



(12) 发明专利申请

(10) 申请公布号 CN 116468846 A

(43) 申请公布日 2023. 07. 21

(21) 申请号 202310094447.9

G06N 3/08 (2023.01)

(22) 申请日 2023.01.17

G06N 3/0464 (2023.01)

(30) 优先权数据

102022200508.2 2022.01.18 DE

102022214336.1 2022.12.22 DE

(71) 申请人 罗伯特·博世有限公司

地址 德国斯图加特

(72) 发明人 D·塔纳纳伊夫 郭泽

(74) 专利代理机构 永新专利商标代理有限公司

72002

专利代理师 郭毅

(51) Int.Cl.

G06T 17/00 (2006.01)

G06V 10/42 (2022.01)

G06V 10/82 (2022.01)

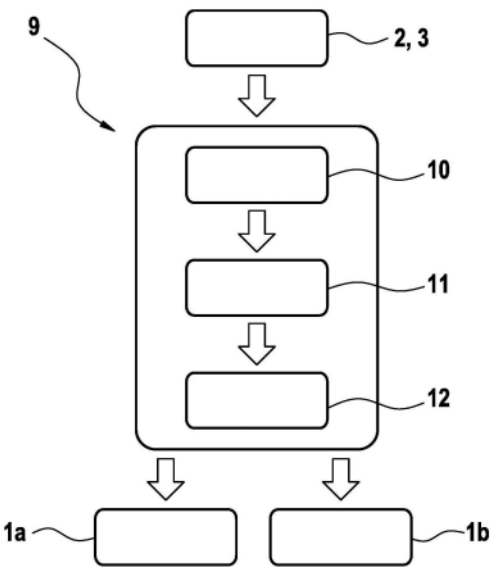
权利要求书1页 说明书10页 附图6页

(54) 发明名称

产生系统环境的鸟瞰视角表示的方法、车辆对象识别系统及存储介质

(57) 摘要

一种用于尤其基于至少一个或者多个数字图像表示 (2) 产生来自对系统周围环境的至少一部分的鸟瞰视角的至少一个表示 (1) 的方法, 所述数字图像表示有利地由该系统、有利地是车辆的至少一个或者多个摄像机获得, 其中, 该方法至少包括以下步骤: a) 获得表示有利地表示唯一的数字图像的数字图像表示 (2), 尤其是连同已拍摄该图像的摄像机的至少一个摄像机参数 (3) 一起, 所述摄像机参数有利地是内在摄像机参数, b) 从该数字图像表示 (2) 中提取至少一个特征 (4), 其中, 有利地以不同的比例尺 (5) 产生特征 (4), c) 将所述至少一个特征 (4) 从图像空间 (6) 变换到鸟瞰视角空间 (7) 中, 有利地以便获得至少一个鸟瞰视角特征 (8)。



1. 一种用于产生来自对系统周围环境的至少一部分的鸟瞰视角的至少一个表示 (1) 的方法, 其中, 所述方法至少包括以下步骤:

- a) 获得数字图像表示 (2),
- b) 从所述数字图像表示 (2) 中提取至少一个特征 (4),
- c) 将所述至少一个特征 (4) 从图像空间 (6) 变换到鸟瞰视角空间 (7) 中。

2. 根据权利要求1所述的方法, 其中, 执行所述方法以用于训练系统和/或深度学习算法, 以便描述围绕系统的3D周围环境的至少一部分。

3. 根据权利要求1或2所述的方法, 其中, 步骤c) 中的变换包括特征压缩。

4. 根据上述权利要求中任一项所述的方法, 其中, 步骤c) 中的变换包括特征膨胀。

5. 根据上述权利要求中任一项所述的方法, 其中, 步骤c) 中的变换包括逆透视投影变换特征产生。

6. 根据上述权利要求中任一项所述的方法, 其中, 步骤c) 中的变换包括对特征的重新采样。

7. 根据上述权利要求中任一项所述的方法, 其中, 步骤c) 中的变换包括特征合并。

8. 根据上述权利要求中任一项所述的方法, 其中, 执行摄像机归一化。

9. 一种计算机程序, 所述计算机程序配置为用于执行根据上述权利要求中任一项所述的方法。

10. 一种机器可读的存储介质, 在所述存储介质上存储有根据权利要求9所述的计算机程序。

11. 一种用于车辆的对象识别系统 (9), 其中, 所述系统配置为用于执行根据权利要求1至8中任一项所述的方法, 和/或所述系统至少包括:

- 多尺度的骨干 (10),
- 鸟瞰视角变换模块 (11),
- 可选的用于特征精细化的模块 (12)。

产生系统环境的鸟瞰视角表示的方法、车辆对象识别系统及存储介质

技术领域

[0001] 本发明涉及一种用于尤其基于至少一个或者多个数字图像表示来产生来自对系统周围环境的至少一部分的鸟瞰视角的至少一个表示的方法,所述数字图像表示有利地由该系统、有利地是车辆的至少一个或者多个摄像机获得。此外,说明一种用于执行该方法的计算机程序和一种具有该计算机程序的机器可读的存储介质。此外,说明一种用于车辆的对象识别系统。

背景技术

[0002] 在先进的驾驶员辅助系统或者自动驾驶系统中,通常使用感知系统,该感知系统提供3D周围环境的表示,并且该表示可以用作用于运动规划系统的输入,该运动规划系统可以决定应如何调度本车辆。该感知系统的关键技术在于,识别车辆可以驶向哪里和围绕该车辆的周围环境看起来如何。使用典型的计算机视觉技术的常规方法是复杂的,因为需要开发许多识别算法并且需要融合步骤,以便获得3D周围环境的概视;该复杂的过程可能也是计算密集的。

