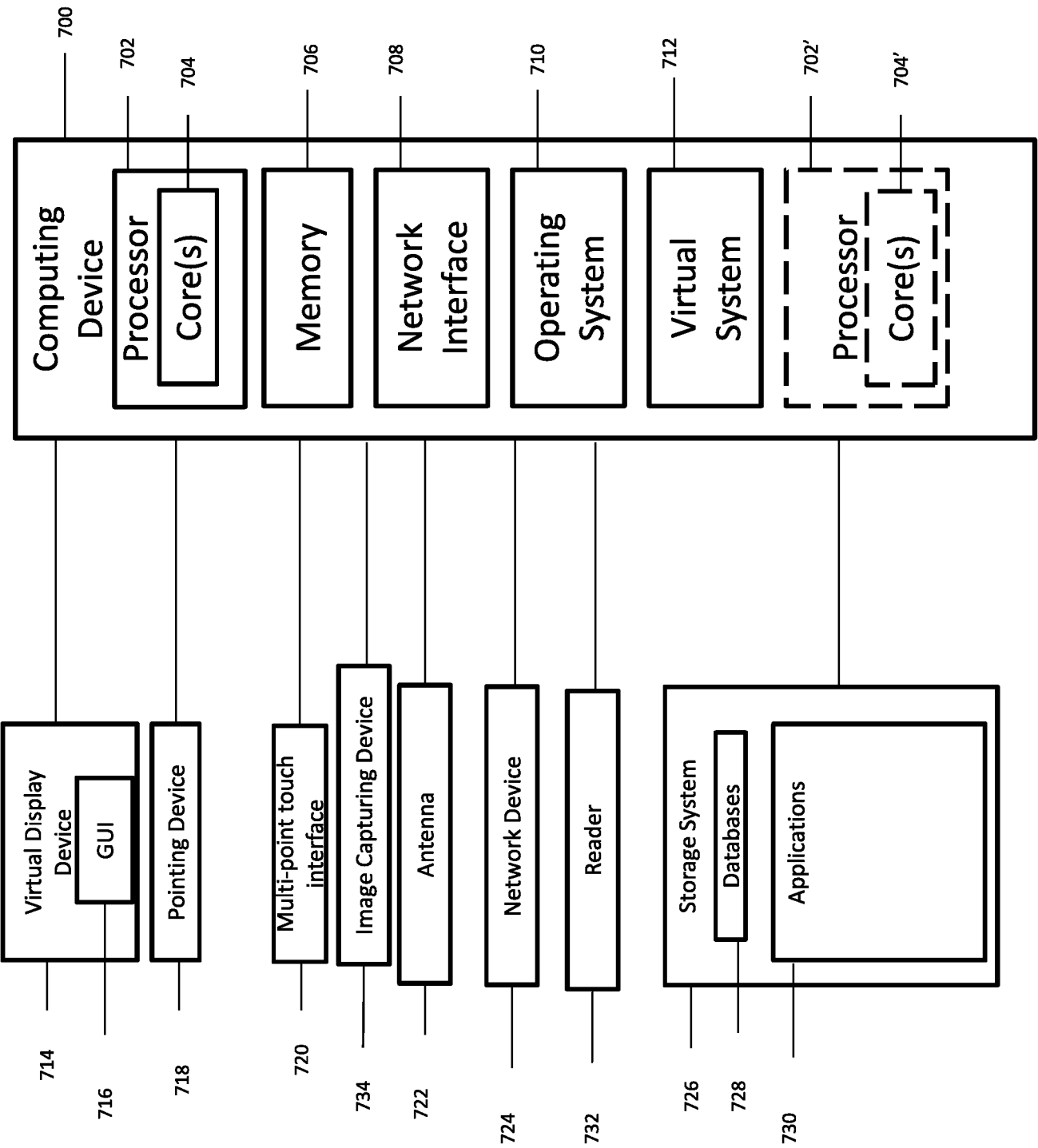


FIG. 7



