(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114270294 B
(45)授权公告日 2025.05.06

(21)申请号 202080036232.6
(22)申请日 2020.12.16
(65)同一申请的已公布的文献号
申请公布号 CN 114270294 A
(43)申请公布日 2022.04.01
(30)优先权数据
62/948,793 2019.12.16 US
16/902,737 2020.06.16 US
(85)PCT国际申请进入国家阶段日
2021.11.15
(86)PCT国际申请的申请数据
PCT/US2020/065350 2020.12.16
(87)PCT国际申请的公布数据
WO2021/127005 EN 2021.06.24

(73)专利权人 辉达公司
地址 美国加利福尼亚州
(72)发明人 N·M·阿拉尔 江海荣 N·普里 N·阿瓦达汗纳姆
(74)专利代理机构 北京市磐华律师事务所 11336
专利代理师 赵楠

(51)Int.Cl.
G06F 3/01(2006.01)
G06V 40/19(2022.01)
G06N 20/00(2019.01)
(56)对比文件
US 2019235236 A1,2019.08.01
US 2019251707 A1,2019.08.15
审查员 刘念

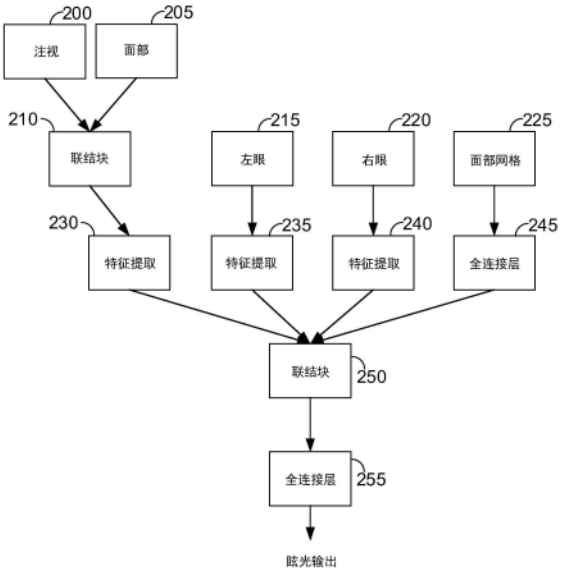
权利要求书2页 说明书26页 附图9页

(54)发明名称

使用眩光作为输入的注视确定

(57)摘要

学习眩光的机器学习系统和方法,从而以对输入图像上的眩光影响更有弹性的方式确定注视方向。除了图像本身之外,机器学习系统还具有独立的眩光表示,例如,关于图像中眩光点位置的信息,作为显式输入。以这种方式,机器学习系统在确定注视方向时明确考虑眩光,从而为包含眩光的图像产生更准确的结果。



CN 114270294 B

1.一种在存在眩光的情况下确定注视方向的方法,所述方法包括:

使用并行处理电路确定图像中对象的注视方向,所述注视方向至少部分地根据一个或更多个机器学习模型的输出来确定,所述机器学习模型将所述图像中的眩光的孤立表示、所述对象的面部的至少一部分的一个或更多个表示、以及对应于所述对象的至少一只眼睛的所述图像的部分作为输入,其中,所述对象的面部的至少一部分的所述一个或更多个表示包括所述对象的面部的第一表示和第二表示,所述对象的面部的所述第一表示包括所述对象的面部的掩码,所述对象的面部的所述第二表示包括所述对象的面部的面部网格,并且,所述一个或更多个机器学习模型包括将所述对象的面部的所述第一表示作为输入的第一机器学习模型,以及将所述对象的面部的所述第二表示作为输入的第二机器学习模型;和

基于确定的注视方向启动操作。

2.根据权利要求1所述的方法,还包括使用处理电路并至少部分地基于所述图像生成所述眩光的表示、所述对象的面部的至少一部分的所述一个或更多个表示以及对应于所述对象的至少一只眼睛的所述图像的部分。

3.如权利要求1所述的方法,其中,所述机器学习模型中的一个将所述眩光的孤立表示和所述对象的面部的至少一部分的表示二者作为输入。

4.如权利要求1所述的方法,其中,所述机器学习模型中的一个将所述眩光的孤立表示作为输入,并且所述机器学习模型中的另一个将所述对象的面部的至少一部分的表示作为输入。

5.如权利要求4所述的方法,其中,所述机器学习模型中的一个包括以所述眩光的孤立表示作为输入的全连接层。

6.如权利要求1所述的方法,其中,使用对应于车辆的传感器设备获得的传感器数据来生成所述图像,其中,所述启动还包括基于所述确定的注视方向启动所述车辆的操作。

7.如权利要求1所述的方法:

其中,对应于所述对象的至少一只眼睛的所述图像的部分还包括对应于所述对象的第一只眼睛的所述图像的第一部分,以及对应于所述对象的第二只眼睛的所述图像的第二部分。

8.一种用于在存在眩光的情况下确定注视方向的系统,所述系统包括:

存储器;和

并行处理电路,其被配置为:

确定图像中对象的注视方向,所述注视方向至少部分地根据一个或更多个机器学习模型的输出来确定,所述机器学习模型将所述图像中眩光的孤立表示、所述对象的面部的至少一部分的一个或更多个表示、以及对应于所述对象的至少一只眼睛的所述图像的部分作为输入,其中,所述对象的面部的至少一部分的所述一个或更多个表示包括所述对象的面部的掩码和所述对象的面部的面部网格,并且,所述一个或更多个机器学习模型包括将所述对象的面部的掩码作为输入的第一机器学习模型,以及将所述对象的面部的面部网格作为输入的第二机器学习模型;和

根据所述注视方向启动操作。

9.根据权利要求8所述的系统,其中,所述并行处理电路进一步被配置为生成所述眩光

的孤立表示、所述对象的面部的至少一部分的所述一个或更多个表示、以及对应于所述对象的所述至少一只眼睛的所述图像的部分，以上都是至少部分地基于所述图像生成。

10.如权利要求8所述的系统，其中，所述机器学习模型中的一个将所述眩光的孤立表示和所述对象的面部的至少一部分的表示两者作为输入。

11.根据权利要求8所述的系统，其中，所述机器学习模型中的一个将所述眩光的孤立表示作为输入，并且所述机器学习模型中的另一个将所述对象的面部的至少一部分的表示作为输入。

12.如权利要求11所述的系统，其中，所述机器学习模型中的一个包括将所述眩光的孤立表示作为输入的全连接层。

13.如权利要求8所述的系统，其中，所述操作还包括至少部分地基于所述注视方向的车辆的操作。

14.如权利要求8所述的系统：

其中，对应于所述对象的至少一只眼睛的所述图像的部分还包括对应于所述对象的第一只眼睛的图像的第一部分，以及对应于所述对象的第二只眼睛的图像的第二部分。

15.一种在存在眩光的情况下确定注视方向的方法，所述方法包括：

接收第一输入数据，所述第一输入数据指示图像中的眩光位置，所述第一输入数据包括：所述图像中的对象的面部的掩码；

接收代表对应于呈现在所述图像中的对象的所述图像的至少一部分的第二输入数据，所述第二输入数据包括所述对象的面部的面部网格；

使用并行处理电路确定所述对象的注视方向，所述注视方向是使用所述第一输入数据、所述第二输入数据、以及对应于所述对象的至少一只眼睛的所述图像的部分的一个或更多个机器学习模型的输出；和

基于所述注视方向启动操作。

16.如权利要求15所述的方法，其中，所述机器学习模型中的一个具有作为输入的第一输入数据和包括所述对象的眼睛在所述图像中的至少一个位置的第三输入数据。

17.如权利要求15所述的方法，其中，所述机器学习模型中的一个具有作为输入的所述第一输入数据，并且所述机器学习模型中的另一个具有作为输入的所述对象的眼睛在所述图像中的至少一个位置。

18.如权利要求15所述的方法，其中，所述操作还包括至少部分地基于所述注视方向的车辆的操作。

19.如权利要求15所述的方法，其中，所述一个或更多个机器学习模型还具有第三输入数据，所述第三输入数据包括对应于所述对象的第一只眼睛的所述图像的一部分，并且其中，所述一个或更多个机器学习模型还具有第四输入数据，所述第四输入数据包括与所述对象的第二只眼相对应的所述图像的一部分。

使用眩光作为输入的注视确定

[0001] 相关申请的交叉引用

[0002] 本申请要求于2020年6月16日提交的第16/902,737号美国非临时专利申请的优先权,该申请要求于2019年12月16日提交的第62/948,793号美国临时专利申请的权益,其中每一项的公开都通过引用全部并入本申请。

背景技术

[0003] 本公开的实施例一般涉及机器学习系统。更具体地说,本公开的实施例涉及使用以眩光作为输入的机器学习系统执行的注视确定。

发明内容

[0004] 最近,卷积神经网络(CNN)已经被开发来估计图像中表示的对象的注视方向。例如,此类CNN可以通过确定有关对象眼睛的某些特征,从对象的输入图像中推理出对象的注视方向。这使得使用此类CNN的系统能够自动确定对象正在看的方向,并实时做出相应的反应。

[0005] 然而,传统的注视确定系统并非没有它们的缺点。当一只或两只眼睛被遮挡或在输入图像中无法清晰显示时,此类系统通常难以确定对象的注视方向。在存在眩光的情况下,传统的CNN系统通常特别难以确定注视方向。来自各种来源的光经常从眼睛、眼镜或其他附近的表面反射,导致对象的图像中的亮点,其距图像传感器(例如,相机)可能至少部分地遮挡特或她的眼睛,从而降低传统注视确定系统的准确性。

[0006] 现有系统已经尝试使用多种方法来补偿或减少眩光的影响,例如通过修改照明中使用的光源、使用各种偏振技术、识别和去除眩光像素,或通过使用其他信息(例如头部姿势)来补偿眼睛信息的不足。然而,这些方法中的每一种的有效性都被证明是有限的。因此,本文针对更稳健的注视估计系统描述了系统和方法,该系统将眩光点合并为注视估计机器学习网络架构的显式输入。示例性注视估计系统可以利用相机或其他图像确定设备,以及能够执行机器学习网络(例如CNN)的推理操作的处理器(例如并行处理器)。在本公开的某些实施例中,系统可以接收由相机拍摄的对象的图像。一个或更多个机器学习网络被构建为接收输入:对象图像中的眩光的孤立表示,对象面部的至少一部分的一个或更多个表示,以及对应于对象的至少一只眼睛的图像的部分。从这些输入中,机器学习网络确定并输出如图像中所示的对象注视方向的估计。该注视方向可以被传输到基于所确定的注视发起操作的各种系统。

[0007] 可以以任何方式确定眩光的表示、对象的面部的表示以及与对象的眼睛对应的图像的部分。例如,在通过上述机器学习网络进行处理之前,系统可以分别确定眩光、对象的面部以及与对象的眼睛对应的图像的部分中的每个表示。眩光的表示可以是以任何方式确定的输入图像中眩光点的二值掩码,而对象面部的表示可以是对象面部至少包括他或她的眼睛的部分的类似确定的二值掩码。对应于眼睛的图像部分可以是眼睛裁剪,其中使用任何形状或对象识别技术来识别眼睛,无论是基于机器学习还是其他方式。

[0008] 在本公开的一个实施例中,一个机器学习模型可以将眩光的孤立表示(例如,眩光点的二值掩码)和对象面部的至少部分的表示(例如,对象面部至少包括他或她的眼睛的部分的二值掩码作为输入。

[0009] 在本公开的另一个实施例中,眩光的孤立表示可以被输入到一个机器学习模型,而对象面部的至少一部分的表示可以被输入到另一个机器学习模型。特别地,眩光的表示可以被输入到一组全连接层,而面部的至少一部分的表示可以被输入到特征提取模型。

[0010] 在本公开的一个实施例中,一个机器学习模型将与对象的一只眼睛对应的图像的部分作为输入,而另一机器学习模型将与对象的另一只眼睛对应的图像的另一部分作为输入。也就是说,两个不同的机器学习模型将取自对象图像的两个眼睛裁剪作为各自的输入。

[0011] 在本公开的另一个实施例中,一个机器学习模型具有作为输入的对象面部的一个表示,例如,从输入图像确定的对象面部的上述二值码,并且另一个机器学习模型具有作为输入对象的面部的另一种表示,例如,也可以从输入图像确定的粗糙面部网格。

[0012] 机器学习模型的最终输出是估计的注视方向或焦点。任何操作都可以响应此输出而启动。例如,机器学习模型的输出可以是车辆驾驶员的注视方向,并且启动的操作可以是响应于确定的注视方向而执行的车辆操作。例如,在确定驾驶员没有注视车辆行驶的方向时,车辆可以向驾驶员发出警告以将他或她的注意力转移回道路。

[0013] 因此,本公开的实施例可以被视为提供了一种机器学习系统,该系统通过显式地学习眩光来确定对象的注视方向。更具体地说,图像中眩光点的位置被用作机器学习模型的单独输入。因此,本公开的实施例的机器学习模型将对象的图像和从图像中提取的眩光点的显式表示作为输入。根据这些(以及可选的其他)输入,确定注视方向。然后可以基于所确定的注视方向启动各种操作。

附图说明

[0014] 结合附图考虑以下详细描述,本公开的上述和其他目的和优点将显而易见,其中相同的附图标记始终指代相同的部分,其中:

[0015] 图1A概念性地说明了在存在眩光的情况下常规注视确定过程的功能;

[0016] 图1B概念性地示出了本公开的实施例的抗眩光注视确定过程的功能;

[0017] 图2是本公开的实施例的示例性抗眩光的机器学习模型的框图表示;

[0018] 图3是本公开的实施例的注视确定系统的框图表示;

[0019] 图4A是根据本公开的一些实施例的示例自主车辆的图示;

[0020] 图4B是根据本公开的一些实施例的图4A示例性自助车辆的相机位置和视野示例;

[0021] 图4C是根据本公开的一下实施例的图4A的示例性自主车辆的示例性系统架构的框图;

[0022] 图4D是根据本公开的一些实施例的基于云的服务器与图4A的示例性自助车辆之间通信的系统图;

[0023] 图5是适用于实现本公开的一些实施例的示例计算设备的框图;

[0024] 图6是示出根据本公开的实施例的用于在存在眩光的情况下确定注视方向的过程步骤的流程图;和

[0025] 图7是本公开的实施例的进一步示例性抗眩光的机器学习模型的框图表示。

具体实施方式

[0026] 在一个实施例中,本公开涉及学习眩光并因此以能对抗眩光的影响的方式确定注视方向的机器学习系统和方法。除了图像本身之外,机器学习系统还具有孤立的眩光表示,例如,关于图像中眩光点位置的信息,作为显式输入。以这种方式,机器学习系统在确定注视方向时显式地考虑眩光,从而为包含眩光的图像产生更准确的结果。

[0027] 图1A概念性地图示了在存在眩光的情况下常规注视确定过程的功能。通常,在概念上,图像100被输入到CNN110,该CNN 110被训练以估计输入图像100中对象的注视方向。眩光的存在通常阻止该过程产生令人满意的结果。在这里,例如,存在于对象眼镜上的眩光点130会遮蔽对象的眼睛,使得CNN 110难以检测到足够多的对象眼睛,从而难以以高度的置信度确定他的注视方向。由于CNN 110没有足够的信息来确定注视,因此CNN 110的输出可能不准确、无意义,或者可能会完全丢弃受影响的输入图像。

[0028] 图1B概念性地图示了本公开的实施例的抗眩光注视确定过程的功能。与图1A的传统CNN 110相比,本公开的实施例描述了专门学习眩光并因此对抗其影响的机器学习模型。因此,当图像100和从图像100中提取的眩光点120的孤立表示被输入到本公开实施例的机器学习模型140时,模型140仍然能够以足够的置信度推理对象的注视方向,导致即使在存在可能部分遮挡对象眼睛的眩光点130的情况下也准确的输出注视向量。图1A和图1B示出了输入到机器学习模型140的图像100以概念性地说明机器学习模型140将提取自对象捕获的图像或与其相关的信息作为输入,注意在实践中,除了图像之外,输入到机器学习模型140可以是不同的表示或信息集。例如,模型140的输入可以是包含眩光点的面部网格,如下文进一步描述的。

[0029] 图2是本公开的实施例的示例性抗眩光的机器学习模型的框图表示。图2的机器学习模型包括如图所示布置的特征提取层230、235和240以及全连接层245和255。特别地,特征提取层230具有两个不同的输入,眩光表示200和面部表示205,它们由联结块210联结。联结的输入然后被传输到特征提取层230。特征提取块230可以根据任何方法提取输入图像100的特征。例如,特征提取块230可以是CNN的特征学习部分,以任何合适的方式构造并使用任何适合提取预期在输入图像中捕获的对象的特征的卷积核和池化层。因此,特征提取块230输出可以表示输入图像中表示的对象的面部的取向的特征。

[0030] 特征提取层235和240可以分别具有单个输入、左眼表示215和右眼表示220。例如,左眼表示215可以是输入图像100中对象左眼的裁剪,而右眼表示220可以是输入图像100中对象右眼的裁剪。类似于特征提取层230,特征提取层235和240各自都可以是CNN的任何特征学习部分,以任何合适的方式构造并且使用任何适于提取对象眼睛的特征的卷积核和池化层。每个特征提取层235、240的输出是一组眼睛特征,其可以分别代表输入215和220的瞳孔的注视方向。由于特征提取层235和240每个都被配置为分析眼睛作为输入,为了效率,它们对应的权重值可以共享或相同。然而,本公开的实施例考虑了特征提取层235和240中的每一个的任何权重值,无论是共享的还是以其他方式。

