



(12) 发明专利申请

(10) 申请公布号 CN 116420096 A

(43) 申请公布日 2023. 07. 11

(21) 申请号 202180072758.4

(74) 专利代理机构 北京市柳沈律师事务所
11105

(22) 申请日 2021.09.20

专利代理师 胡琪

(30) 优先权数据

17/030,669 2020.09.24 US

(51) Int.Cl.

G01S 17/894 (2006.01)

(85) PCT国际申请进入国家阶段日

2023.04.24

(86) PCT国际申请的申请数据

PCT/US2021/051015 2021.09.20

(87) PCT国际申请的公布数据

W02022/066547 EN 2022.03.31

(71) 申请人 埃尔构人工智能有限责任公司

地址 美国宾夕法尼亚州

(72) 发明人 W·L·莱 H·R·伯克-斯威特

W·T·克拉姆佩

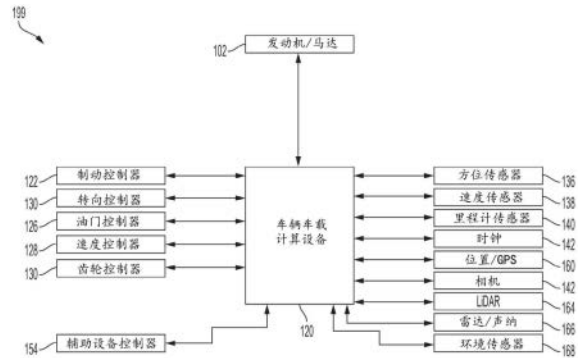
权利要求书3页 说明书11页 附图9页

(54) 发明名称

用于标记LIDAR点云数据的方法和系统

(57) 摘要

公开了用于处理点云数据的系统和方法。该方法包括：接收包括点云数据的3D图像，显示3D图像的2D图像，以及生成包围2D图像中的感兴趣对象的2D边界框。该方法还包括通过在第一方向上投影多个点来生成包括多个投影点的投影图像帧。该方法然后可以包括：显示包括由投影图像帧叠加的2D图像和2D边界框的图像帧，接收包括对投影的多个点中的与感兴趣对象对应的点集的标识的用户输入，标识用于感兴趣对象的标签，以及存储对应于与标签相关联的感兴趣对象的点集。



1. 一种用于处理点云数据的方法,所述方法包括:
 - 接收包括点云数据的3D图像;
 - 在用户界面上显示与所述3D图像相关联的2D图像;
 - 生成包围所述2D图像中感兴趣对象的2D边界框;
 - 通过沿第一方向将所述3D图像中的多个点投影到所述2D图像上来生成包括第一投影多个点的第一投影图像帧;
 - 在所述用户界面上显示包括由所述第一投影图像帧叠加的所述2D图像和所述2D边界框的图像帧;
 - 接收包括对所述第一投影多个点中对应于所述感兴趣对象的点集的标识的第一用户输入;
 - 标识用于所述感兴趣对象的标签;以及
 - 在数据存储中存储对应于与所述标签相关联的所述感兴趣对象的所述点集。
2. 根据权利要求1所述的方法,其中,沿第一方向将所述3D图像中的所述多个点投影到所述2D图像上包括:将所述多个点中的每个点投影到所述2D图像的像素。
3. 根据权利要求1所述的方法,还包括:通过沿第二方向将所述3D图像中的所述多个点投影到所述2D图像上来生成包括第二投影多个点的第二投影图像帧。
4. 根据权利要求3所述的方法,还包括:在所述用户界面上同时显示以下中的至少两个:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。
5. 根据权利要求4所述的方法,其中,接收所述第一用户输入包括:经由与以下中的至少一个的用户交互来接收所述第一用户输入,同时显示在所述用户界面上:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。
6. 根据权利要求5所述的方法,还包括:基于所述第一用户输入自动更新以下中的至少另一个,同时被同时显示在所述用户界面上:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。
7. 根据权利要求3所述的方法,还包括:在所述用户界面上同时显示以下中的至少两个:在所述2D边界框上放大的所述图像帧、在所述边界框上放大的所述第一投影图像帧、或在所述2D边界框上放大的所述第二投影图像帧。
8. 根据权利要求3所述的方法,还包括:标识包围对应于所述感兴趣对象的所述点集的3D边界框。
9. 根据权利要求8所述的方法,其中,标识所述3D边界框包括:经由与以下中的至少一个的用户交互来接收第二用户输入,同时显示在所述用户界面上:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。
10. 根据权利要求9所述的方法,还包括:自动更新以下中的至少另一个来显示所述3D边界框,同时被同时显示在所述用户界面上:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。
11. 根据权利要求8所述的方法,还包括:在所述数据存储中存储与所述标签相关联的所述3D边界框。
12. 根据权利要求8所述的方法,其中,所述3D边界框的尺寸至少基于所述2D图像中所述感兴趣对象的尺寸来确定。

13. 根据权利要求1所述的方法,其中,标识所述标签包括从用户接收所述标签。

14. 根据权利要求1所述的方法,其中,标识所述标签包括:自动将所述标签标识为与所述2D图像中的所述感兴趣对象相关联的所述标签。

15. 根据权利要求1所述的方法,其中,所述2D图像是与所述3D图像包括相同的场景并且在所述3D图像中的所述点云数据的阈值时间内被捕获的相机图像。

16. 根据权利要求1所述的方法,其中,所述点云数据使用LIDAR传感器捕获。

17. 根据权利要求1所述的方法,还包括:使用所存储的对应于与所述标签相关联的所述感兴趣对象的点集来训练机器学习模型,所述机器学习模型被配置用于控制自主车辆。

18. 一种用于处理点云数据的系统,所述系统包括:

显示器;

处理器;以及

非暂时性计算机可读介质,所述非暂时性计算机可读介质包括编程指令,所述编程指令在由所述处理器执行时,使所述处理器:

接收包括点云数据的3D图像;

在所述显示器上显示与所述3D图像相关联的2D图像;

生成包围所述2D图像中感兴趣对象的2D边界框;

通过沿第一方向将所述3D图像中的多个点投影到所述2D图像上来生成包括第一投影多个点的第一投影图像帧;

在所述用户器上显示包括由所述第一投影图像帧叠加的所述2D图像和所述2D边界框的图像帧;

接收包括对所述第一投影多个点中对应于所述感兴趣对象的点集的标识的第一用户输入;

标识用于所述感兴趣对象的标签;以及

在数据存储中存储对应于与所述标签相关联的所述感兴趣对象的所述点集。

19. 根据权利要求18所述的系统,其中,沿第一方向将所述3D图像中的所述多个点投影到所述2D图像上包括:将所述多个点中的每个点投影到所述2D图像的像素。

20. 根据权利要求18所述的系统,还包括编程指令,所述编程指令在被执行时使所述处理器通过在第二方向上将所述3D图像中的所述多个点投影到所述2D图像上来生成包括第二投影多个点的第二投影图像帧。

21. 根据权利要求20所述的系统,还包括编程指令,所述编程指令在被执行时使所述处理器同时显示以下中的至少两个:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。

22. 根据权利要求21所述的系统,其中,使所述处理器接收所述第一用户输入的所述编程指令包括使所述处理器经由与以下中的至少一个的用户交互来接收所述第一用户输入,同时显示在所述用户界面上的编程指令:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。

23. 根据权利要求22所述的系统,还包括:当被执行时使所述处理器基于所述第一用户输入自动更新以下中的至少另一个,同时被同时显示在所述用户界面上的编程指令:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。

24. 根据权利要求20所述的系统,还包括:在所述用户界面上同时显示以下中的至少两个:在所述2D边界框上放大的所述图像帧、在所述边界框上放大的所述第一投影图像帧、或在所述2D边界框上放大的所述第二投影图像帧。

25. 根据权利要求20所述的系统,还包括:当被执行时使所述处理器标识包围对应于所述感兴趣对象的所述点集的3D边界框的编程指令。

26. 根据权利要求25所述的系统,其中,使所述处理器标识所述3D边界框的所述编程指令包括使所述处理器经由与以下中的至少一个的用户交互来接收第二用户输入,同时显示在所述用户界面上的编程指令:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。

27. 根据权利要求26所述的系统,还包括:使所述处理器自动更新以下中的至少另一个来显示所述3D边界框,同时被同时显示在所述用户界面上的编程指令:所述图像帧、所述第一投影图像帧或所述第二投影图像帧。

28. 根据权利要求25所述的系统,还包括:使所述处理器在所述数据存储中存储与所述标签相关联的3D边界框的编程指令。

29. 根据权利要求25所述的系统,其中,所述3D边界框的尺寸至少基于所述2D图像中所述感兴趣对象的尺寸来确定。

30. 根据权利要求18所述的系统,其中,使所述处理器标识所述标签的所述编程指令包括使所述处理器从用户接收所述标签的编程指令。

31. 根据权利要求18所述的系统,其中,使所述处理器标识所述标签的所述编程指令包括使所述处理器自动将所述标签标识为与所述2D图像中的所述感兴趣对象相关联的所述标签的编程指令。

32. 根据权利要求18所述的系统,其中,所述2D图像是与所述3D图像包括相同的场景并且在所述3D图像中的所述点云数据的阈值时间内被捕获的相机图像。

33. 根据权利要求18所述的系统,其中,所述点云数据使用LIDAR传感器捕获。

34. 根据权利要求18所述的系统,还包括:使所述处理器使用所存储的对应于与所述标签相关联的所述感兴趣对象的点集来训练机器学习模型的编程指令,所述机器学习模型被配置用于控制自主车辆。

35. 一种计算机程序产品,所述计算机程序产品包括存储编程指令的存储器,所述编程指令被配置为使处理器通过以下来标记点云数据:

接收包括点云数据的3D图像;

在显示器上显示与所述3D图像相关联的2D图像;

生成包围所述2D图像中感兴趣对象的2D边界框;