发明内容

[0003] 本发明的任务在于,强烈简化相应的方法并且尤其将深度学习的性能用于直接预测可被用于运动规划的最终表示。

[0004] 在本文中根据本发明提出一种用于产生来自对系统周围环境的至少一部分的鸟瞰视角的至少一个表示的方法,其中,所述方法至少包括以下步骤:

[0005] a) 获得数字图像表示,

[0006] b) 从所述数字图像表示中提取至少一个特征,

[0007] c) 将所述至少一个特征从图像空间变换到鸟瞰视角空间中。

[0008] 步骤a)、b)和c)可以按照所说明的顺序如至少一次地和/或重复地被执行以执行该方法。此外,步a)、b)和c)可以至少部分地并行地或同时地执行。该系统例如可以是车辆、例如机动车。该车辆例如可以是汽车。该车辆或者说系统可以设置为用于至少部分自动化的或者自主的(驾驶-)运行。该方法例如可以借助在本文中所描述的系统或者说对象识别系统来执行。

[0009] 该方法尤其用于产生来自对系统周围环境的至少一部分的鸟瞰视角的至少一个图像表示和/或周围环境表示。这尤其基于至少一个或者多个数字图像表示进行。所述数字图像表示有利地能够由该系统的至少一个或者多个摄像机获得。

[0010] 该系统例如可以是车辆、例如机动车。该车辆例如可以是汽车。该车辆或者说系统可以设置为用于至少部分自动化的或者自主的(驾驶-)运行。

[0011] 在步骤a)中,获得数字图像表示。该数字图像表示有利地能够表示或者说是唯一的数字图像。该数字图像表示尤其可以连同至少一个摄像机参数一起或共同被获得。有利

地,所述摄像机参数可以是内在摄像机参数。通常,所述摄像机参数是已拍摄该图像的那些摄像机的参数。

[0012] 在步骤b)中,从该数字图像表示中提取至少一个特征。与此相关地,有利地以不同比例尺产生特征。例如,可以以第一比例尺并且以第二比例尺产生特征,其中,第一比例尺大于或者小于第二比例尺。相同的特征尤其可以以不同比例尺产生。

[0013] 在步骤c)中,将所述至少一个特征从图像空间变换到鸟瞰视角空间中。该图像空间可以是二维的或者三维的空间,该空间可以通过所获得的数字图像表示的光学检测或检测区域来代表。尤其可以是一个或者多个如下摄像机的观察区域或检测区域:该数字图像表示由所述摄像机所获得。该变换优选以如下目标进行:获得至少一个鸟瞰视角特征。该鸟瞰视角特征尤其贡献于从上方描述该周围环境的被观察的场景。鸟瞰视角特征可以包括相对位置元素,用于描述该鸟瞰视角特征相对于该系统的位置。

[0014] 该方法的一种有利实施方式提供用于训练(人工的)贯通式(durchgängige)深度神经网络的新框架,该贯通式深度神经网络的输出可以用于在先进的驾驶员辅助系统中描述围绕本车辆的3D周围环境。贯通式深度神经网络例如也可以被描述为端到端(end-to-end)深度神经网络。

[0015] 根据一种有利构型提出,执行所述方法以用于训练系统和/或深度学习算法,以便描述围绕系统的3D周围环境的至少一部分。例如,可以执行该方法以用于训练贯通式深度神经网络。在此尤其可以是端到端深度神经网络。有利地,这可以是卷积神经网络(简称:CNN)。该方法特别有利地能够被执行用于尤其自动化地生成用于训练人工神经网络或者说算法的训练数据。

[0016] 先进的驾驶员辅助系统的或者自动驾驶系统的感知系统或者说对象识别系统的目标可以是,获得用于进一步的运动规划的所谓的鸟瞰视角(Bird's Eye View,BEV)表示。与此相关地可以有帮助的是,将不同传感器的语义信息和3D信息融合成用于进一步的运动规划的所谓的鸟瞰视角(BEV)表示。与此相关地,根据一种有利的实施方式,可以使用端到端BEV语义地图预测。有利地能够使用编码器-解码器-分割架构,以便直接学习BEV变换。然而,所述方法通常不是一般性的解决方案,因为所述方法通常不能够处理未被看见的摄像机(未出现在训练集中的摄像机图像),所述未被看见的摄像机具有不同摄像机自有的参数。此外,所述方法的性能通常由于架构设计而受限。在本文中所说明的方法能够贡献于解决这些问题。

[0017] 本发明的一种有利实施方式可以包括下述方面中的一个或者多个方面:

[0018] • 可以引入统一的深度神经网络,用于直接预测BEV和对象高度地图/表面高度地图的语义分割。

[0019] • 可以引入新的结构块,用于将特征空间高效地从图像平面变换到BEV平面。

[0020] • 可以引入用于跨越不同摄像机地进行归一化的方法,使得能够将不同摄像机的图像用于训练,并且经训练的模型有利地能够与不同摄像机的图像一起工作。

[0021] 本发明的一种有利实施方式可以包括下述优点中的一个或者多个优点:

[0022] • 尤其是在使用所提出的结构块的情况下,深度神经网络有利地能够高效地从摄像机参数和图像数据中学习视图变换,这有利地为BEV中的语义类别提供良好的预测结果,尤其当在训练阶段中提供足够的数据时。该深度神经网络能够强烈简化感知系统,因为该

深度神经网络能够直接预测BEV地图,这具有如下优点:该系统不具有不同的复杂算法和/或不必要事后将这些算法相互融合。

[0023] • 用于执行本发明的至少一部分的算法有利地能够实现:将具有不同内在参数的不同摄像机的数据组合以用于训练神经网络。该算法使得能够实现将现有的训练数据重新用于新的项目并且有利地节省成本。此外,能够将已经受训练的网络应用到不同摄像机上,而无需重新训练。这有利地降低在开发过程中的开销。

