

(51) International Patent Classification:
G06V 20/59 (2022.01) *G06V 40/20* (2022.01)
G06V 40/10 (2022.01)

(21) International Application Number:
PCT/US2024/036754

(22) International Filing Date:
03 July 2024 (03.07.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
18/349,853 10 July 2023 (10.07.2023) US

(71) Applicant: **NVIDIA CORPORATION** [US/US]; 2788 San Tomas Expressway, Santa Clara, California 95051 (US).

(72) Inventors: **SIVARAMAN, Sakthivel**; c/o NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051 (US). **GURU, Arjun**; c/o NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051 (US). **SHETTY, Rajath**; c/o NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051 (US). **IQBAL, Umar**; c/o NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051 (US). **GALLO, Orazio**; c/o NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051 (US). **SU, Hang**; c/o NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051 (US). **BADKI, Abhishek**; c/o NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051 (US). **HEDAU, Varsha**; c/o NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051 (US).

(54) Title: IMAGE-BASED THREE-DIMENSIONAL OCCUPANT ASSESSMENT FOR IN-CABIN MONITORING SYSTEMS AND APPLICATIONS

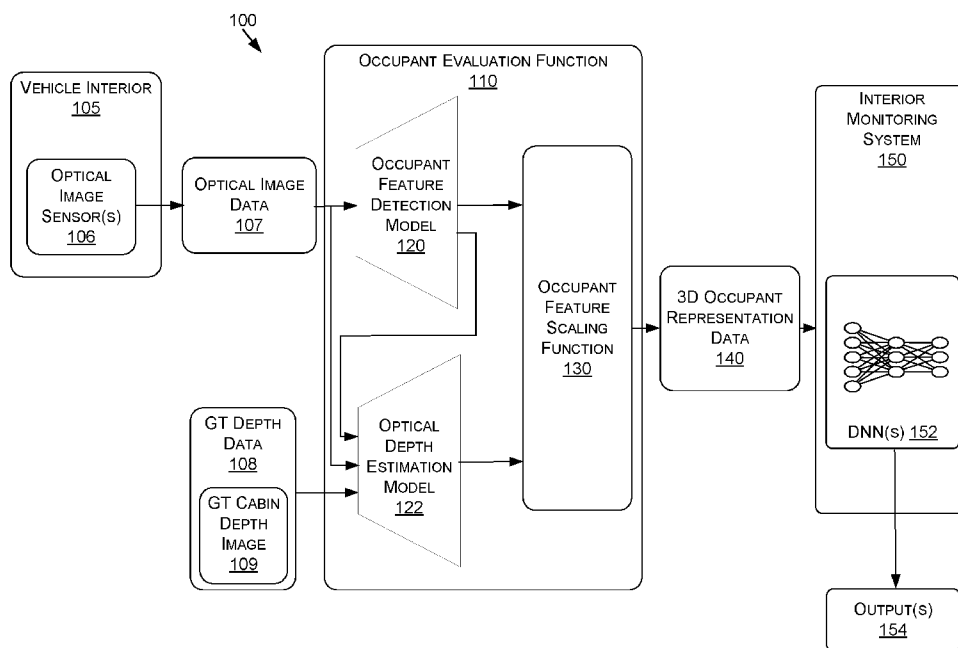


FIGURE 1

(57) Abstract: In various examples, image-based three-dimensional occupant assessment for in-cabin monitoring systems and applications are provided. An evaluation function may determine a 3D representation of an occupant of a machine by evaluating sensor data comprising an image frame from an optical image sensor. The 3D representation may comprise at least one characteristic representative of a size of the occupant, (e.g., a 3D pose and/or 3D shape), which may be used to derive other characteristics such as, but not limited to weight, height, and/or age. A first processing path may generate a representation of one or more features corresponding to at least a portion of the occupant based on optical image data, and a second processing path may determine a depth corresponding to the one or more features based on depth data derived from the optical image data and ground truth depth data corresponding to the interior of the machine.

(74) **Agent: KENDRICK, J. Patrick;** Shook Hardy & Bacon
LLP, 2555 Grand Boulevard, Kansas City, Missouri 64108
(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

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

- 1 -

IMAGE-BASED THREE-DIMENSIONAL OCCUPANT ASSESSMENT FOR IN-CABIN MONITORING SYSTEMS AND APPLICATIONS**CROSS-REFERENCE TO RELATED APPLICATIONS**

- 5 [0001] This application is related to U.S. Patent Application No. 17/935,473, filed on September 26, 2022, titled “MULTI-MODAL SENSOR CALIBRATION FOR IN-CABIN MONITORING SYSTEMS AND APPLICATIONS”, and U.S. Patent Application No. 17/935,465, filed on September 26, 2022, titled “SENSOR CALIBRATION USING FIDUCIAL MARKERS FOR IN-CABIN MONITORING SYSTEMS AND APPLICATIONS”, which are each incorporated herein by reference in its entirety.
- 10 [0002] This application is related to U.S. Patent Application No. 18/349,842, (Attorney Docket No. 22-SC-1598US01/396522), filed on July 10, 2023, titled “THREE-DIMENSIONAL POSE ESTIMATION USING TWO-DIMENSIONAL IMAGES”, U.S. Patent Application No. 18/349,827, (Attorney Docket No. 23-SC-0156US01/396005), filed on July 10, 2023, titled “OCCUPANT EVALUATION USING MULTI-MODAL SENSOR
- 15 FUSION FOR IN-CABIN MONITORING SYSTEMS AND APPLICATIONS”, and U.S. Patent Application No. 18/219,969, (Attorney Docket No. 23-SC-0158US01/396006), filed on July 10, 2023, titled, “CHILD PRESENCE DETECTION FOR IN-CABIN MONITORING SYSTEMS AND APPLICATIONS”, which are each incorporated herein by reference in its entirety.

20 **BACKGROUND**

- [0003] An occupant monitoring system (OMS) may be used within a vehicle cabin to perform real-time assessments of driver and occupant presence, gaze, alertness, or other conditions. For example, an OMS – using data generated or obtained by sensors of the vehicle or machine – may be used to track the direction of a driver’s eye gaze, head pose, or blinking
- 25 (for example, to detect drowsiness, fatigue, and/or distraction), for hand position and/or gesture detection, child and/or pet presence detection, and/or in conjunction with the operation of features such as, but not limited to, seat belt reminders, seat heating, and/or smart airbag deployment. For example, optical sensors may be used to detect motion occurring within a vehicle interior. Optical image sensor data may also be processed to extract image features to
- 30 identify and classify the source of motion. Depth-perception sensors may use radio waves,

- 2 -

laser light, and/or sound waves, for example, to detect the presence or movements of living beings within a vehicle interior (e.g., humans or pets). Such detections may be used within the context of preventing vehicle burglary and/or preventing children or pets from being left alone in the vehicle unintentionally.

5

SUMMARY

[0004] Embodiments of the present disclosure relate to image-based three-dimensional occupant assessment for in-cabin monitoring systems and applications. Systems and methods are disclosed for technologies for an occupant evaluation function for an OMS that may predict a 3-dimensional (3D) pose estimate, and/or a 3D shape estimate, of a vehicle occupant.

10 [0005] In contrast to existing OMS technologies, the occupant evaluation function presented in this disclosure provides, among other things, for an occupant monitoring system that may determine a 3D representation of a vehicle occupant (such as their 3D size, 3D shape and/or 3D pose) using optical image data captured by an OMS optical image sensor (e.g., an in-cabin overhead console fisheye or wide-angle camera). The occupant evaluation function
15 may produce 3D pose estimates and/or 3D shape estimates of a vehicle occupant that may be used as inputs to other systems and functions that further predict other characteristics regarding a vehicle occupant such as a size, weight, and/or age of the vehicle occupant. The occupant evaluation function may determine a 3D representation of a vehicle occupant (such as a 3D shape and/or 3D pose) by evaluating optical image sensor data comprising an image frame
20 captured by an OMS optical image sensor. The image frame may comprise a representation of the vehicle occupant from a perspective that is based at least in part on a position and orientation of the OMS optical image sensor within the vehicle cabin. In some embodiments, the occupant evaluation function executes a first processing path that performs operations to generate a representation of one or more features corresponding to at least a portion of the occupant based
25 on optical image data (e.g., an image frame), and a second processing path that performs operations to determine a depth corresponding to the one or more features based on depth data derived from the optical image data and ground truth depth data (such as a ground truth cabin depth image) corresponding to an interior of the vehicle. In general, the depth information derived from the second processing path may be used by the occupant evaluation function to
30 provide a sense of absolute scale for generating a true-to-scale 3D representation of the vehicle occupant from the representation of the vehicle occupant generated by the first processing path

- 3 -

of the occupant evaluation function. The ground truth cabin depth image corresponding to the interior of the vehicle may be generated in several ways. For example, in some embodiments, vehicle geometry specifications (e.g., CAD model information) on the geometry of the empty cars, combined with an accurate calibration of the vehicle geometry specifications to the OMS optical image sensor, may be used to generate the ground truth cabin depth image.

[0006] The OMS and/or other vehicle system may use the 3D representation for various purposes, such as estimating other characteristics representative of a size of the occupant (e.g., estimating the occupant's height and/or body limb lengths). In some embodiments, the 3D representation of the occupant may be used to control at least one operation of the vehicle based on the estimated occupant characteristic. For example, the characteristic representing the size of the occupant may be used in conjunction with a child presence detection system to estimate an age of the occupant, and/or control an alert system (e.g., alarm and/or notification systems) based on determining that an occupant under an estimated age threshold may have been inadvertently left alone in the vehicle. In some embodiments the characteristic representing the size of the occupant may be used in conjunction with other vehicle safety features. For example, in some embodiments, airbag deployments, driver monitoring systems, HMI applications, and/or other vehicle or machine functions may be controlled based at least on a 3D pose and/or size estimate of the vehicle or machine occupant.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The present systems and methods for image-based three-dimensional occupant assessment for in-cabin monitoring systems and applications are described in detail below with reference to the attached drawing figures, wherein:

[0008] FIG. 1 is an example data flow diagram for a system for image-based three-dimensional occupant assessment, in accordance with some embodiments of the present disclosure;

[0009] FIG. 2 is an example data flow diagram of a system for an example implementation of 3D pose-based three-dimensional occupant assessment, in accordance with some embodiments of the present disclosure;

[0010] FIG. 3 is a diagram illustrating ground truth depth data generation for use with image-based three-dimensional occupant assessment, in accordance with some embodiments of the present disclosure;

- 4 -

[0011] FIG. 4 is a diagram illustrating ray tracing-based determination of absolute depth for one or more kinematic elements, in accordance with some embodiments of the present disclosure;

5 [0012] FIG. 5 is an example data flow diagram of a system for an example implementation of 3D shape-based three-dimensional occupant assessment, in accordance with some embodiments of the present disclosure;

[0013] FIG. 6 is a diagram illustrating using pose detection in combination with 3D shape-based three-dimensional occupant assessment, in accordance with some embodiments of the present disclosure;

10 [0014] FIG. 7 is a diagram illustrating an example method for image-based three-dimensional occupant assessment, in accordance with some embodiments of the present disclosure;

[0015] FIG. 8A is an illustration of an example autonomous vehicle, in accordance with some embodiments of the present disclosure;

15 [0016] FIG. 8B is an example of camera locations and fields of view for the example autonomous vehicle of FIG. 8A, in accordance with some embodiments of the present disclosure;

[0017] FIG. 8C is a block diagram of an example system architecture for the example autonomous vehicle of FIG. 8A, in accordance with some embodiments of the present disclosure;

20 [0018] FIG. 8D is a system diagram for communication between cloud-based server(s) and the example autonomous vehicle of FIG. 8A, in accordance with some embodiments of the present disclosure;

[0019] FIG. 9 is a block diagram of an example computing device suitable for use in implementing some embodiments of the present disclosure; and

25 [0020] FIG. 10 is a block diagram of an example data center suitable for use in implementing some embodiments of the present disclosure.

DETAILED DESCRIPTION

30 [0021] Systems and methods are disclosed related to image-based three-dimensional occupant assessment for in-cabin monitoring systems and applications. Although the present disclosure may be described with respect to an example autonomous or semi-autonomous

- 5 -

vehicle or machine 800 (alternatively referred to herein as “vehicle 800” or “ego-machine 800,” an example of which is described with respect to FIGS. 8A-8D), this is not intended to be limiting. For example, the systems and methods described herein may be used by, without limitation, non-autonomous vehicles or machines, semi-autonomous vehicles or machines
5 (e.g., in one or more advanced driver assistance systems (ADAS)), autonomous vehicles or machines, piloted and un-piloted robots or robotic platforms, warehouse vehicles, off-road vehicles, vehicles coupled to one or more trailers, flying vessels, boats, shuttles, emergency response vehicles, motorcycles, electric or motorized bicycles, aircraft, construction vehicles, trains, underwater crafts, remotely operated vehicles such as drones, and/or other vehicle types.
10 In addition, although the present disclosure may be described with respect to image-based assessments of vehicle occupant characteristics (e.g., estimating occupant size and/or pose for vehicle occupant monitoring systems), this is not intended to be limiting, and the systems and methods described herein may be used in augmented reality, virtual reality, mixed reality, robotics, security and surveillance, autonomous or semi-autonomous machine applications,
15 and/or any other technology spaces where characterizing physical characteristics of a person may be used.

[0022] The present disclosure relates to vehicle occupant detection and monitoring technologies. More specifically, the systems and methods presented in this disclosure provide for technologies for an occupant monitoring system (OMS) that may predict a 3-dimensional
20 (3D) pose estimate, and/or a 3D shape estimate, of a vehicle occupant from an image frame captured by an optical image sensor (e.g., a camera). A size of the vehicle occupant may be derived based on the 3D pose and/or 3D shape estimate. Of the tasks that may be performed by an OMS, child presence detection (CPD) in particular involves, among other things, the task of detecting when a child is an occupant of a vehicle. For example, a child protection system
25 may attempt to assess when an occupant of an idle vehicle is less than a threshold age, and alert the vehicle owner when a child may have been inadvertently left behind inside the vehicle. Size estimation may be an upstream task related to the age estimation task, since a person’s size correlates heavily with age and is one of the most observable characteristics of a person that can be used in assessing the age of a vehicle occupant.

30 [0023] Particularly with respect to detecting characteristics, such as the 3D shape and/or 3D body pose of a vehicle occupant, traditional systems have used in-cabin depth sensors, such as RADAR sensors, that may directly generate 3D data corresponding to detected objects in the vehicle cabin, and that may penetrate structural elements of the vehicle interior

- 6 -

(e.g., car seats) to detect occluded objects not within a line of sight of an optical image sensor. RADAR sensors may produce sensor data that can be used to derive size estimates generally representative of the size of an occupant (e.g., to differentiate a child from an adult occupant) in three dimensions. However, it can be expensive to deploy interior-sensing RADAR sensors in production vehicles. Moreover, the sensor data from a RADAR sensor is of limited resolution that limits its ability to produce 3D body pose estimates that precisely capture the position of body limbs (e.g., for child presence detection and/or occupant age predictions), and primarily rely on sensed motion to sense objects. Other depth-sensing sensors, such as depth-sensing cameras, also referred to as range cameras, produce a two-dimensional (2D) range image, where pixel values of the range image may correspond to a distance from the sensor to sensed images in the sensor's field of view. However, depth-sensing cameras are also expensive to deploy in production vehicles. Moreover, the accuracy of deriving 3D depth data for an occupant from 2D range images produced by a depth-sensing camera may be limited by factors such as relatively low resolution, short sensing distances, and susceptibility to occlusions and/or optical interference.

[0024] Optical image sensor data from monocular optical image sensors, such as a camera that captures standard RGB, IR, and/or RGB-IR image frames, may be obtained using relatively inexpensive devices that may already be deployed in the vehicle cabin for other purposes (e.g., driver gaze detection). That said, monocular optical image sensors, by themselves, do not generate data that conveys a sense of the 3D position of objects in the capture scene, which makes it difficult to train a machine learning model, such as a deep neural network (DNN), to generate accurate 3D shape, 3D size and/or 3D body pose estimates for people captured in the 2D images. Moreover, estimates based on 2D images are vulnerable to producing inaccurate estimates the more the occupant deviates from an upright posture, such as when the occupant is sitting in a slouched or hunched position and/or turned to one side. This is often because 2D to 3D imaging mapping is an ill-posed problem with ambiguous solutions, for example where the same 2D projection may be derived from multiple 3D poses.

[0025] In contrast to existing OMS technologies, the systems and methods presented in this disclosure provide for an occupant monitoring system that may determine a 3D representation of a vehicle occupant (such as their 3D size, 3D shape and/or 3D pose) using one or more image frames captured by at least one OMS optical image sensor (e.g., an in-cabin overhead fisheye or wide-angle camera).

- 7 -

[0026] In some embodiments, an occupant evaluation function as disclosed herein may produce 3D pose estimates and/or 3D shape estimates of a vehicle occupant that may be used as inputs to other systems and functions that further predict other characteristics regarding a vehicle occupant, such as a size, weight, and/or age of the vehicle occupant. The occupant evaluation function may thus be used in conjunction with a child presence detection system designed to protect against children and/or pets from being left alone in a vehicle by accident. In some embodiments the occupant monitoring systems may be used in conjunction with other vehicle safety features. For example, in some embodiments, airbag deployments may be controlled based on a 3D pose estimate of a vehicle occupant. Based on the 3D pose estimate, an airbag deployment system may determine that the vehicle occupant is out of position so that an airbag should not be deployed. In some embodiments, an airbag deployment system may control an airbag deployment pressure based on the 3D pose estimate of a vehicle occupant and/or an estimated size or age of the vehicle occupant derived at least in part from the 3D pose estimate. Moreover, in some embodiments, a 3D shape and/or pose estimate generated by the OMS described herein may be used for occupant posture classification, activity recognition, gaze detection, and human-machine interface (HMI) applications as well. Advantageously, the occupant monitoring system may produce 3D shape and/or 3D pose estimates that are invariant with respect to viewpoint, providing a greater degree of leeway in determining the placement and/or orientation of the OMS optical image sensor that captures the image frames. As such, the OMS may not have to rely on a precise calibration of the orientation of the OMS optical image sensor to an in-cabin coordinate system to predict an occupant's size or pose.

[0027] More specifically, in some embodiments, an occupant evaluation function described herein may determine a 3D representation of a vehicle occupant (such as a 3D shape and/or 3D pose) by evaluating optical image sensor data comprising an image frame captured by an OMS optical image sensor. The image frame may comprise a representation of the vehicle occupant from a perspective that is based at least in part on a position and orientation of the OMS optical image sensor within the vehicle cabin. In some embodiments, the OMS optical image sensor may comprise a monocular optical image sensor, such as a camera that captures standard RGB, IR, and/or RGB-IR image frames of the vehicle interior. In some embodiments, the OMS optical image sensor may comprise a depth-sensing camera that produces a 2D range image of the vehicle interior. The 2D range image may comprise an image frame of pixels where the pixel values of the pixels correspond to a distance from the depth-sensing camera to the elements in the depth-sensing camera's field of view. In some

embodiments, as described in greater detail below, the OMS comprises an occupant evaluation function that may predict a scale-normalized 3D pose estimate corresponding to a vehicle occupant, and then scale that 3D pose estimate to full-scale dimensions of linear measurement (e.g., inches or meters) to determine at least one characteristic representing a size of the occupant. In some embodiments, the OMS comprises an occupant evaluation function that may predict a scale-normalized 3D shape corresponding to a vehicle occupant, and then scale that 3D shape to full-scale dimensions of linear measurement (e.g., inches or meters) to determine at least one characteristic representing a size of the occupant.

[0028] In some embodiments, the occupant evaluation function executes a first processing path that performs operations to generate a representation of one or more features corresponding to at least a portion of the occupant based on optical image data (e.g., an image frame), and a second processing path that performs operations to determine a depth corresponding to the one or more features based on depth data derived from the optical image data and ground truth depth data (such as a ground truth cabin depth image) corresponding to an interior of the vehicle. In general, the depth information derived from the second processing path may be used by the occupant evaluation function to provide a sense of absolute scale for generating a true-to-scale 3D representation of a vehicle occupant from the representation of the vehicle occupant generated by the first processing path of the occupant evaluation function.

[0029] As discussed in greater detail below, the ground truth cabin depth image corresponding to the interior of the vehicle may be generated in several ways. For example, in some embodiments, vehicle geometry specifications (e.g., CAD model information) on the geometry of the empty cars, combined with an accurate calibration of the vehicle geometry specifications to the OMS optical image sensor, may be used to generate the ground truth cabin depth image.

[0030] The occupant evaluation function may generate a 3D representation of the occupant that includes at least one characteristic representative of a size of the occupant, such as a 3D pose and/or a 3D shape. The OMS and/or other vehicle system may use the 3D representation for various purposes, such as estimating other characteristics representative of a size of the occupant (e.g., estimating the occupant's height and/or body limb lengths). In some embodiments, the occupant evaluation function may generate one or more outputs comprising the 3D representation of the occupant that are used to control at least one operation of the vehicle based on the estimated occupant characteristic. For example, the characteristic representing the size of the occupant may be used in conjunction with a child presence detection

- 9 -

system to estimate an age of the occupant and/or control an alert system (e.g., alarm and/or notification systems) based on determining that an occupant under an estimated age threshold may have been inadvertently left alone in the vehicle. In some embodiments, the characteristic representing the size of the occupant may be used in conjunction with other vehicle safety features. For example, in some embodiments, airbag deployments, driver monitoring systems, HMI applications, and/or other vehicle functions may be controlled based at least on a 3D pose and/or size estimate of the vehicle occupant.

[0031] In some embodiments, the occupant evaluation function processes an input optical image frame from an OMS optical image sensor to derive a 3D pose estimate for a vehicle occupant. In such an embodiment, the representation of one or more features corresponding to at least a portion of the occupant may comprise a scale-normalized 3D pose estimate. For example, in some embodiments, the occupant evaluation function may execute a first processing path that includes a person detection model and a 3D pose detection model (e.g., both of which may be implemented using one or more of a Convolutional Neural Network (CNN), Deep Neural Network (DNN), or other machine learning model architecture(s)). The image frame may be processed by the person detection model, which recognizes features of the occupant and crops the image to produce a cropped image (e.g., an image bounded by an outline of the occupant). Based on the cropped image, the 3D pose detection model may generate a scale-normalized 3D pose of the occupant. The scale-normalized 3D pose may comprise a 3D representation of kinematic elements (e.g., body limbs and/or joints) that indicates 3D coordinates for the kinematic elements. In other words, the 3D pose detection model receives the cropped images of the occupant from the person detection model and produces a 3D pose estimate for the vehicle occupant from the captured image frame. The 3D pose detection model may be trained based on synchronized multi-view images of training subjects to produce 3D pose estimates using coordinates that are scale-normalized. That is, the 3D coordinates are scale-normalized in that they may indicate the dimensions and/or relative positions of kinematic elements in relation to each other, rather than in absolute terms (e.g., linear measurement units).

[0032] To map the scale-normalized 3D pose of the occupant to an absolute 3D pose, the second processing path of the occupant evaluation function may determine an absolute 3D depth corresponding to at least one joint from the scale-normalized 3D pose based on depth data derived from at least one image frame and further using a ground truth cabin depth image corresponding to an interior of the vehicle. Using the at least one joint to anchor the scale-

normalized 3D pose to an absolute scale, the occupant evaluation function may determine a set of absolute coordinates for the other kinematic elements of the scale-normalized 3D pose to derive an absolute 3D pose.

[0033] In some embodiments, the second processing path of the occupant evaluation function comprises an optical depth estimation model. The optical depth estimation model may be implemented, for example, using a neural network architecture (such as a Convolutional Neural Network (CNN), Deep Neural Network (DNN), or other machine learning model(s)). The optical depth estimation model may output a range image based on an input comprising an image frame from the optical image sensor used to generate the scale-normalized 3D pose. The pixel values of the range image may correspond to a distance from the optical image sensor to sensed features in the sensor's field of view appearing in the image frame. Therefore, the range image output from the optical depth estimation model provides the occupant evaluation function with depth data corresponding to the pixels of the image frame that represent features corresponding to at least a portion of the occupant, including pixels that represent kinematic elements (such as the at least one body joint) of the occupant captured by the scale-normalized 3D pose. As further discussed below, in order to recover an actual scale of a depth for the kinematic elements in absolute terms, the optical depth estimation model may further use a ground truth cabin depth image corresponding to an interior of the vehicle.

[0034] The occupant evaluation function, having established coordinates for the position of the occupant's joint in three dimensions, may use that joint as an anchor joint to anchor the scale-normalized 3D pose to an absolute scale, from which the occupant evaluation function may determine a set of absolute coordinates for the other kinematic elements of the scale-normalized 3D pose to derive an absolute 3D pose. With an absolute 3D pose for the vehicle occupant determined, the occupant evaluation function may proceed with computing a size of the occupant based on limb length that may be computed, for example, as a function of distance between body joints. Because the 3D coordinates for the kinematic elements in the absolute 3D pose can be mapped directly to absolute distances, body limb lengths of the occupant may be directly computed from the 3D coordinates. For example, a width of the occupant's torso may be estimated by computing a distance between the 3D coordinates of the occupant's left and right shoulder joints. Based, for example, on summing various body limb lengths, the occupant evaluation function may estimate an overall size of the occupant. In some embodiments, an individual limb length may be computed as a distance between two consecutive attached joints. In some embodiments, a compound limb length may be computed

as a function of a combination of kinematic elements. For example, the occupant's wingspan, which goes from the occupant's left wrist to the occupant's right wrist when their arms are extended, may define a kinematic element comprising a combination of individual limbs and a compound limb length for the wingspan computed. In some embodiments, pre-programmed statistical information for human body proportions may be used to estimate body dimensions not directly derivable from the absolute 3D pose. For example, statistically, a person's height is usually about 1.1 times their wingspan. As such, if the absolute 3D pose is missing kinematic information about the occupant below the torso (e.g., if that portion of their body was blocked in the image frame), the occupant evaluation function may refer to statistical information for human body proportions to estimate their height based on deriving the length of their wingspan and multiplying by 1.1, for example.