通过沿第一方向将所述3D图像中的多个点投影到所述2D图像上来生成包括第一投影多个点的第一投影图像帧;

在所述显示器上显示包括由所述第一投影图像帧叠加的所述2D图像和所述2D边界框的图像帧;

接收包括对所述第一投影多个点中对应于所述感兴趣对象的点集的标识的第一用户输入;

标识用于所述感兴趣对象的标签;以及

在数据存储中存储对应于与所述标签相关联的所述感兴趣对象的所述点集。

用于标记LIDAR点云数据的方法和系统

[0001] 交叉引用和优先权的要求

[0002] 本专利文件要求2020年9月24日提交的美国专利申请号17/030,669的优先权,其整体内容通过引用并入本文中。

背景技术

[0003] 成像传感器用于各种应用,包括自主车辆和机器人。例如,自主车辆(AV)使用成像传感器来检测和定位它们的环境中的对象,以实现安全且准确的导航。LIDAR和相机是感知和场景理解的两个基本成像传感器。它们串联构建环境并提供用于检测和定位其他对象的部件,为机器人提供安全导航所需的丰富语义信息。

[0004] 像相机这样的光学传感器长期以来一直用于收集这样的图像数据,并且通常会生成二维(2D)数据。最近,人们对三维(3D)成像数据越来越感兴趣。生成3D点云数据的一个或多个图像帧的3D成像系统的一个示例是LIDAR成像系统。通常,这样的LIDAR系统使用高能激光来确定到目标的距离。在常规的LIDAR系统中,一个或多个激光脉冲用于照射场景,并且测量光脉冲的每个像素通过从激光到目标并返回检测器阵列的往返路径的时间。检测来自目标的反射光并且测量其往返行程时间来确定到目标上的点的距离。针对包括目标的多个点获得计算的范围或距离信息,从而创建3D点云。3D点云可以用于渲染对象的3D形状。

[0005] 为了对象的检测和定位,AV通常将由成像传感器收集的数据输入经过训练的机器学习模型,以基于输入数据标识和定位对象。这样的机器学习模型的训练需要训练数据集,包括来自相机的2D图像、来自LIDAR传感器的3DLIDAR点云以及表示在2D图像和3DLIDAR点云中观察到的车辆和行人等对象的标签。虽然标识和标记2D图像中的对象是很好理解的过程,但准确且全面地向3DLIDAR点云添加标签是一项由受过训练的人员使用计算机工具手动执行来指示和标记对象的困难且劳动密集型的任务。然而,点云中的对象所呈现的特性往往不明显,并且还受到地面和其他杂物的影响,使它们更难标识。当面对大量待标记数据时,人工标记相对较慢,并且容易错误标记具有相似特征的对象。现有的标记工具不是用户友好的,特别是对于与2D图像相比本质上更难可视化的3D点云,导致标记效率低和视觉疲劳。此外,它们要求用户在2D显示器上操纵复杂的3D用户接口,从而导致效率低下和不准确。这样的工具也不会将对象与对应的LIDAR点相关联来在其3D环境中定位对象。

[0006] 本文件描述了旨在解决上述问题和/或其他问题的方法和系统。

发明内容

[0007] 在各种场景中,公开了用于标记LIDAR点云数据的系统和方法。该系统可以包括处理器、显示器和包括编程指令的非暂时性计算机可读介质,编程指令在被执行时使处理器执行下文讨论的各种方法。该系统可以通过接收包括点云数据的3D图像来处理点云数据,在用户界面上显示与3D图像相关联的2D图像,生成包围2D图像中的感兴趣对象的2D边界框,并且通过在第一方向上将3D图像中的多个点投影到2D图像上生成包括第一投影多个点的第一投影图像帧。系统然后可以在用户界面上显示包括由第一投影图像帧叠加的2D图像

和2D边界框的图像帧,并且接收第一用户输入,该第一用户输入包括对第一投影多个图像中与感兴趣对象对应的点集的标识。系统可以标识用于感兴趣对象的标签,并且将对应于与标签相关联的感兴趣对象的点集存储在数据存储中。可选地,通过将多个点中的每一个投影到2D图像的像素,可以在第一方向上将3D图像中的多个点投影到2D图像上。

[0008] 在某些这样的场景中,系统还可以通过在第二方向上将3D图像中的多个点投影到2D图像上来生成包括第二投影多个点的第二投影图像帧。可选地,系统可以同时显示在用户界面上显示图像帧、第一投影图像帧和/或第二投影图像帧中的至少两个。附加地和/或替代地,接收第一用户输入可以包括:经由与以下中的至少一个的用户交互来接收第一用户输入,同时在用户界面上显示:图像帧、第一投影图像帧和/或第二个投影图像帧。系统还可以在同时显示在用户界面上的同时自动更新图像帧、第一投影图像帧和/或第二投影图像帧,以便显示第一用户输入。

[0009] 在一些场景中,系统还可以同时显示在2D边界框上放大的图像帧、在边界框上放大的第一投影图像帧和/或在2D边界框上放大的第二投影图像帧中的至少两个。

[0010] 在一些实现方式中,系统可以标识包围对应于感兴趣对象的点集的3D边界框。可选地,可以通过在用户界面上显示时通过与以下中的至少一个的用户交互接收第二用户输入来标识3D边界:图像帧、第一投影图像帧和/或第二投影图像帧。系统然后可以自动更新图像帧、第一投影图像帧和/或第二投影图像帧以显示3D边界框,同时被同时显示在用户界面上。系统可以可选地在数据存储中存储与标签相关联的3D边界框。3D边界框的尺寸可以至少基于2D图像中感兴趣对象的尺寸来确定。

[0011] 系统可以通过从用户接收标签来标识标签和/或自动将标签标识为与2D图像中的感兴趣对象相关联的标签。

[0012] 2D图像可以是与3D图像包括相同的场景并且与3D图像中的点云数据在阈值时间内捕获的相机图像。可选地,点云数据使用LIDAR传感器来捕获。

[0013] 在各种场景中,系统可以使用存储的对应于与标签相关联的感兴趣对象的点集来训练机器学习模型,该机器学习模型例如被配置用于控制自主车辆。

附图说明

[0014] 图1示出了自主车辆的示例系统和组件。

[0015] 图2示出了用于传感器数据的标记的示例数据流程图。

[0016] 图3是示出用于传感器数据的对象标识和标记的方法的流程图。

[0017] 图4A示出了场景的示例3D图像,并且图4B示出了对应的2D图像。

[0018] 图5示出了显示通过在不同方向上投影LIDAR点云生成的一个或多个投影图像帧的示例显示界面。

[0019] 图6示出了图5的显示界面,其包括在感兴趣对象上放大的图5的图像帧。

[0020] 图7示出了图6的显示界面,其包括用于标识与感兴趣对象相关联的LIDAR点的各种用户动作。

[0021] 图8示出了图7的显示界面,其包括包含标识的LIDAR点的3D边界框。

[0022] 图9是示出车辆和/或外部电子设备的可能电子子系统的各种元件的框图。

具体实施方式

[0023] 除非上下文另有明确规定,否则如本文件中所使用的,单数形式“一”、“一个”和“该”包括复数引用。除非另有定义,否则本文使用的所有技术和科学术语具有与本领域普通技术人员普遍理解的含义相同含义。如本文件中所使用的,术语“包括”是指“包括但不限于”。与本文件相关的附加术语的定义包含在该“具体实施方式”的末尾处。

[0024] 本文件描述了一种方法和系统(例如,下文描述的工具),它们将创建标记界面(例如,图形用户界面),其包括叠加有用于2D可视化的投影LIDAR数据的参考2D图像。标记界面还可以包括包含投影LIDAR数据的其他视图的图像帧。标记界面被配置为允许用户容易地可视化和辨别与参考2D图像关联的3D点云数据,其中用户对标记界面中任何图像帧的动作被反映在所有显示的图像帧中,从而更容易标识与对象关联的LIDAR点并且执行标记。标记界面将3D图像的标记问题分解为2D可视化环境中的一个或多个步骤,并且支持3D图像在2D显示器中的呈现和标记。

[0025] 图1示出了用于诸如AV的车辆的示例系统架构100。车辆包括发动机或电动机102和用于测量车辆和/或其环境的各种参数的各种传感器。两种类型的车辆共有的操作参数传感器包括,例如:方位传感器136,诸如加速度计、陀螺仪和/或惯性测量单元;速度传感器138;和里程计传感器140。车辆还可以具有时钟142,系统使用其来确定操作期间的车辆时间。时钟142可以被编码到车辆车载计算设备中,它可以是单独的设备,或者多个时钟可以是可用的。

[0026] 车辆还将包括操作用于收集关于车辆正在其中行驶的环境的信息的各种传感器。这些传感器可以包括,例如:位置传感器160,诸如全球定位系统(GPS)设备;对象检测传感器,诸如一个或多个相机162;LiDAR传感器系统164;和/或雷达和/或声纳系统166。传感器还可以包括环境传感器168,诸如降水传感器和/或环境温度传感器。对象检测传感器可以使车辆能够检测在任何方向上在车辆100的给定距离范围内的对象,而环境传感器收集关于车辆的行驶区域内的环境条件的数据。

[0027] 在操作期间,信息从传感器传送到车载计算设备120。车载计算设备120分析由传感器捕获的数据并且可选地基于分析的结果控制车辆的操作。例如,车载计算设备120可以经由制动控制器122控制制动;经由转向控制器124控制方向;经由油门控制器126(在汽油动力车辆中)或马达速度控制器128(诸如电动车辆中的电流电平控制器)控制速度和加速度;控制差动齿轮控制器130(在具有变速器的车辆中);和/或其他控制器。