[0024] • 本发明有利地能够对于纯摄像机系统实现与用于包含昂贵的主动式传感器(例如激光雷达、雷达等)的系统相同的自主能力。

[0025] 根据一种优选实施方式,该方法可以包括用于3D周围环境重建和/或运动规划的来自鸟瞰视角的贯通式(端到端)语义地图预测,尤其是在使用深度神经网络的情况下。

[0026] 该方法的一种有利实施方式可以包括下述部分或者步骤中的一个或者多个部分或者步骤:

[0027] • 用于BEV的语义端到端分割和高度预测。

[0028] • 用于BEV预测的唯一且高效的神经网络结构块。

[0029] • 用于使用来自不同摄像机的数据的方法和针对不同摄像机将该算法一般化。

[0030] 该方法的一种有利实施方式可以包括来自鸟瞰视角(BEV)的地面实况的自动化产生。

[0031] 该方法的一种有利实施方式可以包括在鸟瞰视角或者说BEV中的贯通式(端到端)的语义分割和高度预测。

[0032] 根据该方法的产生或者说生成例如可以包括机器的和/或自动化的生成。所述表示可以涉及(所述系统中的)来自鸟瞰视角(英语:Birds-Eye-View;在本文中亦简称为:BEV)的对周围环境的表示。所述表示优选是地面实况表示。替代或累积地,所述表示可以涉及数字的(周围环境-)地图、例如高精度的周围环境地图或者说HD地图(High-Definition map,高清地图)或者用于监测道路基础设施和/或交通基础设施的表示。

[0033] “地面实况”尤其可以包括多个数据组,这些数据组描述用于训练具有机器学习能力的算法和/或具有机器学习能力的系统的基本知识、例如人工神经网络。该基本知识尤其可以涉及充分数量的数据组,以便能够训练相应的算法或相应的系统以用于图像分析处理。

[0034] 在本文中替代或附加地,术语“地面实况”例如可以涉及地面真实、地面实际和/或场地比较(Feldvergleich)。地面实况生成以有利的方式使得能够实现,能够在分析来自该表示的信息时考虑地面实况数据、尤其是地面数据或者说用于描述(周围环境的)表示中的地面(位置和/或走向)的数据。地面实况数据尤其能够提供关于表示中的给定条件和/或维度和/或比例关系的附加信息和/或参考信息。地面实况数据尤其能够贡献于描述:(潜在的)在哪些地点处对象立在地面上或对象与在该表示中可识别的地面接触。地面实况数据例如能够贡献于:能够更具体地检测或者描述表示中的(参考-)对象。地面实况数据尤其能够贡献于:能够更精确地分类来自表示的信息和/或能够在正确性方面检查该分类的结果。因此,地面实况数据特别有利地能够贡献于对具有机器学习能力的算法和/或具有机器学习能力的系统、尤其是人工神经网络的训练。

[0035] 根据另一有利构型提出,在步骤c)中的变换包括特征压缩。尤其可以从所提取的

图像特征中的每个特征中首先沿着高度轴压缩、尤其是通过连续的卷积层压缩特征,所述卷积层有利地沿着高度轴具有步长2 (或者 2^N)。

[0036] 根据另一有利构型提出,在步骤c)中的变换包括特征膨胀。尤其是在经压缩的特征矢量的情况下,接下来的步骤可以在于,沿着高度轴扩展特征,以便在鸟瞰视角中产生相应的特征。为了实现这一点,有利地能够预先将以真实的米为单位的深度范围(高度轴)定义为超参数。

[0037] 根据另一有利构型提出,在步骤c)中的变换包括逆透视变换特征产生(Kartierungs-Merkmalserzeugung)。逆透视变换(Inverses perspektivisches Mapping, IPM)是有利地能够被用于将图像投影到鸟瞰视角上、尤其在假设平坦的地平面的情况下的方法。

[0038] 根据另一有利构型提出,在步骤c)中的变换包括对特征的重新采样。双线性采样尤其可以用于图像栅格或者说图像网格的重新采样。

[0039] 根据另一有利构型提出,在步骤c)中的变换包括特征合并。鸟瞰视角特征尤其可以在像素网格中被重新采样(或者说resampled),并且可以全部具有相同的形状,这些鸟瞰视角特征可以连同IPM特征一起被合并(加和)成最终的鸟瞰视角特征。

[0040] 根据另一有利构型提出,执行摄像机归一化。该摄像机归一化尤其可以根据至少一个摄像机参数来执行。该摄像机归一化尤其可以以如下目的被执行:使得该方法能够与来自(具有不同内在参数的)不同摄像机的图像一起工作。

[0041] 根据另一方面,提出一种用于执行在本文中所提出的方法的计算机程序。换言之,这特别涉及一种计算机程序(产品),该计算机程序(产品)包括指令,所述指令在通过计算机执行该计算机程序(产品)时促使该计算机实施本文所描述的方法。

[0042] 根据另一方面,提出一种机器可读的存储介质,在该存储介质上保存或存储有在本文中所提出的计算机程序。该机器可读的存储介质通常是计算机可读的数据载体。

[0043] 根据另一方面,说明一种用于车辆的对象识别系统,其中,该系统配置为用于执行在本文中所描述的方法,和/或该系统至少包括:

[0044] -多尺度的骨干,和

[0045] -鸟瞰视角变换模块,和

[0046] -可选的用于特征精细化的模块。

[0047] 该系统或者说对象识别系统例如可以包括计算机和/或控制器(Controller),该计算机和/或控制器能够执行指令以便实施该方法。为此,该计算机或者说该控制器例如可以执行所说明的计算机程序。该计算机或者说该控制器例如可以访问所说明的存储介质以便能够执行该计算机程序。