[0031] 全连接层245具有面网格225作为它们的单个输入。全连接层245可以是适合于将面部的输入图像分类为位置分类的任何分类器,该位置分类指示对象的面部在输入图像100中的位置和/或地位。例如,全连接层245可以是多层感知器或CNN的任何其他全连接层,其被配置和训练以将输入对象分类到多个离散位置之一。因此,全连接层245输出面部在输

入图像100中的位置的可能性。

[0032] 特征提取层230、235和240的输出以及全连接层245的输出由联结块250联结在一起,并输入到全连接层255,后者又依次输出对输入图像100中对象的注视方向。全连接层255可以是适合于将对象面部的输入特征分类为指示对象正在注视的方向的方向分类的任何分类器。例如,全连接层255可以是多层感知器或CNN的任何其他全连接层,其被配置和训练以将输入面部和眼睛特征和位置分类到多个离散注视方向之一。全连接层255的输出分类可以是注视方向的任何表示,例如向量、任何预定假想平面上的焦点等。

[0033] 眩光表示200可以是输入图像100中眩光的任何孤立表示,例如眩光表示120。也就是说,眩光表示200可以是传达输入图像100中存在的任何眩光的位置和大小的任何输入。作为一个示例,眩光表示200可以是从输入图像100中提取的眩光点的二值掩码,即,仅包含输入图像100中被确定为表示眩光的那些像素的图像。因此,除了在输入图像100中的对应像素被确定为眩光像素的那些像素之外,二值掩码的每个像素都是黑色像素(例如,在二值表示中,值为0的像素)。作为另一示例,眩光表示200可以是输入图像100中被确定为包含眩光的那些像素的位置的向量。

[0034] 面部表示205可以是输入图像100中面部的任何孤立表示。也就是说,面部表示205可以是仅传达输入图像100中存在的对象的面部的任何输入。例如,面部表示205可以是从输入图像100中提取的面部像素的二值掩码,其中面部像素或与被确定为代表对象的面部的输入图像100的那些像素对应的像素,都是一种颜色(例如,具有一个值),而所有其他像素是黑色像素(例如,具有不同的值)。作为另一示例,面部表示200可以是输入图像100中被确定为表示对象的面部的那些像素的位置的向量。

[0035] 左眼表示215和右眼表示220每个都可以是输入图像100中对象的对应眼睛的任何表示,其可以指示其相应眼睛的注视方向。例如,左眼表示215可以是仅包含对象的左眼的输入图像100的裁剪部分,而右眼表示220可以是仅包含对象的右眼的输入图像100的裁剪部分。

[0036] 面部网格225可以是在输入图像100内传达面部位置的任何输入。作为一个示例,面部网格225可以是对象面部的二值掩码,类似于面部表示205。该面部网格225可以是任何分辨率,例如与输入图像100的分辨率相同的分辨率,或者更粗略的表示,例如25×25平方图像二值掩码,其对应于输入图像100中的面部的各个框使用一种或多种非黑色表示,而剩下的框是黑色的。面网格225可以是任何其他分辨率,例如12×12等。面部网格225中表示的面部可以缩放到任何特定尺寸以帮助计算,或者可以保持与输入图像100相同的比例。此外,面部网格225可以包括传达对象面部在输入图像100中的空间位置的任何信息,无论是图像还是其他。例如,面部网格225可以是以任何方式确定的对象面部的一组标志点。

[0037] 图3是本公开的实施例的一个示例性注视确定系统的框图表示。这里,计算设备300可以是包含能够执行本公开的实施例的注视确定操作的处理电路的任何电子计算设备,与相机310和注视辅助系统320两者进行电子通信。在操作中,相机310,其可对应于如下图4A和图4C的机舱相机441,捕获对象的图像并将其传输到计算设备300,然后计算设备300实现例如图2的机器学习模型。从照相机310的图像确定图2所示的输入,并计算对象的输出注视方向。计算设备300将这个注视方向传输到注视辅助系统320,注视辅助系统320作为响应采取动作或执行一个或更多个操作。计算设备300可以是适合于实现本公开的实施例的

机器学习模型的任何一个或更多个电子计算设备,例如以下更详细描述的计算设备500。

[0038] 注视辅助系统320可以是能够基于它从计算设备300接收的注视方向执行一个或更多个动作的任何系统。可以设想相机310、计算设备300和注视辅助系统320的任何配置。作为一个示例,注视辅助系统320可以是能够确定驾驶员或另一乘客的注视方向并对其做出反应的自主车辆,例如以下更详细描述的自主车辆400。在该示例中,相机310和计算设备300可以位于车辆内,而注视辅助系统320可以代表车辆本身。相机310可以定位在车辆内允许其看到驾驶员或乘客的任何位置。因此,相机310可以捕捉驾驶员或乘客的图像并将它们传输到计算设备300,计算设备300计算输入200、205、215、220和225并确定驾驶员的注视方向。该注视方向然后可以被传输到例如确定车辆可以作为响应采取的动作的另一个软件模块。例如,车辆可以确定注视方向代表分心的驾驶员或未注意道路的驾驶员,并且可以启动任何类型的操作作为响应。此类操作可包括向驾驶员发出的任何类型的警告(例如,视觉或听觉警告、平视显示器上的警告等)、自动驾驶启动、制动或转弯操作或任何其他动作。计算设备300可以作为本地处理器定位在注视辅助系统320的车辆内,或者可以是从相机310接收图像并将注视方向或指令无线传输到注视辅助系统320的车辆的远程处理器。

[0039] 作为另一示例,注视辅助系统320可以是能够响应于用户的运动和注视而显示图像的虚拟现实或增强现实系统。在该示例中,注视辅助系统320包括虚拟现实或增强现实显示器,例如由用户佩戴并被配置为向其投射图像的耳机。相机310和计算设备300可以位于耳机内,相机310捕捉用户的眼睛的图像并且计算设备300确定他或她的注视方向。该注视方向然后可以被传输到虚拟现实或增强现实显示器,其可以作为响应执行任何动作。例如,为了节省计算资源,注视辅助系统320可以仅渲染那些使用确定的注视方向确定的用户视野内的虚拟现实或增强现实元素。类似地,注视辅助系统320可以提醒用户注意被确定为在用户视野之外但用户可能希望避免或可能感兴趣的对象或事件。与上述自主车辆示例一样,虚拟现实或增强现实系统的计算设备300可以位于系统320内,例如,在耳机本身内,或者可以位于远程定位使得图像无线传输到计算设备300并且可以将计算出的注视方向无线返回耳机,耳机又可以执行各种操作作为响应。

[0040] 作为又一示例,注视辅助系统320可以是基于计算机的广告系统,其确定用户正在观看的视觉刺激,例如但不限于广告、警告、对象、人或其他可见区域或兴趣点。更具体地,注视辅助系统可以是任何电子计算系统或设备,例如台式计算机、膝上型计算机、智能电话、服务器计算机等。相机310和计算设备300可以并入该计算设备以指向用户,例如进入或靠近计算设备的显示器。相机310可以捕捉用户的图像并且计算设备300可以确定他或她的注视方向。然后可以将确定的注视方向传输到注视辅助系统320,例如为用户显示广告的计算设备、远程计算设备等。计算设备然后可以使用计算出的注视方向来确定用户关注的主题,提供关于各种广告、警告或其他视觉刺激的有效性的信息。

[0041] 图4A是根据本公开的一些实施例的示例自主车辆400的图示。自主车辆400(在本文中也称为“车辆400”)可包括但不限于乘用车,例如汽车、卡车、公共汽车、急救车、穿梭车、电动或机动自行车、摩托车、消防车、警车、救护车、船、工程车辆、水下航行器、无人机和/或其他类型的车辆(例如,无人驾驶和/或容纳一个或更多个乘客)。自主车辆通常按照美国运输部的一个部门——国家公路交通安全管理局(NHTSA)以及汽车工程师协会(SAE)“Taxonomy and Definitions for Terms Related to Driving Automation Systems for

On-Road Motor Vehicles”(2018年6月15日发布的标准No.J3016-201806,2016年9月30日发布的标准No.J3016-201609,以及该标准的先前和未来的版本)定义的自动化级别进行描述。车辆400可能够实现符合自主驾驶级别的3级-5级中的一个或更多个的功能。例如,取决于实施例,车辆400可能够实现条件自动化(3级)、高自动化(4级)和/或全自动化(5级)。

[0042] 车辆400可以包括诸如底盘、车身、车轮(例如2个、4个、6个、8个、18个等)、轮胎、车轴之类的组件以及车辆的其他组件。车辆400可以包括推进系统450,例如内燃机、混合动力发电厂、全电动发动机和/或另一种推进系统类型。推进系统450可以连接到可以包括变速器的车辆400的传动系以便实现车辆400的推进。可以响应于接收到来自油门/加速器452的信号而控制推进系统450。

[0043] 可以包括方向盘的转向(steering)系统454可以用来在推进系统450操作时(例如在车辆运动时)使车辆400转向(例如沿着希望的路径或路线)。转向系统454可以接收来自转向执行器456的信号。对于全自动(5级)功能而言,方向盘可以是可选的。

[0044] 制动传感器系统446可以用来响应于接收到来自制动执行器448和/或制动传感器的信号而操作车辆制动器。

[0045] 可以包括一个或更多个CPU,一个或更多个片上系统(SoC)404(图4C)和/或一个或更多个GPU的一个或更多个控制器436可以向车辆400的一个或更多个组件和/或系统提供(例如表示命令的)信号。例如,一个或更多个控制器可以发送经由一个或更多个制动执行器448操作车辆制动器、经由一个或更多个转向执行器456操作转向系统454、和/或经由一个或更多个油门/加速器452操作推进系统450的信号。一个或更多个控制器436可以包括一个或更多个板载(例如集成)计算设备(例如超级计算机),所述计算设备处理传感器信号并且输出操作命令(例如表示命令的信号),以实现自主驾驶和/或辅助人类驾驶员驾驶车辆400。一个或更多个控制器436可以包括用于自主驾驶功能的第一控制器436、用于功能性安全功能的第二控制器436、用于人工智能功能(例如计算机视觉)的第三控制器436、用于信息娱乐功能的第四控制器436、用于紧急情况下的冗余的第五控制器436和/或其他控制器。在一些示例中,单个控制器436可以处理上述功能中的两个或更多,两个或更多控制器436可以处理单个功能,和/或其任意组合。

[0046] 一个或更多个控制器436可以响应于接收自一个或更多个传感器的传感器数据(例如传感器输入),提供用于控制车辆400的一个或更多个组件和/或系统的信号。传感器数据可以接收自例如且不限于全球导航卫星系统传感器458(例如全球定位系统传感器)、RADAR传感器460、超声传感器462、LIDAR传感器464、惯性测量单元(IMU)传感器466(例如加速度计、陀螺仪、磁罗盘、磁力计等)、麦克风496、立体相机468、广角相机470(例如鱼眼相机)、红外相机472、环绕相机474(例如360度相机)、远程和/或中程相机498、速度传感器444(例如用于测量车辆400的速率)、振动传感器442、转向传感器440、制动传感器(例如作为制动传感器系统446的部分)和/或其他传感器类型。

[0047] 控制器436中的一个或更多个可以接收来自车辆400的仪表组432的输入(例如由输入数据表示),并且经由人机接口(HMI)显示器434、听觉信号器、扬声器和/或经由车辆400的其他组件提供输出(例如输出数据、显示数据等表示的)。这些输出可以包括诸如车辆速度、速率、时间、地图数据(例如图4C的HD地图422)、位置数据(例如,车辆400例如在地图上的位置)、方向、其他车辆的位置(例如占用网格)之类的信息,如控制器436所感知的关于

对象和对象状态的信息等等。例如,HMI显示器434可以显示关于一个或更多个对象(例如街道指示牌、警示牌、交通灯变化等)的存在性的信息和/或关于车辆已经做出、正在做出或者将会做出的驾驶机动的信息(例如现在变道、两英里后离开34B,等等)。

[0048] 车辆400进一步包括网络接口424,其可以使用一个或更多个无线天线426和/或调制解调器通过一个或更多个网络通信。例如,网络接口424可能够通过LTE、WCDMA、UMTS、GSM、CDMA2000等通信。一个或更多个无线天线426也可以使用诸如蓝牙、蓝牙LE、Z波、ZigBee等等之类的一个或更多个局域网和/或诸如LoRaWAN、SigFox等等之类的一个或更多个低功率广域网(LPWAN)实现环境中的对象(例如车辆、移动设备等等)之间的通信。

[0049] 图4B为根据本公开一些实施例的用于图4A的示例自主车辆400的相机位置和视场的示例。相机和各自的视场是一个示例实施例,并不意图是限制性的。例如,可以包括附加的和/或可替换的相机,和/或这些相机可以位于车辆400上的不同位置。

[0050] 用于相机的相机类型可以包括但不限于可以适于与车辆400的组件和/或系统一起使用的数字相机。所述相机可以在汽车安全完整性级别(ASIL)B下和/或在另一个ASIL下操作。相机类型可以具有任何图像捕获率,例如60帧每秒(fps)、120fps、240fps等等,这取决于实施例。相机可能够使用滚动快门、全局快门、另一种类型的快门或者其组合。在一些示例中,滤色器阵列可以包括红白白白(RCCC)滤色器阵列、红白白蓝(RCCB)滤色器阵列、红蓝绿白(RBGC)滤色器阵列、Foveon X3滤色器阵列、拜耳传感器(RGGB)滤色器阵列、单色传感器滤色器阵列和/或另一种类型的滤色器阵列。在一些实施例中,诸如具有RCCC、RCCB和/或RBGC滤色器阵列的相机之类的清晰像素相机可以用在提高光敏感度的努力中。

[0051] 在一些示例中,所述相机中的一个或更多个可以用来执行高级驾驶员辅助系统(ADAS)功能(例如作为冗余或故障安全设计的部分)。例如,可以安装多功能单目相机以提供包括车道偏离警告、交通指示牌辅助和智能前照灯控制在内的功能。所述相机中的一个或更多个(例如全部相机)可以同时记录和提供图像数据(例如视频)。

[0052] 所述相机中的一个或更多个可以安装在诸如定制设计的(3-D打印的)组件之类的安装组件中,以便切断可能干扰相机的图像数据捕获能力的杂散光和来自汽车内的反射(例如挡风玻璃镜中反射的来自仪表板的反射)。关于翼镜安装组件,翼镜组件可以是定制3-D打印的,使得相机安装板匹配翼镜的形状。在一些示例中,一个或更多个相机可以集成到翼镜中。对于侧视相机而言,一个或更多个相机也可以集成到驾驶室每个拐角的四根柱子内。

[0053] 具有包括车辆400前面的环境部分的视场的相机(例如前置相机)可以用于环视,以帮助识别前向路径和障碍,以及在一个或更多个控制器436和/或控制SoC的帮助下辅助提供对于生成占用网格和/或确定优选车辆路径至关重要的信息。前置相机可以用来执行许多与LIDAR相同的ADAS功能,包括紧急制动、行人检测和碰撞避免。前置相机也可以用于ADAS功能和系统,包括车道偏离警告(“LDW”)、自主巡航控制(“ACC”),和/或诸如交通指示牌识别之类的其他功能。

[0054] 各种各样的相机可以用于前置配置中,包括例如包括CMOS(互补金属氧化物半导体)彩色成像仪在内的单目相机平台。另一个示例可以是广角相机470,其可以用来感知从周边进入视野的对象(例如行人、十字路口交通或者自行车)。尽管图4B中图示出仅仅一个广角相机,但是在车辆400上可以存在任意数量的广角相机470。此外,远程相机498(例如长

视立体相机对)可以用于基于深度的对象检测,尤其是用于尚未针对其训练神经网络的对象。远程相机498也可以用于对象检测和分类以及基本的对象跟踪。

[0055] 一个或更多个立体相机468也可以包括在前置配置中。立体相机468可以包括集成控制单元,该单元包括可扩展处理单元,其可以提供在单个芯片上具有集成的CAN或以太网接口的多核微处理器和可编程逻辑(,例如,FPGA)。这样的单元可以用来生成车辆环境的3-D地图,包括针对图像中的所有点的距离估计。可替代的立体相机468可以包括紧凑型立体视觉传感器,其可以包括两个相机镜头(左右各一个)以及可以测量从车辆到目标对象的距离并且使用生成的信息(例如元数据)激活自主紧急制动和车道偏离警告功能的图像处理芯片。除了本文所描述的那些之外或者可替代地,可以使用其他类型的立体相机468。

[0056] 具有包括车辆400的侧面的环境部分的视场的相机(例如侧视相机)可以用于环视,提供用来创建和更新占用网格以及生成侧撞击碰撞警告的信息。例如,环绕相机474(例如如图4B中所示的四个环绕相机474)可以置于车辆400周围。环绕相机474可以包括广角相机470、鱼眼相机、360度相机和/或类似物。例如,四个鱼眼相机可以置于车辆的前面、后面和侧面。在一种可替代的布置中,车辆可以使用三个环绕相机474(例如左边、右边和后面),并且可以利用一个或更多个其他相机(例如前向相机)作为第四环视相机。

[0057] 具有包括车辆400的后面的环境部分的视场的相机(例如后视相机)可以用于辅助停车、环视、后面碰撞警告以及创建和更新占用网格。可以使用各种各样的相机,包括但不限于也适合作为如本文所描述的前置相机(例如远程和/或中程相机498、立体相机468、红外相机472等等)的相机。

[0058] 具有包括车辆400的内部或舱室的部分的视野的相机可用于监控舱室中的驾驶员、乘客或对象的一个或更多个状态。可以使用任何类型的相机,包括但不限于一个或更多个舱室相机441,其可以是本文所述的任何类型的相机,并且可以放置在车辆400上或内的任何位置,以提供舱室或其内部的视图。例如,一个或更多个舱室相机441可以放置在车辆400仪表板、后视镜、侧视镜、座椅或车门的某些部分之内或之上,并且被定向为获捉任何驾驶员、乘客或任何其他对象或车辆400的一部分的图像。