[0028] 地理位置信息可以从位置传感器160传送到车载计算设备120,其然后可以访问对应于位置信息的环境地图来确定环境的已知固定特征,诸如街道、建筑物、停车标志和/或停止/前进信号。来自相机162的捕获的图像和/或从诸如LIDAR系统164的传感器捕获的对象检测信息从那些传感器)传送到车载计算设备120。对象检测信息和/或捕获的图像可以由车载计算设备120处理以检测接近车辆100的对象。此外或替代地,AV可以将任何数据传输到外部服务器以供处理。用于基于传感器数据和/或捕获的图像进行对象检测的任何已知或将已知的技术都可以用于本文档中公开的实施例中。

[0029] 在本文件中讨论的各种实施例中,描述可以表明车辆或车辆的车载计算设备可以执行编程指令,这些指令使车辆的车载计算设备做出决策并且使用决策来控制一个或多个车辆系统的操作。然而,实施例不限于这种布置,因为在各种实施例中,分析、决策制定和/

或操作控制可以全部或部分地由与车辆的车载计算设备电子通信的其他计算设备来处理。这样的其他计算设备的示例包括与乘坐车辆的人相关联的电子设备(例如智能电话),以及经由无线通信网络与车辆进行电子通信的远程服务器。任何这样的设备的处理器可以执行下文将要讨论的操作。

[0030] 图2示出了用于标记传感器数据来生成供机器人设备(例如,自主车辆)使用的训练数据的示例数据流程图。如图2所描绘的,要被处理以生成用于机器人设备230的训练数据220的传感器数据212被存储在数据存储210中。在一些实施例中,数据存储210可以是自主车辆上的存储器设备,或者是与车辆的车载处理器中的一个或多个进行电子通信的外部存储器设备。

[0031] 传感器数据212可以包括图像数据,其包括2D图像(例如,由一个或多个相机生成的RGB相机图像、热图像等),和/或可以包括3D图像(例如,由一个或多个LIDAR或其他3D成像系统生成的点云)。在实施例中,环境的2D图像和3D图像可能已在相同时间或类似时间捕获,和/或2D图像和3D图像可能以其他方式在时间上相互关联。在一些实施例中,图像数据可以包括环境的2D视频和/或3D视频的多个图像帧。

[0032] 在一些实施例中,包括图像数据的传感器数据212可以由配备有诸如一个或多个相机的图像传感器、一个或多个LIDAR和/或当车辆在环境中移动时将捕获环境的数字2D图像(例如,使用相机)和/或3D图像(即,由LIDAR捕获的点云数据)的其他传感器的AV或其他车辆收集。可选地,多个配备图像传感器的车辆可以捕获传感器数据。在一些实施例中,车辆的图像传感器或在车辆上或在从车辆接收传感器的远程服务器上的另一个处理器将把车辆的位置和姿势与数字2D和3D图像中的每一个相关联。每个数字图像的位置和姿势表示捕获数字图像时车辆的方位(地理位置和方向)。该系统可以从任何适当的数据源接收位置和姿势信息,诸如车辆的GPS和方位传感器(见图1)。传感器数据212因此可以包括2D和/或3D数字图像,连同包括关联位置和姿势的元数据。诸如捕获时间、用于捕获图像的传感器的类型和属性、车辆的类型和属性等的其他元数据也可以与2D和/或3D数字图像相关联。也可以使用其他3D数据集,诸如由立体相机捕获的3D数据、由3D模型生成方法(诸如运动结构(SfM)图像处理)分析2D图像生成的数据或其他3D数据集。

[0033] 传感器数据212可以由工具240处理,其利用本公开中公开的系统、方法和/或技术来生成用于训练一个或多个机器学习模型的标记的训练数据220。如上所讨论的,机器学习模型可以用于自主控制和/或操作机器人设备,诸如自主车辆。在一些实施例中,工具240将标记3D图像的问题分解为2D环境中的一个或多个步骤,并且支持3D图像在2D显示器中的呈现和标记。

[0034] 工具240可以实现为一个或多个计算设备,其包括处理器242、非暂时性存储器244、用户界面246(例如,图形用户界面(GUI))和数据的标记所需的现在或以后已知的其他组件。尽管图2示出了工具240包括处理器242、非暂时性存储器244和用户界面246,但是本公开不限于此,并且这些组件中的一个或多个可以远离工具240定位。如下所述,用户界面246可以被配置为帮助用户标识和/或标记传感器数据212内描绘的特定对象。

[0035] 来自工具240的输出数据214可以包括在源数据212(例如,LIDAR点云)中描绘的标识和标记的对象,并且可以存储在数据存储210中。在一些实施例中,输出数据214可以包括经处理的图像数据,其中每个图像或点云图像帧可以包括相应的源图像,该源图像包括图

像中特定对象的标识和标签。替代地和/或附加地,输出数据214可以包括相应的标识的对象/标签对之间的关联的指示(例如,作为关系数据库)。本公开不限于此,并且输出数据可以包括任何其他合适格式的标记和标识的对象。虽然输出数据214被示出为与传感器数据212一起存储在数据存储实体210中,但这不是限制性的,并且输出数据214可以根据需要附加地或替代地存储在一个或多个其他单独和不同的数据存储实体(未显示)中。

[0036] 输出数据214的一个或多个部分可以并入或包含在训练数据集220中,用于训练一个或多个机器学习模型,机器人设备230可以利用这些模型进行控制和操作(例如,导航期间对自主车辆的控制和操作)。

[0037] 应当理解,图2中所示的数据流程图200和组件可以只是许多可能的实施例中的一个。例如,工具240可以由多个计算设备的网络和/或由云计算系统(例如,协作来标记传感器数据集的多个用户)来实现。在一些实施例中,用户界面246中的至少一个可以在远离处理器242和/或存储器244的计算设备处实现,诸如当用户界面246在客户端设备上实现而处理器242和/或存储器244在一个或多个服务器或后端计算设备处实现(例如,在网络服务或其他类型的客户端/服务器应用中)。在一些实施例中,处理器242、存储器246和/或用户界面246以分布式方式实现。例如,远程客户端设备可以包括处理器存储器的第一部分和/或计算指令,并且通信地连接到远程客户端设备的一个或多个服务器或后端计算设备可以包括处理器存储器的第二部分和/或计算指令。为了便于阅读,一个或多个处理器、一个或多个存储器以及一个或多个用户界面使用单数时态提及;然而,应当理解,这只是为了便于阅读,而不是限制性的。

[0038] 图3是示出用于传感器数据的对象标识和标记的处理的各个步骤的流程图。应当理解,虽然图3和相关描述描述了对从LIDAR传感器获得的点云数据的标记,但是本公开的系统和方法可以用于标记其他类型的3D传感器数据而不背离本公开的原理。

[0039] 在302处,系统可以接收包括使用LIDAR传感器捕获的场景的待标记点云图像帧的3D图像。如本文所使用的,3D图像是指在特定持续时间(例如,0.1秒)内累积的LIDAR传感器数据集。因此,虽然点云数据是连续收集的,但是LIDAR扫描点的3D图像可以以离散时间间隔生成。图4A示出了场景的示例3D图像410,其包括与(包括在对应的椭圆中的)一个或多个对象411、412、413和414相关联的LIDAR扫描点。

[0040] 该系统可以标识304使用相机捕获的并且与3D图像相关联的参考2D图像。如果3D图像包括2D图像中描绘的场景的图像,并且是在捕获2D图像的阈值时间内(例如,同时和/或在几百毫秒内)捕获的,则2D图像可以与3D图像相关联。在实施例中,2D图像可以是通过在与3D图像相同的时刻拍摄相同场景而形成的相机图像。例如,3D图像和2D图像中的每一个可以具有指示其捕获时间的相应时间戳,并且这两个时间戳可以仅相差短的时间间隔(例如,在半秒的时间间隔内、在一秒的时间间隔等)。可选地,系统可以从数据存储中的2D图像中标识与对应于3D图像的姿势和位置相关联的数字图像集。该系统可以通过例如归一化从所标识的数字图像集生成参考2D图像。系统可以使用诸如自动白平衡、自动亮度/对比度和其他校正方法的已知处理来做到这一点。

[0041] 图4B示出与图4A中所示的3D图像相关联的参考2D图像420,并且被示出为包括对应的对象421、422、423和424。

[0042] 如上文所讨论的,接收到的2D和3D图像可以包括诸如但不限于位置、姿势、捕获时

间、用于捕获图像的传感器的类型和属性、环境条件等的元数据和/或可以与元数据相关联。

[0043] 在306处,系统可以在参考2D图像中选择感兴趣对象,其中将使用下文描述的方法在3D图像中标识和标记感兴趣对象。可选地,系统可以从用户接收对感兴趣对象的选择(例如,响应于用户使用用户界面选择参考2D图像中的对象的操作)。例如,系统可以选择和/或接收对图4B中所示的2D图像420中的感兴趣对象421(即,汽车)的选择。可选地,系统可以使用任何现在或以后已知的对象检测算法来自动选择感兴趣对象。

[0044] “感兴趣对象”的示例可以包括但不限于建筑物的组件(例如,门、窗、墙壁、屋顶、楼梯、管路/管道、电气装备、地板材料、装饰方面)、景观组件(例如,树、灌木、车道、水景)、车辆、人、动物等。本公开的系统和方法可以用于标识和标记场景中的多于一个的感兴趣对象,诸如与较大对象(例如,建筑物的整体尺寸)相关联的较小对象(例如,门、窗等)的集合,其中有关这种较小和较大对象的集合的信息可以被处理,并且在一些方面中,一个或多个对象可以从场景中标识。因此,感兴趣对象可以包括与较大对象(例如,建筑物)相关联的一个或多个较小对象(例如,门、窗等)的集合。应当理解,在一些实例中,感兴趣对象可以存在于或部分存在于场景中,而在其他实例中,场景中可能存在多于一个的感兴趣对象。关于感兴趣对象的部分(或至少一部分)指的是在场景中只有一个或多个部分而不是整个对象可见的实例。在一些方面中,本方法可以用于根据用户的期望生成关于可以出现或部分出现在场景中的单个或多个感兴趣对象的信息。对象在场景中部分存在的示例可以包括对象由于遮挡或捕获设备的视角而仅被部分捕获的实例,或者对象的部分适合场景但对象的整体不适合的实例。