[0048] 结合该方法讨论的细节、特征和有利构型也可以相应地出现在本文中所提出的计算机程序和/或存储介质和/或对象识别系统中,反之亦然。就此而言,全面参考那些用于进一步表征这些特征的实施方案。

附图说明

[0049] 下面根据附图更详细地阐述本文所提出的解决方案及其技术领域。应该指出:本发明并不应受所示出的实施例的限制。特别地,只要无明确地另外示出,也可以提取附图中

阐述的事实的部分方面,并将其与出自其他附图和/或本说明书中的其他的组成部分和/或知识组合。示意性示出:

[0050] 图1:在本文中所提出的方法的示例性流程。

[0051] 图2:在本文中所描述的对象识别系统的示例。

[0052] 图3:该方法的一种示例性应用的示图。

[0053] 图4:该方法的一个示例性方面的示图。

[0054] 图5:该方法的一个示例性方面的示图。

[0055] 图6:该方法的一种示例性应用的示图。

[0056] 图7:该方法的一个示例性方面的示图。

[0057] 图8:该方法的一个示例性方面的示图。

[0058] 图9:该方法的一个示例性方面的示图。

具体实施方式

[0059] 图1示意性示出在本文中所提出的方法的示例性流程。该方法用于尤其基于至少一个或者多个数字图像表示2产生来自对系统周围环境的至少一部分的鸟瞰视角的至少一个表示1,所述数字图像表示有利地由该系统、有利地是车辆的至少一个或者多个摄像机获得。步骤a)、b)和c)的借助方框110、120和130示出的顺序是示例性的,并且可以按所示顺序遍历例如至少一次,用以执行该方法。

[0060] 在方框110中,根据步骤a),获得有利地表示唯一的数字图像的数字图像表示2,尤其是与已拍摄该图像的摄像机的至少一个摄像机参数3一起获得,所述摄像机参数有利地是内在的摄像机参数。

[0061] 在方框120中,根据步骤b),从该数字图像表示2中提取至少一个特征4,其中,特征4有利地以不同的比例尺5产生。

[0062] 在方框130中,根据步骤c),至少一个特征4从图像空间6变换到鸟瞰视角空间7中,有利地以便获得至少一个鸟瞰视角特征8。

[0063] 图2示意性示出在本文中所提出的用于车辆的对象识别系统9的一种实施变型的示例性流程。该系统配置为用于执行在图1中描述的方法。该系统包括多尺度骨干10、鸟瞰视角变换模块11和用于特征精细化的模块12。

[0064] 与此相关地,图2示意性示出针对概视具有用于变换到鸟瞰视角中的深度神经网络的系统的示例。

[0065] 例如可以将唯一的数字图像2作为输入提供给系统9。图像2可以与摄像机参数3共同地被如下摄像机提供:图像2借助该摄像机所拍摄。系统9输出周围环境的至少一部分的来自鸟瞰视角的至少一个表示1作为输出。所述输入和输出可以是神经网络的相应的输入和输出。所述输出在本文中例如可以是语义分割地图的表示1a以及具有所估计的对象高度的高度地图的、分别在鸟瞰视角中的表示。

[0066] 图3示出该方法的一种示例性应用的示图。尤其示出针对图2中的系统的输入和输出的真实示例。因此,图3示出针对深度神经网络的输入和输出的示例,该深度神经网络用于根据本发明的方法变换到鸟瞰视角中。

[0067] 尤其当该方法应基于受监控的学习时,通常需要用于深度神经网络的训练阶段的

标记数据或标签数据。如下标记数据是有利的：

[0068] • 在BEV中或在鸟瞰视角中的语义分割地图

[0069] • 在BEV中或在鸟瞰视角中的高度地图

[0070] 用于相应的标签数据的示例同样能够在图3中看到。

[0071] 标签数据有利地能够从贴有语义标签的点云、相应的摄像机图像和/或传感器位置信息中获取。该方法/该算法的输入可以是：单个图像+摄像机参数。该方法/该算法的输出可以是：在BEV中的语义分割地图和/或对象高度地图/表面高度地图。

[0072] 在图2中能够看到示例性架构的概视。在图3中示出针对该系统的结果的示例。

[0073] 在一种优选实施方式中，深度神经网络可以直接从鸟瞰视角来预测语义分割地图1a和/或用于该分割地图中的每个像素的相应的高度地图1b。

[0074] 尤其根据一种优选实施方式的深度神经BEV网络可以包括如下内容：

[0075] • 多尺度骨干10，

[0076] • BEV视图变换模块11，

[0077] • 用于特征精细化的模块12。

[0078] 多尺度骨干10可以是或者包括特征提取器（例如神经卷积网络），该特征提取器可以采纳图像2作为输入并且有利地能够以不同比例尺产生（高级别的）特征，例如输入尺寸的1/8、1/16、1/32、1/64。尤其可以将神经网络架构用作骨干，例如特征金字塔网络（FPN）和/或初始网络（Inception Network）。在图4中、尤其是在图4的左侧示出针对骨干结构的示例。从该数字图像表示2中提取的特征4以不同的比例尺5示出。并行地，在右侧示出针对从特征4中产生的BEV特征8的相应的不同比例尺5。在比例尺5之间可以分别进行上采样（Upsampling）。

[0079] 因此，图4示意性示出针对检测多尺度特征的骨干结构10的示例。

[0080] 尤其可以将多尺度特征4中的每个多尺度特征馈入到BEV视图变换模块11中（其有利的实施方式在较后面的下文详细描述），以便获得BEV特征8。在图5中示出BEV视图变换模块11的示例性概视。

[0081] 所获得的BEV特征可以是用于特征精细化的模块12的输入，该模块可以包括卷积层+堆栈归一化+激活（例如Leaky ReLU）的级联或者ResNet块，所述ResNet块可以进一步精细化BEV特征8。此外，在模块12中，各个鸟瞰视角特征8可以被组合成一个特征（在全鸟瞰视角中的合并BEV特征）。