[0035] In some embodiments, the occupant evaluation function processes the input image frame from the OMS optical image sensor to derive a 3D shape estimate for a vehicle occupant rather than, or in addition to, a 3D pose estimate. For example, in some embodiments, the occupant evaluation function may execute a first processing path that includes a segmentation model that segments the input image frame(s) based on inferring pixels of the image frame that represent the occupant from those that do not represent the occupant. From that inference, the segmentation model may produce an occupant mask (e.g., bounding an outline of pixels corresponding to the occupant) and a vehicle interior mask (e.g., bounding an outline of pixels corresponding to structural elements of the vehicle interior). The second processing path of the occupant evaluation function may comprise a depth estimation model that may be implemented, for example, using a neural network architecture (such as a CNN, DNN or other machine learning model architecture (s)). The optical depth estimation model may output a range image based on an input comprising the image frame from the optical image sensor, where pixel values of the range image may correspond to a distance from the optical image sensor to structural features of the vehicle interior in the sensor's field of view appearing in the optical image frame. The range image output from the optical depth estimation model may provide the occupant evaluation function with depth data corresponding to the pixels of the image frame that represent structural features. The depth data for the structural features may be correlated with pixels that represent kinematic elements (such as the at least one body joint) of the occupant captured within the bounds of the occupant mask.

[0036] In some embodiments, in order to recover an actual scale of a depth for the at least one body joint in absolute terms, the optical depth estimation model may further use a

ground truth cabin depth image corresponding to an interior of the vehicle in combination with the vehicle interior mask to perform affine scaling of the kinematic elements captured within the bounds of the occupant mask. As discussed above, the ground truth cabin depth image may comprise a frame of a range image that represents a field of view of the vehicle interior calibrated to the field of view of the OMS optical image sensor. The optical depth estimation model may learn from the vehicle interior mask which pixels in the optical image data are structural components of the vehicle interior (as opposed to representing the occupant and/or other object(s) located within the vehicle interior) and based on those pixels identified as components of the vehicle interior, refer to the ground truth cabin depth image to determine a distance between various structural interior elements and the OMS optical image sensor. For example, the optical depth estimation model may comprise a structural element depth estimator model that outputs structural element depth data based on an input comprising an image frame of optical image data and the ground truth cabin depth image. The structural element depth data may comprise an image having pixel values that correspond to a distance from the optical image sensor to structural features of the vehicle interior. For example, the structural element depth data produced by the structural element depth estimator model may indicate that a structural component within the vehicle cabin (e.g., a location of a feature on the cabin headliner) is a given distance, d_1 , from the OMS optical image sensor (e.g., in linear measurement units). An occupant feature scaling function may function as a 3D shape estimator to perform affine scaling to generate an occupant 3D shape estimate using the structural element depth data to scale the pixels of the image frame that represent features of the occupant captured within the bounds of the occupant mask.

[0037] To determine a 3D shape of the occupant, the occupant evaluation function may comprise a 3D shape estimator that inputs the occupant mask generated by the segmentation model, and the structural element depth data generated by the optical depth estimation model. The segmentation model and/or the optical depth estimation model may be implemented, for example, using a neural network architecture (such as a CNN, DNN, or other machine learning model architecture(s)). Using the occupant mask and the structural element depth data, the 3D shape estimator may infer that one or more pixels representing a first location corresponding to a surface of the occupant's body, are a relative distance, d_2 , closer to the OMS optical image sensor than pixels corresponding to the structural components of the vehicle interior, and based on a function of the distances d_1 and d_2 , compute an absolute distance of the surface of the occupant's body from the OMS optical image sensor. The 3D shape estimator may infer

relative locations of one or more other surface locations on the occupant's body to the first location, and having established coordinates for the first location of the occupant's body, may determine absolute distances from the first location to the one or more other locations and thus estimate an overall 3D shape of the occupant.

5 **[0038]** In some embodiments, an OMS that performs occupant evaluation may be used in vehicle interior space of a vehicle besides a passenger cabin. For example, the vehicle interior space the embodiments described herein may determine 3D pose and/or shape estimates using optical image data for occupants within a trunk, cargo bed, or other space. While embodiments presented in this disclosure may be implemented in the context of vehicle
10 occupant monitoring systems (including driver monitoring systems) for vehicles such as, but not limited to, non-autonomous vehicles, semi-autonomous vehicles, piloted and un-piloted robots or robotic platforms, warehouse vehicles, off-road vehicles, vehicles coupled to one or more trailers, aircraft, spacecraft, boats, shuttles, emergency response vehicles, construction vehicles, underwater crafts, drones, and/or other vehicle types, other embodiments may include
15 determining extrinsic calibration parameters for sensors that capture image frames of other interior spaces, such as rooms, warehouses, gymnasiums, containers, airport terminals, mines, factories, construction zones, and/or studio sets, as suitable examples and without limitation, in order to monitor occupants of such spaces.

20 **[0039]** The various sensor data processing, pose estimation, size estimation, and/or other models, functions, and algorithms, disclosed herein, may be executed at least in part on one or more processing units, such as one or more graphics processing units that may operate in conjunction with software executed on a central processing unit coupled to a memory. The graphics processing units may be programmed to execute kernels to implement one or more aspects of functions for occupant monitoring described herein. In some embodiments, the
25 execution of some algorithms may be distributed and performed by a combination of processors and cloud computing resources.

30 **[0040]** With reference to FIG. 1, FIG. 1 is an example data flow diagram for a system 100 for image-based three-dimensional occupant assessment, in accordance with some embodiments of the present disclosure. It should be understood that this and other arrangements described herein are set forth only as examples. Other arrangements and elements (e.g., machines, interfaces, functions, orders, groupings of functions, etc.) may be used in addition to or instead of those shown, and some elements may be omitted altogether. Further, many of the elements described herein are functional entities that may be implemented

as discrete or distributed components or in conjunction with other components, and in any suitable combination and location. Various functions described herein as being performed by entities may be carried out by hardware, firmware, and/or software. For instance, various functions may be carried out by a processor executing instructions stored in memory. In some
5 embodiments, the systems, methods, and processes described herein may be executed using similar components, features, and/or functionality to those of example autonomous vehicle 800 of FIGS. 8A-8D, example computing device 900 of FIG. 9, and/or example data center 1000 of FIG. 10.

[0041] As shown in FIG. 1, the image-based three-dimensional occupant assessment
10 system 100 may include an occupant evaluation function 110 that comprises at least one occupant feature detection model 120, at least one optical depth estimation model 122, and an occupant feature scaling function 130. The occupant feature detection model 120 and/or optical depth estimation model 122 may be implemented using a machine learning module, such as implemented using a convolutional neural network (CNN), a deep neural network (DNN),
15 and/or other neural network architecture.

[0042] As shown in FIG. 1, the occupant evaluation function 110 may execute a first processing path that comprises the occupant feature detection model 120 and generates a representation of one or more features of a vehicle occupant based on optical image data 107 that comprises a representation of the vehicle occupant (e.g., an optical image frame) captured
20 by an optical image sensor 106 that may be positioned within the vehicle interior 105 (e.g., the interior of vehicle 800). An optical image sensor 106 may comprise, for example, a camera or other optical sensor that captures RGB, IR, and/or RGB-IR image frames. In some embodiments, the optical image sensor 106 may comprise an OMS sensor such as the OMS sensor(s) 801 described with respect to FIG. 8A.

[0043] The occupant evaluation function 110 may execute a second processing path that comprises the optical depth estimation model 122 and generates depth data corresponding to elements appearing within the vehicle interior 105 (e.g., a depth from the optical image sensor 106 to the respective elements) based on the optical image data 107 and ground truth depth data 108. In some embodiments, ground truth depth data 108 may comprise at least one
30 ground truth cabin depth image 109 that may be used as a ground truth reference of distance between the optical image sensor 106 and various structural elements of the vehicle interior 105 appearing in the optical image data 107. The optical depth estimation model 122 may, in some embodiments, further input the representation of the vehicle occupant generated by the

occupant feature detection model 120 to discern structural elements of the vehicle interior 105 from the feature(s) of the vehicle occupant to better estimate depths for the feature(s) of the vehicle occupant. The occupant evaluation function 110 may input the representation of one or more features of a vehicle occupant from the occupant feature detection model 120 and the
5 depth data from the optical depth estimation model 122 to the occupant feature scaling function 130 to determine an absolute (e.g., true-scale) depth corresponding to the one or more occupant features and/or generate a three-dimensional representation of the occupant that is output as 3D occupant representation data 140. The 3D occupant representation data 140 may include at least one characteristic representative of a size of the occupant (e.g., the occupant's height
10 and/or body limb lengths). Characteristics included in the 3D occupant representation data 140 may comprise a representation such as a 3D pose estimate, a 3D size estimate, and/or a 3D shape estimate of the vehicle occupant. At least one operation of the vehicle may then be controlled based on the characteristic.

[0044] For example, based at least in part on the 3D occupant representation data 140,
15 an interior monitoring system 150 (which may implement one or more components of the OMS) may generate one or more output(s) 154. Output(s) 154 may be generated using one or more machine learning models and/or deep neural networks (DNNs) 152. As an example, the interior monitoring system 150 may use 3D occupant representation data 140 (either alone or in combination with other data such as optical image data 107) to predict the presence and/or
20 location of occupants – such as objects, persons, and/or animals – within the interior space of the vehicle 800. Other systems of the vehicle 800 may determine one or more actions to take based on the predictions and/or may control other tasks or operations. For example, based on output(s) 154, an alarm or warning may be generated, door locks and/or windows may be operated, various functions may be turned on/off, data for a digital assistant, chat bot, digital
25 avatar, and/or the like may be generated, and/or air conditioning or air circulation functions may be operated. In some embodiments, the characteristic representing the size of the occupant from the 3D occupant representation data 140 may be used in conjunction with a child presence detection system to estimate an age of the occupant, and/or control an alert system (e.g., alarm and/or notification systems) based on determining that an occupant under an estimated age
30 threshold may have been inadvertently left alone in the vehicle. In some embodiments, the characteristic representing the size of the occupant may be used in conjunction with other vehicle safety features. For example, in some embodiments, airbag deployments, driver monitoring systems, HMI applications, and/or other vehicle functions may be controlled based

- 16 -

at least on a 3D pose, 3D shape, and/or 3D size estimate of the vehicle occupant provided by the 3D occupant representation data 140.

[0045] With reference to FIG. 2, FIG. 2 is an example data flow diagram 200 of a system 200 for an example 3D pose-based implementation of the image-based three-dimensional occupant assessment such as that shown in FIG. 1, in accordance with some embodiments of the present disclosure.

[0046] In the embodiment shown in FIG. 2, the occupant evaluation function 110 processes optical image data 107 from the optical image sensor 106 to derive 3D occupant representation data 140 that comprises a 3D pose estimate for a vehicle occupant. For example, in the embodiment of FIG. 2, the occupant feature detection model 120 may comprise a person detection model 220 and a 3D pose detection model 222 that use the optical image data 107 to generate a scale-normalized 3D pose estimate 224 for a vehicle occupant. The scale-normalized 3D pose estimate 224 may comprise a representation of one or more features corresponding to at least a portion of the occupant. The person detection model 220 and/or 3D pose detection model 222 may be implemented, for example, using one or more of a Convolutional Neural Network (CNN), Deep Neural Network (DNN), or other machine learning model architecture(s).

[0047] The optical image data 107 may be processed by the person detection model 220 to detect features of the vehicle occupant, for example, to produce a cropped image of the vehicle occupant (e.g., an image bounded by an outline of the vehicle occupant). In some embodiments, when the person detection model 220 detects more than one occupant, separate cropped images of the vehicle occupants may be produced and individually processed for each occupant as described herein.

[0048] Based on the detected features (e.g., the cropped image) of the vehicle occupant produced by the person detection model 220, the 3D pose detection model 222 may generate the scale-normalized 3D pose 224 of the occupant. The scale-normalized 3D pose 224 may comprise a 3D representation of kinematic elements of the vehicle occupant (e.g., one or more body limbs and/or joints), and may indicate relative positions of using 3D coordinates for those kinematic elements. In other words, the 3D pose detection model 222 may receive the cropped images of the occupant from the person detection model 220 and predict the scale-normalized 3D pose 224 for the vehicle occupant, based on the captured optical image data 107. As further discussed herein, the 3D pose detection model 222 may comprise a machine learning model trained based on synchronized multi-view images of training subjects to produce the scale-

normalized 3D pose 224. That is, the 3D coordinates are scale-normalized in that they may indicate the dimensions and/or relative positions of kinematic elements in relation to each other, rather than in absolute terms (e.g., linear measurement units).

[0049] In some embodiments, the scale-normalized 3D pose 224 of the vehicle
5 occupant may be derived temporally based on optical image data 107 that comprises a sequence of image frames from the optical image sensor 106. In such a temporal embodiment, the 3D pose detection model 222 may receive a sequence of cropped images from the person detection model 220 and predict a corresponding sequence of scale-normalized 3D poses 224 with corresponding confidence scores. The occupant evaluation function 110 may implement a
10 voting function to, for example, select from the sequence a scale-normalized 3D pose having the highest confidence score, and/or select a composite scale-normalized 3D pose based on a subset of the sequence having pose estimates that are most in agreement (e.g., within an alignment threshold) to arrive at the scale-normalized 3D pose 224. In some embodiments, the 3D pose detection model 222 may compute scale-normalized 3D pose estimates at different
15 periodicities, such as computing single frame-based scale-normalized 3D pose that may be generated quickly, and a scale-normalized 3D pose derived from optimizing sequences of image frames and/or 3D pose estimates over predefined time windows. Sequences of image frames may be evaluated to better discern moving objects in the scene in order to focus the person detection model 220 and/or the 3D pose detection model 222 on processing a dynamic
20 segment of optical image data 107 over a static region.

[0050] In some embodiments, such as the embodiment of FIG. 2, the second processing path of the occupant evaluation function 110 may comprise an occupant feature depth estimation model 230 that determines an absolute 3D depth of at least one anchor joint appearing in the optical image data 107 that may be used by the scaling function 130 to
25 determine a set of absolute coordinates for the other kinematic elements of the scale-normalized 3D pose 224 to derive an absolute 3D pose estimate 240 of the vehicle occupant to generate the 3D occupant representation data 140. The occupant feature depth estimation model 230 may be implemented, for example, using a neural network architecture (such as a Convolutional Neural Network (CNN), Deep Neural Network (DNN), or other machine
30 learning model(s)).

[0051] To map the scale-normalized 3D pose 224 of the occupant to the absolute 3D pose estimate 240, the occupant feature depth estimation model 230 may input the scale-normalized 3D pose 224 and determine absolute joint depth data 232 corresponding to at least

one joint from the scale-normalized 3D pose 224. Using the joint depth data 232 to correlate the scale-normalized 3D pose to an absolute scale, the occupant feature scaling function 130 may determine a set of absolute coordinates for the other kinematic elements of the scale-normalized 3D pose 224 based on their relative distances and/or positions from the designated anchor joint(s) as indicated by the scale-normalized 3D pose 224.

[0052] In some embodiments, the occupant feature depth estimation model 230 may output the joint depth data 232 in the form of a range image. The pixel values of the range image may correspond to a distance from the optical image sensor 106 to sensed images in the sensor's field of view appearing in the image frame. Therefore, the range image may be correlated to the scale-normalized 3D pose 224 by the occupant feature scaling function 130 to determine depths for the features of the occupant, including the depth of at least one kinematic element for the anchor joint.

[0053] In order to generate the joint depth data 232, the occupant feature depth estimation model 230 may further input the ground truth cabin depth image 109 that represents a field of view of the vehicle interior calibrated to the field of view of the OMS optical image sensor 106. For example, the occupant feature depth estimation model 230 may be informed by the ground truth cabin depth image 109 that a structural component within the vehicle interior 105 (e.g., a location of a feature on the cabin headliner) is a given distance, d_1 , from the OMS optical image sensor 106 (e.g., in linear measurement units). The occupant feature depth estimation model 230, based on the optical image data 107 and the scale-normalized 3D pose 224, may infer that one or more kinematic elements represented by the scale-normalized 3D pose 224, are a relative distance, d_2 , closer to the optical image sensor 106 than pixels corresponding to the structural component that are a known absolute distance of d_1 from the optical image sensor 106, and based on a function of the distances d_1 and d_2 , compute an absolute distance of the kinematic element from the OMS optical image sensor. The occupant feature scaling function 130, based on the joint depth data 232, may use the true-scale absolute coordinates for the position of an anchor joint to anchor the scale-normalized 3D pose 224 to an absolute scale, from which the occupant feature scaling function 130 may determine a set of absolute coordinates for the other kinematic elements of the scale-normalized 3D pose to derive the absolute 3D pose estimate 240.

[0054] As discussed above and also illustrated by FIG. 3 at 300, the ground truth depth data 108 may provide a reference to the depth estimator model 230 from which distances between the optical image sensor 106 and structural elements of the vehicle interior 105 are

known based on known vehicle geometry specifications 310. In some embodiments, the ground truth depth data 108, which may be represented in the form of a ground truth cabin depth image 109, may be generated in several ways. For example, in some embodiments, a depth-rendering function 320 may input a vehicle geometry specification 310 (e.g., CAD model of the vehicle interior 105) that specifies the geometry (e.g., 3D coordinates) of structural elements of the vehicle interior 105, which are translated to the viewpoint of the optical image sensor 106 using sensor calibration parameters 305 to generate the ground truth cabin depth image. For example, in some embodiments, the vehicle geometry specifications 310 may define a mounting location and orientation of the optical image sensor 106 within the vehicle interior 105. Parameters that influence how the 3D volume of the vehicle interior 105 appears when projected onto the 2D coordinate space of the optical image sensor 106 include both extrinsic and intrinsic parameters. Extrinsic parameters may refer to factors that describe the physical orientation of the sensor, such as rotation and translation (also referred to as roll and tilt), and/or other parameters. Intrinsic parameters may refer to factors that describe sensor device optics, such as optical center (also known as the principal point), focal length, skew coefficient, field of view, and/or other parameters. While the intrinsic parameters of a sensor can be established during manufacture, and are expected to remain stable, the extrinsic parameters of rotation and translation instead depend on how the sensor is mounted and oriented within the space of the vehicle cabin. The extrinsic and intrinsic parameters may both play a part in how features of a scene within the 3D coordinate space of the vehicle cabin are mapped to the sensor-captured image frame. In some embodiments, sensor calibration parameters 305 comprise a rotation-translation (RT) transform that describe the extrinsic relationship between the coordinate system of the vehicle geometry specification 310 and the optical image sensor 106. That is, the sensor calibration parameters 305 may comprise an RT transform that accounts for the rotation and translation of the optical image sensor 106 as installed, so that the 3D cabin coordinate system of the vehicle interior 105, as described by the vehicle geometry specification 310, may be mapped to the 2D coordinate space of the field of view of the optical image sensor 106. Since vehicle geometry specification 310 includes data that indicates the 3D coordinate of structural elements of the vehicle interior 105, those 3D coordinates may be mapped, using the RT transform, to the coordinate space of the optical image sensor 106. As an example, a calibration process (e.g., a factory calibration procedure) to produce the ground truth depth data 108 may comprise inputting the vehicle geometry specification 310 and the sensor calibration parameters 305 into the depth-rendering function

320. The depth-rendering function 320 may determine a field of view of the vehicle cabin as it is observed by the optical image sensor 106, which may be based at least on a combination of the sensor's intrinsic parameters, extrinsic parameters, and/or mounting location within the vehicle, and synthesize ground truth depth data 108 (such as a ground truth cabin depth image 109) that specifies depth data for the structural elements of the vehicle interior 105 appearing in the sensor's field of view. A ground truth cabin depth image 109 may comprise a range image, where pixel values of the range image correspond to a distance from the optical image sensor 106 to those structural elements of the vehicle cabin that are predicted by the depth-rendering function 320 as appearing in the field of view of the optical image sensor 106. A representation of the ground truth depth data 108 and/or ground truth cabin depth image 109 may then be programmed into a memory of the vehicle 800 and used by the optical depth estimation model 122 for scaling relative dimensions of structural elements and/or occupant kinematic elements to absolute dimensions, as discussed herein.

[0055] In some embodiments, the occupant feature scaling function 130 may transform the scale-normalized 3D pose 224 to the absolute 3D pose estimate 240 using a ray-tracing algorithm to map the pixel identified in a joint depth data 232 range image back to pixels of an image frame from the optical image data 107, and determine where pixels corresponding with a kinematic element of the scale-normalized 3D pose 224 interact with pixels of the range image. For example, the occupant feature scaling function 130 may select a preferred kinematic element (e.g., a body joint) observable from the optical image data 107 to define as the anchor joint. The occupant feature scaling function 130 may comprise a predetermined prioritized list of body joints, and may select a body joint that is observable from the optical image data 107 based on the prioritized list (e.g., a joint indicated as having a preferred priority) to define as the anchor joint. For example, neck, hip, or other torso body joints typically would serve well as preferred anchor joints because for a seated vehicle occupant, these joints are generally limited in their motion and range of positions (e.g., in contrast to joints such as elbows and/or wrists). As such, it is generally easier for the 3D pose detection model 222 to confidently identify such kinematic elements as opposed to extremities. As shown in FIG. 4 at 400, the occupant feature scaling function 130 may apply a ray-tracing algorithm to compute a ray 410 that extends from a center 415 of the optical image sensor 106 through a 2D pixel coordinate (u, v) of an image frame 425 (e.g., from the optical image data 107) corresponding to a selected anchor joint 420. The ray-tracing algorithm may determine a pixel and/or cluster of pixels 430 from the joint depth data 232 that is closest to the ray 410, and assign a depth

value 435 to the selected anchor joint 420 based on the depth indicated by the value of the pixel(s) closest to the ray 410. Based on the absolute depth of the selected anchor joint 420 and the relative positions of the other kinematic elements indicated by the scale-normalized 3D pose 224, the occupant feature scaling function 130 may scale one or more of the other
5 kinematic elements of the scale-normalized 3D pose 224 as a function of their relative positions and distances with respect to the anchor joint.

[0056] In some embodiments, a plurality of anchor joints 420 may be selected and evaluated by having the ray-tracing algorithm extend a plurality of rays 420 to identify depths for a plurality of different detected body joints observable from the optical image data 107.

10 The occupant feature scaling function 130 may apply an optimization algorithm to optimally correlate distances determined by the ray-tracing algorithm for multiple rays and assign depths to the different body joints. In some embodiments, the joint depth data 232 may be generated by accumulating optical image sensor data 107 over a predefined time duration to obtain a more stable representation of the vehicle occupant.

15 [0057] Referring again to FIG. 2, the occupant evaluation function 110 may proceed with computing the 3D representation data 140 that comprises at least one characteristic representative of a size of the occupant. Because the 3D coordinates for the kinematic elements in the absolute 3D pose 240 can be used directly to determine absolute distances, body limb lengths of the occupant may be directly computed from those 3D coordinates. For example, a
20 width of the occupant's torso may be estimated by computing a distance between the 3D coordinates of the occupant's left and right shoulder joints, as indicated by the absolute 3D pose estimate 240. Accordingly, by summing various body limb lengths, the occupant evaluation function 110 may estimate an overall 3D size of the occupant that may be included in the 3D occupant representation data 140. In some embodiments, an individual limb length
25 may be computed as a distance between two consecutive attached joints. In some embodiments, a compound limb length may be computed as a function of a combination of kinematic elements. For example, the occupant's wingspan, which goes from the occupant's left wrist to the occupant's right wrist when their arms are extended, may define a kinematic element comprising a combination of individual limbs and a compound limb length for the
30 wingspan computed. In some embodiments, the occupant evaluation function 110 uses pre-programmed statistical information for human body proportions to estimate body dimensions not directly derivable from the absolute 3D pose. For example, statistically, a person's height is about 1.1 times their wingspan. As such, if the absolute 3D pose 240 is missing kinematic

- 22 -

information about the occupant below the torso (e.g., if that portion of their body was blocked in the image frame), the occupant evaluation function 110 may refer to statistical information for human body proportions to estimate an occupant height based on deriving the length of their wingspan and multiplying by 1.1, for example.

5 [0058] In some embodiments, the 3D pose detection model 222 may be trained to produce the scale-normalized 3D pose 224 based on synchronized multi-view images of training subjects. Constraining the 3D pose detection model 222 to the task of predicting relative-scale 3D poses allows it to output more accurate predictions by preventing it from learning to predict estimates for the ill-posed problem of directly predicting depth information
10 from a single 2D image. In some embodiments, the 3D pose detection model 222 may be trained based on multi-view training with real world data obtained by capturing synchronized sets of 2D images (e.g., as described in U.S. Patent Application No. 18/349,842, (Attorney Docket No. 22-SC-1598/396522), filed on even date herewith, titled “POSE DETECTION MODEL TRAINING FOR PREDICTING THREE-DIMENSIONAL POSE ESTIMATES
15 USING TWO-DIMENSIONAL IMAGE DATA,” which is incorporated herein by reference in its entirety), to produce a 3D perception model from 2D image data. In some embodiments, the 3D pose detection model 222 may further, or instead, be trained using synthetic 3D pose ground truth data. Synthetic 3D pose ground truth data permits the 3D pose detection model 222 to be trained from data that captures simulated training subjects having a wide diversity of
20 demographics and in a variety of vehicles. Additionally, synthetic data inherently provides full 3D pose ground truth, which aids in the ability to directly supervise the 3D pose detection model 222 training process.