[0045] 系统可以将3D图像中的点(即,3D点云点)投影(308)到参考2D图像上,以获得沿不同投影方向(即,不同视图)的一个或多个投影2D图像帧。这种投影可以包括将3D图像中的每个3D图像点投影或映射到2D图像中的像素。在一些实施例中,系统可以将完整的3D图像投影到参考2D图像上。替代地,系统可以将3D图像的至少包括感兴趣对象的区域投影到参考2D图像上。因此,不是将所有3D图像点投影到2D参考图像中的每个可用像素,系统可以首先过滤掉不是显示感兴趣对象图像的像素的像素。

[0046] 在一些实施例中,系统使用获得投影点的任何现在或以后已知的方法,诸如但不限于使用3D图像和2D图像之间的定义的坐标映射函数来执行投影。坐标映射函数可以从三维坐标到二维坐标/像素的任何映射关系(例如,使用通过从一个向量空间到另一个向量空间的乘法来扭曲点的变换矩阵的线性变换),并且可以根据用于收集3D图像数据的LIDAR传感器的参数(例如,校准、姿势、方向、图像参考系等)和用于拍摄2D图像的相机的参数(例如,相机矩阵、校准、姿势、方向、图像参考系等)来确定,和/或可以根据定义的点云数据和相机图像中的对应对象的坐标预先计算。在一些实施例中,坐标映射函数还可以补偿在记录2D图像的时间和记录与3D图像对应的LIDAR扫描的时间之间的车辆(其上安装了用于收集3D和2D图像的LIDAR和/或相机)的运动。

[0047] 现在参考图5,其示出了显示通过在不同方向将3D图像投影到2D图像上而生成一个或多个投影图像帧的示例显示界面500。可选地,显示界面500可以是用户用来标识和标记3D图像中的感兴趣对象的图形用户界面。例如,图5示出了使用分别在前向LIDAR观察方向(如位于对象前方的相机/旁观者所观察的一由520的左上角的方向所示)、横向LIDAR

观察方向(由位于对象斜前方和一侧的相机/旁观者观察-由530的左上角的方向所示)以及上下LIDAR观察方向上(由位于对象前方和上方的相机(例如,在无人机中)观察一由的左上角的方向所示)3D图像在参考2D图像上的投影而生成的三个投影图像帧520、530、540。这里的方向是基于相机图像帧和LIDAR图像帧,其中相机图像帧位于相机镜头的中心(即x轴指向镜头外的镜筒,z-轴指向上方,y/z平面平行于相机平面,并且坐标系为右手的),并且LIDAR传感器图像帧具有指向上方的z轴,x/y平面取决于LIDAR方位。不同方向上的其他投影图像帧在本公开的范围之内,并且可以作为图5中所示的投影图像帧的附加和/或替代来呈现。

[0048] 显示界面500还可以包括图像帧510,其包括叠加在参考2D图像511之上的投影图像帧512,并且在投影图像帧中示出2D图像以及投影LIDAR点云两者。图5示出了叠加在2D图像上的在前向LIDAR观察方向上生成的投影图像帧(即,图像帧520)。应当注意的是,虽然图5示出了叠加在2D图像上的在前向LIDAR观察方向上生成的投影图像帧,其他投影图像帧也可以叠加在参考2D图像上。系统可以根据感兴趣对象的大小、位置、配置等选择一个或多个投影图像帧叠加在参考2D图像上。在某些实施例中,例如,当对象小且固定(例如,交通灯)时,LIDAR点云的单帧可能不包含与对象对应的许多点。在这些场景中,系统可能会累积在捕获2D图像之前和之后长达几秒的LIDAR点云。这些额外的3D图像可以同时投影到参考2D图像上,以便有更多的LIDAR点对应于感兴趣对象。

[0049] 在310处,系统可以生成包围参考2D图像中的感兴趣对象的2D边界框。系统可以在包括叠加在参考2D图像之上的投影图像帧的图像帧(例如,图5的图像帧510)中生成2D边界框。应当注意,2D边界框可以是完全包围感兴趣对象的任何二维轮廓或边界区域,例如矩形轮廓、正方形轮廓、圆形轮廓、椭圆形轮廓、模仿感兴趣对象的形状的轮廓或任何其他可能的形状。2D边界框可以包括感兴趣对象和轮廓之间的指定距离。指定距离可以被选择为例如阈值,超过该阈值其他对象可以被包括在2D边界框、2D边界框的形状、感兴趣对象的形状和大小等内。

[0050] 在312处,系统可以生成包括在边界框和/或对应于边界框的区域上放大的一个或多个图像帧的显示界面。生成的边界框可以显示在包括2D参考图像的图像帧中。例如,生成的显示界面可以包括具有叠加在参考2D图像和生成的2D边界框之上的投影图像帧的图像帧,以及各种其他投影图像帧,所有这些都放大到生成的2D边界框和/或与边界框对应的区域。图6示出了包括图5中生成的缩放到生成的2D边界框和/或与边界框对应的区域的图像帧的显示600。具体地,图像帧610(对应于图5的图像帧510)包括包围感兴趣对象611的2D图像并且在2D边界框615上被放大的2D边界框615。如图6的图像帧610所示,2D边界框615包括2D参考图像611中的感兴趣对象和投影的LIDAR点612(即,图510中所示的落入2D边界框内的LIDAR点云的子集)。应当注意,如下所述,LIDAR点612不一定与感兴趣对象相关联。此外,系统可以更新投影图像帧510、520、530和540来仅显示对应于LIDAR点612和/或2D边界框的区域的相应投影LIDAR点(分别显示为622、632和642)以生成图6中所示的图像帧620、630和640。系统可以基于3D图像和2D图像之间的坐标映射函数来标识要包括在图像帧620、630和640中的点。

[0051] 缩放的水平可以基于例如感兴趣对象的大小、2D参考图像的分辨率、参考2D图像的大小以及2D边界框相对于参考2D图像的大小的确定。

[0052] 在314处,系统可以接收第一用户输入,其包括在对应于和/或与感兴趣对象相关联的显示的图像帧中的一个或多个中的投影LIDAR点的标识,并且更新剩余的图像帧来反映接收到的输入。例如,图6中所示的显示可以允许用户突出显示、着色(不同的颜色、灰度等)或以其他方式选择图像帧610、620、630和640中的任一个中与感兴趣对象611关联(和/或不关联)的一个或多个LIDAR点。应该注意的是,由于图像帧是同步的(即,使用相同的坐标映射函数创建),所以任何用户对一个图像帧中的投影LIDAR点的动作都自动反映在其他图像帧中。因此,用户能够同时从三个不同的视角(除了边界框视图还包括2D参考图像视图)观看特定对象,并且在四个图像帧中的任何一个中标识LIDAR点,以更好地提高标识精度。不同图像帧的这种同时、同步显示允许用户在以易于可视化的2D格式观察它们的同时,清楚准确地辨别、区分和/或标识与感兴趣对象相关联的3DLIDAR点。

[0053] 例如,图7示出了对应于图6的610、620、630和640并且包括接收的用户输入的图像帧710、720、730和740。如图像帧710中所示,用户可以通过用同与感兴趣对象711相关联的点(例如,具有较深灰颜色/阴影的点713(b))不同的颜色着色它们(与从车辆周围的墙壁、地形或草地反射的点相关联的较浅灰颜色/阴影的点713(a)),来突出他/她辨别为与边界框715中的感兴趣对象711不相关联的点。由于关于图像帧710、720、730和740中的任何图像帧中的LIDAR点的任何用户动作自动反映在其余图像帧中,所以系统可以自动将用户的选择应用到如分别在图像帧720、730和740中以较浅的灰颜色显示的不与感兴趣对象相关联的点723(a)、733(a)和743(a)所示的其他图像帧。类似地,系统可以自动将用户的选择应用到如分别在图像帧720、730和740中以较深灰颜色显示的与感兴趣对象相关联点723(b)、733(b)和743(b)所示的其他图像帧。(图7和本公开的其他图中使用的颜色仅作为示例;实际中可以使用其他颜色。)

[0054] 在与感兴趣对象相关联的LIDAR点的标识后,系统可以接收第二用户输入(316),其包括限制或包围所标识的LIDAR点的3D边界框(例如,长方体、立方体、棱柱等)。例如,系统可以允许用户在如图7中所示的图像帧710、720、730和740中的每个中在与感兴趣对象相关联的LIDAR点周围绘制3D边界框(使用任何现在或以后已知的用户界面绘图工具)。该系统还可以允许用户诸如通过放大、缩小或平移来操纵图像帧,以帮助用户界定标识的LIDAR点。在一些实施例中,系统可以自动生成包围所标识的LIDAR点的3D边界框,该边界框可以可选地由用户操纵以用于准确性或微调。例如,使用图7中所示的图像帧绘制的3D边界框(长方体)816在图8中示出。

[0055] 此外,系统可以自动检测图像帧中一个中的3D边界框的形状、大小和/或方向的变化,并且在其他图像帧中做出对应的变化。即使LIDAR点云不完整(例如,当感兴趣对象的一个或多个表面未反映回LIDAR点时),这也允许用户基于图像帧中的一个或多个图像帧中可见的信息以及2D参考图像调整3D边界框的尺寸、位置和/或方向。例如,在图4至图7中所示的示例中,停放汽车的前侧和乘客侧(即感兴趣对象)没有任何反映的LIDAR点,因为在LIDAR扫描期间,LIDAR传感器正面向驾驶员侧并稍稍靠后)。当前公开的方法和系统允许用户通过基于从2D图像估计的停放汽车的大小操纵/绘制图像帧810中的边界框来调整3D边界框的尺寸,其自动反映在可能丢失LIDAR点的其他图像帧820、830和840中。如图8中所示,图像帧810中的3D边界框816被绘制为适合2D边界框812内部,使得它与感兴趣对象811的尺寸紧密对应。这避免了需要一组完整的LIDAR点来生成3D边界框。在图像帧810中绘制的3D