[0082] 尤其可以从所精细化的BEV特征8中创建两个任务头：

[0083] • 形式为 $h_BEV \times w_BEV \times C$ 的分割头（C是类别的数量）

[0084] • 形式为 $h_BEV \times w_BEV \times 1$ 的高度头

[0085] 因此，图5示意性示出针对从多尺度图像特征4到BEV特征8的变换的示例。

[0086] 该有利的实施方式可以根据单个的（前）摄像机视图的如下示例来描述：如果仅观察一个摄像机视图、例如其前摄像机视图，则BEV地面实况可以覆盖例如宽度为40m且长度为60m的区域，其中，像素网格分辨率为例如0.1m/Pixel，即，BEV地面实况地图可以具有例如以像素为单位的 400×600 （ $40/0.1, 60/0.1$ ）的形状。深度神经网络的输出形状例如可以是用于高度地图的 $400 \times 600 \times 1$ 和用于分割地图的 $400 \times 600 \times C$ ，其中，C是语义类别的数量。为了获得最终的类别索引地图，可以沿着类别轴应用argmax运算。

[0087] 该方法的一种有利实施方式可以包括有利地唯一且高效的用于BEV预测的神经网络结构块。

[0088] 与此相关地的特别有利的结构块可以是BEV视图变换模块11,该BEV视图变换模块11可以将特征从图像特征空间6中变换到鸟瞰视角的特征空间7中。该变换的输入可以是:来自骨干网络10的多尺度图像特征4。该变换的输出可以是:BEV特征8。

[0089] 在图5中示出BEV视图变换模块11的示例性概视。在图6中示出可能的应用示例。

[0090] 图6示出针对产生BEV地面实况结果的示例。

[0091] 图6a示出原始RGB图像,该原始RGB图像可以作为数字图像表示2形成该方法的输入。图6b示出语义分割BEV地图;图6c示出高度地图。所有这些都在前摄像机的视图中。分割地图和高度地图是用于待借助该方法产生的来自鸟瞰视角的表示1的有利示例。

[0092] 图7示意性示出针对BEV视图变换或BEV视图变换模块11的概视的示例。

[0093] 如该模块11的名字所示,该模块的目的在于,将从图像(图像空间6)中获取的特征4变换到鸟瞰视角的空间7中,使得网络优选可以学习更好的特征8,所述更好的特征导致更好的性能。

[0094] 鸟瞰视角变换模块11或者说BEV视图变换模块11和/或BEV变换的一种特别有利的实施方式可以包括如下步骤/部分中的至少一个或者多个或者所有步骤/部分:

- [0095] • 特征压缩
- [0096] • 特种扩展
- [0097] • 逆透视的变换特征产生
- [0098] • 对特征的重新采样
- [0099] • 特征合并

[0100] 变换可以包括特征压缩(英语:feature condensing)。

[0101] 尤其从来自骨干的多尺度特征中的每个特征可以首先沿着高度轴压缩特征,尤其是通过连续的卷积层,所述卷积层有利地沿着高度轴具有步长2(或者 2^N)。在附图6中示出特征压缩的有利概视。该示例的参数如下:特征尺寸: $C \times 64$ (高度) $\times 128$ (宽度)(在步长为4、具有2个卷积层的情况下) $\rightarrow C \times 16 \times 128 \rightarrow C \times 4 \times 128$ 。

[0102] 在图7中在左上方,根据高度H到H1和H2的压缩,借助箭头11a示出针对特征压缩的示例。

[0103] 该变换可以包括特征膨胀(feature splatting,特征抛雪球)。

[0104] 尤其是在经压缩的特征矢量的情况下,接下来的步骤可以在于,沿着高度轴扩展特征,以便在鸟瞰视角中产生相应的特征。为了实现这一点,有利地能够预先将以真实的米为单位的深度范围(高度轴)定义为超参数,例如0-60m。在例如为0.1m/Pixel的预定义的像素网格分辨率的情况下,以像素为单位的深度范围(Z)例如可以计算为 $(range_max - range_min) / pixel_grid_resolution$,在上述示例中即 $(60 - 0) / 0.1 = 600$ 。

[0105] 当定义以像素为单位的深度范围(Z)时,特征抛雪球的目的在于,恢复以Z为单位的经压缩的特征地图的高度维度,其方式是,首先执行 1×1 卷积,然后执行变形运算,例如:

[0106] 目标: $C \times 4 \times 128 \rightarrow C \times Z \times 128$

[0107] 具有滤波器尺寸 $C \times Z \times 1/4$: $(C \times Z \times 1/4) \times 4 \times 128$ 的 1×1 卷积

[0108] 变形: $(C \times Z \times 1/4) \times 4 \times 128 \rightarrow C \times Z \times 128$

[0109] 在图7中在右上方借助箭头11b示出特征抛雪球的示例性概视,例如通过从H或H2向Z膨胀。

[0110] 该变换可以包括逆透视的变换特征产生 (IPM feature generation)。

[0111] 逆透视变换 (IPM) 是一种有利地能够使用的方法,以便将图像投影到鸟瞰视角上,尤其在假设平坦的地平面的情况下。因此能够在 (几乎) 平坦的面的情况下实现合理的结果,但是一旦该面具有有影响的高度 (例如在汽车的情况下), 则该结果可能以强烈畸变的方式显现。

[0112] 在图7中在左下方借助箭头11c示出IPM变换的一种示例性应用。尤其是从维度H、W、C向Z、X、C。

[0113] 在该方法的框架中,IPM有利地能够被应用到每个多尺度特征4上,以便从图像平面6变换到BEV平面7中。然而,地面在现实中不总是平坦的,使得在所得的特征中可能产生误差。因此,可以在产生IPM特征之后添加一个卷积层 (或者多个层)。由于整个过程有利地是可微分的,所以网络可以进行学习以补偿该误差。以这种方式,IPM特征能够如先前的特征那样起作用,并且引导该网络来创建更好的最终的BEV特征。