[0059] With reference to FIG. 5, FIG. 5 is an example data flow diagram 500 of a system 500 for an example 3D shape-based implementation of the image-based three-
25 dimensional occupant assessment such as shown in FIG. 1, in accordance with some embodiments of the present disclosure. In the embodiment shown in FIG. 5, the occupant evaluation function 110 processes optical image data 107 from the optical image sensor 106 to derive 3D occupant representation data 140 that comprises a 3D shape estimate for a vehicle occupant. Here, the occupant feature detection model 120 may comprise a segmentation model
30 520 that uses the optical image data 107 to generate an occupant mask 522 and a vehicle interior mask 524. The segmentation model 520 may segment the optical image data 107, for example, based on inferring pixels of an image frame that represent the vehicle occupant from those that do not represent the vehicle occupant. From that inference, the segmentation model 520 may

produce an occupant mask 520 (e.g., bounding an outline of pixels corresponding to the occupant) and a vehicle interior mask 524 (e.g., bounding an outline of pixels corresponding to structural elements of the vehicle interior 105).

[0060] In such embodiments, the optical depth estimation model 122 may comprise a structural element depth estimator model 530 that outputs structural element depth data 542. The segmentation model 520 and/or the structural element depth estimator model 530 may be implemented, for example, using a neural network architecture (such as a CNN, DNN, or other machine learning model architecture (s)).

[0061] The structural element depth estimator model 530 may output structural element depth data 542 in the form of a range image based on an input comprising an image frame from the optical image data 107. Pixel values of the range image may correspond to a distance from the optical image sensor 106 to structural features of the vehicle interior 105 sensed from the sensor's field of view as appearing in the optical image data 107. In these embodiments, the occupant feature scaling function 130 functions as a 3D shape estimator to generate an occupant 3D shape estimate 550 based on using the structural element depth data 542 to scale the pixels of the image frame that represent features of the occupant captured within the bounds of the occupant mask 522.

[0062] In some embodiments, the structural element depth estimation model 530 may learn from the vehicle interior mask 524 which pixels in the optical image data 107 are components of the vehicle interior 105 (as opposed to pixels representing the occupant and/or other object(s) located within the vehicle interior). Based on those pixels identified as components of the vehicle interior, the structural element depth estimation model 530 may infer distances based on the ground truth depth data 108 to determine distances between various structural interior elements appearing in the optical image data 107. In such embodiments, the structural element depth data 542 output from the structural element depth estimation model 530 may comprise a range image, where pixel values of the range image indicate a distance from the optical image sensor 106 to surfaces of structural elements of the vehicle interior 105 appearing in the optical image data 107. For example, the structural element depth data 542 produced by the structural element depth estimation model 530 may indicate that a structural component within the vehicle cabin (e.g., a location of a feature on the cabin headliner) is a given distance, d_1 , from the OMS optical image sensor (e.g., in linear measurement units).

[0063] To determine a 3D shape of the occupant, the occupant evaluation function the occupant feature scaling function 130 may input the occupant mask 524 generated by the

- 24 -

segmentation model 520, and the structural element depth data 542 generated by the structural element depth estimation model 530. Using the occupant mask 524 and the structural element depth data 542, the occupant feature scaling function 130 performs a 3D shape estimation to infer that one or more pixels representing a first location corresponding to a surface of the occupant's body are a relative distance, d_2 , closer to the optical image sensor 106 than pixels corresponding to one or more structural components of the vehicle interior 105, and based on a function of the distances d_1 and d_2 , compute an absolute distance of the surface of the occupant's body from the optical image sensor 106. The occupant feature scaling function 130 may further infer from the occupant mask 522 relative locations of one or more other locations on the occupant's body to the first location, and having established an absolute 3D position for the first location of the occupant's body, determine absolute distances from the first location to the one or more other locations and thus estimate an overall 3D shape of the occupant. Accordingly, based on the 3D shape estimate 550, the occupant evaluation function 110 may estimate an overall 3D size of the occupant that may be included in the 3D occupant representation data 140.

FIG. 6 is a diagram at 600 illustrating using pose detection in combination with 3D size-based three-dimensional occupant assessment to compute 3D occupant representation data 140, in accordance with some embodiments of the present disclosure. In some embodiments, the 3D shape estimation performed by the occupant feature scaling function 130 may further input a pose estimate (e.g., either a 2D or 3D pose estimate) for the occupant that is generated by further applying the optical image data 107 to a pose detection model 610 to generate the vehicle occupant 3D shape estimate 550. For example, the occupant feature scaling function 130 may input a scale-normalized 3D pose estimate such as discussed above. In such embodiments, the occupant feature scaling function 130 may use kinematic elements represented in the pose estimate (e.g., body limbs and/or joints), and that appear within the occupant mask 522, as locations of the occupant's body from which it may infer relative locations for estimating an overall 3D shape. The occupant feature scaling function 130 may determine that a first location corresponding to a first kinematic element is a relative distance, d_2 , closer to the optical image sensor 106 than pixels corresponding to the structural components of the vehicle interior 105 (e.g., based on the structural element depth data 542) and based on a function of the distances d_1 and d_2 , compute an absolute distance of the kinematic element from the optical image sensor 106. The occupant feature scaling function 130 may infer relative locations of one or more other kinematic elements to the first kinematic

element, and determine absolute distances between kinematic elements and thus estimate an overall 3D shape of the occupant.

[0065] Now referring to FIG. 7, FIG. 7 is a flow diagram showing a method 700 for image-based three-dimensional occupant assessment system, in accordance with some
5 embodiments of the present disclosure. It should be understood that the features and elements described herein with respect to the method 700 of FIG. 7 may be used in conjunction with, in combination with, or substituted for elements of any of the other embodiments discussed herein and *vice versa*. Further, it should be understood that the functions, structures, and other descriptions of elements for embodiments described in FIG. 7 may apply to like or similarly
10 named or described elements across any of the figures and/or embodiments described herein and vice versa.

[0066] Each block of method 700, described herein, comprises a computing process that may be performed using any combination of hardware, firmware, and/or software. For instance, various functions may be carried out by a processor executing instructions stored in
15 memory. The methods may also be embodied as computer-usable instructions stored on computer storage media. The methods may be provided by a standalone application, a service or hosted service (standalone or in combination with another hosted service), or a plug-in to another product, to name a few. In addition, method 700 is described, by way of example, with respect to the image-based three-dimensional occupant assessment systems of any of the
20 figures herein. However, these methods may additionally or alternatively be executed by any one system, or any combination of systems, including, but not limited to, those described herein.

[0067] FIG. 7 illustrates an example method 700 for generating a three-dimensional representation of a size of an occupant of a vehicle based at least on one or more features
25 representing at least a portion of the occupant derived from two-dimensional optical image data and a depth of the one or more features generated from the two-dimensional optical image data and a ground truth cabin depth image corresponding to an interior of the vehicle.

[0068] The method 700, at block B702, includes detecting one or more features representing at least a portion of an occupant of a vehicle based at least on optical image data.
30 In some embodiments, detecting the one or more features may include executing at least one three-dimensional pose detection model to generate a scale-normalized three-dimensional pose estimate of the occupant based at least on the one or more features. For example, an occupant evaluation function may apply the optical image data to at least one person detection model to

define the portion of an image frame representing the occupant and generate a scale-normalized three-dimensional pose estimate based at least on the portion of the image frame representing the occupant. In some embodiments, detecting one or more features may include generating a scale-normalized three-dimensional pose estimate of the occupant based at least on the one or
5 more features, wherein the scale-normalized three-dimensional pose estimate represents a set of kinematic elements of the occupant. In some embodiments, detecting the one or more features may include executing segmentation model, such as the segmentation model 520. In such embodiments, detecting the feature(s) may include generating one or both of a vehicle interior mask and/or an occupant mask based at least on a segmentation of the optical image
10 data. The vehicle interior mask may comprise a first boundary that outlines pixels corresponding to structural elements of the vehicle interior. The occupant mask may comprise a second boundary that outlines pixels corresponding to the one or more features of the occupant.

[0069] The method 700, at block B704, includes determining a depth corresponding to
15 the one or more features based at least on the optical image data and a ground truth cabin depth image corresponding to an interior of the vehicle. For example, the method may generate a range image based at least on the optical image data and the ground truth cabin depth image, wherein the depth corresponding to the one or more features is derived at least from the range image. In some embodiments, determining the depth may include particle transport simulation
20 (e.g., ray tracing, such as illustrated in FIG. 4) to correlate at least a first pixel of the ground truth cabin depth image to the one or more features to determine the depth corresponding to the one or more features. In some embodiments, the range image may be generated at least in part using the vehicle interior mask. In some embodiments, the three-dimensional shape estimate of the occupant may be determined based at least on a pose estimate of the occupant
25 generated from the optical image data.

[0070] The method 700, at block B706, includes generating a three-dimensional representation of the occupant based at least on the one or more features and the depth of the one or more features, wherein the three-dimensional representation comprises at least one characteristic representative of a size of the occupant. The at least one characteristic may
30 comprise either a three-dimensional shape estimate and/or a three-dimensional pose estimate. In some embodiments, the at least one characteristic representative of the size of the occupant is determined based at least on a linear measurement scale three-dimensional pose estimate. In some embodiments, the three-dimensional representation may include a three-dimensional size

- 27 -

estimate derived from the three-dimensional shape estimate and/or the three-dimensional pose estimate. In some embodiments, the characteristic representative of a size of the occupant may be further used to derive other characteristics of the occupant such as, but not limited to weight, height, and/or age.

5 **[0071]** The method 700, at block B708, includes controlling at least one operation of the vehicle based at least on the at least one characteristic. For example, based at least in part on the at least one characteristic representative of a size of the occupant, an interior monitoring system 150 may generate one or more output(s) 154. As an example, the interior monitoring system 150 may use 3D occupant representation data 140 (either alone or in combination with
10 other data such as optical image data 107) to predict the presence and/or location of occupants – such as objects, persons, and/or animals – within the interior space of the vehicle 800. Other systems of the vehicle 800 may determine one or more actions to take based on the predictions and/or control other tasks or operations included in outputs 154. For example, based on output(s) 154, an alarm or warning may be generated, door locks and/or windows may be
15 operated, various functions may be turned on/off, data for a digital assistant, chat bot, digital avatar, and/or the like may be generated, and/or air conditioning or air circulation functions may be operated. In some embodiments, the characteristic representing the size of the occupant from the 3D occupant representation data 140 may be used in conjunction with a child presence detection system to estimate an age of the occupant, and/or control an alert system (e.g., alarm
20 and/or notification systems) based on determining that an occupant under an estimated age threshold may have been inadvertently left alone in the vehicle. In some embodiments, the characteristic representing the size of the occupant may be used in conjunction with other vehicle safety features. For example, in some embodiments, airbag deployments, driver monitoring systems, HMI applications, and/or other vehicle functions may be controlled based
25 at least on a 3D pose, 3D shape, and/or 3D size estimate of the vehicle occupant provided by the 3D occupant representation data 140.

[0072] The systems and methods described herein may be used by, without limitation, non-autonomous vehicles, semi-autonomous vehicles (e.g., in one or more adaptive driver assistance systems (ADAS)), piloted and un-piloted robots or robotic platforms, warehouse
30 vehicles, off-road vehicles, vehicles coupled to one or more trailers, flying vessels, boats, shuttles, emergency response vehicles, motorcycles, electric or motorized bicycles, aircraft, construction vehicles, trains, underwater crafts, remotely operated vehicles such as drones, and/or other vehicle types. Further, the systems and methods described herein may be used for

- 28 -

a variety of purposes, by way of example and without limitation, for machine control, machine locomotion, machine driving, synthetic data generation, model training, perception, augmented reality, virtual reality, mixed reality, robotics, security and surveillance, simulation, autonomous or semi-autonomous machine applications, deep learning, environment
5 simulation, object or actor simulation and/or digital twinning, data center processing, conversational AI, generative AI, light transport simulation (e.g., ray-tracing, path tracing, etc.), collaborative content creation for 3D assets, cloud computing and/or any other suitable applications.

[0073] Disclosed embodiments may be comprised in a variety of different systems such
10 as automotive systems (e.g., a control system for an autonomous or semi-autonomous machine, a perception system for an autonomous or semi-autonomous machine), systems implemented using a robot, aerial systems, medial systems, boating systems, smart area monitoring systems, systems for performing deep learning operations, systems for performing simulation operations, systems for performing digital twin operations, systems implemented using an edge
15 device, systems incorporating one or more virtual machines (VMs), systems for performing generative AI operations using a language model, systems for performing synthetic data generation operations, systems implemented at least partially in a data center, systems for performing conversational AI operations, systems for performing light transport simulation, systems for performing collaborative content creation for 3D assets, systems implemented at
20 least partially using cloud computing resources, and/or other types of systems.

EXAMPLE AUTONOMOUS VEHICLE

[0074] FIG. 8A is an illustration of an example autonomous vehicle 800, in accordance with some embodiments of the present disclosure. The autonomous vehicle 800 (alternatively referred to herein as the “vehicle 800”) may include, without limitation, a passenger vehicle,
25 such as a car, a truck, a bus, a first responder vehicle, a shuttle, an electric or motorized bicycle, a motorcycle, a fire truck, a police vehicle, an ambulance, a boat, a construction vehicle, an underwater craft, a robotic vehicle, a drone, an airplane, a vehicle coupled to a trailer (e.g., a semi-tractor-trailer truck used for hauling cargo), and/or another type of vehicle (e.g., that is unmanned and/or that accommodates one or more passengers). Autonomous vehicles are
30 generally described in terms of automation levels, defined by the National Highway Traffic Safety Administration (NHTSA), a division of the US Department of Transportation, and the

Society of Automotive Engineers (SAE) “Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles” (Standard No. J3016-201806, published on June 15, 2018, Standard No. J3016-201609, published on September 30, 2016, and previous and future versions of this standard). The vehicle 800 may be capable of
5 functionality in accordance with one or more of Level 3 – Level 5 of the autonomous driving levels. The vehicle 800 may be capable of functionality in accordance with one or more of Level 1 – Level 5 of the autonomous driving levels. For example, the vehicle 800 may be capable of driver assistance (Level 1), partial automation (Level 2), conditional automation (Level 3), high automation (Level 4), and/or full automation (Level 5), depending on the
10 embodiment. The term “autonomous,” as used herein, may include any and/or all types of autonomy for the vehicle 800 or other machine, such as being fully autonomous, being highly autonomous, being conditionally autonomous, being partially autonomous, providing assistive autonomy, being semi-autonomous, being primarily autonomous, or other designation.

[0075] The vehicle 800 may include components such as a chassis, a vehicle body,
15 wheels (e.g., 2, 4, 6, 8, 18, etc.), tires, axles, and other components of a vehicle. The vehicle 800 may include a propulsion system 850, such as an internal combustion engine, hybrid electric power plant, an all-electric engine, and/or another propulsion system type. The propulsion system 850 may be connected to a drive train of the vehicle 800, which may include a transmission, to enable the propulsion of the vehicle 800. The propulsion system 850 may
20 be controlled in response to receiving signals from the throttle/accelerator 852.

[0076] A steering system 854, which may include a steering wheel, may be used to steer the vehicle 800 (e.g., along a desired path or route) when the propulsion system 850 is operating (e.g., when the vehicle is in motion). The steering system 854 may receive signals from a steering actuator 856. The steering wheel may be optional for full automation (Level
25 5) functionality.

[0077] The brake sensor system 846 may be used to operate the vehicle brakes in response to receiving signals from the brake actuators 848 and/or brake sensors.

[0078] Controller(s) 836, which may include one or more system on chips (SoCs) 804 (FIG. 8C) and/or GPU(s), may provide signals (e.g., representative of commands) to one or
30 more components and/or systems of the vehicle 800. In some embodiments, one or more operations executed by the controller(s) 836 may be performed in response to the output(s) 154 generated by the interior monitoring system 150. For example, the controller(s) may send signals to operate the vehicle brakes via one or more brake actuators 848, to operate the steering

- 30 -

system 854 via one or more steering actuators 856, to operate the propulsion system 850 via one or more throttle/accelerators 852. The controller(s) 836 may include one or more onboard (e.g., integrated) computing devices (e.g., supercomputers) that process sensor signals, and output operation commands (e.g., signals representing commands) to enable autonomous driving and/or to assist a human driver in driving the vehicle 800. The controller(s) 836 may include a first controller 836 for autonomous driving functions, a second controller 836 for functional safety functions, a third controller 836 for artificial intelligence functionality (e.g., computer vision), a fourth controller 836 for infotainment functionality, a fifth controller 836 for redundancy in emergency conditions, and/or other controllers. In some examples, a single controller 836 may handle two or more of the above functionalities, two or more controllers 836 may handle a single functionality, and/or any combination thereof.

[0079] The controller(s) 836 may provide the signals for controlling one or more components and/or systems of the vehicle 800 in response to sensor data received from one or more sensors (e.g., sensor inputs). The sensor data may be received from, for example and without limitation, global navigation satellite systems (“GNSS”) sensor(s) 858 (e.g., Global Positioning System sensor(s)), RADAR sensor(s) 860, ultrasonic sensor(s) 862, LIDAR sensor(s) 864, inertial measurement unit (IMU) sensor(s) 866 (e.g., accelerometer(s), gyroscope(s), magnetic compass(es), magnetometer(s), etc.), microphone(s) 896, stereo camera(s) 868, wide-view camera(s) 870 (e.g., fisheye cameras), infrared camera(s) 872, surround camera(s) 874 (e.g., 360 degree cameras), long-range and/or mid-range camera(s) 898, speed sensor(s) 844 (e.g., for measuring the speed of the vehicle 800), vibration sensor(s) 842, steering sensor(s) 840, brake sensor(s) (e.g., as part of the brake sensor system 846), one or more occupant monitoring system (OMS) sensor(s) 801 (e.g., one or more interior cameras), and/or other sensor types.

[0080] One or more of the controller(s) 836 may receive inputs (e.g., represented by input data) from an instrument cluster 832 of the vehicle 800 and provide outputs (e.g., represented by output data, display data, etc.) via a human-machine interface (HMI) display 834, an audible annunciator, a loudspeaker, and/or via other components of the vehicle 800. The outputs may include information such as vehicle velocity, speed, time, map data (e.g., the High Definition (“HD”) map 822 of FIG. 8C), location data (e.g., the vehicle’s 800 location, such as on a map), direction, location of other vehicles (e.g., an occupancy grid), information about objects and status of objects as perceived by the controller(s) 836, etc. For example, the HMI display 834 may display information about the presence of one or more objects (e.g., a

- 31 -

street sign, caution sign, traffic light changing, etc.), and/or information about driving maneuvers the vehicle has made, is making, or will make (e.g., changing lanes now, taking exit 34B in two miles, etc.).

[0081] The vehicle 800 further includes a network interface 824 which may use one or
5 more wireless antenna(s) 826 and/or modem(s) to communicate over one or more networks. For example, the network interface 824 may be capable of communication over Long-Term Evolution ("LTE"), Wideband Code Division Multiple Access ("WCDMA"), Universal Mobile Telecommunications System ("UMTS"), Global System for Mobile communication ("GSM"), IMT-CDMA Multi-Carrier ("CDMA2000"), etc. The wireless antenna(s) 826 may
10 also enable communication between objects in the environment (e.g., vehicles, mobile devices, etc.), using local area network(s), such as Bluetooth, Bluetooth Low Energy ("LE"), Z-Wave, ZigBee, etc., and/or low power wide-area network(s) ("LPWANs"), such as LoRaWAN, SigFox, etc.

[0082] FIG. 8B is an example of camera locations and fields of view for the example
15 autonomous vehicle 800 of FIG. 8A, in accordance with some embodiments of the present disclosure. The cameras and respective fields of view are one example embodiment and are not intended to be limiting. For example, additional and/or alternative cameras may be included and/or the cameras may be located at different locations on the vehicle 800.

[0083] The camera types for the cameras may include, but are not limited to, digital
20 cameras that may be adapted for use with the components and/or systems of the vehicle 800. The camera(s) may operate at automotive safety integrity level (ASIL) B and/or at another ASIL. The camera types may be capable of any image capture rate, such as 60 frames per second (fps), 120 fps, 240 fps, etc., depending on the embodiment. The cameras may be capable of using rolling shutters, global shutters, another type of shutter, or a combination
25 thereof. In some examples, the color filter array may include a red clear clear clear (RCCC) color filter array, a red clear clear blue (RCCB) color filter array, a red blue green clear (RBGC) color filter array, a Foveon X3 color filter array, a Bayer sensors (RGGB) color filter array, a monochrome sensor color filter array, and/or another type of color filter array. In some embodiments, clear pixel cameras, such as cameras with an RCCC, an RCCB, and/or an RBGC
30 color filter array, may be used in an effort to increase light sensitivity.

[0084] In some examples, one or more of the camera(s) may be used to perform advanced driver assistance systems (ADAS) functions (e.g., as part of a redundant or fail-safe design). For example, a Multi-Function Mono Camera may be installed to provide functions

including lane departure warning, traffic sign assist and intelligent headlamp control. One or more of the camera(s) (e.g., all of the cameras) may record and provide image data (e.g., video) simultaneously.

[0085] One or more of the cameras may be mounted in a mounting assembly, such as
5 a custom designed (three dimensional (“3D”) printed) assembly, in order to cut out stray light and reflections from within the car (e.g., reflections from the dashboard reflected in the windshield mirrors) which may interfere with the camera’s image data capture abilities. With reference to wing-mirror mounting assemblies, the wing-mirror assemblies may be custom 3D printed so that the camera mounting plate matches the shape of the wing-mirror. In some
10 examples, the camera(s) may be integrated into the wing-mirror. For side-view cameras, the camera(s) may also be integrated within the four pillars at each corner of the cabin.

[0086] Cameras with a field of view that include portions of the environment in front of the vehicle 800 (e.g., front-facing cameras) may be used for surround view, to help identify forward facing paths and obstacles, as well aid in, with the help of one or more controllers 836
15 and/or control SoCs, providing information critical to generating an occupancy grid and/or determining the preferred vehicle paths. Front-facing cameras may be used to perform many of the same ADAS functions as LIDAR, including emergency braking, pedestrian detection, and collision avoidance. Front-facing cameras may also be used for ADAS functions and systems including Lane Departure Warnings (“LDW”), Autonomous Cruise Control (“ACC”),
20 and/or other functions such as traffic sign recognition.

[0087] A variety of cameras may be used in a front-facing configuration, including, for example, a monocular camera platform that includes a complementary metal oxide semiconductor (“CMOS”) color imager. Another example may be a wide-view camera(s) 870
25 that may be used to perceive objects coming into view from the periphery (e.g., pedestrians, crossing traffic or bicycles). Although only one wide-view camera is illustrated in FIG. 8B, there may be any number (including zero) of wide-view cameras 870 on the vehicle 800. In addition, any number of long-range camera(s) 898 (e.g., a long-view stereo camera pair) may be used for depth-based object detection, especially for objects for which a neural network has not yet been trained. The long-range camera(s) 898 may also be used for object detection and
30 classification, as well as basic object tracking.

[0088] Any number of stereo cameras 868 may also be included in a front-facing configuration. In at least one embodiment, one or more of stereo camera(s) 868 may include an integrated control unit comprising a scalable processing unit, which may provide a

programmable logic (“FPGA”) and a multi-core micro-processor with an integrated Controller Area Network (“CAN”) or Ethernet interface on a single chip. Such a unit may be used to generate a 3D map of the vehicle’s environment, including a distance estimate for all the points in the image. An alternative stereo camera(s) 868 may include a compact stereo vision sensor(s) that may include two camera lenses (one each on the left and right) and an image processing chip that may measure the distance from the vehicle to the target object and use the generated information (e.g., metadata) to activate the autonomous emergency braking and lane departure warning functions. Other types of stereo camera(s) 868 may be used in addition to, or alternatively from, those described herein.

10 [0089] Cameras with a field of view that include portions of the environment to the side of the vehicle 800 (e.g., side-view cameras) may be used for surround view, providing information used to create and update the occupancy grid, as well as to generate side impact collision warnings. For example, surround camera(s) 874 (e.g., four surround cameras 874 as illustrated in FIG. 8B) may be positioned to on the vehicle 800. The surround camera(s) 874
15 may include wide-view camera(s) 870, fisheye camera(s), 360-degree camera(s), and/or the like. For example, four fisheye cameras may be positioned on the vehicle’s front, rear, and sides. In an alternative arrangement, the vehicle may use three surround camera(s) 874 (e.g., left, right, and rear), and may leverage one or more other camera(s) (e.g., a forward-facing camera) as a fourth surround view camera.

20 [0090] Cameras with a field of view that include portions of the environment to the rear of the vehicle 800 (e.g., rear-view cameras) may be used for park assistance, surround view, rear collision warnings, and creating and updating the occupancy grid. A wide variety of cameras may be used including, but not limited to, cameras that are also suitable as a front-facing camera(s) (e.g., long-range and/or mid-range camera(s) 898, stereo camera(s) 868),
25 infrared camera(s) 872, etc.), as described herein.