[0059] 图4C为根据本公开一些实施例的用于图4A的示例自主车辆400的示例系统架构的框图。应当理解,这种布置和本文描述的其他布置仅仅作为示例而被阐述。除了所示的那些之外或者代替它们的是,可以使用其他的布置和元素(例如机器、接口、功能、顺序、功能分组等等),并且一些元素可以完全省略。进一步,许多本文描述的元素是功能实体,其可以实现为分立的或分布式组件或者结合其他组件实现,以及以任何适当的组合和位置实现。本文描述为由实体执行的各个功能可以通过硬件、固件和/或软件实现。例如,各个功能可以通过处理器执行存储在存储器中的指令而实现。

[0060] 图4C中车辆400的组件、特征和系统中的每一个被图示为经由总线402连接。总线402可以包括控制器区域网络(CAN)数据接口(可替代地,本文称为“CAN总线”)。CAN可以是车辆400内部的网络,用来辅助控制车辆400的各种特征和功能,例如制动器、加速、制动、转向、挡风玻璃雨刷等等的驱动。CAN总线可以被配置为具有数十或者甚至数百个节点,每个节点具有其自己的唯一标识符(例如CAN ID)。可以读取CAN总线以找到方向盘角度、地速、每分钟发动机转速(RPM)、按钮位置和/或其他车辆状态指示符。CAN总线可以是ASIL B兼容的。

[0061] 尽管这里将总线402描述为CAN总线,但是这并不意图是限制性的。例如,除了CAN总线之外或者可替代地,可以使用FlexRay和/或以太网。此外,尽管用单条线来表示总线402,但是这并不意图是限制性的。例如,可以存在任意数量的总线402,其可以包括一条或更多条CAN总线、一条或更多条FlexRay总线、一条或更多条以太网总线和/或一条或更多条使用不同协议的其他类型的总线。在一些示例中,两条或更多总线402可以用来执行不同的功能,和/或可以用于冗余。例如,第一总线402可以用于碰撞避免功能,并且第二总线402可以用于驱动控制。在任何示例中,每条总线402可以与车辆400的任何组件通信,并且两条或更多总线402可以与相同的组件通信。在一些示例中,车辆内的每个SoC 404、每个控制器436和/或每个计算机可以有权访问相同的输入数据(例如来自车辆400的传感器的输入),并且可以连接到诸如CAN总线之类的公共总线。

[0062] 车辆400可以包括一个或更多个控制器436,例如本文关于图4A所描述的那些控制器。控制器436可以用于各种各样的功能。控制器436可以耦合到车辆400的任何其他不同的组件和系统,并且可以用于车辆400的控制、车辆400的人工智能、用于车辆400的信息娱乐和/或类似物。

[0063] 车辆400可以包括一个或更多个片上系统(SoC)404。SoC 404可以包括CPU 406、GPU 408、处理器410、高速缓存412、加速器414、数据存储415和/或未图示出的其他组件和特征。在各种各样的平台和系统中,SoC 404可以用来控制车辆400。例如,一个或更多个SoC 404可以在系统(例如车辆400的系统)中与HD地图422结合,所述HD地图可以经由网络接口424从一个或更多个服务器(例如图4D的一个或更多个服务器478)获得地图刷新和/或更新。

[0064] CPU 406可以包括CPU集群或者CPU复合体(可替代地,本文称为“CCPLEX”)。CPU 406可以包括多个核和/或L2高速缓存。例如,在一些实施例中,CPU 406在一致性多处理器配置中可以包括八个核。在一些实施例中,CPU 406可以包括四个双核集群,其中每个集群具有专用的L2高速缓存(例如2MB L2高速缓存)。CPU 406(例如CCPLEX)可以被配置为支持同时集群操作,使得CPU 406的集群的任意组合能够在任何给定时间是活动的。

[0065] CPU 406可以实现包括以下特征中的一个或更多个的功率管理能力:各硬件块在空闲时可以自动进行时钟门控以节省动态功率;由于WFI/WFE指令的执行,每个核时钟可以在该核不主动地执行指令时进行门控;每个核可以独立地进行功率门控;当所有核都进行时钟门控或者功率门控时,可以独立地对每个核集群进行时钟门控;以及/或当所有核都进行功率门控时,可以独立地对每个核集群进行功率门控。CPU 406可以进一步实现用于管理功率状态的增强算法,其中指定允许的功率状态和期望的唤醒时间,并且硬件/微代码为所述核、集群和CCPLEX确定要进入的最佳的功率状态。处理核可以在软件中支持简化的功率状态进入序列,该工作被卸载到微代码。

[0066] GPU 408可以包括集成的GPU(可替代地,本文称为“iGPU”)。GPU 408可以是可编程的,并且对于并行工作负载而言是高效的。在一些示例中,GPU 408可以使用增强张量指令集。GPU 408可以包括一个或更多个流式微处理器,其中每个流式微处理器可以包括L1高速缓存(例如具有至少96KB存储能力的L1高速缓存),并且这些流式微处理器中的两个或更多可以共享L2高速缓存(例如具有512KB存储能力的L2高速缓存)。在一些实施例中,GPU 408可以包括至少八个流式微处理器。GPU 408可以使用基于计算机的应用编程接口(API)。此

外,GPU 408可以使用一个或更多个并行计算平台和/或编程模型(例如NVIDIA的CUDA)。

[0067] 在汽车和嵌入式使用的情况下,可以对GPU 408进行功率优化以实现最佳性能。例如,可以在鳍式场效应晶体管(FinFET)上制造GPU 408。然而,这并不意图是限制性的,并且GPU 408可以使用其他半导体制造工艺来制造。每个流式微处理器可以合并划分成多个块的若干混合精度处理核。例如且非限制性地,可以将64个PF32核和32个PF64核划分成四个处理块。在这样的示例中,每个处理块可以分配16个FP32核、8个FP64核、16个INT32核、用于深层学习矩阵算术的两个混合精度NVIDIA张量核、L0指令高速缓存、线程束(warp)调度器、分派单元和/或64KB寄存器文件。此外,流式微处理器可以包括独立的并行整数和浮点数据路径,以利用计算和寻址计算的混合提供工作负载的高效执行。流式微处理器可以包括独立线程调度能力,以允许实现并行线程之间的更细粒度的同步和协作。流式微处理器可以包括组合的L1数据高速缓存和共享存储器单元,以便在简化编程的同时提高性能。

[0068] GPU 408可以包括在一些示例中提供大约900GB/s的峰值存储器带宽的高带宽存储器(HBM)和/或16GB HBM2存储器子系统。在一些示例中,除了HBM存储器之外或者可替代地,可以使用同步图形随机存取存储器(SGRAM),例如第五代图形双倍数据速率同步随机存取存储器(GDDR5)。

[0069] GPU 408可以包括统一存储器技术,其包括访问计数器以允许存储器页面更精确地迁移到最频繁地访问它们的处理器,从而提高处理器之间共享的存储器范围的效率。在一些示例中,地址转换服务(ATS)支持可以用来允许GPU 408直接访问CPU 406页表。在这样的示例中,当GPU 408存储器管理单元(MMU)经历遗漏时,可以将地址转换请求传输至CPU 406。作为响应,CPU 406可以在其页表中寻找用于地址的虚拟-物理映射,并且将转换传输回GPU 408。这样,统一存储器技术可以允许单个统一虚拟地址空间用于CPU 406和GPU 408二者的存储器,从而简化了GPU 408编程和将应用程序移植(port)到GPU 408。

[0070] 此外,GPU 408可以包括访问计数器,其可以跟踪GPU 408访问其他处理器的存储器的频率。访问计数器可以帮助确保存储器页面移至最频繁地访问这些页面的处理器的物理存储器。

[0071] SoC 404可以包括任意数量的高速缓存412,包括本文描述的那些高速缓存。例如,高速缓存412可以包括CPU 406和GPU 408二者可用的L3高速缓存(例如,其连接到CPU 406和GPU 408二者)。高速缓存412可以包括回写高速缓存,其可以例如通过使用高速缓存一致性协议(例如MEI、MESI、MSI等)跟踪行的状态。取决于实施例,L3高速缓存可以包括4MB或者更多,但是也可以使用更小的高速缓存大小。

[0072] SoC 404可以包括算术逻辑单元(ALU),其可以用于执行关于车辆400的各种任务或操作中的任一个的处理——例如处理DNN。此外,SoC 404可以包括浮点单元(FPU)——或其他数学协处理器或数字协处理器类型——用于在系统内执行数学运算。例如,SoC 104可以包括集成为CPU 406和/或GPU 408内的执行单元的一个或更多个FPU。

[0073] SoC 404可以包括一个或更多个加速器414(例如硬件加速器、软件加速器或者其组合)。例如,SoC 404可以包括硬件加速集群,其可以包括优化的硬件加速器和/或大型片上存储器。该大型片上存储器(例如4MB SRAM)可以使得硬件加速集群能够加速神经网络和其他计算。硬件加速集群可以用来补充GPU 408,并且卸载GPU 408的一些任务(例如释放GPU 408的更多周期以用于执行其他任务)。作为一个示例,加速器414可以用于足够稳定以

易于控制加速的有针对性的工作负载(例如感知、卷积神经网络(CNN)等等)。当在本文中使用时,术语“CNN”可以包括所有类型的CNN,包括基于区域的或者区域卷积神经网络(RCNN)和快速RCNN(例如用于对象检测)。

[0074] 加速器414(例如硬件加速集群)可以包括深度学习加速器(DLA)。DLA可以包括可以被配置成为深度学习应用和推理提供额外的每秒10万亿次操作的一个或更多个张量处理单元(TPU)。TPU可以是被配置为执行图像处理功能(例如用于CNN、RCNN等)且针对执行图像处理功能而优化的加速器。DLA可以进一步针对特定的一组神经网络类型和浮点运算以及推理进行优化。DLA的设计可以比通用GPU提供每毫米更高的性能,并且远远超过CPU的性能。TPU可以执行若干功能,包括单实例卷积函数,支持例如用于特征和权重二者的INT8、INT16和FP16数据类型,以及后处理器功能。

[0075] DLA可以在处理的或者未处理的数据上针对各种各样的功能中的任何功能快速且高效地执行神经网络,尤其是CNN,例如且不限于:用于使用来自相机传感器的数据进行对象识别和检测的CNN;用于使用来自相机传感器的数据进行距离估计的CNN;用于使用来自麦克风的数据进行应急车辆检测和识别与检测的CNN;用于使用来自相机传感器的数据进行面部识别和车主识别的CNN;以及/或用于安全和/或安全相关事件的CNN。

[0076] DLA可以执行GPU 408的任何功能,并且通过使用推理加速器,例如,设计者可以使DLA或GPU 408针对任何功能。例如,设计者可以将CNN的处理和浮点运算聚焦在DLA上,并且将其他功能留给GPU 408和/或其他加速器414。

[0077] 加速器414(例如硬件加速集群)可以包括可编程视觉加速器(PVA),其在本文中可以可替代地称为计算机视觉加速器。PVA可以被设计和配置为加速用于高级驾驶员辅助系统(ADAS)、自主驾驶和/或增强现实(AR)和/或虚拟现实(VR)应用的计算机视觉算法。PVA可以提供性能与灵活性之间的平衡。例如,每个PVA可以包括例如且不限于任意数量的精简指令集计算机(RISC)核、直接存储器访问(DMA)和/或任意数量的向量处理器。

[0078] RISC核可以与图像传感器(例如本文描述的任何相机的图像传感器)、图像信号处理器和/或类似物交互。这些RISC核中的每一个可以包括任意数量的存储器。取决于实施例,RISC核可以使用若干协议中的任何协议。在一些示例中,RISC核可以执行实时操作系统(RTOS)。RISC核可以使用一个或更多个集成电路设备、专用集成电路(ASIC)和/或存储设备实现。例如,RISC核可以包括指令高速缓存和/或紧密耦合的RAM。

[0079] DMA可以使得PVA的组件能够独立于CPU 406访问系统存储器。DMA可以支持用来向PVA提供优化的任意数量的特征,包括但不限于支持多维寻址和/或循环寻址。在一些示例中,DMA可以支持高达六个或更多维度的寻址,其可以包括块宽度、块高度、块深度、水平块步进、竖直块步进和/或深度步进。

[0080] 向量处理器可以是可编程处理器,其可以被设计为高效且灵活地执行用于计算机视觉算法的编程并且提供信号处理能力。在一些示例中,PVA可以包括PVA核和两个向量处理子系统分区。PVA核可以包括处理器子系统、一个或更多个DMA引擎(例如两个DMA引擎)和/或其他外围设备。向量处理子系统可以作为PVA的主处理引擎而操作,并且可以包括向量处理单元(VPU)、指令高速缓存和/或向量存储器(例如VMEM)。VPU核可以包括数字信号处理器,诸如例如单指令多数据(SIMD)、超长指令字(VLIW)数字信号处理器。SIMD和VLIW的组合可以增强吞吐量和速率。

[0081] 向量处理器中的每一个可以包括指令高速缓存并且可以耦合到专用存储器。结果,在一些示例中,向量处理器中的每一个可以被配置为独立于其他向量处理器执行。在其他示例中,包括在特定PVA中的向量处理器可以被配置为采用数据并行化。例如,在一些实施例中,包括在单个PVA中的多个向量处理器可以执行相同的计算机视觉算法,但是在图像的不同区域上执行。在其他示例中,包括在特定PVA中的向量处理器可以在相同的图像上同时执行不同的计算机视觉算法,或者甚至在序列图像或者图像的部分上执行不同的算法。除其他的以外,任意数量的PVA可以包括在硬件加速集群中,并且任意数量的向量处理器可以包括在这些PVA中的每一个中。此外,PVA可以包括附加的纠错码(ECC)存储器,以增强总体系统安全性。

[0082] 加速器414(例如硬件加速集群)可以包括片上计算机视觉网络和SRAM,以提供用于加速器414的高带宽、低延迟SRAM。在一些示例中,片上存储器可以包括由例如且不限于八个现场可配置的存储器块组成的至少4MB SRAM,其可以由PVA和DLA二者访问。每对存储器块可以包括高级外围总线(APB)接口、配置电路系统、控制器和复用器。可以使用任何类型的存储器。PVA和DLA可以经由向PVA和DLA提供高速存储器访问的主干(backbone)访问存储器。主干可以包括(例如使用APB)将PVA和DLA互连到存储器的片上计算机视觉网络。

[0083] 片上计算机视觉网络可以包括在传输任何控制信号/地址/数据之前确定PVA和DLA二者都提供就绪且有效的信号的接口。这样的接口可以提供用于传输控制信号/地址/数据的单独相位和单独信道,以及用于连续数据传输的突发式通信。这种类型的接口可以符合ISO 26262或者IEC 61508标准,但是也可以使用其他标准和协议。

[0084] 在一些示例中,SoC 404可以包括例如在2018年8月10日提交的美国专利申请No.16/101,232中描述的实时光线追踪硬件加速器。该实时光线追踪硬件加速器可以用来快速且高效地确定(例如世界模型内的)对象的位置和范围,以便生成实时可视化仿真,以用于RADAR信号解释、用于声音传播合成和/或分析、用于SONAR系统仿真、用于一般波传播仿真、用于为了定位和/或其他功能的目的与LIDAR数据相比较和/或用于其他用途。在一些实施例中,一个或更多个树遍历单元(TTU)可用于执行一个或更多个光线追踪相关操作。

[0085] 加速器414(例如硬件加速器集群)具有广泛的自主驾驶用途。PVA可以是可编程视觉加速器,其可以用于ADAS和自主车辆中的关键处理阶段。PVA的能力是需要可预测处理、低功率和低延迟的算法域的良好匹配。换言之,PVA在半密集或者密集规则计算上,甚至在需要具有低延迟和低功率的可预测运行时间的小数据集上都表现良好。因此,在用于自主车辆的平台的背景下,PVA被设计为运行经典计算机视觉算法,因为它们在对象检测和整数数学运算方面很有效。

[0086] 例如,根据该技术的一个实施例,PVA用来执行计算机立体视觉。在一些示例中,可以使用基于半全局匹配的算法,但是这并不意图是限制性的。许多用于3-5级自主驾驶的应用都需要即时运动估计/立体匹配(例如来自运动的结构、行人识别、车道检测等等)。PVA可以在来自两个单目相机的输入上执行计算机立体视觉功能。

[0087] 在一些示例中,PVA可以用来执行密集的光流。例如,PVA可用于处理原始RADAR数据(例如,使用4D快速傅立叶变换)以在发射下一个RADAR脉冲之前提供处理后的RADAR信号。在其他示例中,PVA用于飞行时间深度处理,其例如通过处理原始飞行时间数据以提供经处理的飞行时间数据。

[0088] DLA可以用来运行任何类型的网络以增强控制和驾驶安全性,包括例如输出用于每个对象检测的置信度度量的神经网络。这样的置信度值可以解释为概率,或者解释为提供每个检测与其他检测相比的相对“权重”。该置信度值使得系统能够做出关于哪些检测应当被认为是真阳性检测而不是假阳性检测的进一步决策。例如,系统可以为置信度设置阈值,并且仅仅将超过阈值的检测看作真阳性检测。在自动紧急制动(AEB)系统中,假阳性检测会使得车辆自动地执行紧急制动,这显然是不希望的。因此,只有最确信的检测才应当被认为是AEB的触发因素。DLA可以运行用于回归置信度值的神经网络。该神经网络可以将至少一些参数子集作为其输入,例如边界框维度,(例如从另一个子系统)获得的地平面估计,与车辆400取向、距离相关的惯性测量单元(IMU)传感器466输出,从神经网络和/或其他传感器(例如LIDAR传感器464或RADAR传感器460)获得的对象的3D位置估计等。