FIG. 8

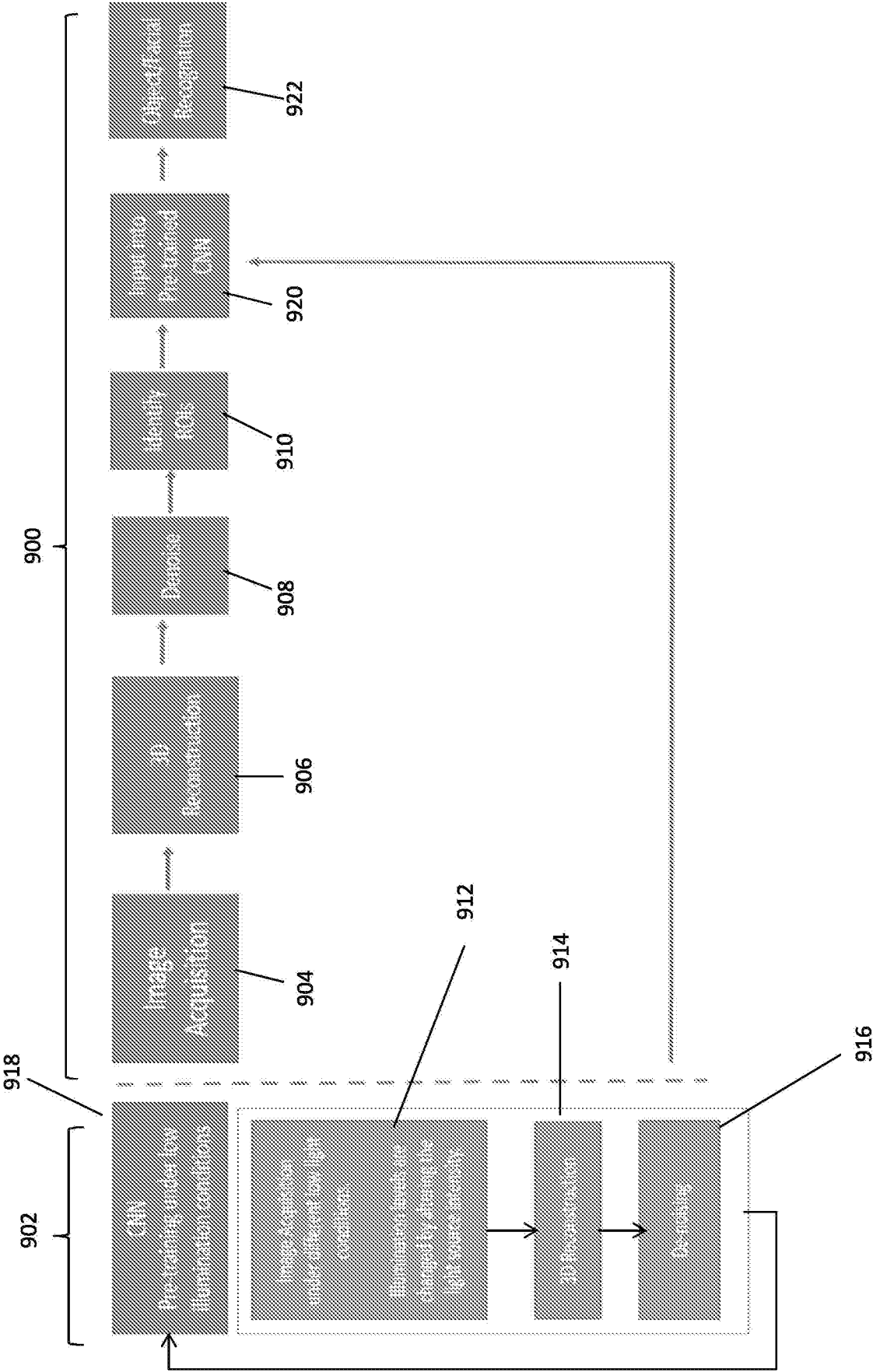


FIG. 9