边界框816的尺寸反映在其他投影图像帧820、830和840的边界框826、836和846中。

[0056] 3D边界框可以提供关于空间位置、方向以及它包含的对象的大小的信息。3D边界框可以提供关于空间位置、方向以及它包含的对象的大小的信息。

[0057] 在318处,系统可以将标签分配给标识的LIDAR点和3D边界框。通常,标签是多个类别和/或分类中的一个,机器学习模型可以利用其来了解什么类型的数据点分组代表不同的类别/分类。标签的示例集合包括“汽车”、“大型车辆”、“行人”、“骑自行车的人”、“摩托车手”、“交通标志”、“道路”、“交通灯”等。更细化的标签示例包括“公共汽车”、“半挂卡车”、“动物”、“冰块”、“坑洼”等。在一些实施例中,标签可以是参考2D图像中的感兴趣对象标识的标签。例如,系统可以使用任何现在或以后已知的对象标识方法(例如,使用经过训练的机器学习模型使用对象库来标识和标记2D图像中的对象)为参考2D图像中的一个或多个对象执行对象标识和计算标签。在一些实施例中,可以使用例如包括2D图像和与2D图像中的一个或多个对象相关联的标签的显示将标签呈现给用户以供验证。替代地和/或附加地,用户可以为感兴趣对象分配标签。在实施例中,用户可以使用下拉菜单来选择下拉菜单中显示的标签之一以与3D边界框相关联。在另一实施例中,用户可以输入自由形式的文本来为3D边界框标识或区分的对象的类别或分类生成新标签。标记可以是语义的并且可以被机器学习模型用来理解世界或环境中标记的对象的性质。

[0058] 系统然后将与标签相关联的LIDAR点和3D边界框存储(320)到数据存储,用于机器学习模型的训练(上面关于图2描述),和/或用于机器人设备(例如,自主车辆)的控制和操作。

[0059] 虽然这里没有示出,但是上文讨论的各种显示和用户界面可以包括许多可选择的用户控件,这些控件分别对应于许多特征,用户可以经由这些特征辨别、区分和/或标识对象和相关联的LIDAR点,并且执行上文描述用户功能中的一个或多个(例如,绘制3D边界框)。本领域技术人员应当理解,本公开的附图中所示的图像帧的布置和数量仅作为示例;在实践中可以使用其他布置/数字。

[0060] 图9描绘了可以包括在诸如车辆的内部处理系统、外部监视和报告系统或远程服务器的系统的任何电子组件中的内部硬件的示例。电气总线900用作互连硬件的其他图示组件的信息高速公路。处理器905是系统的中央处理设备,其被配置为进行执行编程指令所需的计算和逻辑运算。如本文件和权利要求中所使用的,术语“处理器”和“处理设备”可以指单个处理器或共同执行操作的集合的处理器集合中的任何数量的处理器,诸如中央处理单元(CPU)、图形处理单元(GPU)、远程服务器或这些的组合。只读存储器(ROM)、随机存取存储器(RAM)、闪存存储器、硬盘驱动和其他能够存储电子数据的设备构成存储器设备925的示例。存储器设备可以包括存储数据和/或指令的单个设备或设备集合。本发明的各种实施例可以包括包含编程指令的计算机可读介质,这些编程指令被配置为使一个或多个处理器、印刷设备和/或扫描设备执行在前面附图的上下文中描述的功能。

[0061] 可选的显示接口930可以允许来自总线900的信息以视觉、图形或字母数字格式显示在显示设备935上,诸如在车辆的仪表盘内显示系统上。还可以提供音频接口和音频输出(诸如扬声器)。与外部设备的通信可以使用诸如无线天线、射频标识(RFID)标签和/或短程或近场通信收发器的各种通信设备940发生,其中的每一个都可以可选地经由一个或多个通信系统与设备的其他组件通信地连接。通信设备940可以被配置成通信地连接到通信网

络,诸如因特网、局域网或蜂窝电话数据网络。

[0062] 硬件还可以包括允许从诸如键盘或小键盘、操纵杆、触摸屏、触摸板、遥控器、指示设备和/或麦克风的输入设备950接收数据的用户接口传感器945。还可以从可以捕获视频和/或静止图像的相机920接收数字图像。该系统还可以从诸如加速度计、陀螺仪或惯性测量单元的运动和/或方位传感器970接收数据。该系统还可以从LiDAR系统960接收数据,例如本文档前面所述的数据。

[0063] 上文公开的特征和功能以及替代方案可以组合到许多其他不同的系统或应用中。各种组件可以以硬件或软件或嵌入式软件来实现。本领域的技术人员可以做出各种目前无法预见或未预料到的替代、修改、变化或改进,其中每一个也旨在被所公开的实施例所涵盖。

[0064] 与上文提供的公开内容相关的术语包括:

[0065] 术语“车辆”是指能够运载一名或多名乘员和/或货物并由任何形式的能量提供动力的任何移动形式的交通工具。术语“车辆”包括但不限于汽车、卡车、货车、火车、自主车辆、飞机、无人机等。“自主车辆”是具有处理器、编程指令和可以由处理器控制而无需人类操作员的动力传动系统组件的车辆。自主车辆可以是完全自主的,因为对于大部分或所有驾驶条件和功能,它不需要人类操作员来完成,或者它可以是半自主的,因为在某些条件下或某些操作可能需要人类操作员,或者人类操作员可以凌驾于车辆的自主系统之上并且可以控制车辆。自主车辆还包括自主系统增强了车辆的人类操作的车辆,诸如具有驾驶员辅助转向、速度控制、制动、停车和其他系统的车辆。

[0066] 在本文档中,术语“安装”可以用于指附接到车辆、固定在车辆内或定位在车辆上的传感器或其他组件。

[0067] “电子设备”或“计算设备”是指包括处理器和存储器的设备。每个设备可以具有其自己的处理器和/或存储器,或者处理器和/或存储器可以与虚拟机或容器布置中的其他设备共享。存储器将包含或接收编程指令,其在处理器执行时,使电子设备根据编程指令执行一个或多个操作。

[0068] 术语“存储器”、“存储器设备”、“数据存储”、“数据存储设施”等均指计算机可读数据、编程指令或两者都存储在其上的非暂时性设备。除非另有明确说明,否则术语“存储器”、“存储器设备”、“数据存储”、“数据存储设施”等旨在包括单个设备实施例、多个存储器设备一起或共同存储数据或指令集的实施例以及这样的设备内的各个扇区。

[0069] 术语“处理器”和“处理设备”是指被配置为执行编程指令的电子设备的硬件组件。除非另有明确说明,否则单数术语“处理器”或“处理设备”旨在包括单个处理设备实施例和其中多个处理设备一起或共同执行处理的实施例。

[0070] 术语“标记”是指标识图像中不同对象区并为每个对象区赋予唯一标记值的过程。

[0071] “场景”或“环境”是可以包含其中一个或多个对象可以是感兴趣的一个或多个对象的位置,以便系统可以被配置为在3D传感器数据集中提供感兴趣对象的标识、测量、标记等。例如,场景可以包括一个或多个建筑物、道路、景观元素、车辆等。在另一个非限制性示例中,场景可以包括位于建筑物内部的一个或多个对象,其中一个或多个对象可以包括墙壁、地板、机械装备、窗户、门、门口、家具、固定装置、计算机和相关装备等,所有这些或其中之一可以包括符合本公开的“一个或多个感兴趣对象”。

[0072] “自动化设备”或“机器人设备”是指包括处理器、编程指令和基于来自处理器的命令可以在最少或没有人为干预的情况下执行至少一些操作或任务的一个或多个组件的电子设备。例如,自动化设备可以执行一个或多个自动功能或功能集。这样的操作、功能或任务的示例可以包括但不限于导航、运输、驾驶、交付、装载、卸载、医疗相关处理、建筑相关处理等。示例自动化设备可以包括但不限于自主车辆、无人机和其他自主机器人设备。

[0073] 在本文档中,术语“通信链路”和“通信路径”表示第一设备经由其向一个或多个其他设备发送通信信号和/或从一个或多个其他设备接收通信信号的有线或无线路径。如果设备能够经由通信链路发送和/或接收数据,则设备是“通信连接的”。“电子通信”是指在两个或更多个电子设备之间,无论是通过有线网络还是无线网络以及无论是直接还是经由一个或多个中间设备间接,经由一个或多个信号传输数据。

[0074] 在本文档中,除非特别说明,否则当使用诸如“第一”和“第二”等顺序的相对术语来修饰名词时,这种用法只是为了区分一个项目与另一个项目,而不是要求有顺序的顺序。此外,当项目被描述为与另一项目“相对”不同(诸如相对较短的距离或相对较长的距离)时,相对性的项目表示项目将被相互比较。

[0075] 此外,诸如“垂直”和“水平”或“前”和“后”等相对方位的术语在使用时旨在是指彼此相对而不是绝对的,并且仅指取决于设备的取向的与这些术语相关联的设备的一种可能的方位。当本文档使用术语“前”、“后”和“侧面”来指车辆的区域时,它们指的是相对于车辆的默认行驶区域的车辆区域。例如,汽车的“前”是与靠近车辆的尾灯相比更靠近车辆的前照灯的区域,而汽车的“后”是与靠近车辆的前照灯相比更靠近车辆的尾灯的区域。此外,术语“前”和“后”不一定限于前向或后向区域,而且还分别包括与靠近后部相比更靠近前部的侧面区域,或反之亦然。车辆的“侧面”旨在指车辆最前部和最后部之间的面向侧面的部分。