[0089] SoC 404可以包括一个或更多个数据存储416(例如存储器)。数据存储416可以是SoC 404的片上存储器,其可以存储要在GPU和/或DLA上执行的神经网络。在一些示例中,为了冗余和安全,数据存储416可以容量足够大以存储神经网络的多个实例。数据存储412可以包括L2或L3高速缓存412。对数据存储416的引用可以包括对与如本文所描述的PVA、DLA和/或其他加速器414关联的存储器的引用。

[0090] SoC 404可以包括一个或更多个处理器410(例如嵌入式处理器)。处理器410可以包括启动和功率管理处理器,其可以是用于处理启动功率和管理功能以及有关安全实施的专用处理器和子系统。启动和功率管理处理器可以是SoC 404启动序列的一部分,并且可以提供运行时间功率管理服务。启动功率和管理处理器可以提供时钟和电压编程、辅助系统低功率状态转换、SoC 404热和温度传感器管理和/或SoC 404功率状态管理。每个温度传感器可以实现为环形振荡器,其输出频率与温度成比例,并且SoC 404可以使用环形振荡器检测CPU 406、GPU 408和/或加速器414的温度。如果确定温度超过阈值,那么启动和功率管理处理器可以进入温度故障例程并且将SoC 404置于较低功率状态和/或将车辆400置于司机安全停车模式(例如使车辆400安全停车)。

[0091] 处理器410可以进一步包括可以用作音频处理引擎的一组嵌入式处理器。音频处理引擎可以是一种音频子系统,其允许实现对于通过多个接口的多声道音频的完全硬件支持以及一系列广泛而灵活的音频I/O接口。在一些示例中,音频处理引擎是具有带有专用RAM的数字信号处理器的专用处理器核。

[0092] 处理器410可以进一步包括始终在处理器上的引擎,其可以提供必要的硬件特征以支持低功率传感器管理和唤醒用例。该始终在处理器上的引擎可以包括处理器核、紧密耦合的RAM、支持外围设备(例如定时器和中断控制器)、各种I/O控制器外围设备和路由逻辑。

[0093] 处理器410可以进一步包括安全集群引擎,其包括处理汽车应用的安全管理的专用处理器子系统。安全集群引擎可以包括两个或更多处理器核、紧密耦合的RAM、支持外围设备(例如定时器、中断控制器等等)和/或路由逻辑。在安全模式下,所述两个或更多核可以操作于锁步模式下,并且用作具有检测它们的操作之间的任何差异的比较逻辑的单核。

[0094] 处理器410可以进一步包括实时相机引擎,其可以包括用于处理实时相机管理的专用处理器子系统。

[0095] 处理器410可以进一步包括高动态范围信号处理器,其可以包括图像信号处理器,

该图像信号处理器是一种硬件引擎,该硬件引擎是相机处理管线的部分。

[0096] 处理器410可以包括可以是(例如微处理器上实现的)处理块的视频图像复合器,其实现视频回放应用程序生成用于播放器窗口的最终图像所需的视频后处理功能。视频图像复合器可以对广角相机470、环绕相机474和/或对驾驶室内监控相机传感器执行镜头畸变校正。驾驶室内监控相机传感器优选地由运行在高级SoC的另一个实例上的神经网络监控,被配置为识别驾驶室内事件并且相对应地做出响应。驾驶室内系统可以执行唇读,以激活移动电话服务并拨打电话、口述电子邮件、改变车辆目的地、激活或改变车辆的信息娱乐系统和设置或者提供语音激活的网上冲浪。某些功能仅在车辆操作于自主模式下时对于驾驶员可用,并且在其他情况下被禁用。

[0097] 视频图像复合器可以包括用于空间和时间降噪的增强时间降噪。例如,在视频中出现运动的情况下,降噪适当地对空间信息加权,降低邻近帧提供的信息的权重。在图像或者图像的部分不包括运动的情况下,视频图像复合器执行的时间降噪可以使用来自先前的图像的信息以降低当前图像中的噪声。

[0098] 视频图像复合器也可以被配置为对输入立体镜头帧执行立体校正。当操作系统桌面正在使用并且GPU 408无需连续地渲染(render)新的表面时,视频图像复合器可以进一步用于用户接口组成。甚至在GPU 408上电并且激活,主动执行3D渲染时,视频图像复合器可以用来减轻GPU 408的负担以提高性能和响应能力。

[0099] SoC 404可以进一步包括用于从相机接收视频和输入的移动行业处理器接口(MIPI)相机串行接口、高速接口和/或可以用于相机和有关像素输入功能的视频输入块。SoC 404可以进一步包括可以由软件控制并且可以用于接收未提交到特定角色的I/O信号的输入/输出控制器。SoC 404可以进一步包括大范围的外围设备接口,以使能与外围设备、音频编解码器、功率管理和/或其他设备通信。SoC 404可以用来处理来自(通过千兆多媒体串行链路和以太网连接的)相机、传感器(例如可以通过以太网连接的LIDAR传感器464、RADAR传感器460等等)的数据,来自总线402的数据(例如车辆400的速率、方向盘位置等等),来自(通过以太网或CAN总线连接的)GNSS传感器458的数据。SoC 404可以进一步包括专用高性能大容量存储控制器,其可以包括它们自己的DMA引擎,并且其可以用来从日常数据管理任务中释放CPU 406。

[0100] SoC 404可以是具有灵活架构的端到端平台,该架构跨越自动化3-5级,从而提供利用和高效使用计算机视觉和ADAS技术以实现多样性和冗余、连同深度学习工具一起提供用于灵活可靠驾驶软件堆栈的平台的综合功能安全架构。SoC 404可以比常规的系统更快、更可靠,甚至更加能量高效和空间高效。例如,当与CPU 406、GPU 408和数据存储45结合时,加速器414可以提供用于3-5级自主车辆的快速高效平台。

[0101] 因此该技术提供了不能通过常规系统实现的能力和功能。例如,计算机视觉算法可以在CPU上执行,这些CPU可以使用诸如C编程语言之类的高级编程语言配置为跨各种各样的视觉数据执行各种各样的处理算法。然而,CPU常常不能满足许多计算机视觉应用的性能要求,诸如与例如执行时间和功耗有关的那些要求。特别地,许多CPU不能实时地执行复杂的对象检测算法,这是车载ADAS应用的要求和实用3-5级自主车辆的要求。

[0102] 与常规系统形成对比的是,通过提供CPU复合体、GPU复合体和硬件加速集群,本文描述的技术允许同时和/或顺序地执行多个神经网络,并且将结果组合在一起以实现3-5级

自主驾驶功能。例如,在DLA或dGPU(例如GPU 420)上执行的CNN可以包括文本和单词识别,允许超级计算机读取和理解交通指示牌,包括尚未针对其特别地训练神经网络的指示牌。DLA可以进一步包括能够识别、解释和提供对指示牌的语义理解,并且将该语义理解传递给运行在CPU复合体上的路径规划模块的神经网络。

[0103] 作为另一个示例,如3、4或5级驾驶所需的,多个神经网络可以同时运行。例如,由"注意:闪烁的灯指示结冰条件"组成的警告指示牌连同电灯可以由若干神经网络独立地或者共同地进行解释。指示牌本身可以由部署的第一神经网络(例如经过训练的神经网络)识别为交通指示牌,文本"闪烁的灯指示结冰条件"可以由部署的第二神经网络解释,该部署的第二神经网络告知车辆的路径规划软件(优选地在CPU复合体上执行)当检测到闪烁的灯时,存在结冰条件。闪烁的灯可以通过在多个帧上操作部署的第三神经网络而识别,该神经网络告知车辆的路径规划软件闪烁的灯的存在(或不存在)。所有三个神经网络可以例如在DLA内和/或在GPU 408上同时运行。

[0104] 在一些示例中,用于面部识别和车主识别的CNN可以使用来自相机传感器的数据识别车辆400的授权的驾驶员和/或车主的存在。始终在传感器上的处理引擎可以用来在车主接近驾驶员车门时解锁车辆并且打开灯,并且在安全模式下,在车主离开车辆时禁用车辆。按照这种方式,SoC 404提供了防范盗窃和/或劫车的安全性。

[0105] 在另一个示例中,用于应急车辆检测和识别的CNN可以使用来自麦克风496的数据来检测并且识别应急车辆警报(siren)。与使用通用分类器检测警报并且手动地提取特征的常规系统形成对比的是,SoC 404使用CNN以对环境和城市声音分类以及对视觉数据分类。在优选的实施例中,运行在DLA上的CNN被训练为识别应急车辆的相对关闭速率(例如通过使用多普勒效应)。CNN也可以被训练为识别如GNSS传感器458所识别的特定于车辆在其中操作的局部区域的应急车辆。因此,例如,当在欧洲操作时,CNN将寻求检测欧洲警报,并且当在美国时,CNN将寻求识别仅仅北美的警报。一旦检测到应急车辆,在超声传感器462的辅助下,控制程序可以用来执行应急车辆安全例程,使车辆放慢速度,开到路边,停下车辆,和/或使车辆空转,直到应急车辆通过。

[0106] 车辆可以包括可以经由高速互连(例如PCIe)耦合到SoC 404的CPU 418(例如分立的CPU或dCPU)。CPU 418可以包括例如X86处理器。CPU 418可以用来执行各种各样的功能中的任何功能,包括例如仲裁ADAS传感器与SoC 404之间潜在地不一致的结果,和/或监控控制器436和/或信息娱乐SoC 430的状态和健康状况。

[0107] 车辆400可以包括可以经由高速互连(例如NVIDIA的NVLINK)耦合到SoC 404的GPU 420(例如分立的GPU或dGPU)。GPU 420可以例如通过执行冗余的和/或不同的神经网络而提供附加的人工智能功能,并且可以用来基于来自车辆400的传感器的输入(例如传感器数据)来训练和/或更新神经网络。

[0108] 车辆400可以进一步包括网络接口424,该网络接口可以包括一个或更多个无线天线426(例如用于不同通信协议的一个或更多个无线天线,例如蜂窝天线、蓝牙天线等等)。网络接口424可以用来使能通过因特网与云(例如与服务器478和/或其他网络设备)、与其他车辆和/或与计算设备(例如乘客的客户端设备)的无线连接。为了与其他车辆通信,可以在这两辆车之间建立直接链接,和/或可以建立间接链接(例如跨网络以及通过因特网)。直接链接可以使用车对车通信链路提供。车对车通信链路可以向车辆400提供关于接近车辆

400的车辆(例如车辆400前面、侧面和/或后面的车辆)的信息。该功能可以是车辆400的协作自适应巡航控制功能的部分。

[0109] 网络接口424可以包括提供调制和解调功能并且使得控制器436能够通过无线网络通信的SoC。网络接口424可以包括用于从基带到射频的上转换以及从射频到基带的下转换的射频前端。频率转换可以通过公知的过程执行,和/或可以使用超外差(super-heterodyne)过程执行。在一些示例中,射频前端功能可以由单独的芯片提供。网络接口可以包括用于通过LTE、WCDMA、UMTS、GSM、CDMA2000、蓝牙、蓝牙LE、Wi-Fi、Z波、ZigBee、LoRaWAN和/或其他无线协议通信的无线功能。

[0110] 车辆400可以进一步包括可包括片外(例如SoC 404外)存储装置的数据存储428。数据存储428可以包括一个或更多个存储元件,包括RAM、SRAM、DRAM、VRAM、闪存、硬盘和/或可以存储至少一个比特的数据的其他组件和/或设备。

[0111] 车辆400可以进一步包括GNSS传感器458(例如GPS和/或辅助GPS传感器),用于辅助映射、感知、占用网格生成和/或路径规划功能。可以使用任意数量的GNSS传感器458,包括例如且不限于使用带有以太网到串行(RS-232)网桥的USB连接器的GPS。

[0112] 车辆400可以进一步包括RADAR传感器460。RADAR传感器460可以甚至在黑暗和/或恶劣天气条件下也由车辆400用于远程车辆检测。RADAR功能安全级别可以是ASIL B。RADAR传感器460可以使用CAN和/或总线402(例如以传输RADAR传感器460生成的数据)以用于控制以及访问对象跟踪数据,在一些示例中接入以太网以访问原始数据。可以使用各种各样的RADAR传感器类型。例如且非限制性地,RADAR传感器460可以适合前面、后面和侧面RADAR使用。在一些示例中,使用脉冲多普勒RADAR传感器。

[0113] RADAR传感器460可以包括不同的配置,例如具有窄视场的远程、具有宽视场的短程、短程侧覆盖等等。在一些示例中,远程RADAR可以用于自适应巡航控制功能。远程RADAR系统可以提供通过两个或更多独立扫描实现的广阔视场(例如250m范围内)。RADAR传感器460可以帮助区分静态对象和运动对象,并且可以由ADAS系统用于紧急制动辅助和前方碰撞警告。远程RADAR传感器可以包括具有多根(例如六根或更多)固定RADAR天线以及高速CAN和FlexRay接口的单站多模RADAR。在具有六根天线的示例中,中央四根天线可以创建聚焦的波束图案,其被设计为在更高速率下以来自邻近车道的最小交通干扰记录车辆400的周围环境。其他两根天线可以扩展视场,使得快速地检测进入或离开车辆400的车道的车辆成为可能。

[0114] 作为一个示例,中程RADAR系统可以包括高达460m(前面)或80m(后面)的范围以及高达42度(前面)或450度(后面)的视场。短程RADAR系统可以包括但不限于被设计为安装在后保险杠两端的RADAR传感器。当安装在后保险杠两端时,这样的RADAR传感器系统可以创建持续地监控后方和车辆旁边的盲点的两个波束。

[0115] 短程RADAR系统可以在ADAS系统中用于盲点检测和/或变道辅助。

[0116] 车辆400可以进一步包括超声传感器462。可以置于车辆400的前面、后面和/或侧面的超声传感器462可以用于停车辅助和/或创建和更新占用网格。可以使用各种各样的超声传感器462,并且不同的超声传感器462可以用于不同的检测范围(例如2.5m、4m)。超声传感器462可以操作于功能安全级别的ASIL B。

[0117] 车辆400可以包括LIDAR传感器464。LIDAR传感器464可以用于对象和行人检测、紧

急制动、碰撞避免和/或其他功能。LIDAR传感器464可以为功能安全级别的ASIL B。在一些示例中,车辆400可以包括可以使用以太网(例如以将数据提供给千兆以太网交换机)的多个LIDAR传感器464(例如两个、四个、六个等等)。

[0118] 在一些示例中,LIDAR传感器464可能够对360度视场提供对象列表及其距离。商业上可用的LIDAR传感器464可以具有例如近似100m的广告范围,精度为2cm-3cm,支持100Mbps以太网连接。在一些示例中,可以使用一个或更多个非突出的LIDAR传感器464。在这样的示例中,LIDAR传感器464可以实现为可以嵌入到车辆400的前面、后面、侧面和/或拐角的小设备。在这样的示例中,LIDAR传感器464可以甚至对于低反射率对象提供高达120度水平的和35度竖直的视场,具有200m的范围。前面安装的LIDAR传感器464可以被配置用于45度与135度之间的水平视场。

[0119] 在一些示例中,也可以使用诸如3D闪光LIDAR之类的LIDAR技术。3D闪光LIDAR使用激光的闪光作为发射源,以照亮高达约200m的车辆周围环境。闪光LIDAR单元包括接受器,该接受器将激光脉冲传输时间和反射光记录在每个像素上,其进而与从车辆到对象的范围相对应。闪光LIDAR可以允许利用每个激光闪光生成周围环境的高度精确且无失真的图像。在一些示例中,可以部署四个闪光LIDAR传感器,车辆400的每一侧一个。可用的3D闪光LIDAR系统包括没有风扇以外的运动组件(moving part)的固态3D注视阵列LIDAR相机(例如非扫描LIDAR设备)。闪光LIDAR设备可以使用每帧5纳秒I类(眼睛安全)激光脉冲,并且可以以3D范围点云和共同寄存的强度数据的形式捕获反射的激光。通过使用闪光LIDAR,并且因为闪光LIDAR是没有运动组件的固态设备,LIDAR传感器464可以不太容易受到运动模糊、振动和/或震动的影响。

[0120] 该车辆可以进一步包括IMU传感器466。在一些示例中,IMU传感器466可以位于车辆400的后轴的中心。IMU传感器466可以包括例如且不限于加速度计、磁力计、陀螺仪、磁罗盘和/或其他传感器类型。在一些示例中,例如在六轴应用中,IMU传感器466可以包括加速度计和陀螺仪,而在九轴应用中,IMU传感器466可以包括加速度计、陀螺仪和磁力计。

[0121] 在一些实施例中,IMU传感器466可以实现为微型高性能GPS辅助惯性导航系统(GPS/INS),其结合微机电系统(MEMS)惯性传感器、高灵敏度GPS接收器和高级卡尔曼滤波算法以提供位置、速度和姿态的估计。这样,在一些示例中,IMU传感器466可以使得车辆400能够在无需来自磁传感器的输入的情况下通过直接观察从GPS到IMU传感器466的速度变化并且将其相关来估计方向(heading)。在一些示例中,IMU传感器466和GNSS传感器458可以结合到单个集成单元中。

[0122] 该车辆可以包括置于车辆400中和/或车辆400周围的麦克风496。除别的以外,麦克风496可以用于应急车辆检测和识别。

[0123] 该车辆可以进一步包括任意数量的相机类型,包括立体相机468、广角相机470、红外相机472、环绕相机474、远程和/或中程相机498和/或其他相机类型。这些相机可以用来捕获车辆400整个外围周围的图像数据。使用的相机类型取决于实施例和车辆400的要求,并且相机类型的任意组合可以用来提供车辆400周围的必要覆盖。此外,相机的数量可以根据实施例而不同。例如,该车辆可以包括六个相机、七个相机、十个相机、十二个相机和/或另一数量的相机。作为一个示例且非限制性地,这些相机可以支持千兆多媒体串行链路(GMSL)和/或千兆以太网。所述相机中的每一个在本文关于图4A和图4B更详细地进行了描