[0091] Cameras with a field of view that include portions of the interior environment within the cabin of the vehicle 800 (e.g., one or more OMS sensor(s) 801) may be used as part of an occupant monitoring system (OMS) such as, but not limited to, a driver monitoring system (DMS). For example, OMS sensors (e.g., the OMS sensor(s) 801) may be used (e.g.,
30 by the controller(s) 836) to track an occupant’s and/or driver’s gaze direction, head pose, and/or blinking. This gaze information may be used to determine a level of attentiveness of the occupant or driver (e.g., to detect drowsiness, fatigue, and/or distraction), and/or to take responsive action to prevent harm to the occupant or operator. In some embodiments, data

from OMS sensors may be used to enable gaze-controlled operations triggered by driver and/or non-driver occupants such as, but not limited to, adjusting cabin temperature and/or airflow, opening and closing windows, controlling cabin lighting, controlling entertainment systems, adjusting mirrors, adjusting seat positions, and/or other operations. In some embodiments, an OMS may be used for applications such as determining when objects and/or occupants have been left behind in a vehicle cabin (e.g., by detecting occupant presence after the driver exits the vehicle).

[0092] FIG. 8C is a block diagram of an example system architecture for the example autonomous vehicle 800 of FIG. 8A, in accordance with some embodiments of the present disclosure. It should be understood that this and other arrangements described herein are set forth only as examples. Other arrangements and elements (e.g., machines, interfaces, functions, orders, groupings of functions, etc.) may be used in addition to or instead of those shown, and some elements may be omitted altogether. Further, many of the elements described herein are functional entities that may be implemented as discrete or distributed components or in conjunction with other components, and in any suitable combination and location. Various functions described herein as being performed by entities may be carried out by hardware, firmware, and/or software. For instance, various functions may be carried out by a processor executing instructions stored in memory.

[0093] Each of the components, features, and systems of the vehicle 800 in FIG. 8C are illustrated as being connected via bus 802. The bus 802 may include a Controller Area Network (CAN) data interface (alternatively referred to herein as a “CAN bus”). A CAN may be a network inside the vehicle 800 used to aid in control of various features and functionality of the vehicle 800, such as actuation of brakes, acceleration, braking, steering, windshield wipers, etc. A CAN bus may be configured to have dozens or even hundreds of nodes, each with its own unique identifier (e.g., a CAN ID). The CAN bus may be read to find steering wheel angle, ground speed, engine revolutions per minute (RPMs), button positions, and/or other vehicle status indicators. The CAN bus may be ASIL B compliant.

[0094] Although the bus 802 is described herein as being a CAN bus, this is not intended to be limiting. For example, in addition to, or alternatively from, the CAN bus, FlexRay and/or Ethernet may be used. Additionally, although a single line is used to represent the bus 802, this is not intended to be limiting. For example, there may be any number of busses 802, which may include one or more CAN busses, one or more FlexRay busses, one or more Ethernet busses, and/or one or more other types of busses using a different protocol. In

some examples, two or more busses 802 may be used to perform different functions, and/or may be used for redundancy. For example, a first bus 802 may be used for collision avoidance functionality and a second bus 802 may be used for actuation control. In any example, each bus 802 may communicate with any of the components of the vehicle 800, and two or more
5 busses 802 may communicate with the same components. In some examples, each SoC 804, each controller 836, and/or each computer within the vehicle may have access to the same input data (e.g., inputs from sensors of the vehicle 800), and may be connected to a common bus, such the CAN bus.

[0095] The vehicle 800 may include one or more controller(s) 836, such as those
10 described herein with respect to FIG. 8A. The controller(s) 836 may be used for a variety of functions. The controller(s) 836 may be coupled to any of the various other components and systems of the vehicle 800, and may be used for control of the vehicle 800, artificial intelligence of the vehicle 800, infotainment for the vehicle 800, and/or the like.

[0096] The vehicle 800 may include a system(s) on a chip (SoC) 804. The SoC 804
15 may include CPU(s) 806, GPU(s) 808, processor(s) 810, cache(s) 812, accelerator(s) 814, data store(s) 816, and/or other components and features not illustrated. The SoC(s) 804 may be used to control the vehicle 800 in a variety of platforms and systems. For example, the SoC(s) 804 may be combined in a system (e.g., the system of the vehicle 800) with an HD map 822 which may obtain map refreshes and/or updates via a network interface 824 from one or more
20 servers (e.g., server(s) 878 of FIG. 8D).

[0097] The CPU(s) 806 may include a CPU cluster or CPU complex (alternatively referred to herein as a “CCPLEX”). The CPU(s) 806 may include multiple cores and/or L2 caches. For example, in some embodiments, the CPU(s) 806 may include eight cores in a coherent multi-processor configuration. In some embodiments, the CPU(s) 806 may include
25 four dual-core clusters where each cluster has a dedicated L2 cache (e.g., a 2 MB L2 cache). The CPU(s) 806 (e.g., the CCPLEX) may be configured to support simultaneous cluster operation enabling any combination of the clusters of the CPU(s) 806 to be active at any given time.

[0098] The CPU(s) 806 may implement power management capabilities that include
30 one or more of the following features: individual hardware blocks may be clock-gated automatically when idle to save dynamic power; each core clock may be gated when the core is not actively executing instructions due to execution of WFI/WFE instructions; each core may be independently power-gated; each core cluster may be independently clock-gated when all

cores are clock-gated or power-gated; and/or each core cluster may be independently power-gated when all cores are power-gated. The CPU(s) 806 may further implement an enhanced algorithm for managing power states, where allowed power states and expected wakeup times are specified, and the hardware/microcode determines the best power state to enter for the core, cluster, and CCPLX. The processing cores may support simplified power state entry sequences in software with the work offloaded to microcode.

[0099] The GPU(s) 808 may include an integrated GPU (alternatively referred to herein as an “iGPU”). The GPU(s) 808 may be programmable and may be efficient for parallel workloads. The GPU(s) 808, in some examples, may use an enhanced tensor instruction set.

10 The GPU(s) 808 may include one or more streaming microprocessors, where each streaming microprocessor may include an L1 cache (e.g., an L1 cache with at least 96KB storage capacity), and two or more of the streaming microprocessors may share an L2 cache (e.g., an L2 cache with a 512 KB storage capacity). In some embodiments, the GPU(s) 808 may include at least eight streaming microprocessors. The GPU(s) 808 may use compute application programming interface(s) (API(s)). In addition, the GPU(s) 808 may use one or more parallel computing platforms and/or programming models (e.g., NVIDIA’s CUDA).

15 [00100] The GPU(s) 808 may be power-optimized for best performance in automotive and embedded use cases. For example, the GPU(s) 808 may be fabricated on a Fin field-effect transistor (FinFET). However, this is not intended to be limiting and the GPU(s) 808 may be fabricated using other semiconductor manufacturing processes. Each streaming microprocessor may incorporate a number of mixed-precision processing cores partitioned into multiple blocks. For example, and without limitation, 64 PF32 cores and 32 PF64 cores may be partitioned into four processing blocks. In such an example, each processing block may be allocated 16 FP32 cores, 8 FP64 cores, 16 INT32 cores, two mixed-precision NVIDIA
25 TENSOR COREs for deep learning matrix arithmetic, an L0 instruction cache, a warp scheduler, a dispatch unit, and/or a 64 KB register file. In addition, the streaming microprocessors may include independent parallel integer and floating-point data paths to provide for efficient execution of workloads with a mix of computation and addressing calculations. The streaming microprocessors may include independent thread scheduling capability to enable finer-grain synchronization and cooperation between parallel threads. The streaming microprocessors may include a combined L1 data cache and shared memory unit in
30 order to improve performance while simplifying programming.

[00101] The GPU(s) 808 may include a high bandwidth memory (HBM) and/or a 16 GB HBM2 memory subsystem to provide, in some examples, about 900 GB/second peak memory bandwidth. In some examples, in addition to, or alternatively from, the HBM memory, a synchronous graphics random-access memory (SGRAM) may be used, such as a graphics double data rate type five synchronous random-access memory (GDDR5).

[00102] The GPU(s) 808 may include unified memory technology including access counters to allow for more accurate migration of memory pages to the processor that accesses them most frequently, thereby improving efficiency for memory ranges shared between processors. In some examples, address translation services (ATS) support may be used to allow the GPU(s) 808 to access the CPU(s) 806 page tables directly. In such examples, when the GPU(s) 808 memory management unit (MMU) experiences a miss, an address translation request may be transmitted to the CPU(s) 806. In response, the CPU(s) 806 may look in its page tables for the virtual-to-physical mapping for the address and transmits the translation back to the GPU(s) 808. As such, unified memory technology may allow a single unified virtual address space for memory of both the CPU(s) 806 and the GPU(s) 808, thereby simplifying the GPU(s) 808 programming and porting of applications to the GPU(s) 808.

[00103] In addition, the GPU(s) 808 may include an access counter that may keep track of the frequency of access of the GPU(s) 808 to memory of other processors. The access counter may help ensure that memory pages are moved to the physical memory of the processor that is accessing the pages most frequently.

[00104] The SoC(s) 804 may include any number of cache(s) 812, including those described herein. For example, the cache(s) 812 may include an L3 cache that is available to both the CPU(s) 806 and the GPU(s) 808 (e.g., that is connected both the CPU(s) 806 and the GPU(s) 808). The cache(s) 812 may include a write-back cache that may keep track of states of lines, such as by using a cache coherence protocol (e.g., MEI, MESI, MSI, etc.). The L3 cache may include 4 MB or more, depending on the embodiment, although smaller cache sizes may be used.

[00105] The SoC(s) 804 may include an arithmetic logic unit(s) (ALU(s)) which may be leveraged in performing processing with respect to any of the variety of tasks or operations of the vehicle 800 – such as processing DNNs. In addition, the SoC(s) 804 may include a floating point unit(s) (FPU(s)) – or other math coprocessor or numeric coprocessor types – for performing mathematical operations within the system. For example, the SoC(s) 804 may

include one or more FPUs integrated as execution units within a CPU(s) 806 and/or GPU(s) 808.

[00106] The SoC(s) 804 may include one or more accelerators 814 (e.g., hardware accelerators, software accelerators, or a combination thereof). For example, the SoC(s) 804
5 may include a hardware acceleration cluster that may include optimized hardware accelerators and/or large on-chip memory. The large on-chip memory (e.g., 4MB of SRAM), may enable the hardware acceleration cluster to accelerate neural networks and other calculations. The hardware acceleration cluster may be used to complement the GPU(s) 808 and to off-load some of the tasks of the GPU(s) 808 (e.g., to free up more cycles of the GPU(s) 808 for performing
10 other tasks). As an example, the accelerator(s) 814 may be used for targeted workloads (e.g., perception, convolutional neural networks (CNNs), etc.) that are stable enough to be amenable to acceleration. The term “CNN,” as used herein, may include all types of CNNs, including region-based or regional convolutional neural networks (RCNNs) and Fast RCNNs (e.g., as used for object detection).

[00107] The accelerator(s) 814 (e.g., the hardware acceleration cluster) may include a
15 deep learning accelerator(s) (DLA). The DLA(s) may include one or more Tensor processing units (TPUs) that may be configured to provide an additional ten trillion operations per second for deep learning applications and inferencing. The TPUs may be accelerators configured to, and optimized for, performing image processing functions (e.g., for CNNs, RCNNs, etc.). The
20 DLA(s) may further be optimized for a specific set of neural network types and floating point operations, as well as inferencing. The design of the DLA(s) may provide more performance per millimeter than a general-purpose GPU, and vastly exceeds the performance of a CPU. The TPU(s) may perform several functions, including a single-instance convolution function, supporting, for example, INT8, INT16, and FP16 data types for both features and weights, as
25 well as post-processor functions.

[00108] The DLA(s) may quickly and efficiently execute neural networks, especially CNNs, on processed or unprocessed data for any of a variety of functions, including, for example and without limitation: a CNN for object identification and detection using data from camera sensors; a CNN for distance estimation using data from camera sensors; a CNN for
30 emergency vehicle detection and identification and detection using data from microphones; a CNN for facial recognition and vehicle owner identification using data from camera sensors; and/or a CNN for security and/or safety related events.

[00109] The DLA(s) may perform any function of the GPU(s) 808, and by using an inference accelerator, for example, a designer may target either the DLA(s) or the GPU(s) 808 for any function. For example, the designer may focus processing of CNNs and floating point operations on the DLA(s) and leave other functions to the GPU(s) 808 and/or other accelerator(s) 814. In some embodiments, one or more of the models disclosed above of the occupant evaluation function 110 may be executed on a SoC 804, CPU(s) 806, GPU(s) 808 and/or accelerators 814.

[00110] The accelerator(s) 814 (e.g., the hardware acceleration cluster) may include a programmable vision accelerator(s) (PVA), which may alternatively be referred to herein as a computer vision accelerator. The PVA(s) may be designed and configured to accelerate computer vision algorithms for the advanced driver assistance systems (ADAS), autonomous driving, and/or augmented reality (AR) and/or virtual reality (VR) applications. The PVA(s) may provide a balance between performance and flexibility. For example, each PVA(s) may include, for example and without limitation, any number of reduced instruction set computer (RISC) cores, direct memory access (DMA), and/or any number of vector processors.

[00111] The RISC cores may interact with image sensors (e.g., the image sensors of any of the cameras described herein), image signal processor(s), and/or the like. Each of the RISC cores may include any amount of memory. The RISC cores may use any of a number of protocols, depending on the embodiment. In some examples, the RISC cores may execute a real-time operating system (RTOS). The RISC cores may be implemented using one or more integrated circuit devices, application specific integrated circuits (ASICs), and/or memory devices. For example, the RISC cores may include an instruction cache and/or a tightly coupled RAM.

[00112] The DMA may enable components of the PVA(s) to access the system memory independently of the CPU(s) 806. The DMA may support any number of features used to provide optimization to the PVA including, but not limited to, supporting multi-dimensional addressing and/or circular addressing. In some examples, the DMA may support up to six or more dimensions of addressing, which may include block width, block height, block depth, horizontal block stepping, vertical block stepping, and/or depth stepping.

[00113] The vector processors may be programmable processors that may be designed to efficiently and flexibly execute programming for computer vision algorithms and provide signal processing capabilities. In some examples, the PVA may include a PVA core and two vector processing subsystem partitions. The PVA core may include a processor subsystem,

DMA engine(s) (e.g., two DMA engines), and/or other peripherals. The vector processing subsystem may operate as the primary processing engine of the PVA, and may include a vector processing unit (VPU), an instruction cache, and/or vector memory (e.g., VMEM). A VPU core may include a digital signal processor such as, for example, a single instruction, multiple data (SIMD), very long instruction word (VLIW) digital signal processor. The combination of the SIMD and VLIW may enhance throughput and speed.

[00114] Each of the vector processors may include an instruction cache and may be coupled to dedicated memory. As a result, in some examples, each of the vector processors may be configured to execute independently of the other vector processors. In other examples, the vector processors that are included in a particular PVA may be configured to employ data parallelism. For example, in some embodiments, the plurality of vector processors included in a single PVA may execute the same computer vision algorithm, but on different regions of an image. In other examples, the vector processors included in a particular PVA may simultaneously execute different computer vision algorithms, on the same image, or even execute different algorithms on sequential images or portions of an image. Among other things, any number of PVAs may be included in the hardware acceleration cluster and any number of vector processors may be included in each of the PVAs. In addition, the PVA(s) may include additional error correcting code (ECC) memory, to enhance overall system safety.

[00115] The accelerator(s) 814 (e.g., the hardware acceleration cluster) may include a computer vision network on-chip and SRAM, for providing a high-bandwidth, low latency SRAM for the accelerator(s) 814. In some examples, the on-chip memory may include at least 4MB SRAM, consisting of, for example and without limitation, eight field-configurable memory blocks that may be accessible by both the PVA and the DLA. Each pair of memory blocks may include an advanced peripheral bus (APB) interface, configuration circuitry, a controller, and a multiplexer. Any type of memory may be used. The PVA and DLA may access the memory via a backbone that provides the PVA and DLA with high-speed access to memory. The backbone may include a computer vision network on-chip that interconnects the PVA and the DLA to the memory (e.g., using the APB).

[00116] The computer vision network on-chip may include an interface that determines, before transmission of any control signal/address/data, that both the PVA and the DLA provide ready and valid signals. Such an interface may provide for separate phases and separate channels for transmitting control signals/addresses/data, as well as burst-type communications

for continuous data transfer. This type of interface may comply with ISO 26262 or IEC 61508 standards, although other standards and protocols may be used.

[00117] In some examples, the SoC(s) 804 may include a real-time ray-tracing hardware accelerator, such as described in U.S. Patent Application No. 16/101,232, filed on August 10, 5 2018. The real-time ray-tracing hardware accelerator may be used to quickly and efficiently determine the positions and extents of objects (e.g., within a world model), to generate real-time visualization simulations, for RADAR signal interpretation, for sound propagation synthesis and/or analysis, for simulation of SONAR systems, for general wave propagation simulation, for comparison to LIDAR data for purposes of localization and/or other functions, 10 and/or for other uses. In some embodiments, one or more tree traversal units (TTUs) may be used for executing one or more ray-tracing related operations.

[00118] The accelerator(s) 814 (e.g., the hardware accelerator cluster) have a wide array of uses for autonomous driving. The PVA may be a programmable vision accelerator that may be used for key processing stages in ADAS and autonomous vehicles. The PVA's capabilities 15 are a good match for algorithmic domains needing predictable processing, at low power and low latency. In other words, the PVA performs well on semi-dense or dense regular computation, even on small data sets, which need predictable run-times with low latency and low power. Thus, in the context of platforms for autonomous vehicles, the PVAs are designed to run classic computer vision algorithms, as they are efficient at object detection and operating 20 on integer math.

[00119] For example, according to one embodiment of the technology, the PVA is used to perform computer stereo vision. A semi-global matching-based algorithm may be used in some examples, although this is not intended to be limiting. Many applications for Level 3-5 autonomous driving require motion estimation/stereo matching on-the-fly (e.g., structure from 25 motion, pedestrian recognition, lane detection, etc.). The PVA may perform computer stereo vision function on inputs from two monocular cameras.

[00120] In some examples, the PVA may be used to perform dense optical flow. According to process raw RADAR data (e.g., using a 4D Fast Fourier Transform) to provide Processed RADAR. In other examples, the PVA is used for time of flight depth processing, 30 by processing raw time of flight data to provide processed time of flight data, for example.

[00121] The DLA may be used to run any type of network to enhance control and driving safety, including for example, a neural network that outputs a measure of confidence for each object detection. Such a confidence value may be interpreted as a probability, or as providing

a relative “weight” of each detection compared to other detections. This confidence value enables the system to make further decisions regarding which detections should be considered as true positive detections rather than false positive detections. For example, the system may set a threshold value for the confidence and consider only the detections exceeding the
5 threshold value as true positive detections. In an automatic emergency braking (AEB) system, false positive detections would cause the vehicle to automatically perform emergency braking, which is obviously undesirable. Therefore, only the most confident detections should be considered as triggers for AEB. The DLA may run a neural network for regressing the confidence value. The neural network may take as its input at least some subset of parameters,
10 such as bounding box dimensions, ground plane estimate obtained (e.g., from another subsystem), inertial measurement unit (IMU) sensor 866 output that correlates with the vehicle 800 orientation, distance, 3D location estimates of the object obtained from the neural network and/or other sensors (e.g., LIDAR sensor(s) 864 or RADAR sensor(s) 860), among others.

[00122] The SoC(s) 804 may include data store(s) 816 (e.g., memory). The data store(s)
15 816 may be on-chip memory of the SoC(s) 804, which may store neural networks to be executed on the GPU and/or the DLA. In some examples, the data store(s) 816 may be large enough in capacity to store multiple instances of neural networks for redundancy and safety. The data store(s) 816 may comprise L2 or L3 cache(s) 812. Reference to the data store(s) 816 may include reference to the memory associated with the PVA, DLA, and/or other
20 accelerator(s) 814, as described herein.

[00123] The SoC(s) 804 may include one or more processor(s) 810 (e.g., embedded processors). The processor(s) 810 may include a boot and power management processor that may be a dedicated processor and subsystem to handle boot power and management functions and related security enforcement. The boot and power management processor may be a part
25 of the SoC(s) 804 boot sequence and may provide runtime power management services. The boot power and management processor may provide clock and voltage programming, assistance in system low power state transitions, management of SoC(s) 804 thermals and temperature sensors, and/or management of the SoC(s) 804 power states. Each temperature sensor may be implemented as a ring-oscillator whose output frequency is proportional to
30 temperature, and the SoC(s) 804 may use the ring-oscillators to detect temperatures of the CPU(s) 806, GPU(s) 808, and/or accelerator(s) 814. If temperatures are determined to exceed a threshold, the boot and power management processor may enter a temperature fault routine

and put the SoC(s) 804 into a lower power state and/or put the vehicle 800 into a chauffeur to safe stop mode (e.g., bring the vehicle 800 to a safe stop).

[00124] The processor(s) 810 may further include a set of embedded processors that may serve as an audio processing engine. The audio processing engine may be an audio subsystem
5 that enables full hardware support for multi-channel audio over multiple interfaces, and a broad and flexible range of audio I/O interfaces. In some examples, the audio processing engine is a dedicated processor core with a digital signal processor with dedicated RAM.

[00125] The processor(s) 810 may further include an always on processor engine that may provide necessary hardware features to support low power sensor management and wake
10 use cases. The always on processor engine may include a processor core, a tightly coupled RAM, supporting peripherals (e.g., timers and interrupt controllers), various I/O controller peripherals, and routing logic.

[00126] The processor(s) 810 may further include a safety cluster engine that includes a dedicated processor subsystem to handle safety management for automotive applications. The
15 safety cluster engine may include two or more processor cores, a tightly coupled RAM, support peripherals (e.g., timers, an interrupt controller, etc.), and/or routing logic. In a safety mode, the two or more cores may operate in a lockstep mode and function as a single core with comparison logic to detect any differences between their operations.

[00127] The processor(s) 810 may further include a real-time camera engine that may
20 include a dedicated processor subsystem for handling real-time camera management.

[00128] The processor(s) 810 may further include a high-dynamic range signal processor that may include an image signal processor that is a hardware engine that is part of the camera processing pipeline.

[00129] The processor(s) 810 may include a video image compositor that may be a
25 processing block (e.g., implemented on a microprocessor) that implements video post-processing functions needed by a video playback application to produce the final image for the player window. The video image compositor may perform lens distortion correction on wide-view camera(s) 870, surround camera(s) 874, and/or on in-cabin monitoring camera sensors (such as OMS sensor(s) 801. In-cabin monitoring camera sensor is preferably monitored by a
30 neural network running on another instance of the Advanced SoC, configured to identify in cabin events and respond accordingly. An in-cabin system may perform lip reading to activate cellular service and place a phone call, dictate emails, change the vehicle's destination, activate or change the vehicle's infotainment system and settings, or provide voice-activated web

surfing. Certain functions are available to the driver only when the vehicle is operating in an autonomous mode, and are disabled otherwise.

[00130] The video image compositor may include enhanced temporal noise reduction for both spatial and temporal noise reduction. For example, where motion occurs in a video,
5 the noise reduction weights spatial information appropriately, decreasing the weight of information provided by adjacent frames. Where an image or portion of an image does not include motion, the temporal noise reduction performed by the video image compositor may use information from the previous image to reduce noise in the current image.

[00131] The video image compositor may also be configured to perform stereo
10 rectification on input stereo lens frames. The video image compositor may further be used for user interface composition when the operating system desktop is in use, and the GPU(s) 808 is not required to continuously render new surfaces. Even when the GPU(s) 808 is powered on and active doing 3D rendering, the video image compositor may be used to offload the GPU(s) 808 to improve performance and responsiveness.

[00132] The SoC(s) 804 may further include a mobile industry processor interface (MIPI) camera serial interface for receiving video and input from cameras, a high-speed interface, and/or a video input block that may be used for camera and related pixel input functions. The SoC(s) 804 may further include an input/output controller(s) that may be controlled by software and may be used for receiving I/O signals that are uncommitted to a
20 specific role.

[00133] The SoC(s) 804 may further include a broad range of peripheral interfaces to enable communication with peripherals, audio codecs, power management, and/or other devices. The SoC(s) 804 may be used to process data from cameras (e.g., connected over Gigabit Multimedia Serial Link and Ethernet), sensors (e.g., LIDAR sensor(s) 864, RADAR
25 sensor(s) 860, etc. that may be connected over Ethernet), data from bus 802 (e.g., speed of vehicle 800, steering wheel position, etc.), data from GNSS sensor(s) 858 (e.g., connected over Ethernet or CAN bus). The SoC(s) 804 may further include dedicated high-performance mass storage controllers that may include their own DMA engines, and that may be used to free the CPU(s) 806 from routine data management tasks.

[00134] The SoC(s) 804 may be an end-to-end platform with a flexible architecture that spans automation levels 3-5, thereby providing a comprehensive functional safety architecture that leverages and makes efficient use of computer vision and ADAS techniques for diversity and redundancy, provides a platform for a flexible, reliable driving software stack, along with

deep learning tools. The SoC(s) 804 may be faster, more reliable, and even more energy-efficient and space-efficient than conventional systems. For example, the accelerator(s) 814, when combined with the CPU(s) 806, the GPU(s) 808, and the data store(s) 816, may provide for a fast, efficient platform for level 3-5 autonomous vehicles.