[0114] 在图8中示出逆透视变换特征产生 (IPM) 在真实情况中的应用的示例。

[0115] 该变换可以包括对特征的重新采样 (英语:feature re-sampling)。

[0116] 如上文针对特征膨胀或“特征抛雪球”所提到的那样,BEV像素网格可以基于以米为单位的宽度 (X) 和深度 (Z) 和像素网格分辨率 ($r, m/Pixel$) 来定义。以像素为单位的栅格尺寸可以是 ($X/r, Z/r$)。

[0117] 在摄像机的示例性的内在矩阵 ($M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$) 的情况下,可以执行重新采样,

以便将特征值从BEV特征空间 ($Z \times W \times C$) 映射到BEV网格空间或者说鸟瞰视角网格空间 ($Z \times X \times C$) 中。

[0118] $\begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Z \end{bmatrix}, x = \frac{f_x X}{Z} + c_x$

[0119] 双线性采样可以被用于栅格或网格的重新采样。

[0120] 在图7中间借助箭头11d示出对特征的重新采样的示例。

[0121] 变换可以包括特征合并 (feature merging)。

[0122] BEV特征可以在像素网格中重新采样或者说重采样 (resampled), 并且可以全部具有相同的形状,所述BEV特征可以连同IPM特征一起被合并 (相加) 成最终的BEV特征8。在图7中借助箭头11e示出针对这一点的示例。

[0123] 所合并的BEV特征8可以作为用于任务头的分割和高度估计的输入用于最终的预测。

[0124] 图8示意性示出逆透视变换特征产生 (IPM) 的在真实情况中的应用的示例。箭头11c1示出真正的IPM变换可以包括变换步骤 (在本文中借助箭头11c1示出) 和重新采样步骤 (在本文中借助箭头11c2) 示出。

[0125] 例如,该方法可以包括摄像机归一化,尤其是根据至少一个摄像机参数3。

[0126] 该方法的一个特别有利的方面,该方法可以与来自 (具有不同内在参数的) 不同摄

像机的图像一起进行训练/工作。

[0127] 在不同的自主的移动式机器人系统或者自动驾驶的汽车的情况下CNN (Convolutional Neural Network, 卷积神经网络)的可能的性能下降的主要起因可能是训练数据与来自现场的传感器数据之间的空缺。即使训练数据已被移动式机器人系统的传感器收集,但是由于传感器位置的误差和不准确的安装,在相似的机器人的情况下该性能可能降低。该摄像机的为止可以与其外在参数连接,所述外在参数是x位置、y位置和z位置以及滚动角、俯仰角和偏航角。在内在的系数和/或畸变系数方面的轻微区别和/或在摄像机的投影模型(例如鱼眼、针孔)方面的区别,可以提高CNN的复杂性,以便该CNN可以在所有这些情况中良好地一般化。

[0128] 该方法能够有助于降低多摄像机系统的复杂性。尤其可以引入虚拟摄像机,该虚拟摄像机具有例如固定的内在模型、畸变模型、外在模型和/或摄像机模型,和/或可以将所有传感器摄像机重新投影到给定的虚拟摄像机上。

[0129] 一个有利方面可以是,应对不同的摄像机内部的或内在的参数3。

[0130] 如在上文的算法中提到的那样,尤其是摄像机的焦距可以影响BEV视图中的深度范围。这意味着,可以根据摄像机的图像来训练的网络通常不能够在输入图像上产生正确深度,所述输入图像来源于具有不同焦距的其他摄像机。在一种有利的扩展方案中,该方法的目的尤其在于,解决该问题并且有利地实现如下要点中的至少一个或者两个要点:

[0131] • 借助不同摄像机的图像来训练

[0132] • 在不同摄像机的图像上预测有说服力的结果

[0133] 在附图9中示出该方法的示例性概视。与此相关地,图9示出归一的焦距和特征的形状的定向的使用的概视,其中,归一的焦距: $f_c = f_2$,重构形因数: f_c/f 。

[0134] 在该示例中,在方框910中,可以获得具有维度 $H \times W$ (图像表示2)和焦距 f_1 (摄像机参数3)的第一图像。在方框920中,可以获得具有维度 $H \times W$ 和焦距 $f_2 = f_1/2$ 的第二图像。在方框930中,第一图像可以变形或重构形为维度 $H/2 \times W/2$,其具有归一化的焦距 f_c 。在方框940中,第二图像可以保留其维度 $H \times W$,并且给第二图像配属归一化的焦距 f_c 。在方框950中,可以在骨干中对两个图像进行特征提取。此外,在方框950中也可以借助滚动定向层将图像定向。在方框960中,针对第一图像,输出具有维度 $h_f \times w_f$ 的特征。在方框970中,针对第二图像,输出具有维度 $h_f \times w_f$ 的特征。

[0135] 尤其可以使用归一的焦距(f_c),并且可以相对于该焦距将输入图像归一化,即,所述输入图像的尺寸以因数 f_c/f 改变,其中, f 是相应使用的摄像机的焦距。尺寸变化可以导致用于网络的不同的输入图像形状。为了补偿比例尺区别,可以使用滚动-定向层或者说滚动定向的层,以便补偿特征形状,即,尽管输入图像形状不同,最终所提取的特征地图或特征表示有利地能够总是具有相同的形状。

[0136] 一个有利的方面可以是应对不同的摄像机转动。相应的方法可以包括如下所述的步骤:

[0137] 该方法可以包括对旋转补偿的计算。

[0138] 尤其是在给定的初始摄像机转动 $roll_{raw}$ 、 $pitch_{raw}$ 、 yaw_{raw} 的情况下,能够补偿该摄像机的旋转,以便在训练数据集中获得该摄像机的精确旋转 $roll_{correct}$ 、 $pitch_{correct}$ 、 $yaw_{correct}$ 。原始摄像机的定向尤其可以作为旋转矩阵 $world_T_raw_cam \in R^{3 \times 3}$ 示出,并且正

确的定向可以作为 $\text{world_T_correct_cam} \in \mathbb{R}^{3 \times 3}$ 示出,然后从原始摄像机到正确的摄像机的旋转可以如下所述地进行:

[0139]	$\text{correct_cam_T_raw_cam} = \text{inv}(\text{world_T_correct_cam}) * \text{world_T_raw_cam}$	(1)
--------	--	-----

[0140] 在此, $\text{correct_cam_T_raw_cam} \in \mathbb{R}^{3 \times 3}$ 表示该摄像机从原始取向到正确取向的变换, $\text{inv}()$ 表示相应于逆矩阵运算,*表示点乘运算。

[0141] 该方法可以包括如下射束的求取:所述射束相当于任意原始摄像机。

[0142] 尤其可以将原始的摄像机畸变模块表示为 $\text{raw_distortion_model}$ 。该模型可以从未畸变的图像中获得归一化的图像坐标($z=1$)作为输入,并且为畸变图像提供相应的坐标。逆畸变模型 $\text{inv_raw_distortion_model}$ 尤其是同时可以获得用于畸变图像的归一化图像坐标($z=1$),并且提供在未畸变的图像上的相应位置。投影模型尤其是可以表示为 $\text{raw_projection_model}$ 。该模型可以将射束从3D空间投影到2D图像上。逆投影模型尤其是同时可以被称为 $\text{inv_raw_projection_model}$,该逆投影模型能够获得2D图像坐标并将这些2D图像坐标投影到3D空间中。原始摄像机内在属性($\text{Rohkameraintrinsik}$)可以被称为 raw_intrinsic :

[0143] 为了找到3D射束,可以执行如下内容:

[0144]	$\text{raw_3d_rays} = \text{inv_raw_projection_model}(\text{inv_raw_distortion_model}(\text{inv}(\text{raw_intrinsic}) * \text{pixels_coordinates}))$	(2)
--------	---	-----

[0145] 该方法可以包括旋转补偿。

[0146]	$\text{3d_rays_correct} = \text{correct_cam_T_raw_cam} * \text{raw_3d_rays}$	(3)
--------	--	-----

[0147] 该方法可以包括到虚拟正确摄像机上的投影。

[0148] 尤其可以将正确的摄像机畸变的模型称为 $\text{correct_distortion_model}$ 。该模型可以从未畸变的图像中获得归一化的图像坐标($z=1$)作为输入,并且提供畸变图像的相应坐标。该投影模型尤其可以表示为 $\text{correct_projection_model}$ 。该模型可以将射束从3D空间投影到2D单位射束($z=1$)上。正确的摄像机内在属性可以被称为 correct_intrinsic 。正确虚拟摄像机图像可以以如下方式创建:

[0149]	$\text{correct_image} = \text{correct_intrinsic} * \text{correct_distortion_model}(\text{correct_projection_model}(\text{3d_rays_correct}))$	(4)
--------	--	-----

[0150] 经修正的图像有利地能够具有如在训练时间中的摄像机那样精确的内在的和外在的畸变模型和投影模型,因此,有利地能够减少域空缺(domain gap),尤其是不仅在相同的摄像机类型(例如针眼)的情况下,而且有利地跨越不同的摄像机几何形状类型(例如鱼眼、全向摄像机等)。