述。

[0124] 车辆400可以进一步包括振动传感器442。振动传感器442可以测量车辆的诸如车轴之类的组件的振动。例如,振动的变化可以指示道路表面的变化。在另一个示例中,当使用两个或更多振动传感器442时,振动之间的差异可以用来确定道路表面的摩擦或滑移(例如当动力驱动轴与自由旋转轴之间存在振动差异时)。

[0125] 车辆400可以包括ADAS系统438。在一些示例中,ADAS系统438可以包括SoC。ADAS系统438可以包括自主/自适应/自动巡航控制(ACC)、协作自适应巡航控制(CACC)、前方撞车警告(FCW)、自动紧急制动(AEB)、车道偏离警告(LDW)、车道保持辅助(LKA)、盲点警告(BSW)、后方穿越交通警告(RCTW)、碰撞警告系统(CWS)、车道居中(LC)和/或其他特征和功能。

[0126] ACC系统可以使用RADAR传感器460、LIDAR传感器464和/或相机。ACC系统可以包括纵向ACC和/或横向ACC。纵向ACC监控并控制到紧接在车辆400前方的车辆的距离,并且自动地调节车速以维持离前方车辆的安全距离。横向ACC执行距离保持,并且在必要时建议车辆400改变车道。横向ACC与诸如LC和CWS之类的其他ADAS应用程序有关。

[0127] CACC使用来自其他车辆的信息,该信息可以经由网络接口424和/或无线天线426经由无线链路或者通过网络连接(例如通过因特网)间接地从其他车辆接收。直接链接可以由车对车(V2V)通信链路提供,而间接链接可以是基础设施到车辆(I2V)的通信链路。通常,V2V通信概念提供关于紧接在前的车辆(例如紧接在车辆400前方且与其处于相同车道的车辆)的信息,而I2V通信概念提供关于前方更远处的交通的信息。CACC系统可以包括I2V和V2V信息源中的任一个或者二者。给定车辆400前方车辆的信息,CACC可以更加可靠,并且它有可能提高交通流的畅通性且降低道路拥堵。

[0128] FCW系统被设计为提醒驾驶员注意危险,使得驾驶员可以采取纠正措施。FCW系统使用耦合到专用处理器、DSP、FPGA和/或ASIC的前置相机和/或RADAR传感器460,该专用处理器、DSP、FPGA和/或ASIC电耦合至诸如显示器、扬声器和/或振动组件之类的驾驶员反馈。FCW系统可以提供例如声音、视觉警告、振动和/或快速制动脉冲形式的警告。

[0129] AEB系统检测即将发生的与另一车辆或其他对象的前方碰撞,并且可以在驾驶员在指定的时间或距离参数内没有采取纠正措施的情况下自动地应用制动器。AEB系统可以使用耦合到专用处理器、DSP、FPGA和/或ASIC的前置相机和/或RADAR传感器460。当AEB系统检测到危险时,它典型地首先提醒(alert)驾驶员采取纠正措施以避免碰撞,并且如果驾驶员没有采取纠正措施,那么AEB系统可以自动地应用制动器以努力防止或者至少减轻预测的碰撞的影响。AEB系统可以包括诸如动态制动支持和/或碰撞迫近制动之类的技术。

[0130] LDW系统提供了诸如方向盘或座位振动之类的视觉、听觉和/或触觉警告,以在车辆400穿过车道标记时提醒驾驶员。当驾驶员指示有意偏离车道时,通过激活转弯信号,不激活LDW系统。LDW系统可以使用耦合到专用处理器、DSP、FPGA和/或ASIC的前侧朝向相机,该专用处理器、DSP、FPGA和/或ASIC电耦合至诸如显示器、扬声器和/或振动组件之类的驾驶员反馈。

[0131] LKA系统是LDW系统的变型。如果车辆400开始离开车道,那么LKA系统提供纠正车辆400的转向输入或制动。BSW系统检测并向驾驶员警告汽车盲点中的车辆。BSW系统可以提供视觉、听觉和/或触觉警报以指示合并或改变车道是不安全的。系统可以在驾驶员使用转

弯信号时提供附加的警告。BSW系统可以使用耦合到专用处理器、DSP、FPGA和/或ASIC的后侧朝向相机和/或RADAR传感器460,该专用处理器、DSP、FPGA和/或ASIC电耦合至诸如显示器、扬声器和/或振动组件之类的驾驶员反馈。

[0132] RCTW系统可以在车辆400倒车时在后置相机范围之外检测到对象时提供视觉、听觉和/或触觉通知。一些RCTW系统包括AEB以确保应用车辆制动器以避免撞车。RCTW系统可以使用耦合到专用处理器、DSP、FPGA和/或ASIC的一个或更多个后置RADAR传感器460,该专用处理器、DSP、FPGA和/或ASIC电耦合至诸如显示器、扬声器和/或振动组件之类的驾驶员反馈。

[0133] 常规的ADAS系统可能易于出现假阳性结果,这可能会让驾驶员烦恼并分散注意力,但是典型地不是灾难性的,因为ADAS系统提醒驾驶员并且允许驾驶员决定安全条件是否真正存在并且相对应地采取行动。然而,在自主车辆400中,在冲突结果的情况下,车辆400本身必须决定是否注意(heed)来自主计算机或者辅助计算机(例如第一控制器436或第二控制器436)的结果。例如,在一些实施例中,ADAS系统438可以是用于向备用计算机合理性模块提供感知信息的备用和/或辅助计算机。备用计算机合理性监视器可以在硬件组件上运行冗余多样的软件,以检测感知和动态驾驶任务中的故障。来自ADAS系统438的输出可以提供给监督MCU。如果来自主计算机和辅助计算机的输出冲突,那么监督MCU必须确定如何协调该冲突以确保安全操作。

[0134] 在一些示例中,主计算机可以被配置为向监督MCU提供置信度评分,指示主计算机对所选结果的置信度。如果置信度评分超过阈值,那么监督MCU可以遵循主计算机的方向,而不管辅助计算机是否提供冲突或不一致的结果。在置信度评分不满足阈值的情况下并且在主计算机和辅助计算机指示不同的结果(例如冲突)的情况下,监督MCU可以在这些计算机之间进行仲裁以确定适当的结果。

[0135] 监督MCU可以被配置为运行神经网络,所述神经网络被训练并且被配置为基于来自主计算机和辅助计算机的输出,确定辅助计算机提供假警报的条件。因此,监督MCU中的神经网络可以了解何时可以信任辅助计算机的输出以及何时不能。例如,当辅助计算机为基于RADAR的FCW系统时,监督MCU中的神经网络可以了解FCW系统何时正在识别事实上不是危险的金属对象,例如触发警报的排水栅格或井盖。类似地,当辅助计算机是基于相机的LDW系统时,监督MCU中的神经网络可以学习在骑车者或行人在场并且车道偏离实际上是最安全的策略时无视该LDW。在包括运行在监督MCU上的神经网络的实施例中,监督MCU可以包括适合于利用关联的存储器运行神经网络的DLA或GPU中的至少一个。在优选的实施例中,监督MCU可以包括SoC 404的组件和/或作为SoC 404的组件而被包括。

[0136] 在其他示例中,ADAS系统438可以包括使用传统计算机视觉规则执行ADAS功能的辅助计算机。这样,辅助计算机可以使用经典的计算机视觉规则(如果-那么),并且在监督MCU中存在神经网络可以提高可靠性、安全性和性能。例如,多样化的实现方式和有意的非完全相同(non-identity)使得整个系统更加容错,对于软件(或者软件-硬件接口)功能造成的故障而言尤其如此。例如,如果在主计算机上运行的软件中存在软件漏洞或错误并且运行在辅助计算机上的非完全相同的软件代码提供相同的总体结果,那么监督MCU可以更加确信总体结果是正确的,并且由主计算机使用的软件或硬件中的漏洞不造成实质性的错误。

[0137] 在一些示例中,ADAS系统438的输出可以馈送至主计算机的感知块和/或主计算机的动态驾驶任务块。例如,如果ADAS系统438由于对象紧接在前的原因而指示前方碰撞警告,那么感知块可以在识别对象时使用该信息。在其他示例中,辅助计算机可以具有它自己的神经网络,其被训练并且因此如本文所描述的降低假阳性的风险。

[0138] 车辆400可以进一步包括信息娱乐SoC 430(例如车载信息娱乐系统(IVI))。尽管被图示和描述为SoC,但是信息娱乐系统可以不是SoC,并且可以包括两个或更多分立的组件。信息娱乐SoC 430可以包括可以用来向车辆400提供音频(例如音乐、个人数字助理、导航指令、新闻、广播等等)、视频(例如TV、电影、流媒体等等)、电话(例如免提呼叫)、网络连接(例如LTE、WiFi等等)和/或信息服务(例如导航系统,后停车援助,无线电数据系统,诸如燃油水平、覆盖的总距离、制动燃油水平、油位、车门开/关、空气过滤器信息之类的车辆有关信息,等等)的硬件和软件的组合。例如,信息娱乐SoC 430可以包括收音机、盘播放器、导航系统、视频播放器、USB和蓝牙连接、车载电脑、车载娱乐、WiFi、方向盘音频控件、免提语音控件、平视显示器(HUD)、HMI显示器434、远程信息处理设备、控制面板(例如用于控制各种组件、特征和/或系统,和/或与其交互)和/或其他组件。信息娱乐SoC 430可以进一步用来向车辆的用户提供信息(例如视觉的和/或听觉的),例如来自ADAS系统438的信息,诸如规划的车辆机动、轨迹、周围环境信息(例如交叉路口信息、车辆信息、道路信息等等)之类的自主驾驶信息,和/或其他信息。

[0139] 信息娱乐SoC 430可以包括GPU功能。信息娱乐SoC 430可以通过总线402(例如CAN总线、以太网等)与车辆400的其他设备、系统和/或组件通信。在一些示例中,信息娱乐SoC 430可以耦合至监督MCU,使得在主控制器436(例如车辆400的主和/或备用计算机)出现故障的情况下,信息娱乐系统的GPU可以执行一些自驾驶功能。在这样的示例中,信息娱乐SoC 430可以如本文所描述的将车辆400置于司机安全停车模式。

[0140] 车辆400可以进一步包括仪表组432(例如数字仪表板、电子仪表组、数字仪表面板等等)。仪表组432可以包括控制器和/或超级计算机(例如分立的控制器或超级计算机)。仪表组432可以包括一套仪器,例如车速表、燃油水平、油压、转速表、里程表、转弯指示器、换档位置指示器、安全带警告灯、停车制动警告灯、发动机故障灯、安全气囊(SRS)系统信息、照明控件、安全系统控件、导航信息等等。在一些示例中,信息可以被显示和/或在信息娱乐SoC 430和仪表组432之间共享。换言之,仪表组432可以作为信息娱乐SoC 430的部分而被包括,或者反之亦然。

[0141] 图4D为根据本公开一些实施例的基于云的服务器与图4A的示例自主车辆400之间的通信的系统示意图。系统476可以包括服务器478、网络490以及包括车辆400在内的车辆。服务器478可以包括多个GPU 484(A)-1284(H)(这里统称为GPU 484)、PCIe交换机482(A)-482(H)(这里统称为PCIe交换机482)和/或CPU 480(A)-480(B)(这里统称为CPU 480)。GPU 484、CPU 480和PCIe交换机可以与诸如例如且不限于NVIDIA开发的NVLink接口488之类的高速互连和/或PCIe连接486互连。在一些示例中,GPU 484经由NVLink和/或NVSwitch SoC连接,并且GPU 484和PCIe交换机482经由PCIe互连连接。尽管图示出八个GPU 484、两个CPU 480和两个PCIe交换机,但是这并不意图是限制性的。取决于实施例,服务器478中的每一个可以包括任意数量的GPU 484、CPU 480和/或PCIe交换机。例如,服务器478中的每一个可以包括八个、十六个、三十二个和/或更多GPU 484。

[0142] 服务器478可以通过网络490并且从车辆接收图像数据,该图像数据表示示出诸如最近开始的道路工程之类的意外或改变的道路状况的图像。服务器478可以通过网络490并且向车辆传输神经网络492、更新的神经网络492和/或地图信息494,包括关于交通和道路状况的信息。对地图信息494的更新可以包括对于HD地图422的更新,例如关于建筑工地、坑洼、弯道、洪水或其他障碍物的信息。在一些示例中,神经网络492、更新的神经网络492和/或地图信息494可以已从新的训练和/或从环境中的任意数量的车辆接收的数据中表示和/或基于数据中心处执行的训练(例如使用服务器478和/或其他服务器)的经验生成。

[0143] 服务器478可以用来基于训练数据训练机器学习模型(例如神经网络)。训练数据可以由车辆生成,和/或可以在仿真中生成(例如使用游戏引擎)。在一些示例中,训练数据被标记(例如在神经网络受益于有监督学习的情况下)和/或经历其他预处理,而在其他示例中,训练数据不被标记和/或预处理(例如在神经网络无需有监督学习的情况下)。训练可以根据任何一类或更多类机器学习技术进行,包括但不限于以下类:监督训练、半监督训练、无监督训练、自学习、强化学习、联合学习、迁移学习、特征学习(包括主组件和集群分析)、多线性子空间学习、流形学习、表示学习(包括备用字典学习)、基于规则的机器学习、异常检测及其任何变体或组合。一旦机器学习模型被训练,机器学习模型可以由车辆使用(例如通过网络490传输至车辆),和/或机器学习模型可以由服务器478用来远程地监控车辆。

[0144] 在一些示例中,服务器478可以接收来自车辆的数据,并且将该数据应用到最新的实时神经网络以用于实时智能推理。服务器478可以包括由GPU 484供电的深度学习超级计算机和/或专用AI计算机,例如NVIDIA开发的DGX和DGX站机器。然而,在一些示例中,服务器478可以包括仅使用CPU供电的数据中心的深度学习基础设施。

[0145] 服务器478的深度学习基础设施可能够快速实时推理,并且可以使用该能力来评估并验证车辆400中的处理器、软件和/或关联硬件的健康状况。例如,深度学习基础设施可以接收来自车辆400的定期更新,例如图像序列和/或车辆400已经定位的位于该图像序列中的对象(例如经由计算机视觉和/或其他机器学习对象分类技术)。深度学习基础设施可以运行它自己的神经网络以识别对象并且将它们与车辆400识别的对象相比较,如果结果不匹配并且该基础设施得出车辆400中的AI发生故障的结论,那么服务器478可以向车辆400传输信号,指示车辆400的故障保护计算机进行控制,通知乘客,并且完成安全停车操作。

[0146] 为了推理,服务器478可以包括GPU 484和一个或更多个可编程推理加速器(例如NVIDIA的TensorRT 3)。GPU供电的服务器和推理加速的组合可以使得实时响应成为可能。在其他示例中,例如在性能不那么重要的情况下,CPU、FPGA和其他处理器供电的服务器可以用于推理。

[0147] 图5为适合用于实现本公开一些实施例的示例计算设备500的框图。计算设备500可以包括直接或间接耦合下列设备的互连系统502:存储器504,一个或更多个中央处理单元(CPU)506,一个或更多个图形处理单元(GPU)508,通信接口510,I/O端口512,输入/输出组件514,电源516,一个或更多个呈现组件518(例如显示器),以及一个或更多个逻辑单元520。

[0148] 尽管图5的各个框被示为经由具有线路的互连系统502连接,但是这并不意图是限

制性的,并且仅仅为了清楚起见。例如,在一些实施例中,诸如显示设备之类的呈现组件518可以被认为是I/O组件514(例如如果显示器为触摸屏)。作为另一个示例,CPU 506和/或GPU 508可以包括存储器(例如,存储器504可以表示除了GPU 508、CPU 506和/或其他组件的存储器以外的存储设备)。换言之,图5的计算设备仅仅是说明性的。在诸如“工作站”、“服务器”、“膝上型电脑”、“台式机”、“平板电脑”、“客户端设备”、“移动设备”、“手持式设备”、“游戏控制台”、“电子控制单元(ECU)”、“虚拟现实系统”、“增强现实系统”和/或其他设备或系统类型之类的类别之间不进行区分,因为所有这些都被考虑在图5的计算设备的范围内。

[0149] 互连系统502可以表示一条或更多条链路或总线,例如地址总线、数据总线、控制总线或者其组合。互连系统502可以包括一种或更多种总线类型,例如行业标准架构(ISA)总线、扩展行业标准架构(EISA)总线、视频电子标准协会(VESA)总线、外围组件互连(PCI)总线、外围组件互连快速(PCIe)总线,和/或另一种类型的总线或链路。在一些实施例中,组件之间存在直接连接。作为示例,CPU 506可以直接连接到存储器504。此外,CPU 506可以直接连接到GPU 508。在组件之间存在直接或点对点连接的情况下,互连系统502可以包括PCIe链路以执行连接。在这些示例中,计算设备500中不需要包括PCI总线。

[0150] 存储器504可以包括各种各样的计算机可读介质中的任何介质。计算机可读介质可以是可以由计算设备500访问的任何可用介质。计算机可读介质可以包括易失性和非易失性介质以及可移除和不可移除介质。举例而言且非限制性地,计算机可读介质可以包括计算机存储介质和通信介质。