5 [00135] The technology thus provides capabilities and functionality that cannot be achieved by conventional systems. For example, computer vision algorithms may be executed on CPUs, which may be configured using high-level programming language, such as the C programming language, to execute a wide variety of processing algorithms across a wide variety of visual data. However, CPUs are oftentimes unable to meet the performance
10 requirements of many computer vision applications, such as those related to execution time and power consumption, for example. In particular, many CPUs are unable to execute complex object detection algorithms in real-time, which is a requirement of in-vehicle ADAS applications, and a requirement for practical Level 3-5 autonomous vehicles.

[00136] In contrast to conventional systems, by providing a CPU complex, GPU
15 complex, and a hardware acceleration cluster, the technology described herein allows for multiple neural networks to be performed simultaneously and/or sequentially, and for the results to be combined together to enable Level 3-5 autonomous driving functionality. For example, a CNN executing on the DLA or dGPU (e.g., the GPU(s) 820) may include a text and word recognition, allowing the supercomputer to read and understand traffic signs, including
20 signs for which the neural network has not been specifically trained. The DLA may further include a neural network that is able to identify, interpret, and provides semantic understanding of the sign, and to pass that semantic understanding to the path planning modules running on the CPU Complex.

[00137] As another example, multiple neural networks may be run simultaneously, as is
25 required for Level 3, 4, or 5 driving. For example, a warning sign consisting of “Caution: flashing lights indicate icy conditions,” along with an electric light, may be independently or collectively interpreted by several neural networks. The sign itself may be identified as a traffic sign by a first deployed neural network (e.g., a neural network that has been trained), the text “Flashing lights indicate icy conditions” may be interpreted by a second deployed neural
30 network, which informs the vehicle’s path planning software (preferably executing on the CPU Complex) that when flashing lights are detected, icy conditions exist. The flashing light may be identified by operating a third deployed neural network over multiple frames, informing the

vehicle's path-planning software of the presence (or absence) of flashing lights. All three neural networks may run simultaneously, such as within the DLA and/or on the GPU(s) 808.

[00138] In some examples, a CNN for facial recognition and vehicle owner identification may use data from camera sensors to identify the presence of an authorized driver and/or owner of the vehicle 800. The always on sensor processing engine may be used to unlock the vehicle when the owner approaches the driver door and turn on the lights, and, in security mode, to disable the vehicle when the owner leaves the vehicle. In this way, the SoC(s) 804 provide for security against theft and/or carjacking.

[00139] In another example, a CNN for emergency vehicle detection and identification may use data from microphones 896 to detect and identify emergency vehicle sirens. In contrast to conventional systems that use general classifiers to detect sirens and manually extract features, the SoC(s) 804 use the CNN for classifying environmental and urban sounds, as well as classifying visual data. In a preferred embodiment, the CNN running on the DLA is trained to identify the relative closing speed of the emergency vehicle (e.g., by using the Doppler Effect). The CNN may also be trained to identify emergency vehicles specific to the local area in which the vehicle is operating, as identified by GNSS sensor(s) 858. Thus, for example, when operating in Europe the CNN will seek to detect European sirens, and when in the United States the CNN will seek to identify only North American sirens. Once an emergency vehicle is detected, a control program may be used to execute an emergency vehicle safety routine, slowing the vehicle, pulling over to the side of the road, parking the vehicle, and/or idling the vehicle, with the assistance of ultrasonic sensors 862, until the emergency vehicle(s) passes.

[00140] The vehicle may include a CPU(s) 818 (e.g., discrete CPU(s), or dCPU(s)), that may be coupled to the SoC(s) 804 via a high-speed interconnect (e.g., PCIe). The CPU(s) 818 may include an X86 processor, for example. The CPU(s) 818 may be used to perform any of a variety of functions, including arbitrating potentially inconsistent results between ADAS sensors and the SoC(s) 804, and/or monitoring the status and health of the controller(s) 836 and/or infotainment SoC 830, for example.

[00141] The vehicle 800 may include a GPU(s) 820 (e.g., discrete GPU(s), or dGPU(s)), that may be coupled to the SoC(s) 804 via a high-speed interconnect (e.g., NVIDIA's NVLINK). The GPU(s) 820 may provide additional artificial intelligence functionality, such as by executing redundant and/or different neural networks, and may be used to train and/or update neural networks based on input (e.g., sensor data) from sensors of the vehicle 800.

[00142] The vehicle 800 may further include the network interface 824 which may include one or more wireless antennas 826 (e.g., one or more wireless antennas for different communication protocols, such as a cellular antenna, a Bluetooth antenna, etc.). The network interface 824 may be used to enable wireless connectivity over the Internet with the cloud (e.g.,
5 with the server(s) 878 and/or other network devices), with other vehicles, and/or with computing devices (e.g., client devices of passengers). To communicate with other vehicles, a direct link may be established between the two vehicles and/or an indirect link may be established (e.g., across networks and over the Internet). Direct links may be provided using a vehicle-to-vehicle communication link. The vehicle-to-vehicle communication link may
10 provide the vehicle 800 information about vehicles in proximity to the vehicle 800 (e.g., vehicles in front of, on the side of, and/or behind the vehicle 800). This functionality may be part of a cooperative adaptive cruise control functionality of the vehicle 800.

[00143] The network interface 824 may include a SoC that provides modulation and demodulation functionality and enables the controller(s) 836 to communicate over wireless
15 networks. The network interface 824 may include a radio frequency front-end for up-conversion from baseband to radio frequency, and down conversion from radio frequency to baseband. The frequency conversions may be performed through well-known processes, and/or may be performed using super-heterodyne processes. In some examples, the radio frequency front end functionality may be provided by a separate chip. The network interface
20 may include wireless functionality for communicating over LTE, WCDMA, UMTS, GSM, CDMA2000, Bluetooth, Bluetooth LE, Wi-Fi, Z-Wave, ZigBee, LoRaWAN, and/or other wireless protocols.

[00144] The vehicle 800 may further include data store(s) 828 which may include off-chip (e.g., off the SoC(s) 804) storage. The data store(s) 828 may include one or more storage
25 elements including RAM, SRAM, DRAM, VRAM, Flash, hard disks, and/or other components and/or devices that may store at least one bit of data.

[00145] The vehicle 800 may further include GNSS sensor(s) 858. The GNSS sensor(s) 858 (e.g., GPS, assisted GPS sensors, differential GPS (DGPS) sensors, etc.), to assist in mapping, perception, occupancy grid generation, and/or path planning functions. Any number
30 of GNSS sensor(s) 858 may be used, including, for example and without limitation, a GPS using a USB connector with an Ethernet to Serial (RS-232) bridge.

[00146] The vehicle 800 may further include RADAR sensor(s) 860. The RADAR sensor(s) 860 may be used by the vehicle 800 for long-range vehicle detection, even in darkness

and/or severe weather conditions. RADAR functional safety levels may be ASIL B. The RADAR sensor(s) 860 may use the CAN and/or the bus 802 (e.g., to transmit data generated by the RADAR sensor(s) 860) for control and to access object tracking data, with access to Ethernet to access raw data in some examples. A wide variety of RADAR sensor types may be used. For example, and without limitation, the RADAR sensor(s) 860 may be suitable for front, rear, and side RADAR use. In some example, Pulse Doppler RADAR sensor(s) are used.

[00147] The RADAR sensor(s) 860 may include different configurations, such as long range with narrow field of view, short range with wide field of view, short range side coverage, etc. In some examples, long-range RADAR may be used for adaptive cruise control functionality. The long-range RADAR systems may provide a broad field of view realized by two or more independent scans, such as within a 250m range. The RADAR sensor(s) 860 may help in distinguishing between static and moving objects, and may be used by ADAS systems for emergency brake assist and forward collision warning. Long-range RADAR sensors may include monostatic multimodal RADAR with multiple (e.g., six or more) fixed RADAR antennae and a high-speed CAN and FlexRay interface. In an example with six antennae, the central four antennae may create a focused beam pattern, designed to record the vehicle's 800 surroundings at higher speeds with minimal interference from traffic in adjacent lanes. The other two antennae may expand the field of view, making it possible to quickly detect vehicles entering or leaving the vehicle's 800 lane.

[00148] Mid-range RADAR systems may include, as an example, a range of up to 860m (front) or 80m (rear), and a field of view of up to 42 degrees (front) or 850 degrees (rear). Short-range RADAR systems may include, without limitation, RADAR sensors designed to be installed at both ends of the rear bumper. When installed at both ends of the rear bumper, such a RADAR sensor systems may create two beams that constantly monitor the blind spot in the rear and next to the vehicle.

[00149] Short-range RADAR systems may be used in an ADAS system for blind spot detection and/or lane change assist.

[00150] The vehicle 800 may further include ultrasonic sensor(s) 862. The ultrasonic sensor(s) 862, which may be positioned at the front, back, and/or the sides of the vehicle 800, may be used for park assist and/or to create and update an occupancy grid. A wide variety of ultrasonic sensor(s) 862 may be used, and different ultrasonic sensor(s) 862 may be used for different ranges of detection (e.g., 2.5m, 4m). The ultrasonic sensor(s) 862 may operate at functional safety levels of ASIL B.

[00151] The vehicle 800 may include LIDAR sensor(s) 864. The LIDAR sensor(s) 864 may be used for object and pedestrian detection, emergency braking, collision avoidance, and/or other functions. The LIDAR sensor(s) 864 may be functional safety level ASIL B. In some examples, the vehicle 800 may include multiple LIDAR sensors 864 (e.g., two, four, six, etc.) that may use Ethernet (e.g., to provide data to a Gigabit Ethernet switch).

[00152] In some examples, the LIDAR sensor(s) 864 may be capable of providing a list of objects and their distances for a 360-degree field of view. Commercially available LIDAR sensor(s) 864 may have an advertised range of approximately 800m, with an accuracy of 2cm-3cm, and with support for a 800Mbps Ethernet connection, for example. In some examples, one or more non-protruding LIDAR sensors 864 may be used. In such examples, the LIDAR sensor(s) 864 may be implemented as a small device that may be embedded into the front, rear, sides, and/or corners of the vehicle 800. The LIDAR sensor(s) 864, in such examples, may provide up to a 120-degree horizontal and 35-degree vertical field-of-view, with a 200m range even for low-reflectivity objects. Front-mounted LIDAR sensor(s) 864 may be configured for a horizontal field of view between 45 degrees and 135 degrees.

[00153] In some examples, LIDAR technologies, such as 3D flash LIDAR, may also be used. 3D Flash LIDAR uses a flash of a laser as a transmission source, to illuminate vehicle surroundings up to approximately 200m. A flash LIDAR unit includes a receptor, which records the laser pulse transit time and the reflected light on each pixel, which in turn corresponds to the range from the vehicle to the objects. Flash LIDAR may allow for highly accurate and distortion-free images of the surroundings to be generated with every laser flash. In some examples, four flash LIDAR sensors may be deployed, one at each side of the vehicle 800. Available 3D flash LIDAR systems include a solid-state 3D staring array LIDAR camera with no moving parts other than a fan (e.g., a non-scanning LIDAR device). The flash LIDAR device may use a 5 nanosecond class I (eye-safe) laser pulse per frame and may capture the reflected laser light in the form of 3D range point clouds and co-registered intensity data. By using flash LIDAR, and because flash LIDAR is a solid-state device with no moving parts, the LIDAR sensor(s) 864 may be less susceptible to motion blur, vibration, and/or shock.

[00154] The vehicle may further include IMU sensor(s) 866. The IMU sensor(s) 866 may be located at a center of the rear axle of the vehicle 800, in some examples. The IMU sensor(s) 866 may include, for example and without limitation, an accelerometer(s), a magnetometer(s), a gyroscope(s), a magnetic compass(es), and/or other sensor types. In some examples, such as in six-axis applications, the IMU sensor(s) 866 may include accelerometers

and gyroscopes, while in nine-axis applications, the IMU sensor(s) 866 may include accelerometers, gyroscopes, and magnetometers.

[00155] In some embodiments, the IMU sensor(s) 866 may be implemented as a miniature, high performance GPS-Aided Inertial Navigation System (GPS/INS) that combines
5 micro-electro-mechanical systems (MEMS) inertial sensors, a high-sensitivity GPS receiver, and advanced Kalman filtering algorithms to provide estimates of position, velocity, and attitude. As such, in some examples, the IMU sensor(s) 866 may enable the vehicle 800 to estimate heading without requiring input from a magnetic sensor by directly observing and correlating the changes in velocity from GPS to the IMU sensor(s) 866. In some examples, the
10 IMU sensor(s) 866 and the GNSS sensor(s) 858 may be combined in a single integrated unit.

[00156] The vehicle may include microphone(s) 896 placed in and/or around the vehicle 800. The microphone(s) 896 may be used for emergency vehicle detection and identification, among other things.

[00157] The vehicle may further include any number of camera types, including stereo
15 camera(s) 868, wide-view camera(s) 870, infrared camera(s) 872, surround camera(s) 874, long-range and/or mid-range camera(s) 898, and/or other camera types. The cameras may be used to capture image data around an entire periphery of the vehicle 800. The types of cameras used depends on the embodiments and requirements for the vehicle 800, and any combination of camera types may be used to provide the necessary coverage around the vehicle 800. In
20 addition, the number of cameras may differ depending on the embodiment. For example, the vehicle may include six cameras, seven cameras, ten cameras, twelve cameras, and/or another number of cameras. The cameras may support, as an example and without limitation, Gigabit Multimedia Serial Link (GMSL) and/or Gigabit Ethernet. Each of the camera(s) is described with more detail herein with respect to FIG. 8A and FIG. 8B.

[00158] The vehicle 800 may further include vibration sensor(s) 842. The vibration sensor(s) 842 may measure vibrations of components of the vehicle, such as the axle(s). For example, changes in vibrations may indicate a change in road surfaces. In another example, when two or more vibration sensors 842 are used, the differences between the vibrations may be used to determine friction or slippage of the road surface (e.g., when the difference in
30 vibration is between a power-driven axle and a freely rotating axle).

[00159] The vehicle 800 may include an ADAS system 838. The ADAS system 838 may include a SoC, in some examples. The ADAS system 838 may include autonomous/adaptive/automatic cruise control (ACC), cooperative adaptive cruise control

(CACC), forward crash warning (FCW), automatic emergency braking (AEB), lane departure warnings (LDW), lane keep assist (LKA), blind spot warning (BSW), rear cross-traffic warning (RCTW), collision warning systems (CWS), lane centering (LC), and/or other features and functionality.

5 **[00160]** The ACC systems may use RADAR sensor(s) 860, LIDAR sensor(s) 864, and/or a camera(s). The ACC systems may include longitudinal ACC and/or lateral ACC. Longitudinal ACC monitors and controls the distance to the vehicle immediately ahead of the vehicle 800 and automatically adjust the vehicle speed to maintain a safe distance from vehicles ahead. Lateral ACC performs distance keeping, and advises the vehicle 800 to change lanes
10 when necessary. Lateral ACC is related to other ADAS applications such as LCA and CWS.

[00161] CACC uses information from other vehicles that may be received via the network interface 824 and/or the wireless antenna(s) 826 from other vehicles via a wireless link, or indirectly, over a network connection (e.g., over the Internet). Direct links may be provided by a vehicle-to-vehicle (V2V) communication link, while indirect links may be
15 infrastructure-to-vehicle (I2V) communication link. In general, the V2V communication concept provides information about the immediately preceding vehicles (e.g., vehicles immediately ahead of and in the same lane as the vehicle 800), while the I2V communication concept provides information about traffic further ahead. CACC systems may include either or both I2V and V2V information sources. Given the information of the vehicles ahead of the
20 vehicle 800, CACC may be more reliable and it has potential to improve traffic flow smoothness and reduce congestion on the road.

[00162] FCW systems are designed to alert the driver to a hazard, so that the driver may take corrective action. FCW systems use a front-facing camera and/or RADAR sensor(s) 860, coupled to a dedicated processor, DSP, FPGA, and/or ASIC, that is electrically coupled to
25 driver feedback, such as a display, speaker, and/or vibrating component. FCW systems may provide a warning, such as in the form of a sound, visual warning, vibration and/or a quick brake pulse.

[00163] AEB systems detect an impending forward collision with another vehicle or other object, and may automatically apply the brakes if the driver does not take corrective
30 action within a specified time or distance parameter. AEB systems may use front-facing camera(s) and/or RADAR sensor(s) 860, coupled to a dedicated processor, DSP, FPGA, and/or ASIC. When the AEB system detects a hazard, it typically first alerts the driver to take corrective action to avoid the collision and, if the driver does not take corrective action, the

AEB system may automatically apply the brakes in an effort to prevent, or at least mitigate, the impact of the predicted collision. AEB systems, may include techniques such as dynamic brake support and/or crash imminent braking.

[00164] LDW systems provide visual, audible, and/or tactile warnings, such as steering
5 wheel or seat vibrations, to alert the driver when the vehicle 800 crosses lane markings. A LDW system does not activate when the driver indicates an intentional lane departure, by activating a turn signal. LDW systems may use front-side facing cameras, coupled to a dedicated processor, DSP, FPGA, and/or ASIC, that is electrically coupled to driver feedback, such as a display, speaker, and/or vibrating component.

10 [00165] LKA systems are a variation of LDW systems. LKA systems provide steering input or braking to correct the vehicle 800 if the vehicle 800 starts to exit the lane.

[00166] BSW systems detects and warn the driver of vehicles in an automobile's blind spot. BSW systems may provide a visual, audible, and/or tactile alert to indicate that merging or changing lanes is unsafe. The system may provide an additional warning when the driver
15 uses a turn signal. BSW systems may use rear-side facing camera(s) and/or RADAR sensor(s) 860, coupled to a dedicated processor, DSP, FPGA, and/or ASIC, that is electrically coupled to driver feedback, such as a display, speaker, and/or vibrating component.

[00167] RCTW systems may provide visual, audible, and/or tactile notification when an object is detected outside the rear-camera range when the vehicle 800 is backing up. Some
20 RCTW systems include AEB to ensure that the vehicle brakes are applied to avoid a crash. RCTW systems may use one or more rear-facing RADAR sensor(s) 860, coupled to a dedicated processor, DSP, FPGA, and/or ASIC, that is electrically coupled to driver feedback, such as a display, speaker, and/or vibrating component.

[00168] Conventional ADAS systems may be prone to false positive results which may
25 be annoying and distracting to a driver, but typically are not catastrophic, because the ADAS systems alert the driver and allow the driver to decide whether a safety condition truly exists and act accordingly. However, in an autonomous vehicle 800, the vehicle 800 itself must, in the case of conflicting results, decide whether to heed the result from a primary computer or a secondary computer (e.g., a first controller 836 or a second controller 836). For example, in
30 some embodiments, the ADAS system 838 may be a backup and/or secondary computer for providing perception information to a backup computer rationality module. The backup computer rationality monitor may run a redundant diverse software on hardware components to detect faults in perception and dynamic driving tasks. Outputs from the ADAS system 838

may be provided to a supervisory MCU. If outputs from the primary computer and the secondary computer conflict, the supervisory MCU must determine how to reconcile the conflict to ensure safe operation.

[00169] In some examples, the primary computer may be configured to provide the
5 supervisory MCU with a confidence score, indicating the primary computer's confidence in the chosen result. If the confidence score exceeds a threshold, the supervisory MCU may follow the primary computer's direction, regardless of whether the secondary computer provides a conflicting or inconsistent result. Where the confidence score does not meet the threshold, and where the primary and secondary computer indicate different results (e.g., the
10 conflict), the supervisory MCU may arbitrate between the computers to determine the appropriate outcome.

[00170] The supervisory MCU may be configured to run a neural network(s) that is trained and configured to determine, based on outputs from the primary computer and the secondary computer, conditions under which the secondary computer provides false alarms.
15 Thus, the neural network(s) in the supervisory MCU may learn when the secondary computer's output may be trusted, and when it cannot. For example, when the secondary computer is a RADAR-based FCW system, a neural network(s) in the supervisory MCU may learn when the FCW system is identifying metallic objects that are not, in fact, hazards, such as a drainage grate or manhole cover that triggers an alarm. Similarly, when the secondary computer is a
20 camera-based LDW system, a neural network in the supervisory MCU may learn to override the LDW when bicyclists or pedestrians are present and a lane departure is, in fact, the safest maneuver. In embodiments that include a neural network(s) running on the supervisory MCU, the supervisory MCU may include at least one of a DLA or GPU suitable for running the neural network(s) with associated memory. In preferred embodiments, the supervisory MCU may
25 comprise and/or be included as a component of the SoC(s) 804.

[00171] In other examples, ADAS system 838 may include a secondary computer that performs ADAS functionality using traditional rules of computer vision. As such, the secondary computer may use classic computer vision rules (if-then), and the presence of a neural network(s) in the supervisory MCU may improve reliability, safety and performance.
30 For example, the diverse implementation and intentional non-identity makes the overall system more fault-tolerant, especially to faults caused by software (or software-hardware interface) functionality. For example, if there is a software bug or error in the software running on the primary computer, and the non-identical software code running on the secondary computer

provides the same overall result, the supervisory MCU may have greater confidence that the overall result is correct, and the bug in software or hardware on primary computer is not causing material error.

[00172] In some examples, the output of the ADAS system 838 may be fed into the primary computer's perception block and/or the primary computer's dynamic driving task block. For example, if the ADAS system 838 indicates a forward crash warning due to an object immediately ahead, the perception block may use this information when identifying objects. In other examples, the secondary computer may have its own neural network which is trained and thus reduces the risk of false positives, as described herein.

[00173] The vehicle 800 may further include the infotainment SoC 830 (e.g., an in-vehicle infotainment system (IVI)). Although illustrated and described as a SoC, the infotainment system may not be a SoC, and may include two or more discrete components. The infotainment SoC 830 may include a combination of hardware and software that may be used to provide audio (e.g., music, a personal digital assistant, navigational instructions, news, radio, etc.), video (e.g., TV, movies, streaming, etc.), phone (e.g., hands-free calling), network connectivity (e.g., LTE, Wi-Fi, etc.), and/or information services (e.g., navigation systems, rear-parking assistance, a radio data system, vehicle related information such as fuel level, total distance covered, brake fuel level, oil level, door open/close, air filter information, etc.) to the vehicle 800. For example, the infotainment SoC 830 may include radios, disk players, navigation systems, video players, USB and Bluetooth connectivity, carputers, in-car entertainment, Wi-Fi, steering wheel audio controls, hands free voice control, a heads-up display (HUD), an HMI display 834, a telematics device, a control panel (e.g., for controlling and/or interacting with various components, features, and/or systems), and/or other components. The infotainment SoC 830 may further be used to provide information (e.g., visual and/or audible) to a user(s) of the vehicle, such as information from the ADAS system 838, autonomous driving information such as planned vehicle maneuvers, trajectories, surrounding environment information (e.g., intersection information, vehicle information, road information, etc.), and/or other information.

[00174] The infotainment SoC 830 may include GPU functionality. The infotainment SoC 830 may communicate over the bus 802 (e.g., CAN bus, Ethernet, etc.) with other devices, systems, and/or components of the vehicle 800. In some examples, the infotainment SoC 830 may be coupled to a supervisory MCU such that the GPU of the infotainment system may perform some self-driving functions in the event that the primary controller(s) 836 (e.g., the

- 55 -

primary and/or backup computers of the vehicle 800) fail. In such an example, the infotainment SoC 830 may put the vehicle 800 into a chauffeur to safe stop mode, as described herein.

[00175] The vehicle 800 may further include an instrument cluster 832 (e.g., a digital dash, an electronic instrument cluster, a digital instrument panel, etc.). The instrument cluster
5 832 may include a controller and/or supercomputer (e.g., a discrete controller or supercomputer). The instrument cluster 832 may include a set of instrumentation such as a speedometer, fuel level, oil pressure, tachometer, odometer, turn indicators, gearshift position indicator, seat belt warning light(s), parking-brake warning light(s), engine-malfunction light(s), airbag (SRS) system information, lighting controls, safety system controls, navigation
10 information, etc. In some examples, information may be displayed and/or shared among the infotainment SoC 830 and the instrument cluster 832. In other words, the instrument cluster 832 may be included as part of the infotainment SoC 830, or vice versa.

[00176] FIG. 8D is a system diagram for communication between cloud-based server(s) and the example autonomous vehicle 800 of FIG. 8A, in accordance with some embodiments
15 of the present disclosure. The system 876 may include server(s) 878, network(s) 890, and vehicles, including the vehicle 800. The server(s) 878 may include a plurality of GPUs 884(A)-884(H) (collectively referred to herein as GPUs 884), PCIe switches 882(A)-882(H) (collectively referred to herein as PCIe switches 882), and/or CPUs 880(A)-880(B) (collectively referred to herein as CPUs 880). The GPUs 884, the CPUs 880, and the PCIe
20 switches may be interconnected with high-speed interconnects such as, for example and without limitation, NVLink interfaces 888 developed by NVIDIA and/or PCIe connections 886. In some examples, the GPUs 884 are connected via NVLink and/or NVSwitch SoC and the GPUs 884 and the PCIe switches 882 are connected via PCIe interconnects. Although eight GPUs 884, two CPUs 880, and two PCIe switches are illustrated, this is not intended to be
25 limiting. Depending on the embodiment, each of the server(s) 878 may include any number of GPUs 884, CPUs 880, and/or PCIe switches. For example, the server(s) 878 may each include eight, sixteen, thirty-two, and/or more GPUs 884.

[00177] The server(s) 878 may receive, over the network(s) 890 and from the vehicles, image data representative of images showing unexpected or changed road conditions, such as
30 recently commenced road-work. The server(s) 878 may transmit, over the network(s) 890 and to the vehicles, neural networks 892, updated neural networks 892, and/or map information 894, including information regarding traffic and road conditions. The updates to the map information 894 may include updates for the HD map 822, such as information regarding

construction sites, potholes, detours, flooding, and/or other obstructions. In some examples, the neural networks 892, the updated neural networks 892, and/or the map information 894 may have resulted from new training and/or experiences represented in data received from any number of vehicles in the environment, and/or based on training performed at a datacenter (e.g.,
5 using the server(s) 878 and/or other servers).