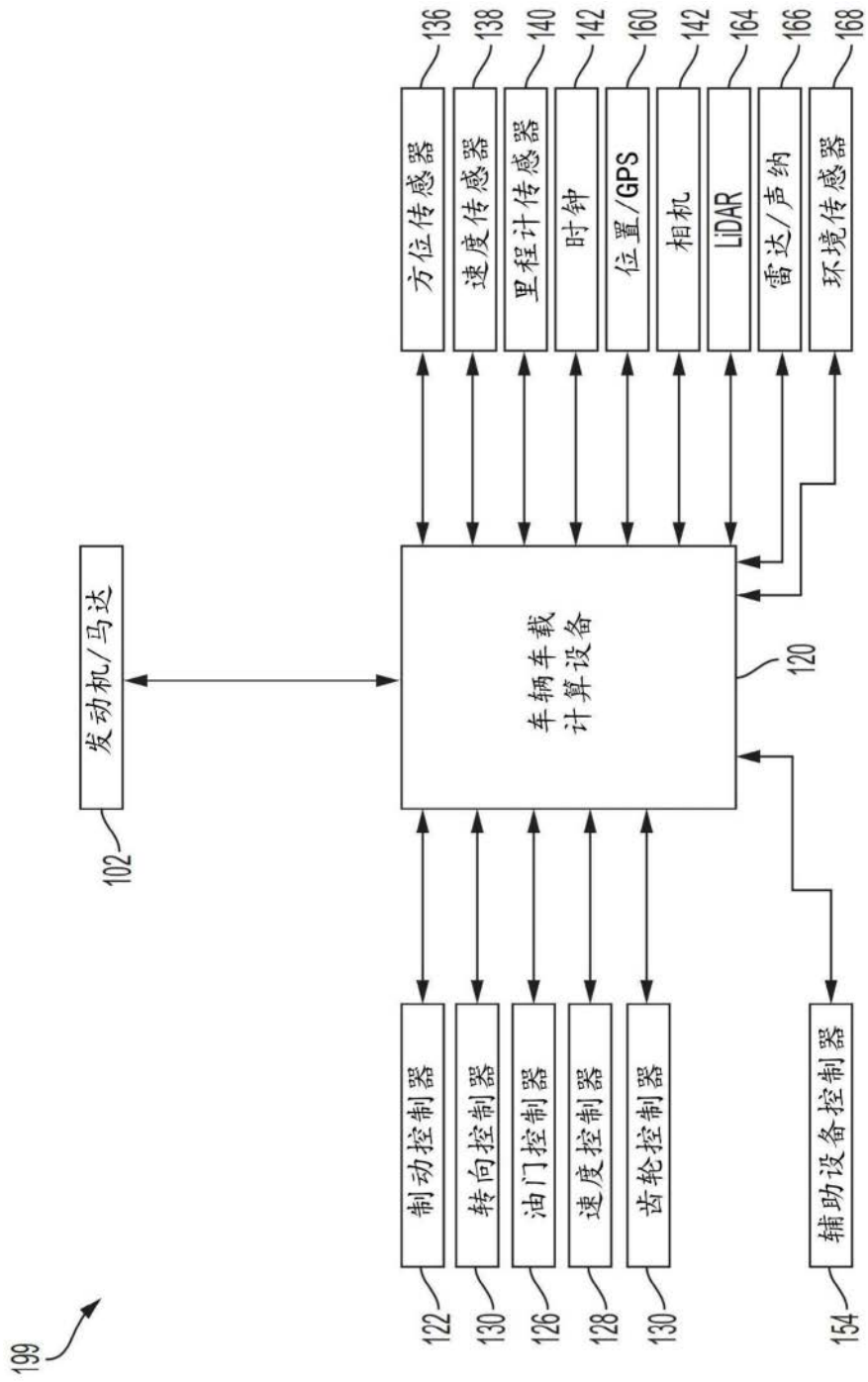


图1

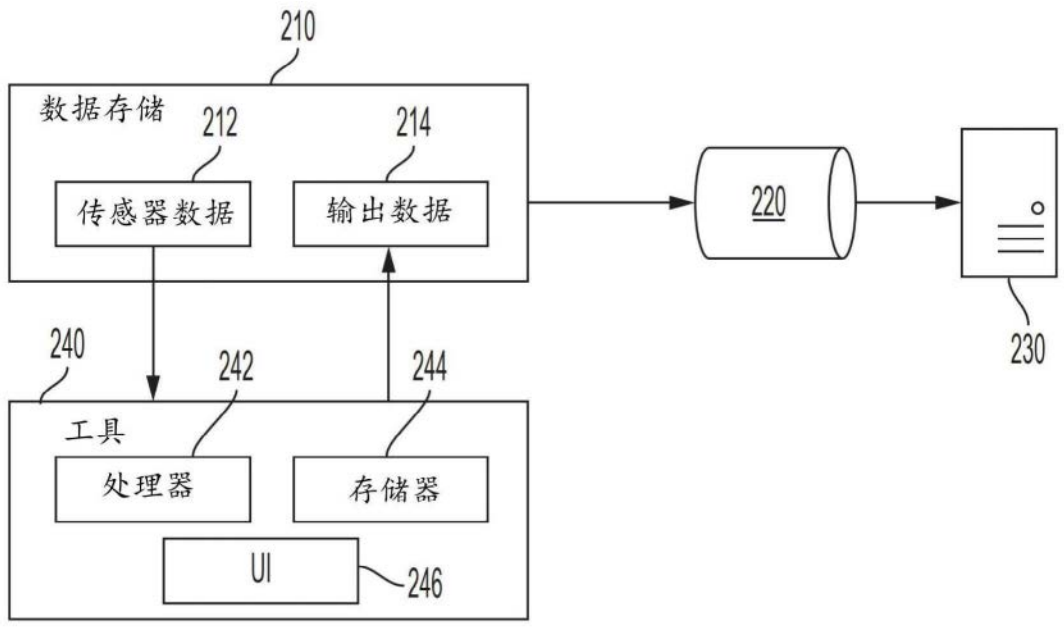


图2

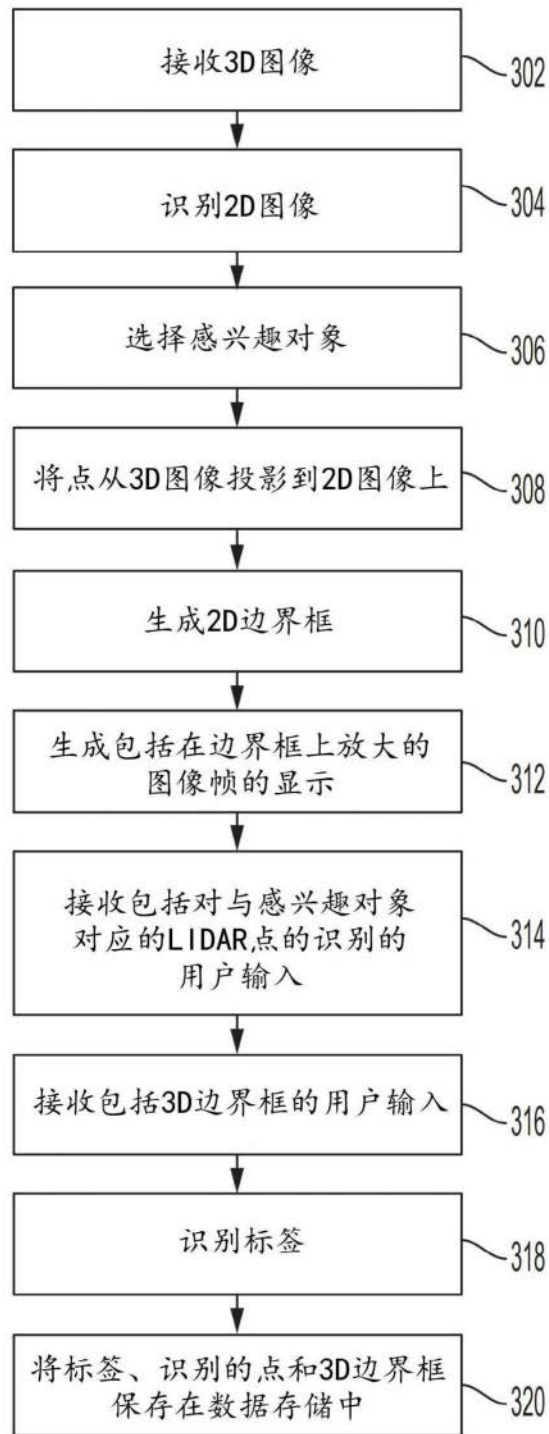


图3