[0151] 计算机存储介质可以包括易失性和非易失性介质和/或可移除和不可移除介质,其以用于存储诸如计算机可读指令、数据结构、程序模块和/或其他数据类型之类的信息的任何方法或技术实现。例如,存储器504可以存储计算机可读指令(例如其表示程序和/或程序元素,例如操作系统)。计算机存储介质可以包括但不限于RAM、ROM、EEPROM、闪存或者其他存储技术、CD-ROM、数字多功能盘(DVD)或其他光盘存储装置、磁带盒、磁带、磁盘存储装置或其他磁存储设备,或者可以用来存储期望的信息且可以由计算设备500访问的任何其他介质。当在本文使用时,计算机存储介质并不包括信号本身。

[0152] 计算机存储介质可以在诸如载波之类的调制数据信号或其他传输机制中包含计算机可读指令、数据结构、程序模块和/或其他数据类型,并且包括任何信息输送介质。术语“调制数据信号”可以指这样的信号,该信号使它的特性中的一个或更多个以这样的将信息编码到该信号中的方式设置或改变。举例而言且非限制性地,计算机存储介质可以包括诸如有线网络或直接有线连接之类的有线介质,以及诸如声音、RF、红外和其他无线介质之类的无线介质。任何以上所述的组合也应当包含在计算机可读介质的范围内。

[0153] CPU 506可以被配置为执行至少一些计算机可读指令以便控制计算设备500的一个或更多个组件执行本文描述的方法和/或过程中的一个或更多个。CPU 506中的每一个可以包括能够同时处理大量软件线程的一个或更多个核(例如一个、两个、四个、八个、二十八个、七十二个等等)。CPU 506可以包括任何类型的处理器,并且可以包括不同类型的处理器,这取决于实现的计算设备500的类型(例如具有用于移动设备的较少核的处理器以及具有用于服务器的更多核的处理器)。例如,取决于计算设备500的类型,处理器可以是使用精简指令集计算(RISC)实现的高级RISC机器(ARM)处理器或者使用复杂指令集计算(CISC)实现的x86处理器。除了一个或更多个微处理器或者诸如数学协处理器之类的补充协处理器

之外,计算设备500还可以包括一个或更多个CPU 506。

[0154] 除了或替代CPU 506,GPU 508可以被配置为执行至少一些计算机可读指令以控制计算设备500的一个或更多个组件来执行本文描述的一种或更多种方法和/或过程。一个或更多个GPU 508可以是集成GPU(例如,具有一个或更多个CPU 506和/或一个或更多个GPU 508可以是分立的GPU。在实施例中,一个或更多个GPU 508可以是一个或更多个CPU 506的协处理器。计算设备500可以使用GPU 508来渲染图形(例如,3D图形)或执行通用计算。例如,GPU 508可用于GPU上的通用计算(GPGPU)。GPU 508可以包括能够同时处理成百上千个软件线程的成百上千个核心。GPU 508可以响应于渲染命令(例如,通过主机接口从CPU 506接收的渲染命令)生成用于输出图像的像素数据。GPU 508可以包括图形存储器,例如显示存储器,用于存储像素数据或任何其他合适的数据,例如GPGPU数据。显示存储器可以被包括作为存储器504的一部分。GPU 508可以包括并行操作(例如,经由链路)的两个或更多个GPU。链路可以直接连接GPU(例如,使用NVLINK),也可以通过交换机(例如,使用NVSwitch)连接GPU。当组合在一起时,每个GPU 508可以为输出的不同部分或不同的输出(例如,用于第一图像的第一GPU和用于第二图像的第二GPU)生成像素数据或GPGPU数据。每个GPU可包括自己的存储器,也可与其他GPU共享存储器。

[0155] 除了或替代CPU 506和/或GPU 508,逻辑单元520可以被配置为执行至少一些计算机可读指令以控制计算设备500的一个或更多个组件来执行本文所述的一种或更多种方法和/或过程。在实施例中,CPU 506、GPU 508和/或逻辑单元520可以分立地或联合地执行方法、过程和/或其部分的任何组合。一个或更多个逻辑单元520可以是CPU 506和/或GPU 508中的一个或更多个的一部分和/或与CPU 506和/或GPU 508中的一个或更多个集成,和/或一个或更多个逻辑单元520可以是分立的组件或以其他方式在CPU 506和/或GPU 508的外部。在实施例中,一个或更多个逻辑单元520可以是一个或更多个CPU 506和/或一个或更多个GPU 508的协处理器。

[0156] 逻辑单元520的示例包括一个或更多个处理核心和/或其组件,例如张量核心(TC)、张量处理单元(TPU)、像素视觉核心(PVC)、视觉处理单元(VPU)、图形处理集群(GPC)、纹理处理集群(TPC)、流式多处理器(SM)、树遍历单元(TTU)、人工智能加速器(AIA)、深度学习加速器(DLA)、算术逻辑单元(ALU))、专用集成电路(ASIC)、浮点单元(FPU)、I/O元件、外围组件互连(PCI)或外围组件互连快速(PCIe)元件等。

[0157] 通信接口510可以包括一个或更多个接收器、发送器和/或收发器,其使得计算设备500能够经由电子通信网络与其他计算设备通信,包括有线和/或无线通信。通信接口510可以包括使能通过若干不同网络中的任何网络进行通信的组件和功能,所述网络例如无线网络(例如Wi-Fi、Z波、蓝牙、蓝牙LE、ZigBee等等)、有线网络(例如通过以太网或无限带宽通信)、低功率广域网(例如LoRaWAN、SigFox等等)和/或因特网。

[0158] I/O端口512可以使得计算设备500能够逻辑地耦合到包括I/O组件514、呈现组件518和/或其他组件在内的其他设备,其中一些可以内置到(例如集成到)计算设备500中。说明性I/O组件514包括麦克风、鼠标、键盘、操纵杆、游戏垫、游戏控制器、碟形卫星天线、扫描仪、打印机、无线设备等等。I/O组件514可以提供处理用户生成的空中手势、语音或其他生理输入的自然用户接口(NUI)。在一些实例中,输入可以传输至适当的网络元件以便进一步处理。NUI可以实现语音识别、手写笔识别、面部识别、生物特征识别、屏幕上和邻近屏幕的

手势识别、空中手势、头部和眼睛跟踪以及与计算设备500的显示器关联的触摸识别(如下文更详细地描述的)的任意组合。计算设备500可以包括诸如立体相机系统之类的深度相机、红外相机系统、RGB相机系统、触摸屏技术以及这些的组合,以用于手势检测和识别。此外,计算设备500可以包括使能运动检测的加速度计或陀螺仪(例如作为惯性测量单元(IMU)的部分)。在一些示例中,加速度计或陀螺仪的输出可以由计算设备500用来渲染沉浸式增强现实或者虚拟现实。

[0159] 电源516可以包括硬接线电源、电池电源或者其组合。电源516可以向计算设备500供电以使得计算设备500的组件能够操作。呈现组件518可以包括显示器(例如监视器、触摸屏、电视屏幕、平视显示器(HUD)、其他显示器类型或者其组合)、扬声器和/或其他呈现组件。呈现组件518可以接收来自其他组件(例如GPU 508、CPU 506等等)的数据,并且输出该数据(例如作为图像、视频、声音等等)。

[0160] 本公开可以在由计算机或者诸如个人数字助理或其他手持式设备之类的其他机器执行的、包括诸如程序模块之类的计算机可执行指令的机器可使用指令或者计算机代码的一般背景下进行描述。通常,包括例程、程序、对象、组件、数据结构等等的程序模块指的是执行特定任务或者实现特定抽象数据类型的代码。本公开可以在各种各样的系统配置中实践,这些配置包括手持式设备、消费电子器件、通用计算机、更专业的计算设备等等。本公开也可以在其中任务由通过通信网络链接的远程处理设备执行的分布式计算环境中实践。

[0161] 如在本文中使用的,“和/或”关于两个或更多元素的叙述应当解释为仅指一个元素或者元素组合。例如,“元素A、元素B和/或元素C”可以包括仅仅元素A,仅仅元素B,仅仅元素C,元素A和元素B,元素A和元素C,元素B和元素C,或者元素A、B和C。此外,“元素A或元素B中的至少一个”可以包括元素A中的至少一个,元素B中的至少一个,或者元素A中的至少一个和元素B中的至少一个。进一步,“元素A和元素B中的至少一个”可以包括元素A中的至少一个,元素B中的至少一个,或者元素A中的至少一个和元素B中的至少一个。

[0162] 这里详细地描述了本公开的主题以满足法定要求。然而,描述本身并非意在限制本公开的范围。相反地,本发明人已经设想到,要求保护的主题也可以以其他的方式具体化,以包括与本文中结合其他当前或未来技术描述的步骤不同的步骤或者相似的步骤的组合。而且,尽管术语“步骤”和/或“块”在本文中可以用来隐含采用的方法的不同元素,但是这些术语不应当被解释为暗示本文公开的各个步骤之中或之间的任何特定顺序,除非明确描述了各步骤的顺序。

[0163] 下面结合图6描述的方法的每个块包括可以使用硬件、固件和/或软件的任何组合来执行的计算过程。例如,各种功能可以由执行存储在存储器中的指令的处理器执行。所述方法还可以实现为存储在计算机存储介质上的计算机可用指令。这些方法可以由独立应用程序、服务或托管服务(独立或与另一个托管服务组合)或另一个产品的插件提供。此外,作为示例,针对图4A-4D的示例性自主车辆系统描述了图6的方法。然而,这些方法可附加地或替代地由任何一个系统或系统的任何组合执行,包括但不限于本文所述的那些系统。

[0164] 图6示出了根据本公开的实施例的用于在存在眩光的情况下确定注视方向的过程步骤的流程图。图6的处理在计算设备300接收到相机310拍摄的对象的图像时开始(步骤600)。计算设备然后在接收到的图像中识别对象的面部和眼睛(步骤610)。可使用任何方法或过程在图像内定位对象面部,包括不使用神经网络检测面部的已知的基于计算机视觉的

面部检测过程,例如边缘检测方法、特征搜索方法、概率面部模型、图形匹配、定向梯度直方图(HOG),其被馈送至输入分类器,如支持向量机、HaarCascade分类器等。还可以使用基于神经网络的面部识别方法来确定面部位置,例如采用深度神经网络(DNN)面部识别方案的方法以及任何其他方法。本公开的实施例还考虑了来自同一图像的对象眼睛的位置。眼睛定位可以以任何方式执行,例如通过已知的基于计算机视觉的眼睛检测过程,包括任何上述非基于神经网络的技术、基于神经网络的眼睛识别方法等。

[0165] 一旦对象的面部和眼睛位于接收到的图像内,计算设备300从接收到的图像中提取眩光点并生成面部表示(步骤620)。可使用任何方法或过程识别眩光。例如,眩光像素可以被识别为接收到的图像中亮度值高于预定阈值的那些像素。作为另一示例,可以根据捕获图像中的已知或特征眩光形状(例如,圆形或星形斑点等)来识别眩光像素。例如,可以通过将眩光像素设置为白色或饱和像素,并将剩余像素设置为黑色像素来形成眩光掩码。

[0166] 可使用任何方法或过程确定面部表示,例如,已识别面部的二值掩码。例如,可以通过将与所识别的面部相对应的像素设置为白色或另一均匀颜色,而剩余像素设置为黑色像素来形成这样的二值掩码。

[0167] 然后选择眼睛裁剪(步骤630),例如通过以已知方式在步骤610中识别的每只眼睛周围绘制边界框,并相应地裁剪图像。类似地,然后计算面网格(步骤640)。面部网格225可以例如通过根据可以以任何已知方式绘制的边界框裁剪所识别的面部并将裁剪后的图像投影到预定的粗网格上来形成,其中包含分配一个或更多个面部像素的网格元素颜色(或第一个值),以及不包含设置为黑色(或与第一个值不同的值)的面部像素的网格元素。

[0168] 然后,将步骤620的掩码、步骤630的眼睛裁剪和步骤640的面部网格输入到本公开的实施例的机器学习模型(例如,特征提取层230、235和240,以及全连接层245和255),以如上所述计算注视方向(步骤650)。然后将该注视方向输出到注视辅助系统320(步骤660),其可基于其接收到的注视方向执行任何一个或更多个动作(步骤670)。

[0169] 本公开的实施例考虑以除图2中所示的方式以外的其他方式配置的机器学习模型。图7是一个这样的替代示例的框图表示。这里,图7的机器学习模型包括全连接层710、740和755,以及特征提取层715、745和750。特征提取层745和750以及全连接层755具有左眼表示725、右眼表示730和面部表示735作为输入。图7的模型的最右边的三个分支基本上与图2的分支相同。也就是说,特征提取层745和750以及全连接层755可以分别与特征提取层235和240以及全连接层245基本相同。类似地,输入725、730和735可分别与输入215、220和225基本相同。

[0170] 眩光表示700和面部表示705可以与图2的输入200和205基本相同。因此,眩光表示700可以是在输入图像100中识别的眩光点的二值掩码,并且面部表示705可以是从输入图像100提取的面部像素的二值掩码。然而,与图2不同,眩光表示700被输入到全连接层710,并且面部表示705被输入到特征提取层715。全连接层710输出描述输入图像100中眩光的空间位置的特征,并且特征提取层715输出描述输入图像100中脸部的取向的特征。这两个输出通过联结块720联结,并输入到全连接层740,其输出描述输入图像100中面部取向的一组特征。然后,如上文结合图2所述,处理全连接层740、特征提取层745和750以及全连接层755的输出。具体地,输出通过联结块760联结并输入到输出注视方向的全连接层765。通过这种方式,眩光的空间位置用于确定面部方向,而不是眩光点本身。

[0171] 本公开实施例的任何全连接层和特征提取层的训练可以以任何合适的方式进行。在一些实施例中,本公开实施例的机器学习模型,例如,图2和7,可以进行端到端的训练。在某些实施例中,本公开实施例的全连接层和特征提取层可以使用其注视方向已知的对象的标记输入图像以有监督的方式进行训练。在一些实施例中,可将眩光点添加到这些输入图像以模拟眩光的存在。可以以任何方式将添加的眩光点添加到其各自的图像中,例如通过在各种图像中随机放置、在输入图像中的预定位置放置等。

[0172] 本领域普通技术人员将注意到,本公开的实施例不限于注视的估计。具体而言,本公开的实施例的方法和系统可用于在确定任何动眼神经变量时考虑眩光。例如,图2和/或图7的机器学习模型可以被构造和训练,使得FC层255、765的输出是除注视之外的任何其他动眼神经变量的估计,例如瞳孔大小、瞳孔运动、认知负荷、疲劳、例如由于药物或酒精导致的丧失能力、通过虹膜的人或物体识别、任何其他生理反应等。这可以通过例如使用标记有感兴趣的任何动眼神经变量的值的输入数据集(例如,眩光掩码、掩码、眼睛修剪、面部网格等)来训练本公开实施例的机器学习模型来实现。机器学习模型也可以被构造来提取在确定这种动眼神经变量时有用的任何特征。以这种方式,可以构造和训练本公开实施例的机器学习模型,以便以抗眩光的方式估计任何动眼神经变量,即,在确定任何动眼神经变量时考虑眩光。

[0173] 为了解释的目的,上述描述使用了特定的术语来提供对本公开的透彻理解。然而,对于本领域技术人员来说,实施本发明的方法和系统不需要具体细节是显而易见的。因此,出于说明和描述的目的,给出了本发明特定实施例的前述描述。它们并非旨在详尽无遗或将本发明限制于所公开的精确形式。鉴于上述教导,许多修改和变化是可能的。例如,眩光点可以以任何方式输入到各种机器学习模型,例如通过输入眩光点的二元掩码、输入表格或眩光位置集或眩光的任何其他表示。选择和描述实施例是为了最好地解释本发明的原理及其实际应用,从而使本领域的其他技术人员能够最好地利用本公开的方法和系统以及经过各种修改的各种实施例,以适合预期的特定用途。此外,可以混合和匹配或以其他方式组合所公开的或以其他方式公开的各种实施例的不同特征,以创建本发明所设想的其他实施例。

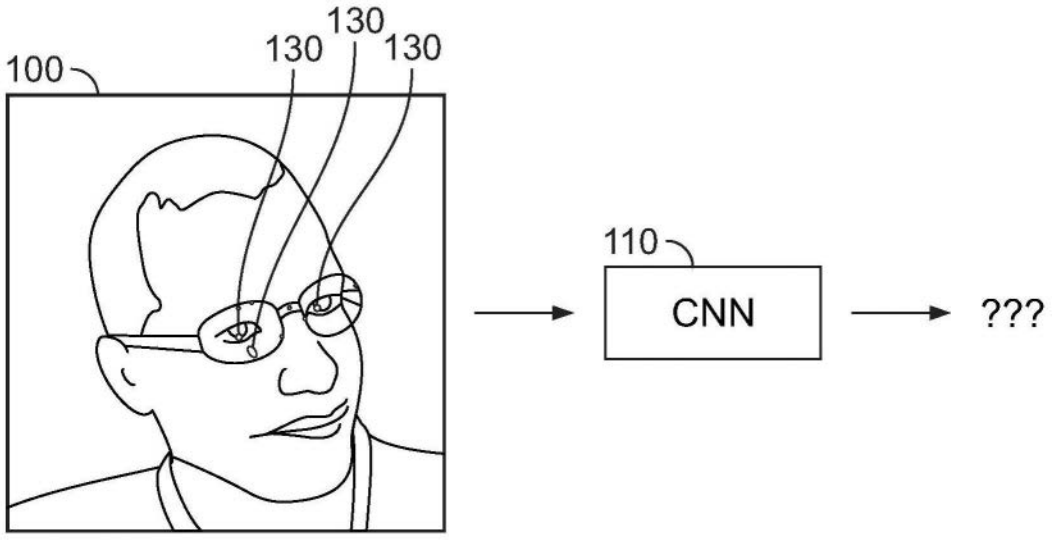


图1A(现有技术)