[00178] The server(s) 878 may be used to train machine learning models (e.g., neural networks) based on training data. The training data may be generated by the vehicles, and/or may be generated in a simulation (e.g., using a game engine). In some examples, the training data is tagged (e.g., where the neural network benefits from supervised learning) and/or
10 undergoes other pre-processing, while in other examples the training data is not tagged and/or pre-processed (e.g., where the neural network does not require supervised learning). Training may be executed according to any one or more classes of machine learning techniques, including, without limitation, classes such as: supervised training, semi-supervised training, unsupervised training, self-learning, reinforcement learning, federated learning, transfer
15 learning, feature learning (including principal component and cluster analyses), multi-linear subspace learning, manifold learning, representation learning (including sparse dictionary learning), rule-based machine learning, anomaly detection, and any variants or combinations therefor. Once the machine learning models are trained, the machine learning models may be used by the vehicles (e.g., transmitted to the vehicles over the network(s) 890, and/or the
20 machine learning models may be used by the server(s) 878 to remotely monitor the vehicles.

[00179] In some examples, the server(s) 878 may receive data from the vehicles and apply the data to up-to-date real-time neural networks for real-time intelligent inferencing. The server(s) 878 may include deep-learning supercomputers and/or dedicated AI computers powered by GPU(s) 884, such as a DGX and DGX Station machines developed by NVIDIA.
25 However, in some examples, the server(s) 878 may include deep learning infrastructure that use only CPU-powered datacenters.

[00180] The deep-learning infrastructure of the server(s) 878 may be capable of fast, real-time inferencing, and may use that capability to evaluate and verify the health of the processors, software, and/or associated hardware in the vehicle 800. For example, the deep-
30 learning infrastructure may receive periodic updates from the vehicle 800, such as a sequence of images and/or objects that the vehicle 800 has located in that sequence of images (e.g., via computer vision and/or other machine learning object classification techniques). The deep-learning infrastructure may run its own neural network to identify the objects and compare

- 57 -

them with the objects identified by the vehicle 800 and, if the results do not match and the infrastructure concludes that the AI in the vehicle 800 is malfunctioning, the server(s) 878 may transmit a signal to the vehicle 800 instructing a fail-safe computer of the vehicle 800 to assume control, notify the passengers, and complete a safe parking maneuver.

- 5 **[00181]** For inferencing, the server(s) 878 may include the GPU(s) 884 and one or more programmable inference accelerators (e.g., NVIDIA's TensorRT). The combination of GPU-powered servers and inference acceleration may make real-time responsiveness possible. In other examples, such as where performance is less critical, servers powered by CPUs, FPGAs, and other processors may be used for inferencing.

10 **EXAMPLE COMPUTING DEVICE**

- [00182]** FIG. 9 is a block diagram of an example computing device(s) 900 suitable for use in implementing some embodiments of the present disclosure. Computing device 900 may include an interconnect system 902 that directly or indirectly couples the following devices: memory 904, one or more central processing units (CPUs) 906, one or more graphics processing units (GPUs) 908, a communication interface 910, input/output (I/O) ports 912, input/output components 914, a power supply 916, one or more presentation components 918 (e.g., display(s)), and one or more logic units 920. In at least one embodiment, the computing device(s) 900 may comprise one or more virtual machines (VMs), and/or any of the components thereof may comprise virtual components (e.g., virtual hardware components). For non-limiting examples, one or more of the GPUs 908 may comprise one or more vGPUs, one or more of the CPUs 906 may comprise one or more vCPUs, and/or one or more of the logic units 920 may comprise one or more virtual logic units. As such, a computing device(s) 900 may include discrete components (e.g., a full GPU dedicated to the computing device 900), virtual components (e.g., a portion of a GPU dedicated to the computing device 900), or a combination thereof. In some embodiments, one or more aspects of the occupant evaluation function 110 may be implemented using code executed on the CPUs 906 and/or GPUs 908.

- [00183]** Although the various blocks of FIG. 9 are shown as connected via the interconnect system 902 with lines, this is not intended to be limiting and is for clarity only. For example, in some embodiments, a presentation component 918, such as a display device, may be considered an I/O component 914 (e.g., if the display is a touch screen). As another

example, the CPUs 906 and/or GPUs 908 may include memory (e.g., the memory 904 may be representative of a storage device in addition to the memory of the GPUs 908, the CPUs 906, and/or other components). In other words, the computing device of FIG. 9 is merely illustrative. Distinction is not made between such categories as “workstation,” “server,”
5 “laptop,” “desktop,” “tablet,” “client device,” “mobile device,” “hand-held device,” “game console,” “electronic control unit (ECU),” “virtual reality system,” and/or other device or system types, as all are contemplated within the scope of the computing device of FIG. 9.

[00184] The interconnect system 902 may represent one or more links or busses, such as an address bus, a data bus, a control bus, or a combination thereof. The interconnect system
10 902 may include one or more bus or link types, such as an industry standard architecture (ISA) bus, an extended industry standard architecture (EISA) bus, a video electronics standards association (VESA) bus, a peripheral component interconnect (PCI) bus, a peripheral component interconnect express (PCIe) bus, and/or another type of bus or link. In some embodiments, there are direct connections between components. As an example, the CPU 906
15 may be directly connected to the memory 904. Further, the CPU 906 may be directly connected to the GPU 908. Where there is direct, or point-to-point connection between components, the interconnect system 902 may include a PCIe link to carry out the connection. In these examples, a PCI bus need not be included in the computing device 900.

[00185] The memory 904 may include any of a variety of computer-readable media. The
20 computer-readable media may be any available media that may be accessed by the computing device 900. The computer-readable media may include both volatile and nonvolatile media, and removable and non-removable media. By way of example, and not limitation, the computer-readable media may comprise computer-storage media and communication media.

[00186] The computer-storage media may include both volatile and nonvolatile media
25 and/or removable and non-removable media implemented in any method or technology for storage of information such as computer-readable instructions, data structures, program modules, and/or other data types. For example, the memory 904 may store computer-readable instructions (e.g., that represent a program(s) and/or a program element(s), such as an operating system. Computer-storage media may include, but is not limited to, RAM, ROM, EEPROM,
30 flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which may be used to store the desired information and

which may be accessed by computing device 900. As used herein, computer storage media does not comprise signals per se.

[00187] The computer storage media may embody computer-readable instructions, data structures, program modules, and/or other data types in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media. The term “modulated data signal” may refer to a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, the computer storage media may include wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media. Combinations of any of the above should also be included within the scope of computer-readable media.

[00188] The CPU(s) 906 may be configured to execute at least some of the computer-readable instructions to control one or more components of the computing device 900 to perform one or more of the methods and/or processes described herein. The CPU(s) 906 may each include one or more cores (e.g., one, two, four, eight, twenty-eight, seventy-two, etc.) that are capable of handling a multitude of software threads simultaneously. The CPU(s) 906 may include any type of processor, and may include different types of processors depending on the type of computing device 900 implemented (e.g., processors with fewer cores for mobile devices and processors with more cores for servers). For example, depending on the type of computing device 900, the processor may be an Advanced RISC Machines (ARM) processor implemented using Reduced Instruction Set Computing (RISC) or an x86 processor implemented using Complex Instruction Set Computing (CISC). The computing device 900 may include one or more CPUs 906 in addition to one or more microprocessors or supplementary co-processors, such as math co-processors.

[00189] In addition to or alternatively from the CPU(s) 906, the GPU(s) 908 may be configured to execute at least some of the computer-readable instructions to control one or more components of the computing device 900 to perform one or more of the methods and/or processes described herein. One or more of the GPU(s) 908 may be an integrated GPU (e.g., with one or more of the CPU(s) 906 and/or one or more of the GPU(s) 908 may be a discrete GPU. In embodiments, one or more of the GPU(s) 908 may be a coprocessor of one or more of the CPU(s) 906. The GPU(s) 908 may be used by the computing device 900 to render graphics (e.g., 3D graphics) or perform general purpose computations. For example, the GPU(s) 908 may be used for General-Purpose computing on GPUs (GPGPU). The GPU(s)

908 may include hundreds or thousands of cores that are capable of handling hundreds or thousands of software threads simultaneously. The GPU(s) 908 may generate pixel data for output images in response to rendering commands (e.g., rendering commands from the CPU(s) 906 received via a host interface). The GPU(s) 908 may include graphics memory, such as display memory, for storing pixel data or any other suitable data, such as GPGPU data. The display memory may be included as part of the memory 904. The GPU(s) 908 may include two or more GPUs operating in parallel (e.g., via a link). The link may directly connect the GPUs (e.g., using NVLINK) or may connect the GPUs through a switch (e.g., using NVSwitch). When combined together, each GPU 908 may generate pixel data or GPGPU data for different portions of an output or for different outputs (e.g., a first GPU for a first image and a second GPU for a second image). Each GPU may include its own memory, or may share memory with other GPUs.

[00190] In addition to or alternatively from the CPU(s) 906 and/or the GPU(s) 908, the logic unit(s) 920 may be configured to execute at least some of the computer-readable instructions to control one or more components of the computing device 900 to perform one or more of the methods and/or processes described herein. In embodiments, the CPU(s) 906, the GPU(s) 908, and/or the logic unit(s) 920 may discretely or jointly perform any combination of the methods, processes and/or portions thereof. One or more of the logic units 920 may be part of and/or integrated in one or more of the CPU(s) 906 and/or the GPU(s) 908 and/or one or more of the logic units 920 may be discrete components or otherwise external to the CPU(s) 906 and/or the GPU(s) 908. In embodiments, one or more of the logic units 920 may be a coprocessor of one or more of the CPU(s) 906 and/or one or more of the GPU(s) 908.

[00191] Examples of the logic unit(s) 920 include one or more processing cores and/or components thereof, such as Data Processing Units (DPUs), Tensor Cores (TCs), Tensor Processing Units (TPUs), Pixel Visual Cores (PVCs), Vision Processing Units (VPUs), Graphics Processing Clusters (GPCs), Texture Processing Clusters (TPCs), Streaming Multiprocessors (SMs), Tree Traversal Units (TTUs), Artificial Intelligence Accelerators (AIAs), Deep Learning Accelerators (DLAs), Arithmetic-Logic Units (ALUs), Application-Specific Integrated Circuits (ASICs), Floating Point Units (FPUs), input/output (I/O) elements, peripheral component interconnect (PCI) or peripheral component interconnect express (PCIe) elements, and/or the like.

[00192] The communication interface 910 may include one or more receivers, transmitters, and/or transceivers that enable the computing device 900 to communicate with

other computing devices via an electronic communication network, included wired and/or wireless communications. The communication interface 910 may include components and functionality to enable communication over any of a number of different networks, such as wireless networks (e.g., Wi-Fi, Z-Wave, Bluetooth, Bluetooth LE, ZigBee, etc.), wired
5 networks (e.g., communicating over Ethernet or InfiniBand), low-power wide-area networks (e.g., LoRaWAN, SigFox, etc.), and/or the Internet. In one or more embodiments, logic unit(s) 920 and/or communication interface 910 may include one or more data processing units (DPUs) to transmit data received over a network and/or through interconnect system 902 directly to (e.g., a memory of) one or more GPU(s) 908.

10 **[00193]** The I/O ports 912 may enable the computing device 900 to be logically coupled to other devices including the I/O components 914, the presentation component(s) 918, and/or other components, some of which may be built in to (e.g., integrated in) the computing device 900. Illustrative I/O components 914 include a microphone, mouse, keyboard, joystick, game pad, game controller, satellite dish, scanner, printer, wireless device, etc. The I/O components
15 914 may provide a natural user interface (NUI) that processes air gestures, voice, or other physiological inputs generated by a user. In some instances, inputs may be transmitted to an appropriate network element for further processing. An NUI may implement any combination of speech recognition, stylus recognition, facial recognition, biometric recognition, gesture recognition both on screen and adjacent to the screen, air gestures, head and eye tracking, and
20 touch recognition (as described in more detail below) associated with a display of the computing device 900. The computing device 900 may include depth cameras, such as stereoscopic camera systems, infrared camera systems, RGB camera systems, touchscreen technology, and combinations of these, for gesture detection and recognition. Additionally, the computing device 900 may include accelerometers or gyroscopes (e.g., as part of an inertia
25 measurement unit (IMU)) that enable detection of motion. In some examples, the output of the accelerometers or gyroscopes may be used by the computing device 900 to render immersive augmented reality or virtual reality.

[00194] The power supply 916 may include a hard-wired power supply, a battery power supply, or a combination thereof. The power supply 916 may provide power to the computing
30 device 900 to enable the components of the computing device 900 to operate.

[00195] The presentation component(s) 918 may include a display (e.g., a monitor, a touch screen, a television screen, a heads-up-display (HUD), other display types, or a combination thereof), speakers, and/or other presentation components. The presentation

component(s) 918 may receive data from other components (e.g., the GPU(s) 908, the CPU(s) 906, DPUs, etc.), and output the data (e.g., as an image, video, sound, etc.).

EXAMPLE DATA CENTER

[00196] FIG. 10 illustrates an example data center 1000 that may be used in at least one
5 embodiments of the present disclosure. The data center 1000 may include a data center infrastructure layer 1010, a framework layer 1020, a software layer 1030, and/or an application layer 1040.

[00197] As shown in FIG. 10, the data center infrastructure layer 1010 may include a resource orchestrator 1012, grouped computing resources 1014, and node computing resources
10 (“node C.R.s”) 1016(1)-1016(N), where “N” represents any whole, positive integer. In at least one embodiment, node C.R.s 1016(1)-1016(N) may include, but are not limited to, any number of central processing units (CPUs) or other processors (including DPUs, accelerators, field programmable gate arrays (FPGAs), graphics processors or graphics processing units (GPUs), etc.), memory devices (e.g., dynamic read-only memory), storage devices (e.g., solid state or
15 disk drives), network input/output (NW I/O) devices, network switches, virtual machines (VMs), power modules, and/or cooling modules, etc. In some embodiments, one or more node C.R.s from among node C.R.s 1016(1)-1016(N) may correspond to a server having one or more of the above-mentioned computing resources. In addition, in some embodiments, the node C.R.s 1016(1)-1016(N) may include one or more virtual components, such as vGPUs, vCPUs,
20 and/or the like, and/or one or more of the node C.R.s 1016(1)-1016(N) may correspond to a virtual machine (VM).

[00198] In at least one embodiment, grouped computing resources 1014 may include separate groupings of node C.R.s 1016 housed within one or more racks (not shown), or many racks housed in data centers at various geographical locations (also not shown). Separate
25 groupings of node C.R.s 1016 within grouped computing resources 1014 may include grouped compute, network, memory or storage resources that may be configured or allocated to support one or more workloads. In at least one embodiment, several node C.R.s 1016 including CPUs, GPUs, DPUs, and/or other processors may be grouped within one or more racks to provide compute resources to support one or more workloads. The one or more racks may also include
30 any number of power modules, cooling modules, and/or network switches, in any combination.

[00199] The resource orchestrator 1012 may configure or otherwise control one or more node C.R.s 1016(1)-1016(N) and/or grouped computing resources 1014. In at least one embodiment, resource orchestrator 1012 may include a software design infrastructure (SDI) management entity for the data center 1000. The resource orchestrator 1012 may include hardware, software, or some combination thereof.

[00200] In at least one embodiment, as shown in FIG. 10, framework layer 1020 may include a job scheduler 1033, a configuration manager 1034, a resource manager 1036, and/or a distributed file system 1038. The framework layer 1020 may include a framework to support software 1032 of software layer 1030 and/or one or more application(s) 1042 of application layer 1040. The software 1032 or application(s) 1042 may respectively include web-based service software or applications, such as those provided by Amazon Web Services, Google Cloud and Microsoft Azure. The framework layer 1020 may be, but is not limited to, a type of free and open-source software web application framework such as Apache Spark™ (hereinafter “Spark”) that may utilize distributed file system 1038 for large-scale data processing (e.g., “big data”). In at least one embodiment, job scheduler 1033 may include a Spark driver to facilitate scheduling of workloads supported by various layers of data center 1000. The configuration manager 1034 may be capable of configuring different layers such as software layer 1030 and framework layer 1020 including Spark and distributed file system 1038 for supporting large-scale data processing. The resource manager 1036 may be capable of managing clustered or grouped computing resources mapped to or allocated for support of distributed file system 1038 and job scheduler 1033. In at least one embodiment, clustered or grouped computing resources may include grouped computing resource 1014 at data center infrastructure layer 1010. The resource manager 1036 may coordinate with resource orchestrator 1012 to manage these mapped or allocated computing resources. In some embodiments, one or more aspects of the occupant evaluation function 110 may be implemented by software 1032 or application(s) 1042.

[00201] In at least one embodiment, software 1032 included in software layer 1030 may include software used by at least portions of node C.R.s 1016(1)-1016(N), grouped computing resources 1014, and/or distributed file system 1038 of framework layer 1020. One or more types of software may include, but are not limited to, Internet web page search software, e-mail virus scan software, database software, and streaming video content software.

[00202] In at least one embodiment, application(s) 1042 included in application layer 1040 may include one or more types of applications used by at least portions of node C.R.s

1016(1)-1016(N), grouped computing resources 1014, and/or distributed file system 1038 of framework layer 1020. One or more types of applications may include, but are not limited to, any number of a genomics application, a cognitive compute, and a machine learning application, including training or inferencing software, machine learning framework software
5 (e.g., PyTorch, TensorFlow, Caffe, etc.), and/or other machine learning applications used in conjunction with one or more embodiments.

[00203] In at least one embodiment, any of configuration manager 1034, resource manager 1036, and resource orchestrator 1012 may implement any number and type of self-modifying actions based on any amount and type of data acquired in any technically feasible
10 fashion. Self-modifying actions may relieve a data center operator of data center 1000 from making possibly bad configuration decisions and possibly avoiding underutilized and/or poor performing portions of a data center.

[00204] The data center 1000 may include tools, services, software or other resources to train one or more machine learning models or predict or infer information using one or more
15 machine learning models according to one or more embodiments described herein. For example, a machine learning model(s) may be trained by calculating weight parameters according to a neural network architecture using software and/or computing resources described above with respect to the data center 1000. In at least one embodiment, trained or
20 deployed machine learning models corresponding to one or more neural networks may be used to infer or predict information using resources described above with respect to the data center 1000 by using weight parameters calculated through one or more training techniques, such as but not limited to those described herein.

[00205] In at least one embodiment, the data center 1000 may use CPUs, application-specific integrated circuits (ASICs), GPUs, FPGAs, and/or other hardware (or virtual compute
25 resources corresponding thereto) to perform training and/or inferencing using above-described resources. Moreover, one or more software and/or hardware resources described above may be configured as a service to allow users to train or performing inferencing of information, such as image recognition, speech recognition, or other artificial intelligence services.

EXAMPLE NETWORK ENVIRONMENTS

30 [00206] Network environments suitable for use in implementing embodiments of the disclosure may include one or more client devices, servers, network attached storage (NAS),

other backend devices, and/or other device types. The client devices, servers, and/or other device types (e.g., each device) may be implemented on one or more instances of the computing device(s) 900 of FIG. 9 – e.g., each device may include similar components, features, and/or functionality of the computing device(s) 900. In addition, where backend devices (e.g., servers, NAS, etc.) are implemented, the backend devices may be included as part of a data center 1000, an example of which is described in more detail herein with respect to FIG. 10.

[00207] Components of a network environment may communicate with each other via a network(s), which may be wired, wireless, or both. The network may include multiple networks, or a network of networks. By way of example, the network may include one or more Wide Area Networks (WANs), one or more Local Area Networks (LANs), one or more public networks such as the Internet and/or a public switched telephone network (PSTN), and/or one or more private networks. Where the network includes a wireless telecommunications network, components such as a base station, a communications tower, or even access points (as well as other components) may provide wireless connectivity.

[00208] Compatible network environments may include one or more peer-to-peer network environments – in which case a server may not be included in a network environment – and one or more client-server network environments – in which case one or more servers may be included in a network environment. In peer-to-peer network environments, functionality described herein with respect to a server(s) may be implemented on any number of client devices.

[00209] In at least one embodiment, a network environment may include one or more cloud-based network environments, a distributed computing environment, a combination thereof, etc. A cloud-based network environment may include a framework layer, a job scheduler, a resource manager, and a distributed file system implemented on one or more of servers, which may include one or more core network servers and/or edge servers. A framework layer may include a framework to support software of a software layer and/or one or more application(s) of an application layer. The software or application(s) may respectively include web-based service software or applications. In embodiments, one or more of the client devices may use the web-based service software or applications (e.g., by accessing the service software and/or applications via one or more application programming interfaces (APIs)). The framework layer may be, but is not limited to, a type of free and open-source software web application framework such as that may use a distributed file system for large-scale data processing (e.g., “big data”).

[00210] A cloud-based network environment may provide cloud computing and/or cloud storage that carries out any combination of computing and/or data storage functions described herein (or one or more portions thereof). Any of these various functions may be distributed over multiple locations from central or core servers (e.g., of one or more data centers that may be distributed across a state, a region, a country, the globe, etc.). If a connection to a user (e.g., a client device) is relatively close to an edge server(s), a core server(s) may designate at least a portion of the functionality to the edge server(s). A cloud-based network environment may be private (e.g., limited to a single organization), may be public (e.g., available to many organizations), and/or a combination thereof (e.g., a hybrid cloud environment).

[00211] The client device(s) may include at least some of the components, features, and functionality of the example computing device(s) 900 described herein with respect to FIG. 9. By way of example and not limitation, a client device may be embodied as a Personal Computer (PC), a laptop computer, a mobile device, a smartphone, a tablet computer, a smart watch, a wearable computer, a Personal Digital Assistant (PDA), an MP3 player, a virtual reality headset, a Global Positioning System (GPS) or device, a video player, a video camera, a surveillance device or system, a vehicle, a boat, a flying vessel, a virtual machine, a drone, a robot, a handheld communications device, a hospital device, a gaming device or system, an entertainment system, a vehicle computer system, an embedded system controller, a remote control, an appliance, a consumer electronic device, a workstation, an edge device, any combination of these delineated devices, or any other suitable device.

[00212] The disclosure may be described in the general context of computer code or machine-useable instructions, including computer-executable instructions such as program modules, being executed by a computer or other machine, such as a personal data assistant or other handheld device. Generally, program modules including routines, programs, objects, components, data structures, etc., refer to code that perform particular tasks or implement particular abstract data types. The disclosure may be practiced in a variety of system configurations, including hand-held devices, consumer electronics, general-purpose computers, more specialty computing devices, etc. The disclosure may also be practiced in distributed computing environments where tasks are performed by remote-processing devices that are linked through a communications network.

[00213] As used herein, a recitation of “and/or” with respect to two or more elements should be interpreted to mean only one element, or a combination of elements. For example, “element A, element B, and/or element C” may include only element A, only element B, only

- 67 -

element C, element A and element B, element A and element C, element B and element C, or elements A, B, and C. In addition, “at least one of element A or element B” may include at least one of element A, at least one of element B, or at least one of element A and at least one of element B. Further, “at least one of element A and element B” may include at least one of
5 element A, at least one of element B, or at least one of element A and at least one of element B.

[00214] The subject matter of the present disclosure is described with specificity herein to meet statutory requirements. However, the description itself is not intended to limit the scope of this disclosure. Rather, the inventors have contemplated that the claimed subject
10 matter might also be embodied in other ways, to include different steps or combinations of steps similar to the ones described in this document, in conjunction with other present or future technologies. Moreover, although the terms “step” and/or “block” may be used herein to connote different elements of methods employed, the terms should not be interpreted as implying any particular order among or between various steps herein disclosed unless and
15 except when the order of individual steps is explicitly described.

CLAIMS

What is claimed is:

1. A system comprising:
5 one or more processing units to:
detect one or more features representing at least a portion of an occupant of a machine based at least on optical image data;
determine a depth corresponding to the one or more features based at least on the optical image data and a ground truth cabin depth image corresponding to an interior
10 of the machine;
generate a three-dimensional representation of the occupant based at least on the one or more features and the depth of the one or more features, wherein the three-dimensional representation comprises at least one characteristic representative of a size of the occupant; and
15 control at least one operation of the machine based at least on the at least one characteristic.
2. The system of claim 1, wherein the at least one characteristic comprises either a three-dimensional shape estimate or a three-dimensional pose estimate.
20
3. The system of claim 1, wherein the one or more processing units are further to:
execute at least one three-dimensional pose detection model to generate a scale-normalized three-dimensional pose estimate of the occupant based at least on the one or more features; and
25 scale the scale-normalized three-dimensional pose estimate to a linear measurement scale three-dimensional pose estimate based at least on the depth corresponding to the one or more features.
4. The system of claim 3, wherein the one or more processing units are further to:
30 generate the at least one characteristic representative of the size of the occupant based at least on the linear measurement scale three-dimensional pose estimate.