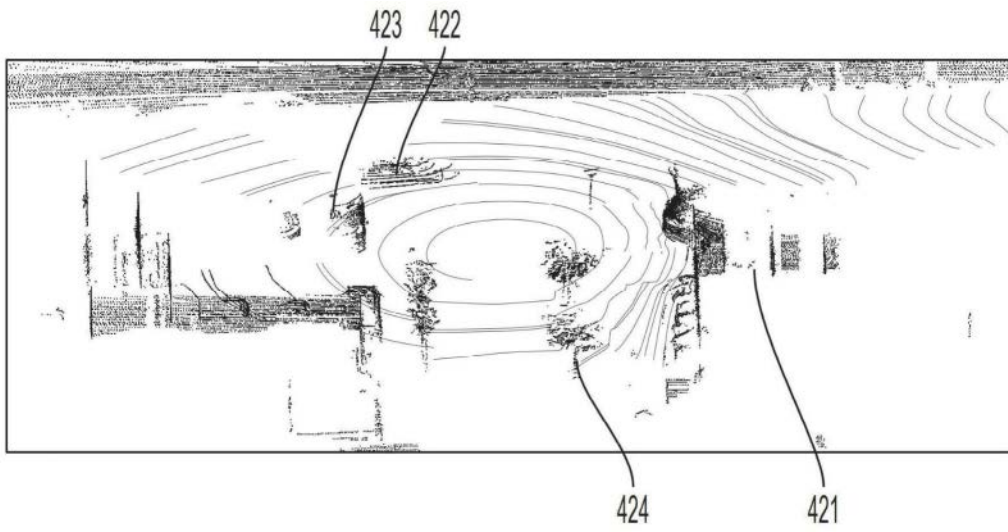


图4A

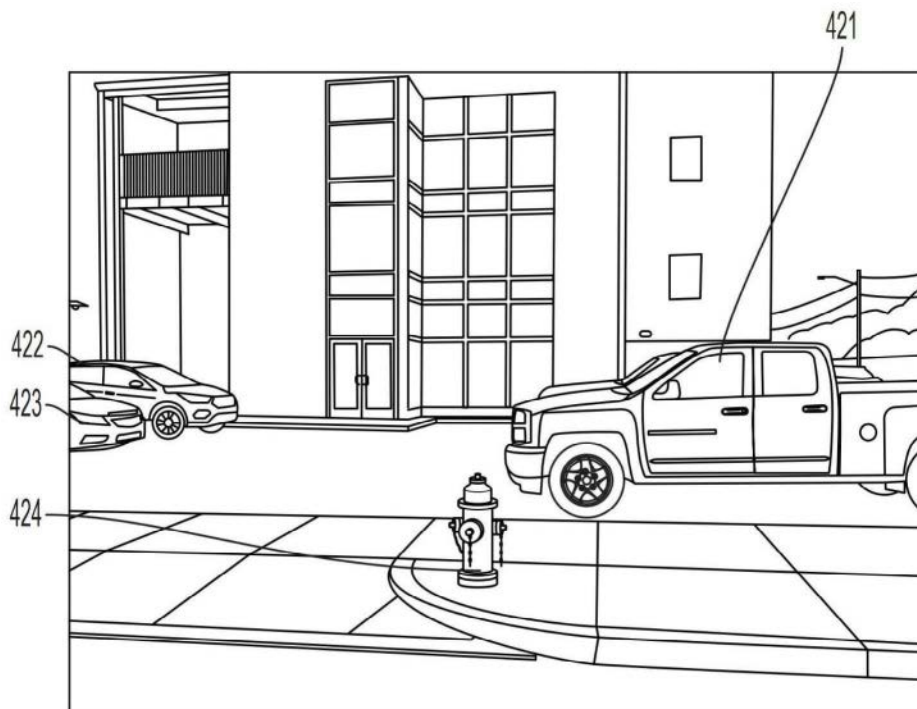


图4B

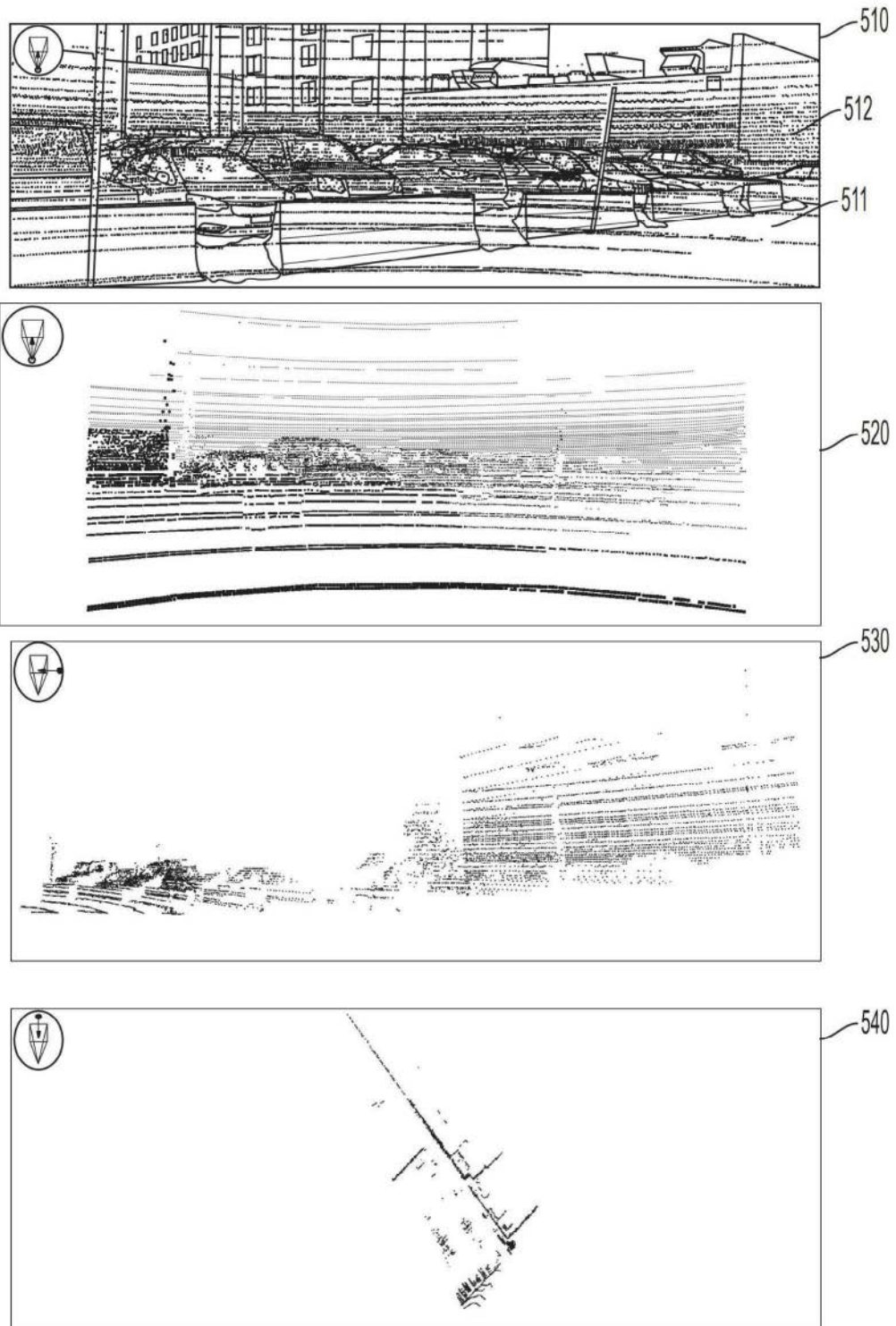


图5

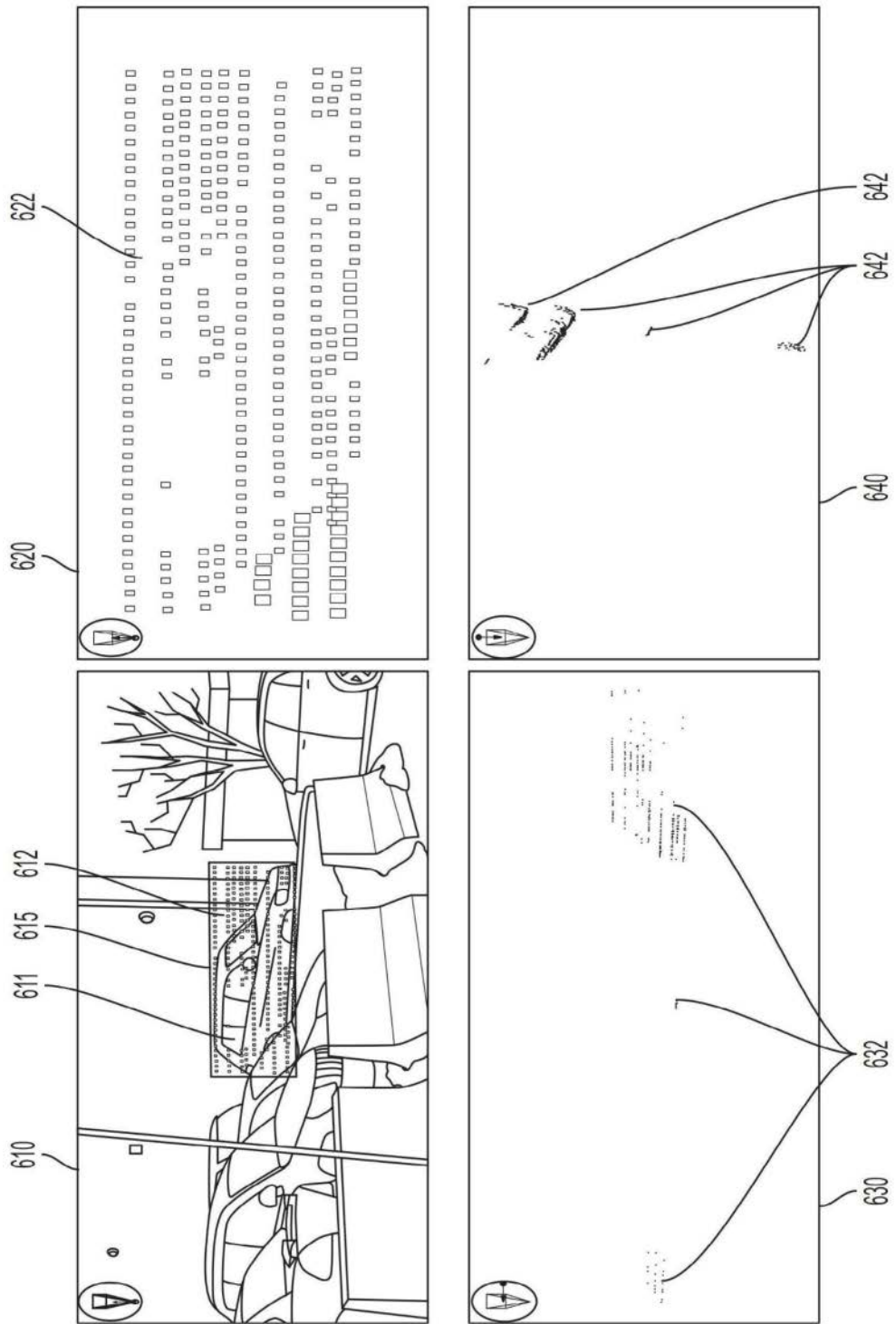