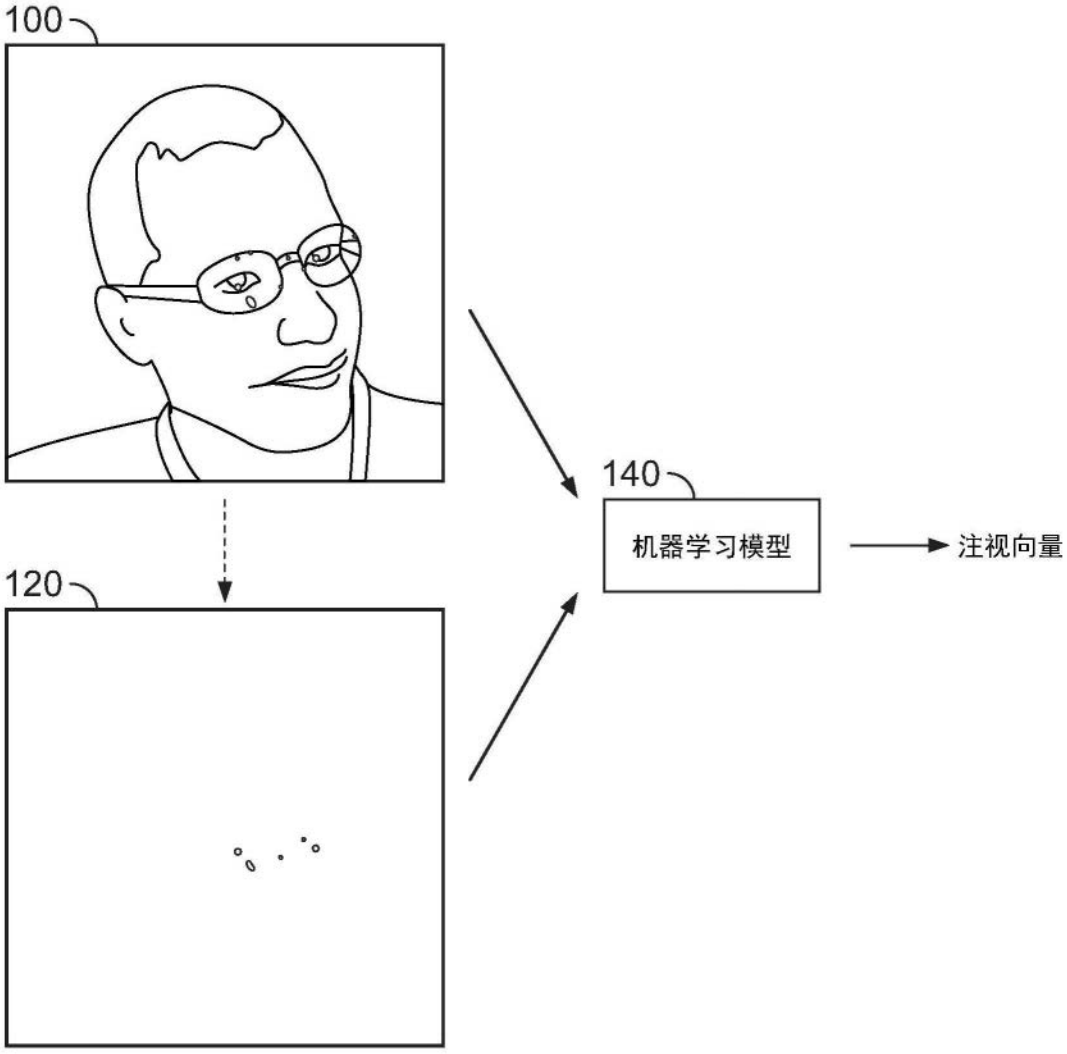


图1B

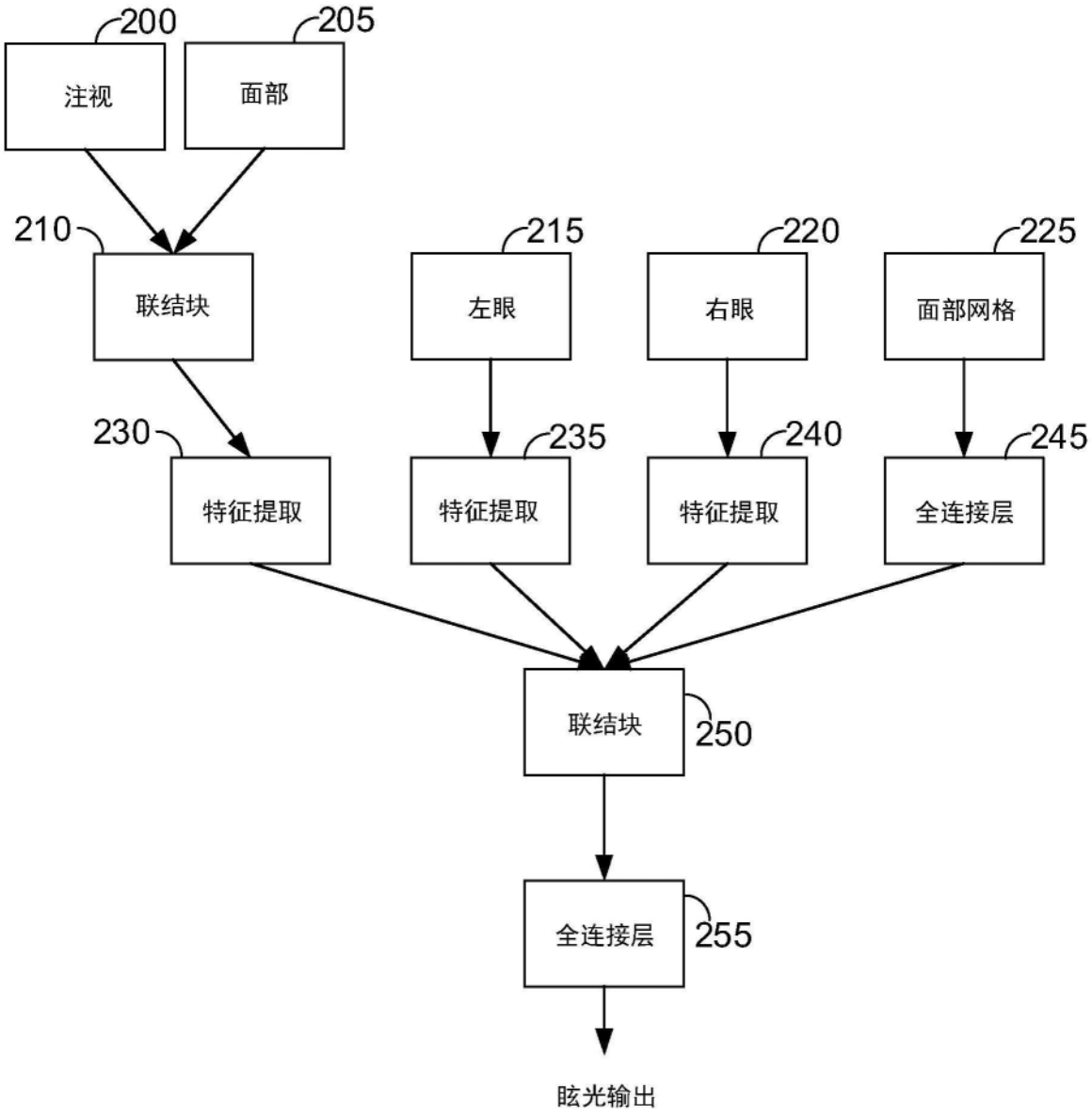

200
205
注视
面部
210
215
220
225
联结块
左眼
右眼
面部网格
230
235
240
245
特征提取
特征提取
特征提取
全连接层
联结块
250
全连接层
255
眩光输出

图2


310
300
320
相机
计算设备
注视辅助系统

图3

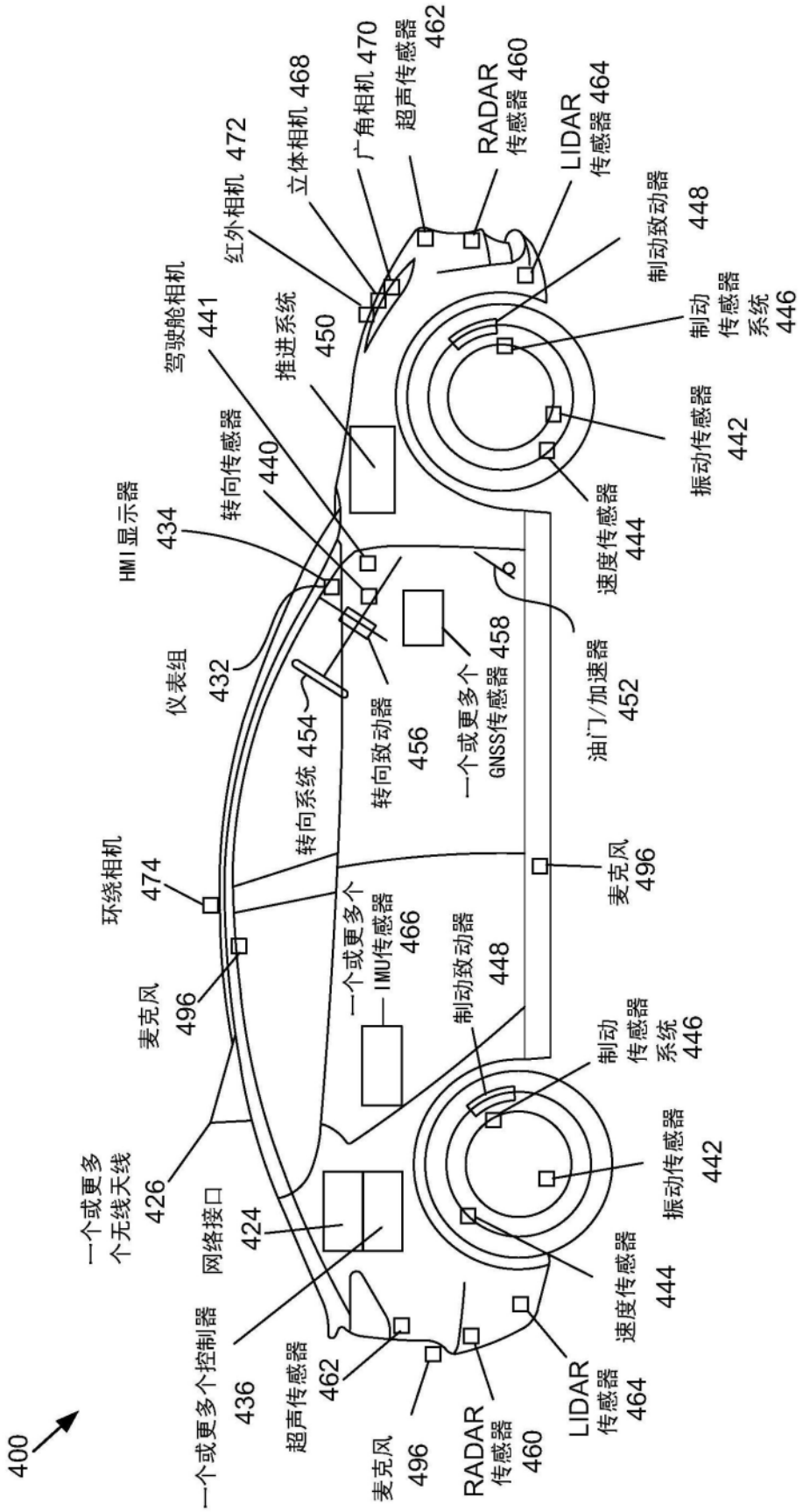

400
一个或更多个控制器 436
一个或更多个无线天线 426
网络接口 424
超声传感器 462
麦克风 496
RADAR 传感器 460
LIDAR 传感器 464
速度传感器 444
振动传感器 442
制动传感器系统 446
制动致动器 448
麦克风 496
环绕相机 474
一个或更多个 IMU传感器 466
转向系统 454
转向致动器 456
仪表组 432
HMI显示器 434
转向传感器 440
驾驶舱相机 441
推进系统 450
红外相机 472
立体相机 468
广角相机 470
超声传感器 462
RADAR 传感器 460
LIDAR 传感器 464
制动致动器 448
制动传感器系统 446
振动传感器 442
速度传感器 444
一个或更多个 GNSS传感器 458
油门/加速器 452
麦克风 496

图4A

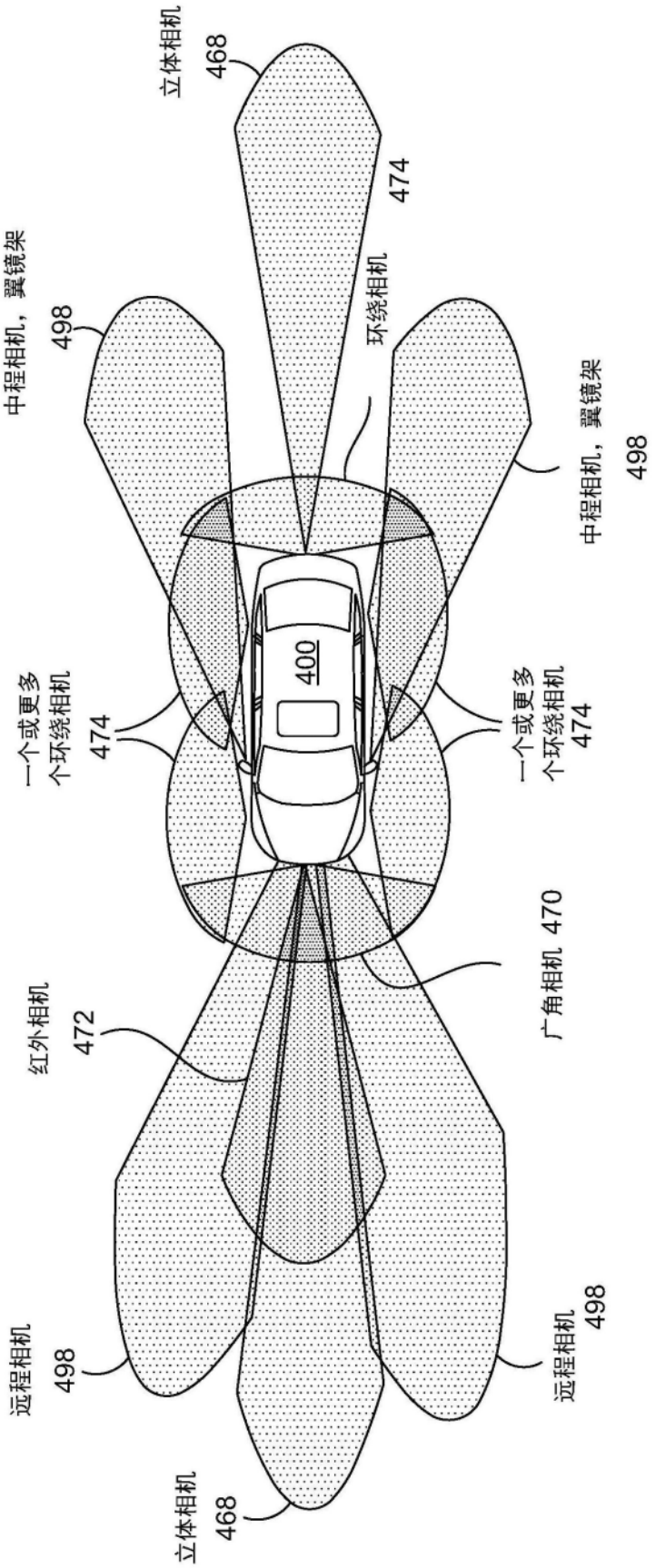

立体相机
468
中程相机，翼镜架
498
环绕相机
474
中程相机，翼镜架
498
一个或更多
个环绕相机
474
400
一个或更多
个环绕相机
474
红外相机
472
广角相机 470
远程相机
498
远程相机
498
立体相机
468