- 69 -

5. The system of claim 3, wherein the one or more processing units are further to:
generate the scale-normalized three-dimensional pose estimate based at least on
a set of kinematic elements of the occupant detected using the one or more features by
the at least one three-dimensional pose detection model.
- 5
6. The system of claim 3, wherein the one or more processing units are further to:
apply the optical image data to at least one person detection model to define the
portion of an image frame representing the occupant; and
generate the scale-normalized three-dimensional pose estimate based at least on
10 the portion of the image frame representing the occupant.
7. The system of claim 1, wherein the one or more processing units are further to:
perform a ray tracing to correlate at least a first pixel of the ground truth cabin
depth image to the one or more features to determine the depth corresponding to the
one or more features.
- 15
8. The system of claim 1, wherein the one or more processing units are further to:
generate a scale-normalized three-dimensional pose estimate of the occupant
based at least on the one or more features, wherein the scale-normalized three-
dimensional pose estimate represents a set of kinematic elements of the occupant.
- 20
9. The system of claim 1, wherein the one or more processing units are further to:
generate a range image based at least on the optical image data and the ground truth
cabin depth image, wherein the depth corresponding to the one or more features is derived at
25 least from the range image.
10. The system of claim 9, wherein the one or more processing units are further to:
generate a vehicle interior mask based at least on a segmentation of the optical
image data, wherein the vehicle interior mask comprises a boundary that outlines pixels
30 that correspond to structural elements of the interior; and
generate the range image further based on the vehicle interior mask.
11. The system of claim 1, wherein the one or more processing units are further to:

- 70 -

generate a vehicle interior mask based at least on a segmentation of the optical image data, wherein the vehicle interior mask comprises a first boundary that outlines pixels corresponding to structural elements of the interior;

5 generate an occupant mask based at least on the segmentation of the optical image data, wherein the occupant mask comprises a second boundary that outlines pixels corresponding to the one or more features; and

generate a three-dimensional shape estimate of the occupant based at least on the occupant mask and a range image derived based at least on the optical image data, the ground truth cabin depth image, and the vehicle interior mask.

10

12. The system of claim 11, wherein the one or more processing units are further to:

generate the three-dimensional shape estimate of the occupant further based at least on a pose estimate of the occupant generated from the optical image data.

15

13. The system of claim 1, wherein the system is comprised in at least one of:

a control system for an autonomous or semi-autonomous machine;

a perception system for an autonomous or semi-autonomous machine;

a system for performing simulation operations;

20 a system for performing digital twin operations;

a system for performing light transport simulation;

a system for performing collaborative content creation for three-dimensional assets;

25 a system for generating or presenting at least one of virtual reality content, augmented reality content, or mixed reality content;

a system for performing deep learning operations;

a system for performing real-time streaming;

a system implemented using an edge device;

a system implemented using a robot;

30 a system for performing conversational AI operations;

a system for generating synthetic data;

a system incorporating one or more virtual machines (VMs);

a system implemented at least partially in a data center;

- 71 -

a system for performing generative AI operations;
a system implemented at least partially using a language model; or
a system implemented at least partially using cloud computing resources.

- 5 14. A processor comprising:
one or more processing units to:
 generate a first representation of at least a portion of an occupant of a machine
based at least on two-dimensional optical image data;
 determine a depth corresponding to one or more surface locations of the
10 occupant based at least on the two-dimensional optical image data and a ground truth
cabin depth image corresponding to an interior of the machine;
 scale the first representation of at least the portion of the occupant to a second
representation of at least the portion of the occupant based at least on the depth
corresponding to the one or more surface locations of the occupant, wherein the second
15 representation is a three-dimensional representation;
 determine at least one characteristic representative of a size of the occupant
based at least on the second representation; and
 control at least one operation of the machine based at least on the at least one
characteristic of the size of the occupant.

- 20 15. The processor of claim 14, wherein the at least one characteristic comprises
either a three-dimensional shape estimate or a three-dimensional pose estimate.

- 25 16. The processor of claim 14, wherein the one or more processing units are further
to:
 execute at least one three-dimensional pose detection model to generate a scale-
normalized three-dimensional pose estimate of the occupant based at least on one or
more features representing at least the portion of the occupant of the machine based at
least on the two-dimensional optical image data; and
30 scale the scale-normalized three-dimensional pose estimate to a linear
measurement scale three-dimensional pose estimate based at least on the depth
corresponding to the one or more features.

17. The processor of claim 14, wherein the one or more processing units are further to:

5 generate a range image based at least on the two-dimensional optical image data and the ground truth cabin depth image, wherein the depth corresponding to the one or more surface locations is derived at least from the range image.

18. The processor of claim 14, wherein the one or more processing units are further to:

10 generate a vehicle interior mask based at least on a segmentation of the two-dimensional optical image data, wherein the vehicle interior mask comprises a first boundary that outlines pixels corresponding to structural elements of the interior;

generate an occupant mask based at least on the segmentation of the two-dimensional optical image data, wherein the occupant mask comprises a second boundary that outlines pixels corresponding to the one or more surface locations; and

15 generate a three-dimensional shape estimate of the occupant based at least on the occupant mask and a range image derived based at least on the two-dimensional optical image data, the ground truth cabin depth image, and the vehicle interior mask.

19. The processor of claim 14, wherein the processor is comprised in at least one of:

a control system for an autonomous or semi-autonomous machine;

a perception system for an autonomous or semi-autonomous machine;

a system for performing simulation operations;

a system for performing digital twin operations;

25 a system for performing light transport simulation;

a system for performing collaborative content creation for three-dimensional assets;

a system for generating or presenting at least one of virtual reality content, augmented reality content, or mixed reality content;

30 a system for performing deep learning operations;

a system for performing real-time streaming;

a system implemented using an edge device;

a system implemented using a robot;

- 73 -

- 5 a system for performing conversational AI operations;
 a system for generating synthetic data;
 a system incorporating one or more virtual machines (VMs);
 a system implemented at least partially in a data center;
10 a system for performing generative AI operations;
 a system implemented at least partially using a language model; or
 a system implemented at least partially using cloud computing resources.
20. A method comprising:
- 10 generating a three-dimensional representation of a size of an occupant of a
 machine based at least on one or more features representing at least a portion of the
 occupant derived from two-dimensional optical image data and a depth of the one or
 more features generated from the two-dimensional optical image data and a ground
 truth cabin depth image corresponding to an interior of the machine.

15

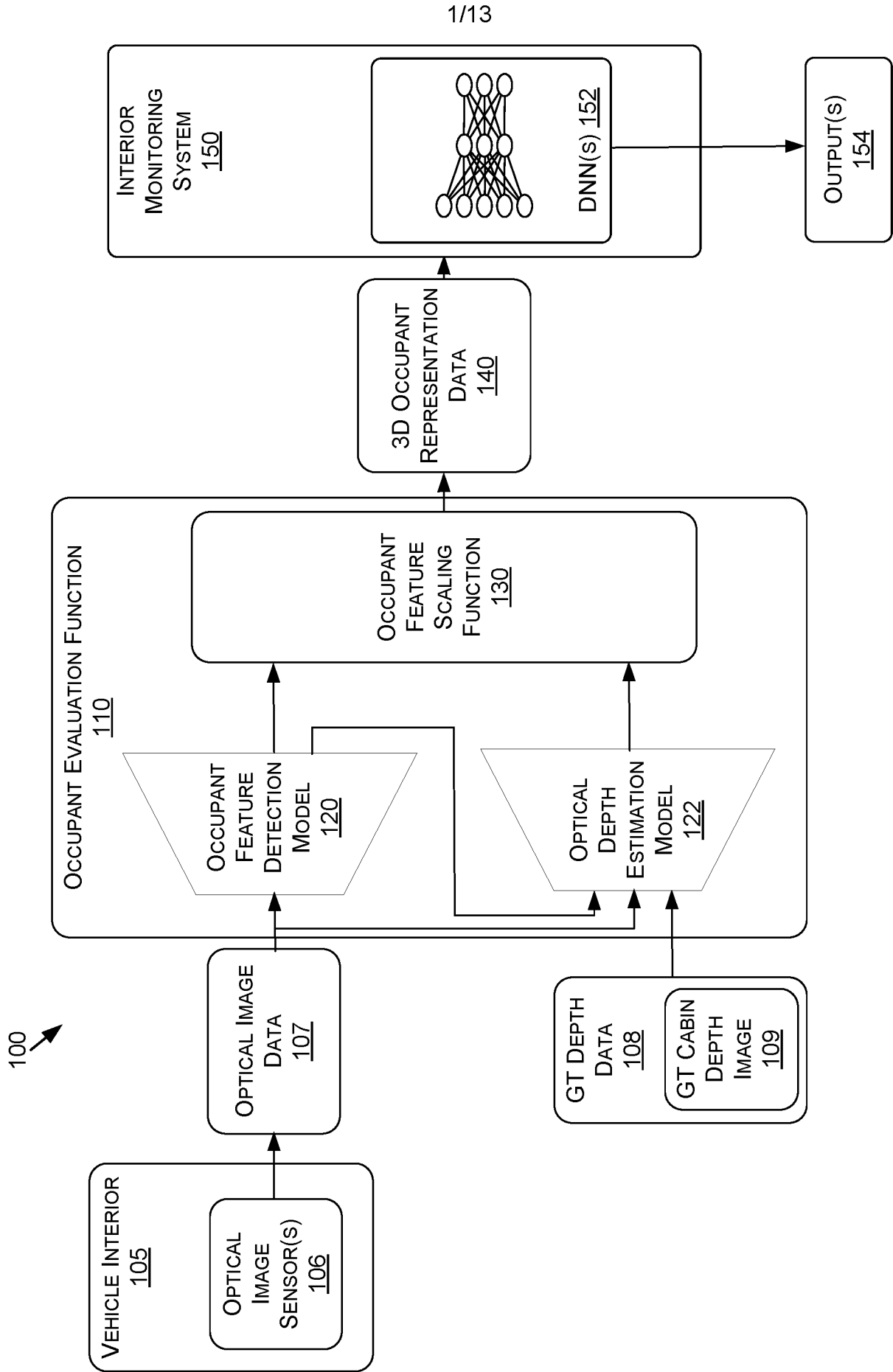


FIGURE 1

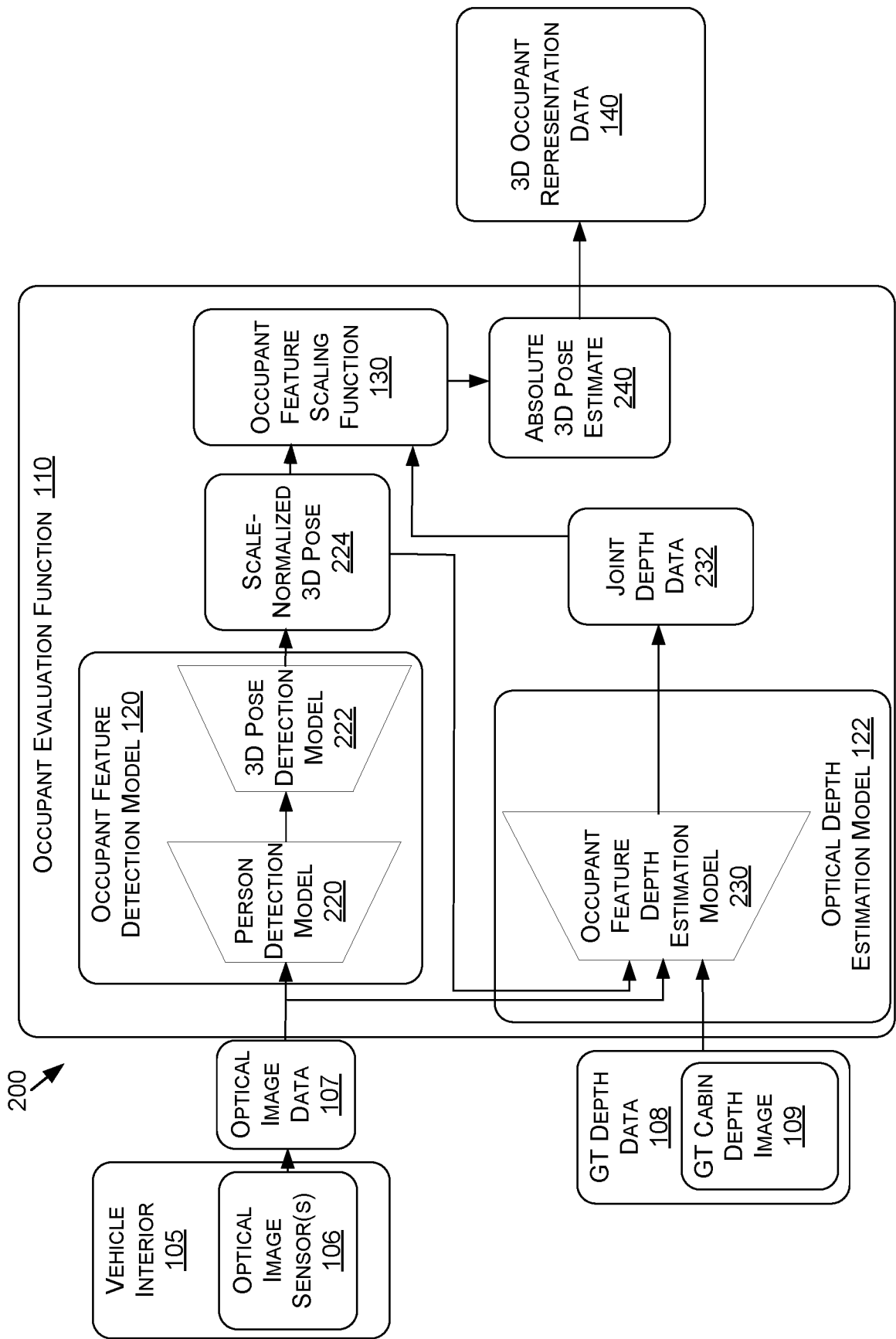


FIGURE 2

300 ↗

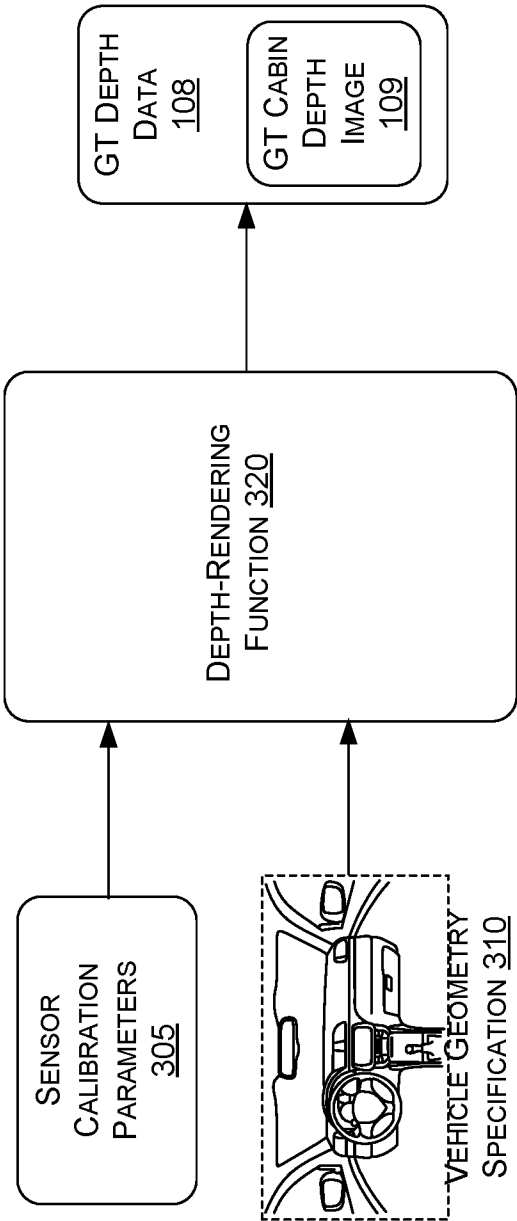


FIGURE 3

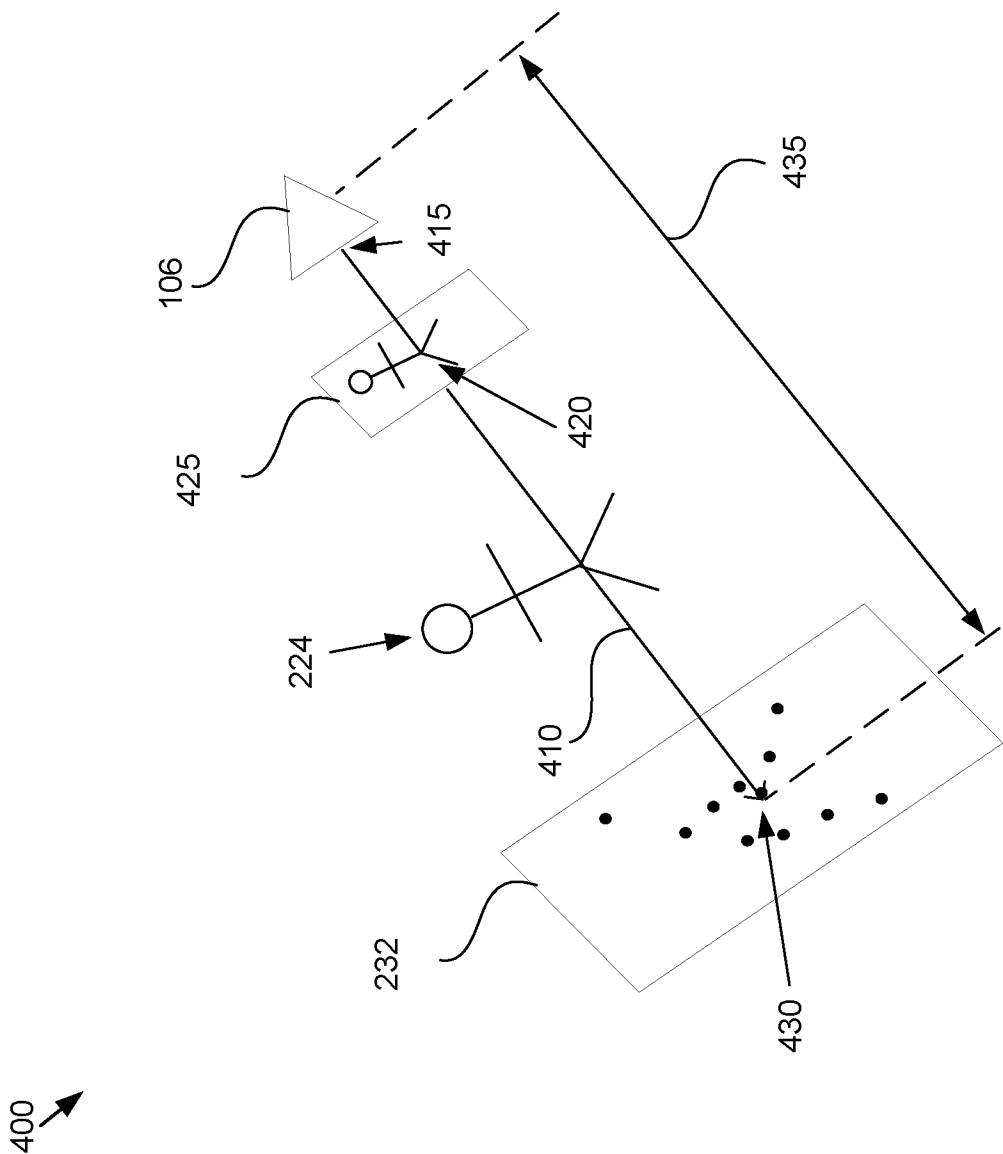


FIGURE 4

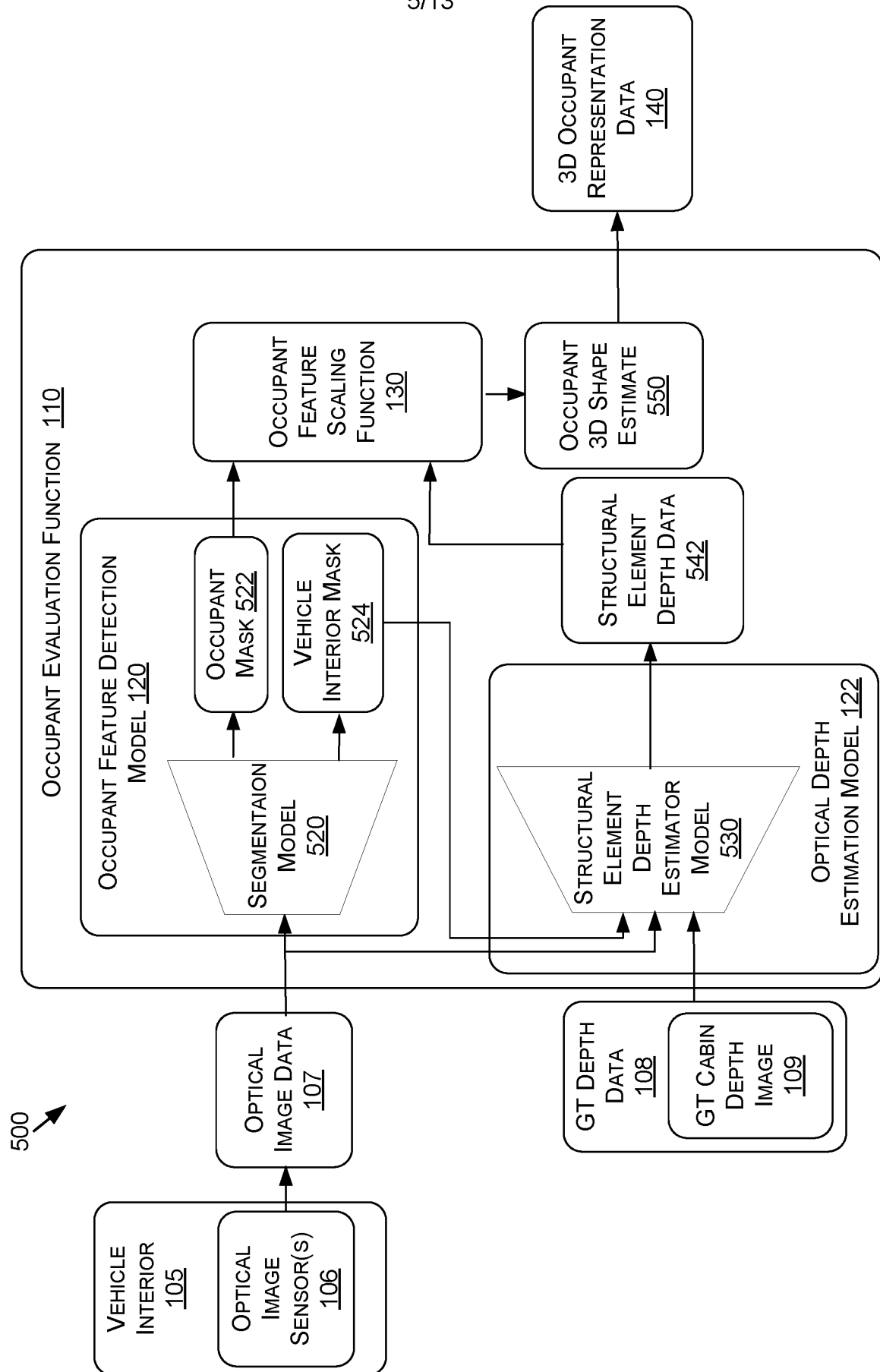


FIGURE 5

600 ↗

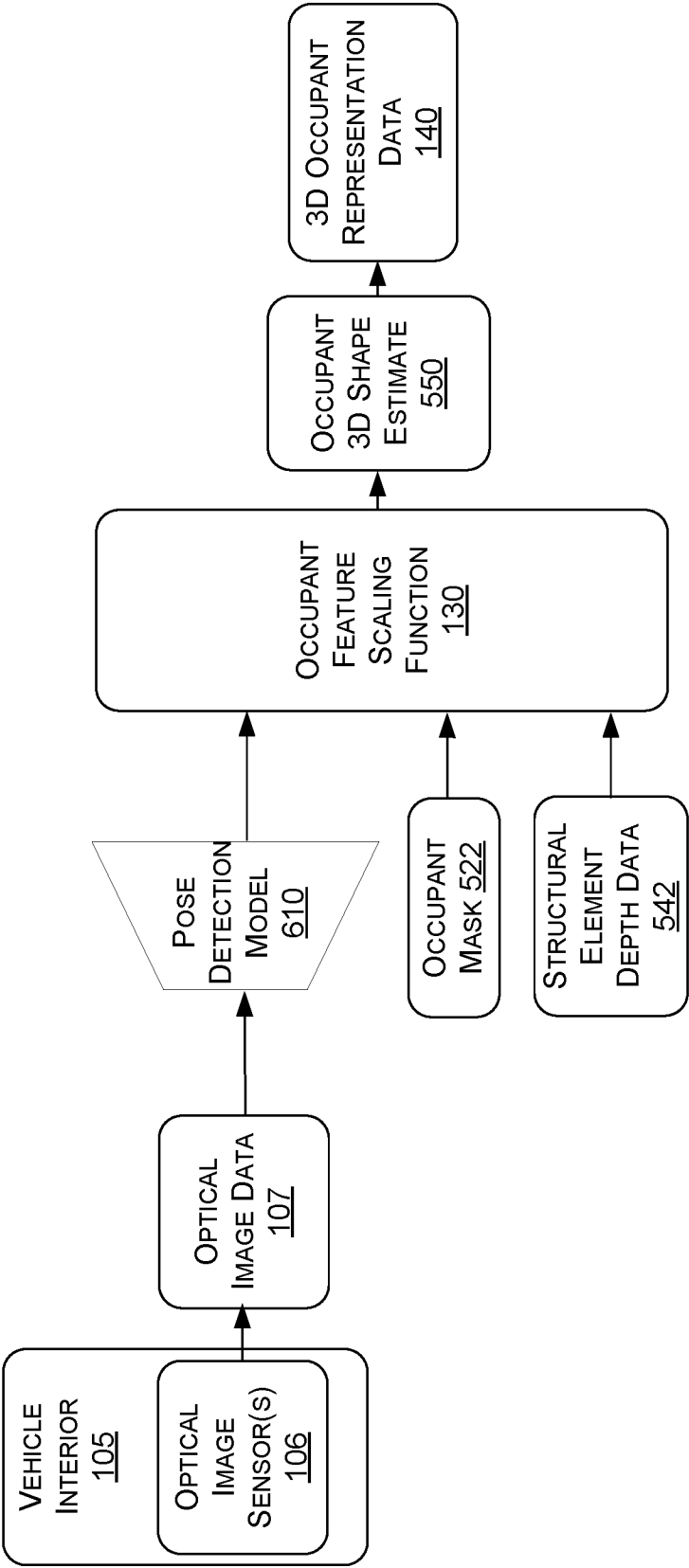


FIGURE 6

7/13

700
↓

DETECT ONE OR MORE FEATURES REPRESENTING AT LEAST
A PORTION OF AN OCCUPANT OF A MACHINE BASED AT
LEAST ON OPTICAL IMAGE DATA