图6

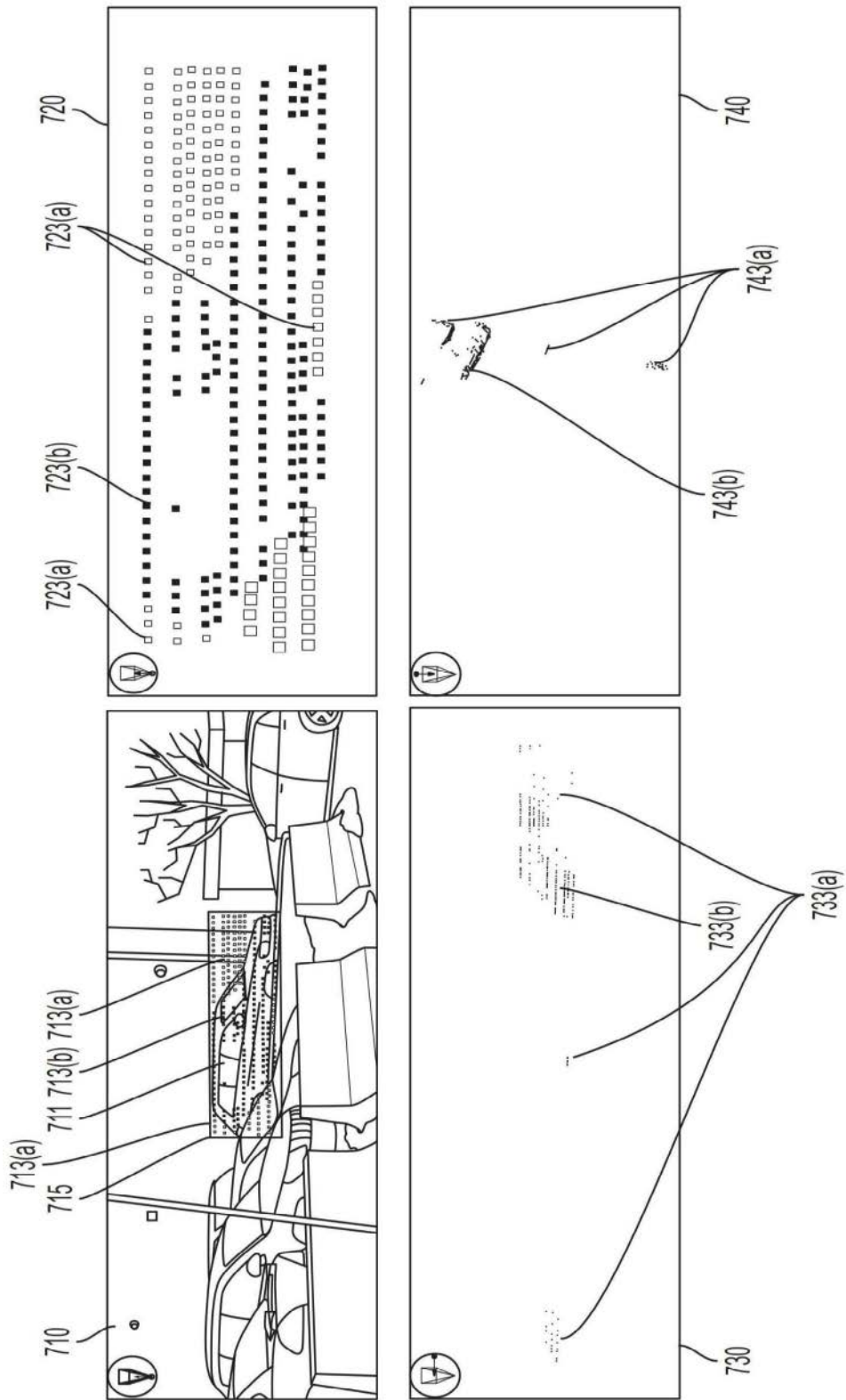


图7

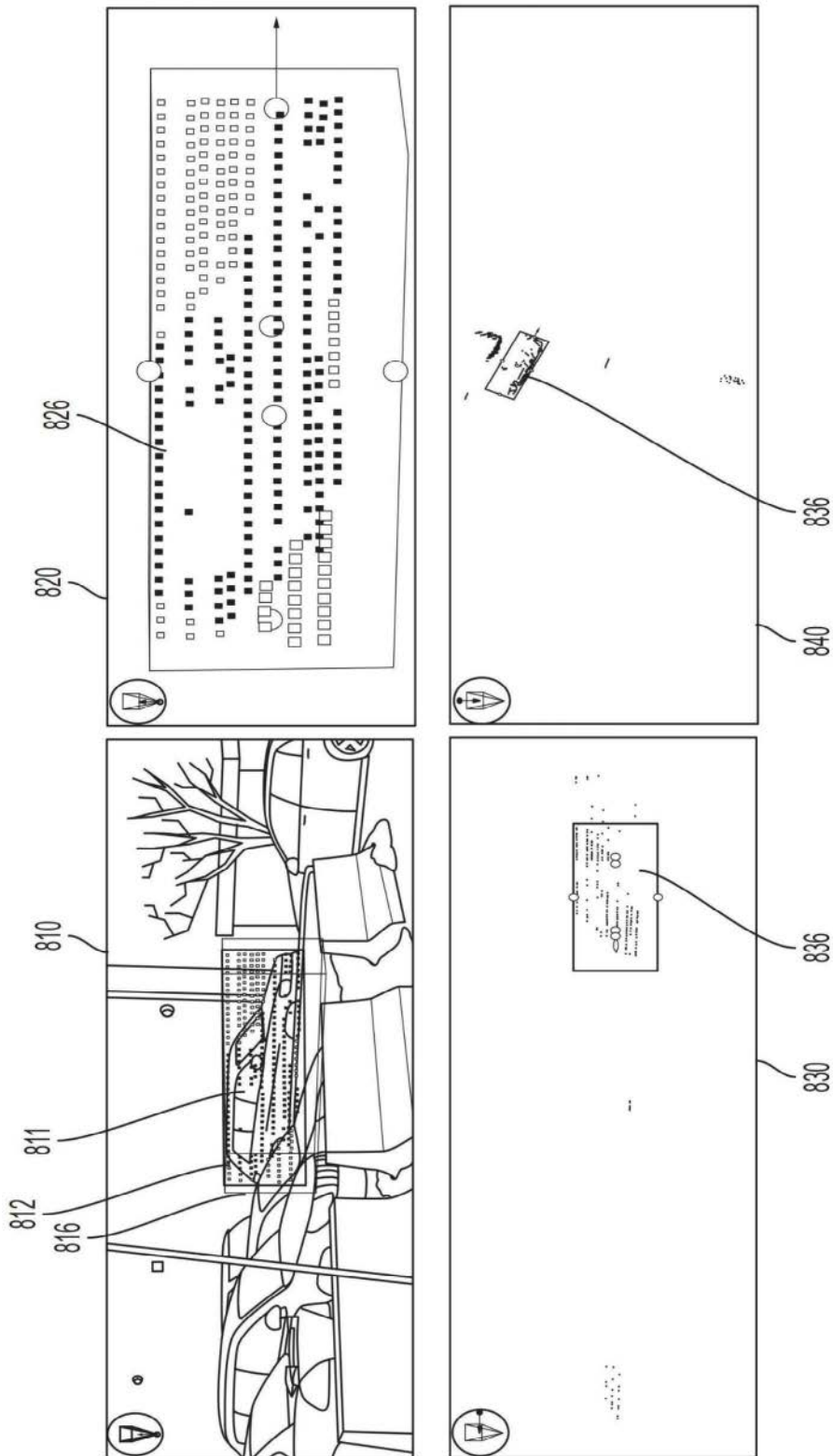


图8

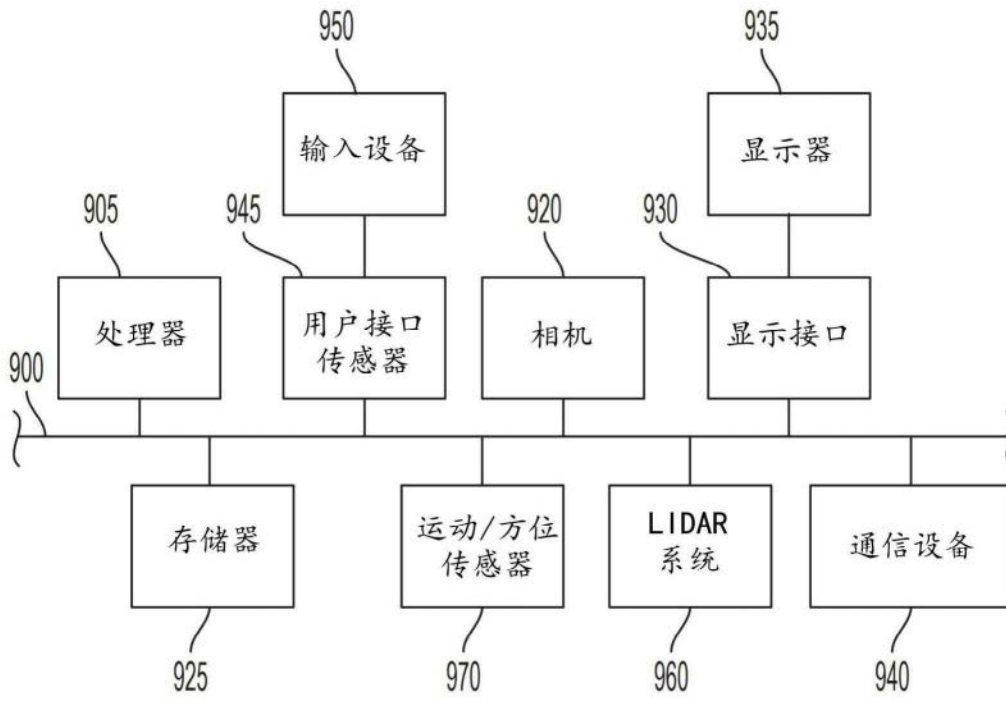


图9