图4B

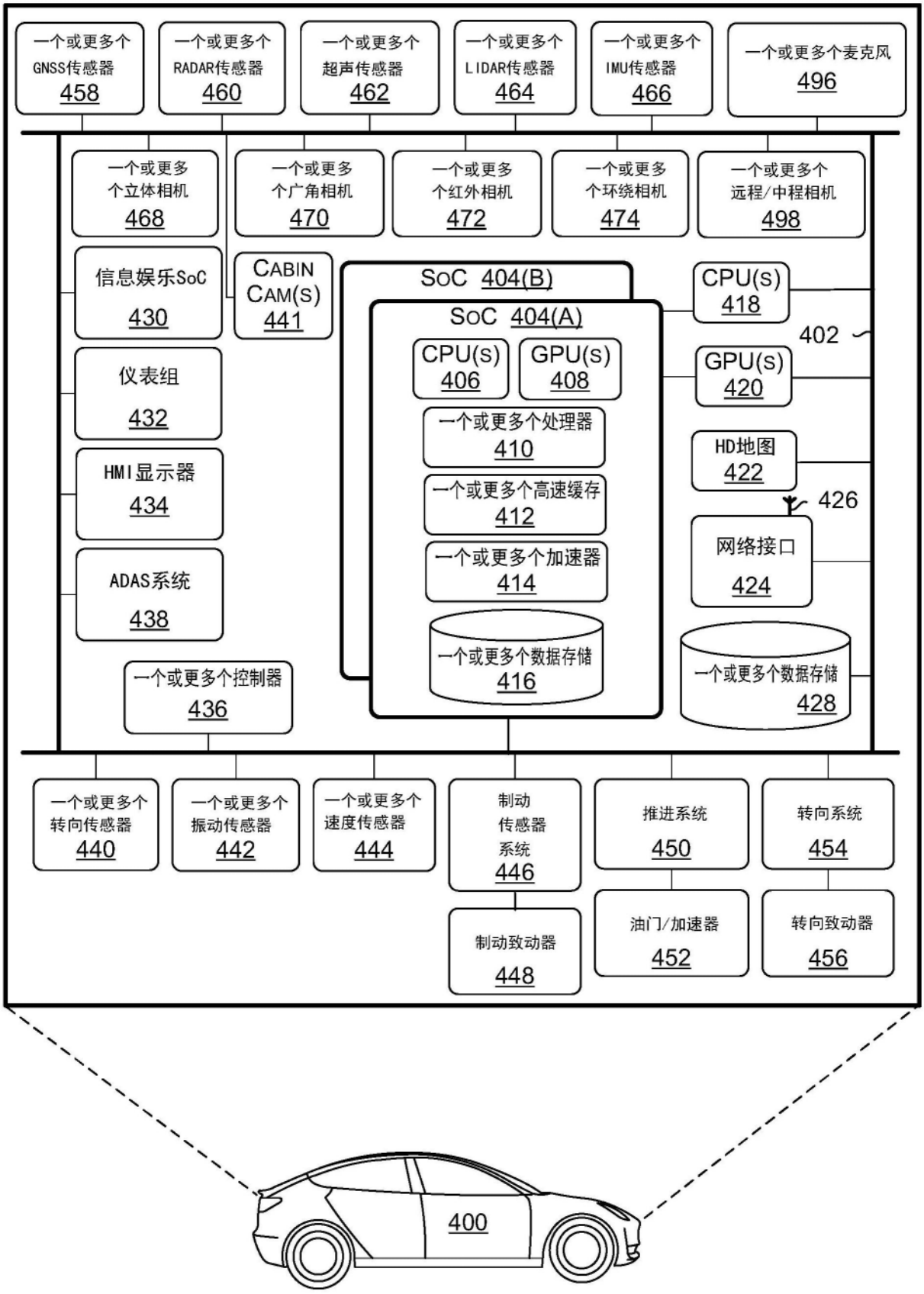

一个或更多个 GNSS传感器 458
一个或更多个 RADAR传感器 460
一个或更多个 超声传感器 462
一个或更多个 LIDAR传感器 464
一个或更多个 IMU传感器 466
一个或更多个麦克风 496
一个或更多个立体相机 468
一个或更多个广角相机 470
一个或更多个红外相机 472
一个或更多个环绕相机 474
一个或更多个远程/中程相机 498
信息娱乐SoC 430
CABIN CAM(S) 441
SoC 404(B)
SoC 404(A)
CPU(S) 406
GPU(S) 408
一个或更多个处理器 410
一个或更多个高速缓存 412
一个或更多个加速器 414
一个或更多个数据存储 416
CPU(S) 418
402
GPU(S) 420
HD地图 422
426
网络接口 424
一个或更多个数据存储 428
仪表组 432
HMI显示器 434
ADAS系统 438
一个或更多个控制器 436
一个或更多个转向传感器 440
一个或更多个振动传感器 442
一个或更多个速度传感器 444
制动传感器系统 446
制动致动器 448
推进系统 450
油门/加速器 452
转向系统 454
转向致动器 456
400

图4C

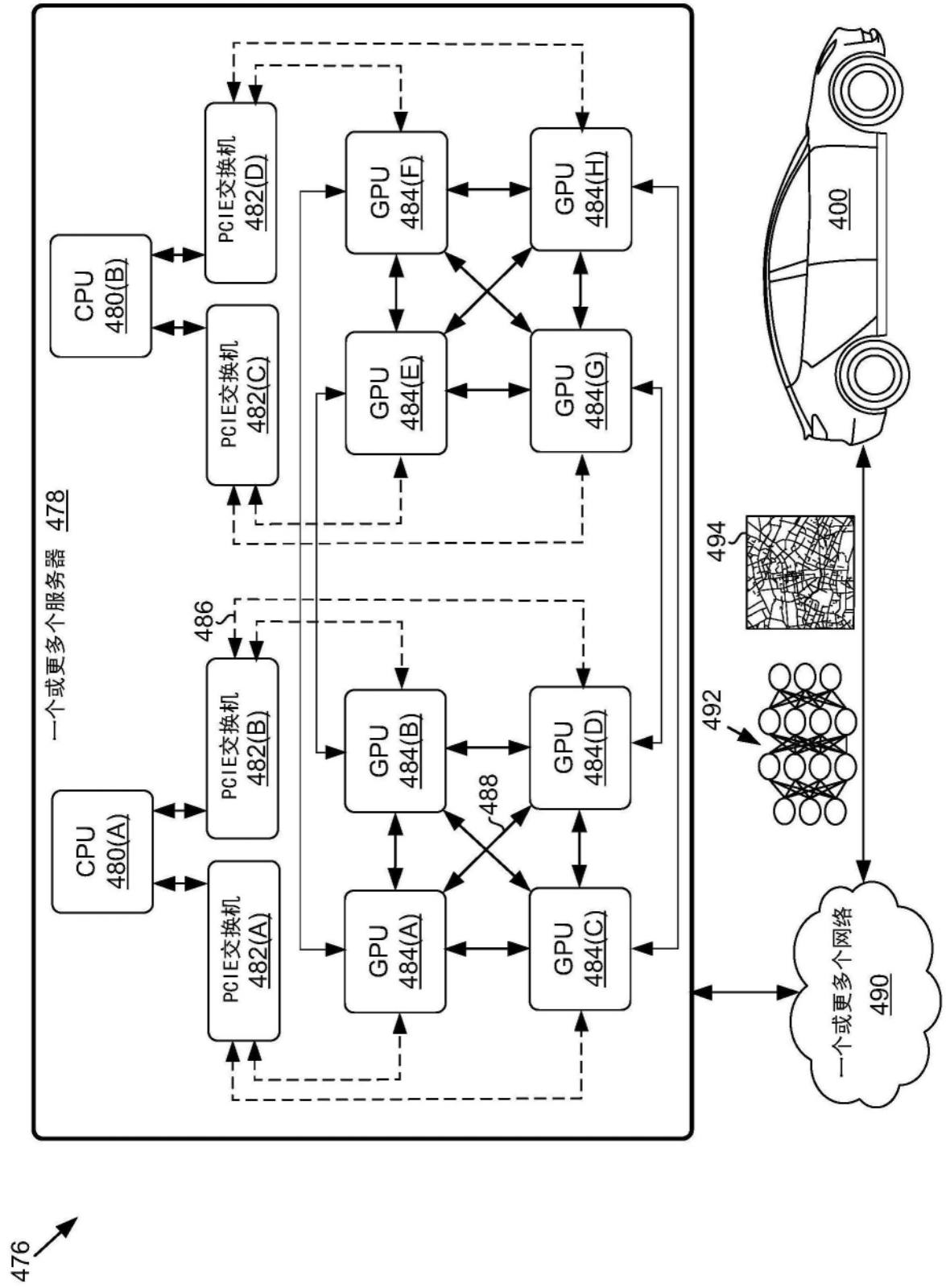

476
一个或更多个服务器 478
CPU 480(A)
CPU 480(B)
PCIE交换机 482(A)
PCIE交换机 482(B)
PCIE交换机 482(C)
PCIE交换机 482(D)
486
488
GPU 484(A)
GPU 484(B)
GPU 484(C)
GPU 484(D)
GPU 484(E)
GPU 484(F)
GPU 484(G)
GPU 484(H)
400
494
492
一个或更多个网络 490

图4D

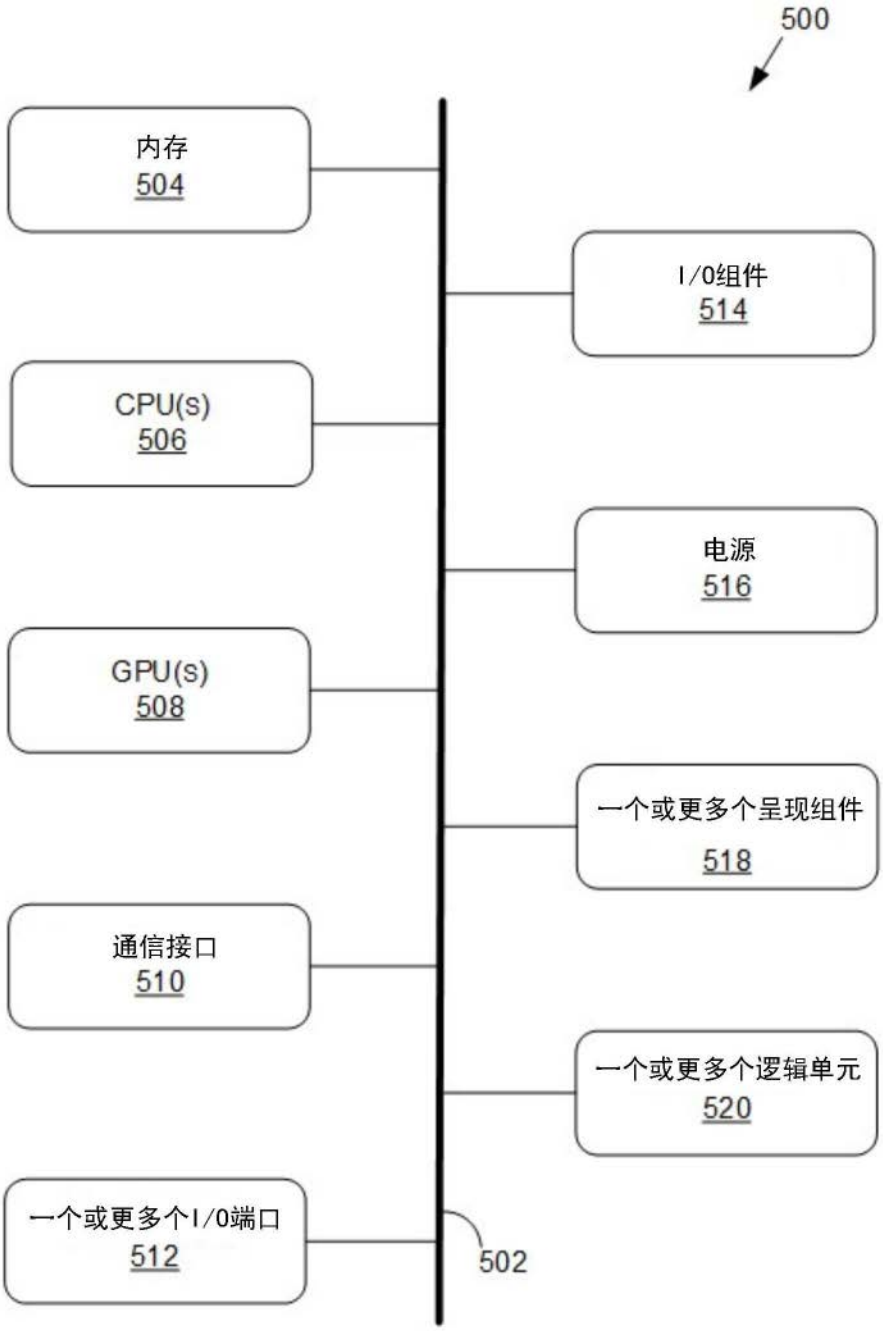

500
内存
504
I/O组件
514
CPU(s)
506
电源
516
GPU(s)
508
一个或更多个呈现组件
518
通信接口
510
一个或更多个逻辑单元
520
一个或更多个I/O端口
512
502

图5

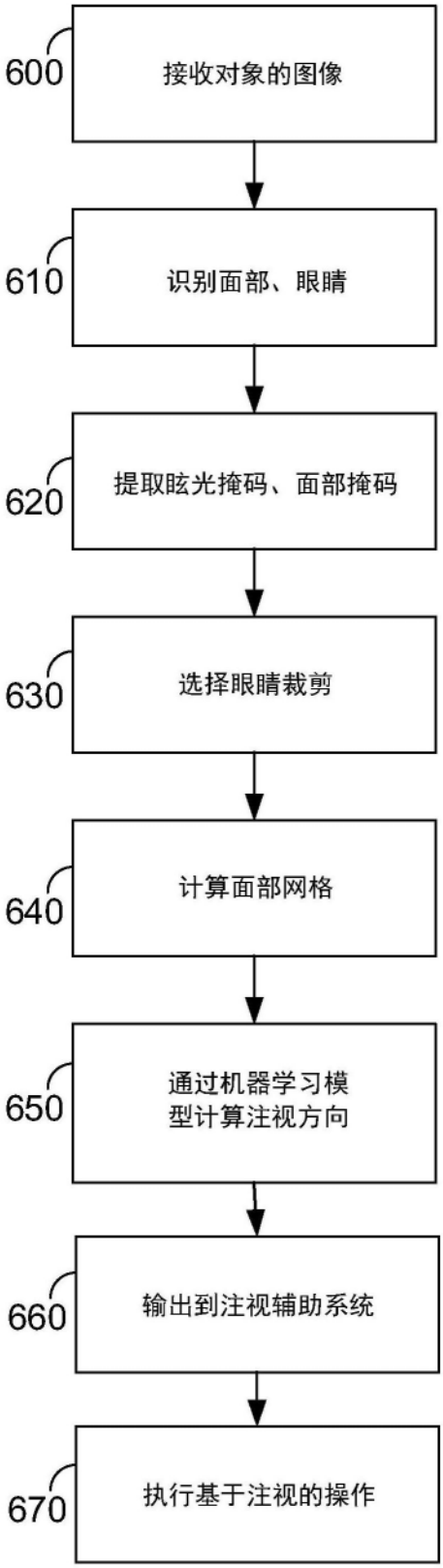

600
接收对象的图像
610
识别面部、眼睛
620
提取眩光掩码、面部掩码
630
选择眼睛裁剪
640
计算面部网格
650
通过机器学习模型计算注视方向
660
输出到注视辅助系统
670
执行基于注视的操作

图6

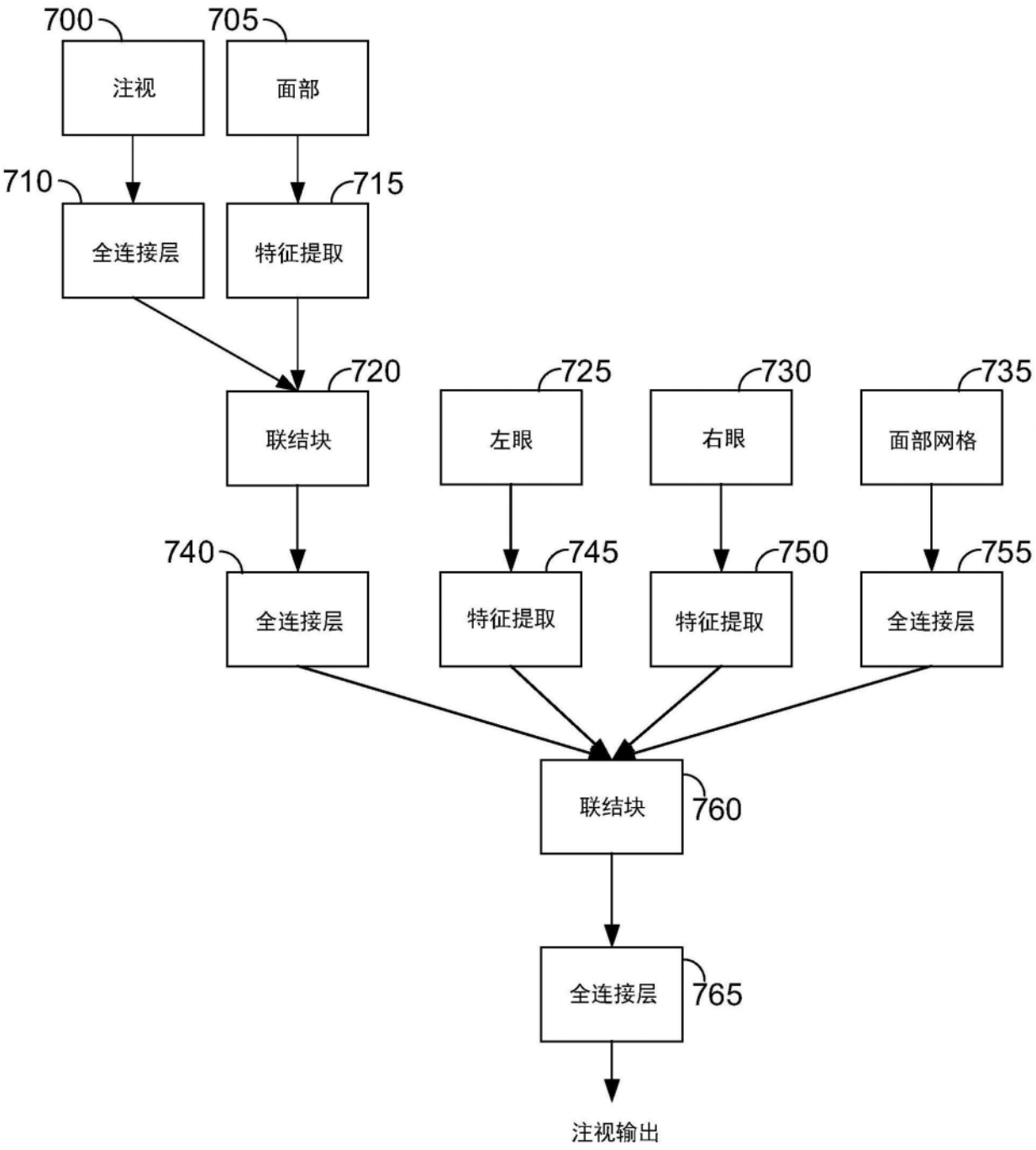

700
注视
705
面部
710
全连接层
715
特征提取
720
联结块
725
左眼
730
右眼
735
面部网格
740
全连接层
745
特征提取
750
特征提取
755
全连接层
联结块
760
全连接层
765
注视输出

图7