B702



DETERMINE A DEPTH CORRESPONDING TO THE ONE OR
MORE FEATURES BASED AT LEAST ON THE OPTICAL IMAGE
DATA AND A GROUND TRUTH DEPTH IMAGE
CORRESPONDING TO AN INTERIOR OF THE MACHINE

B704



GENERATE A THREE-DIMENSIONAL REPRESENTATION OF
THE OCCUPANT BASED AT LEAST ON THE ONE OR MORE
FEATURES AND THE DEPTH OF THE ONE OR MORE
FEATURES, WHEREIN THE THREE-DIMENSIONAL
REPRESENTATION COMPRISES AT LEAST ONE
CHARACTERISTIC REPRESENTATIVE OF A SIZE OF THE
OCCUPANT

B706



CONTROL AT LEAST ONE OPERATION OF THE MACHINE
BASED AT LEAST ON THE AT LEAST ONE CHARACTERISTIC

B708

FIGURE 7

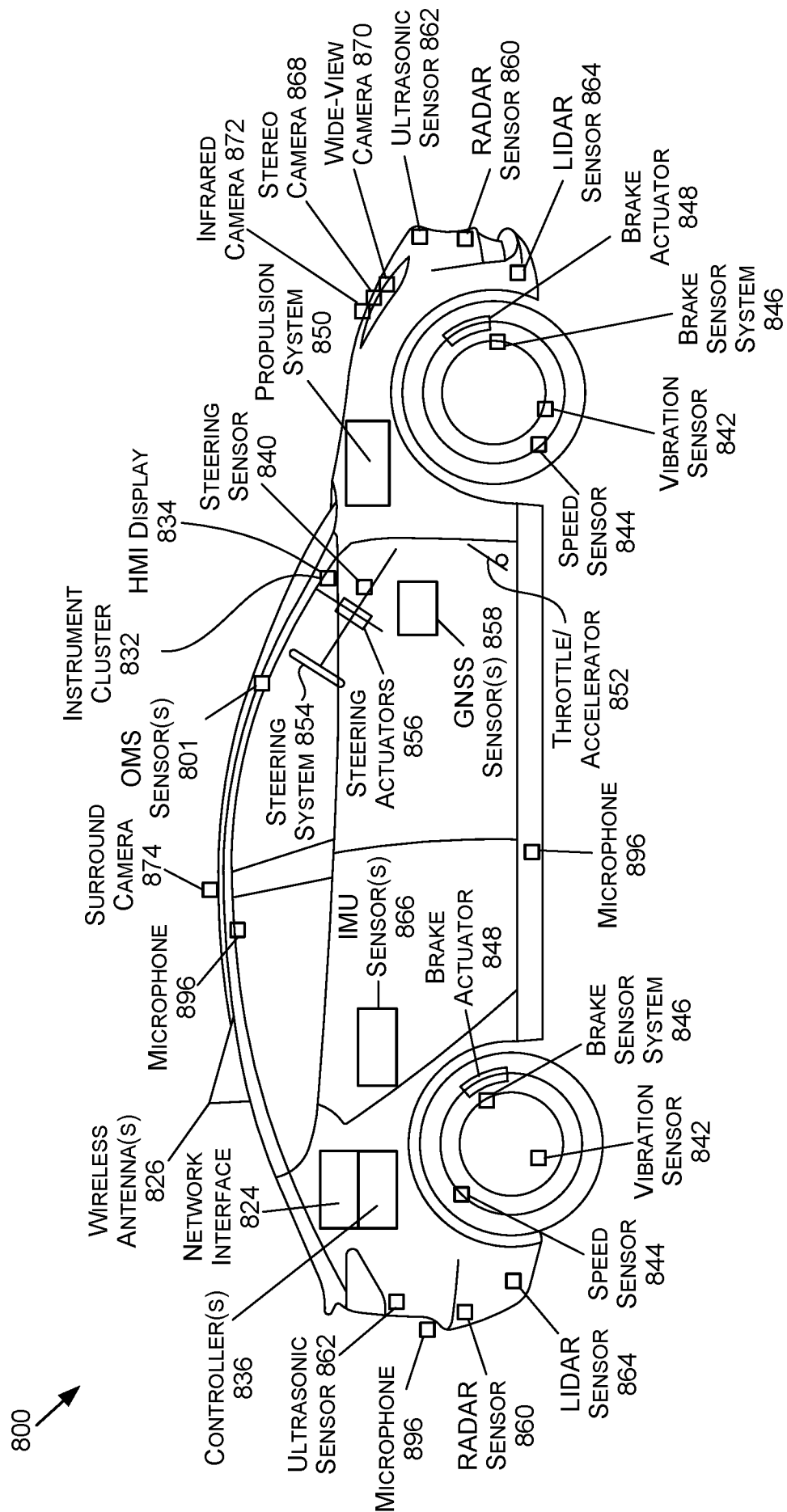


FIGURE 8A

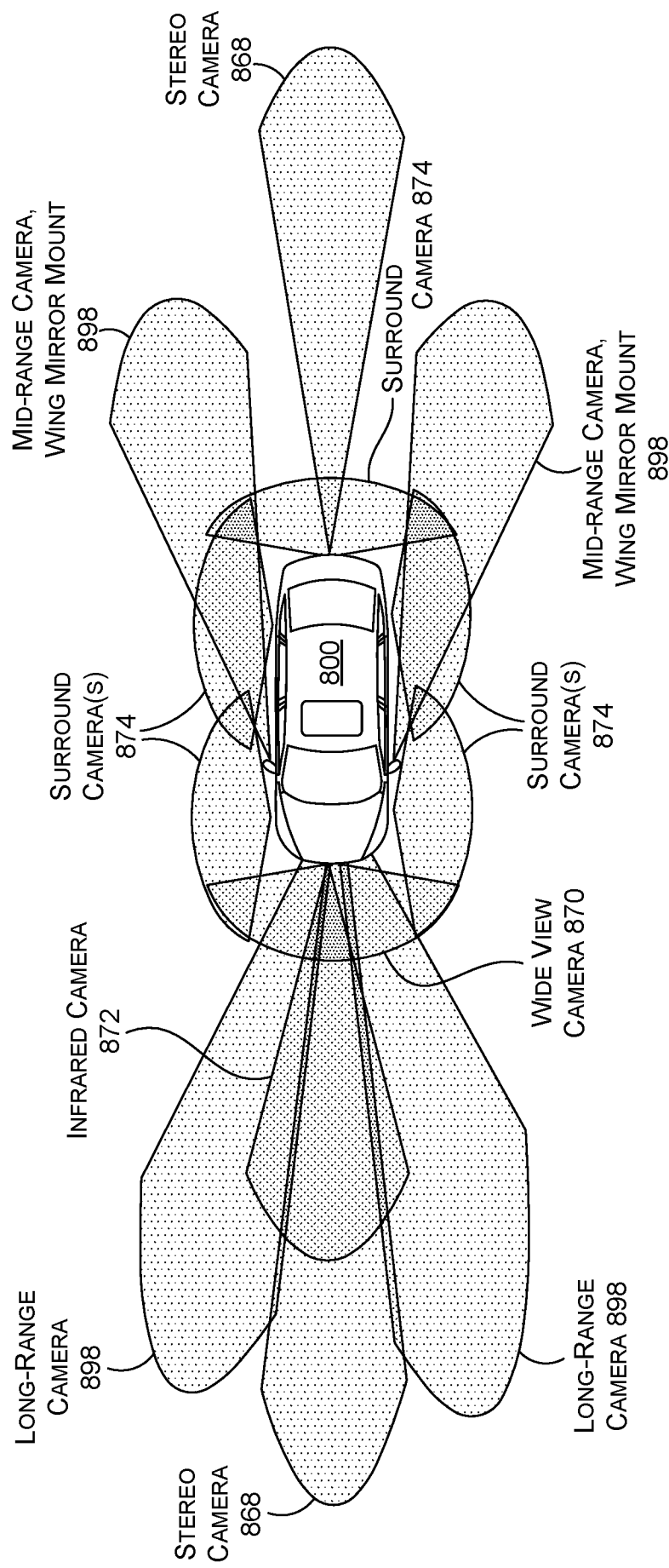


FIGURE 8B

10/13

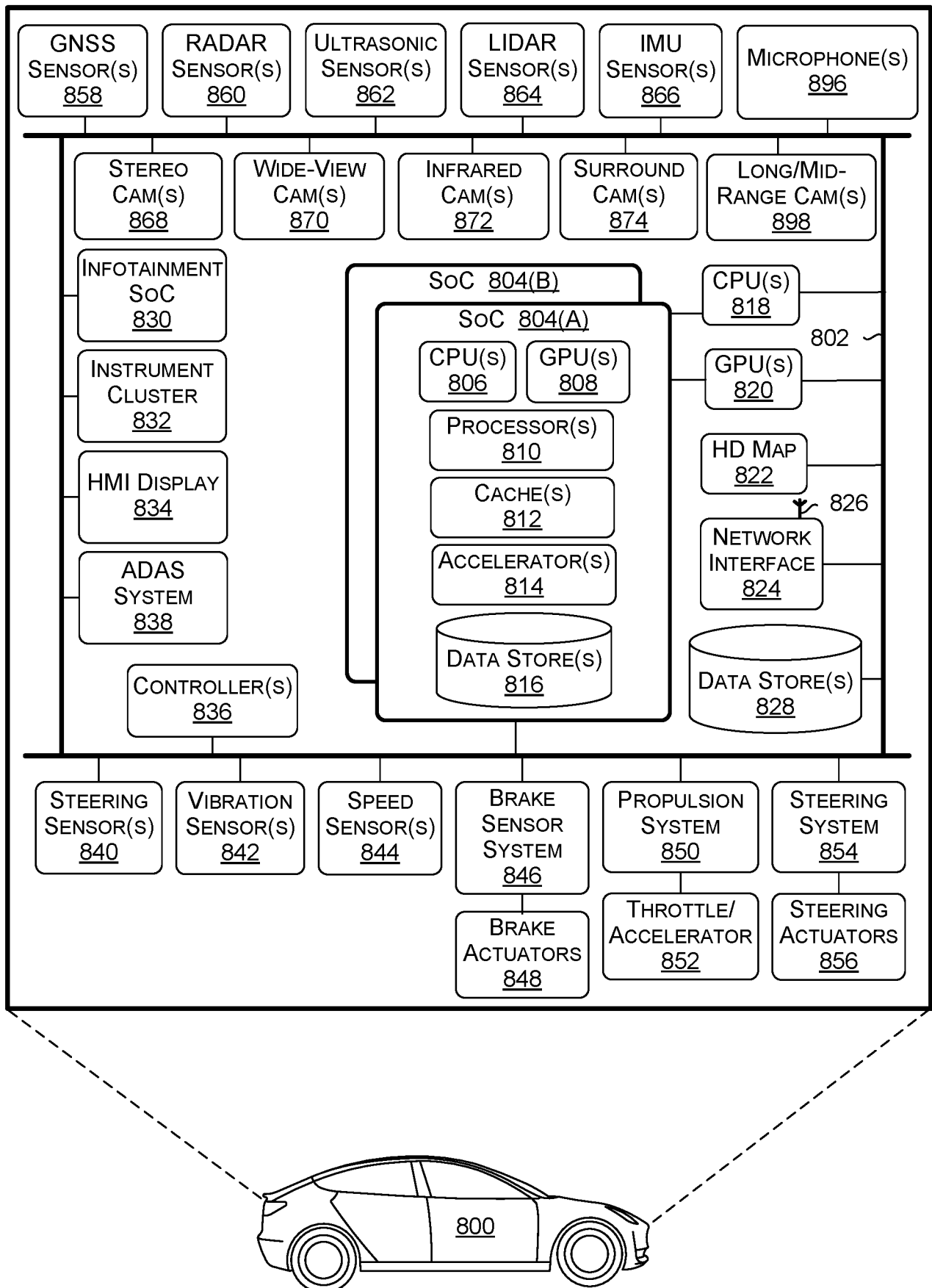


FIGURE 8C

876 ↗

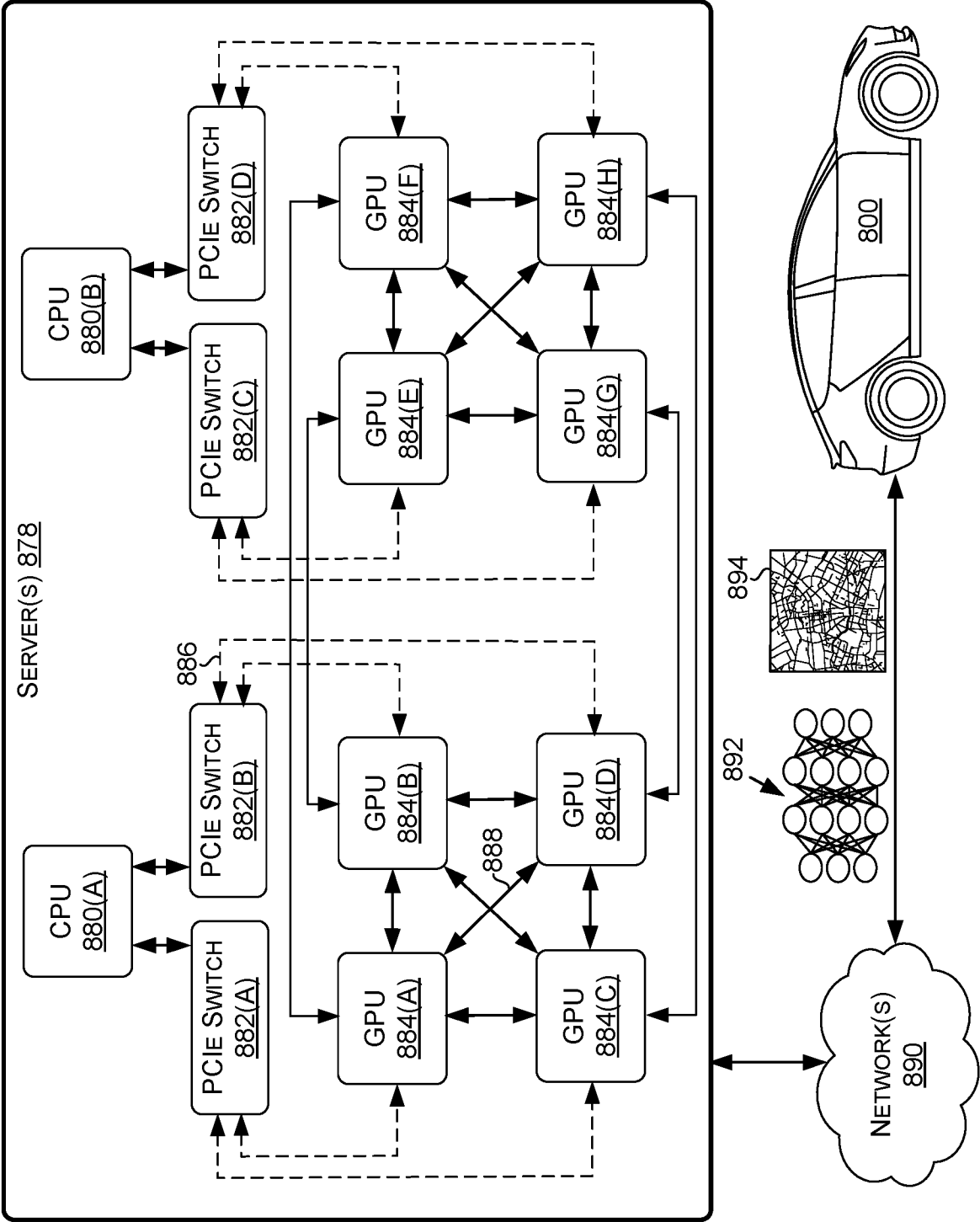


FIGURE 8D

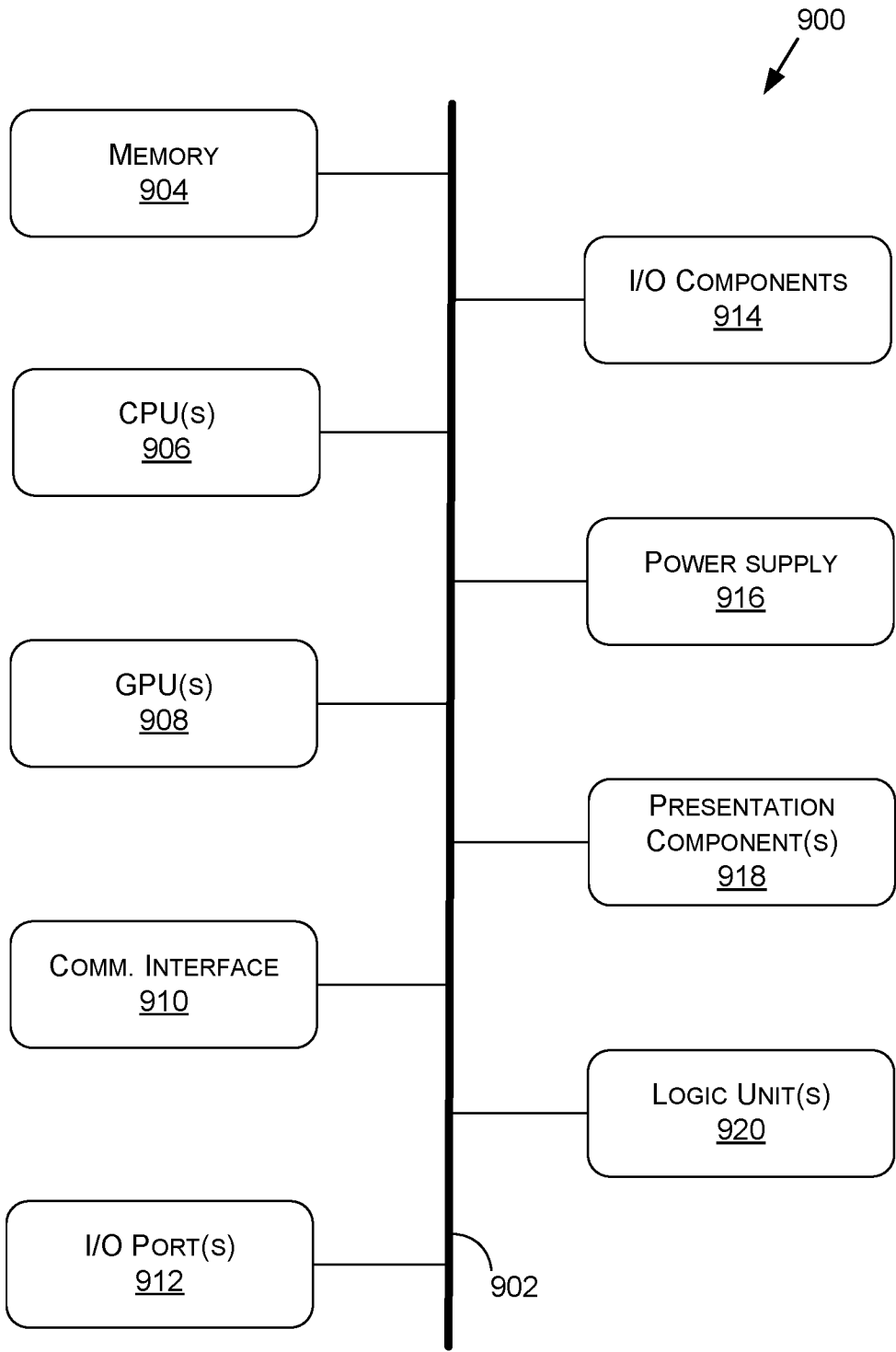


FIGURE 9

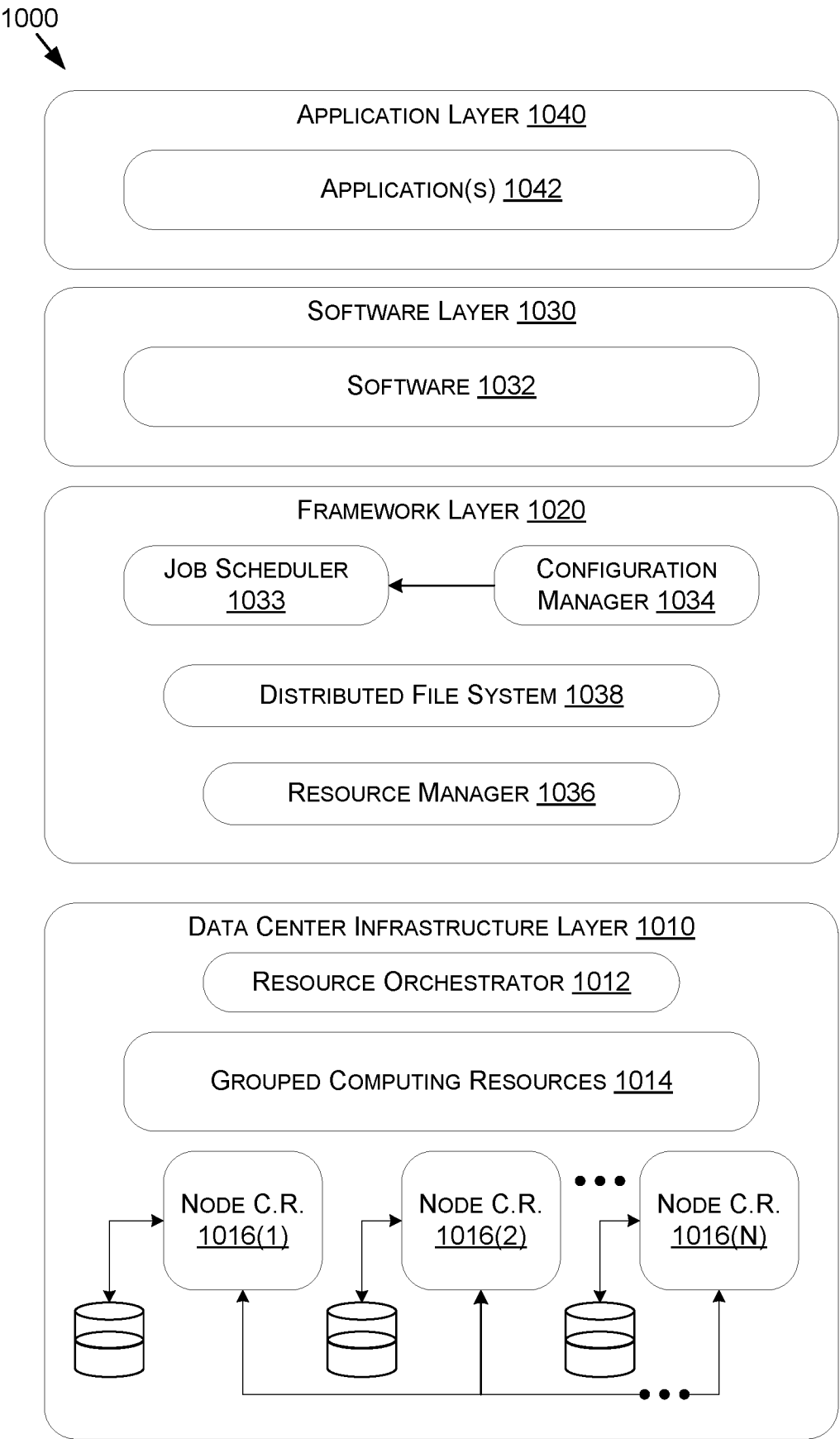


FIGURE 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/036754

A. CLASSIFICATION OF SUBJECT MATTER INV. G06V20/59 G06V40/10 G06V40/20 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06V Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2022/292705 A1 (FRIEDMAN YANIV [IL] ET AL) 15 September 2022 (2022-09-15) paragraph [0190] - paragraph [0205]; figure 8 -----	1 - 20
X	US 2022/388527 A1 (BARTH ALEXANDER [DE] ET AL) 8 December 2022 (2022-12-08) paragraph [0047] - paragraph [0053] paragraph [0064] - paragraph [0065] -----	1 - 3, 5, 8, 12 - 14, 16, 19, 20
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 11 October 2024		Date of mailing of the international search report 25/10/2024
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Rajadell Rojas, Olga

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2024/036754

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2022292705 A1	15-09-2022	CN 114144814 A	04-03-2022
		EP 3966740 A1	16-03-2022
		US 2022292705 A1	15-09-2022
		US 2024087165 A1	14-03-2024
		WO 2021005603 A1	14-01-2021

US 2022388527 A1	08-12-2022	CN 115439832 A	06-12-2022
		EP 4099214 A1	07-12-2022
		US 2022388527 A1	08-12-2022