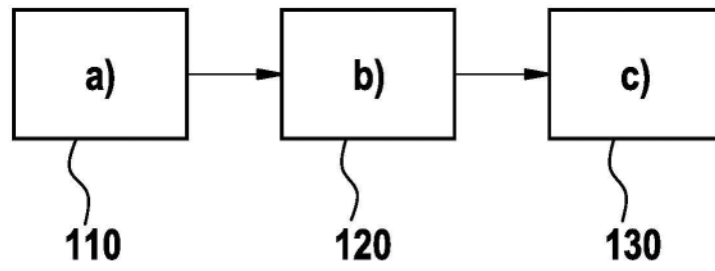


图1

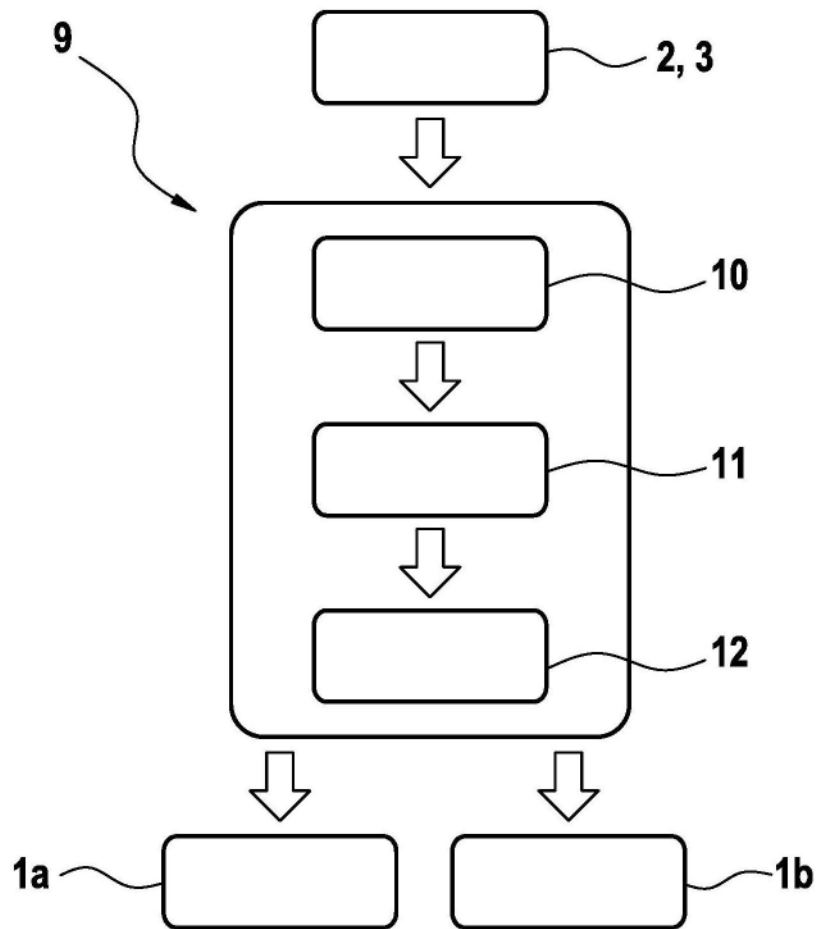


图2

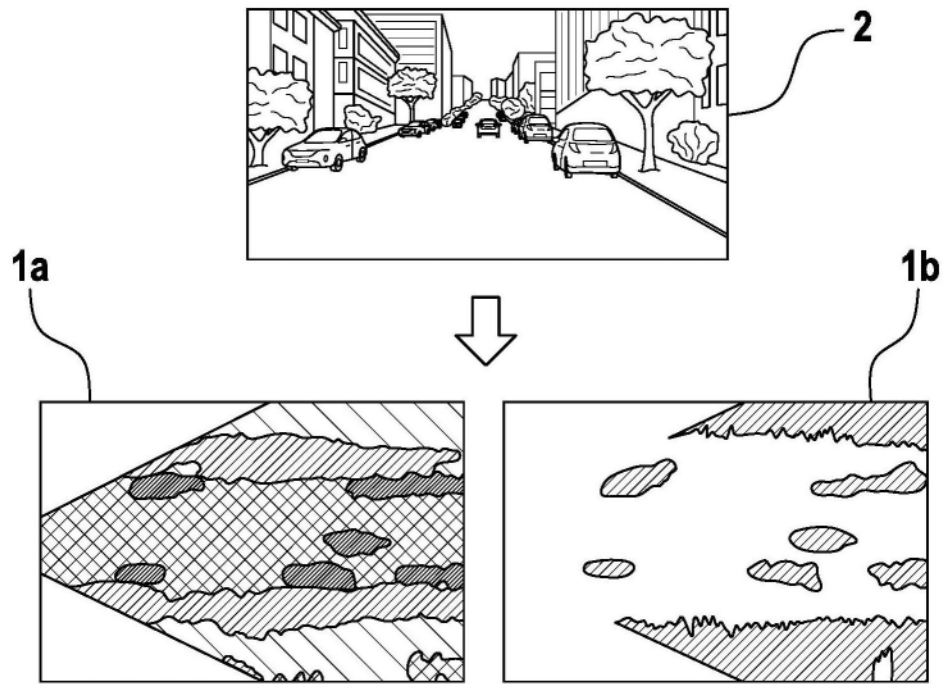


图3

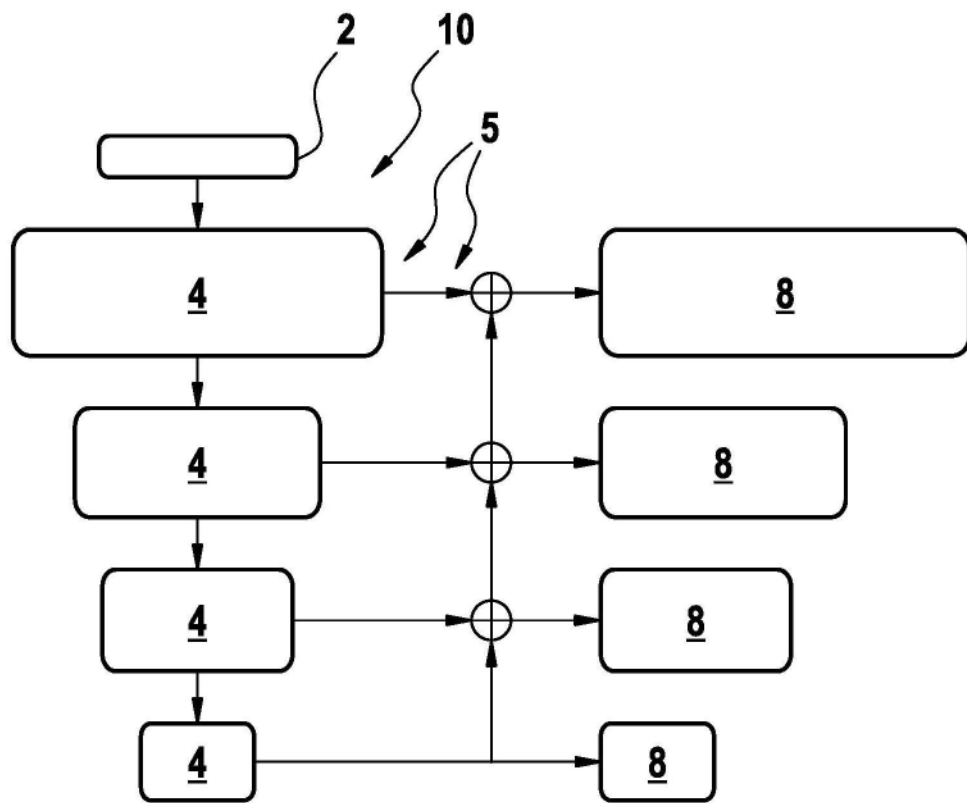


图4

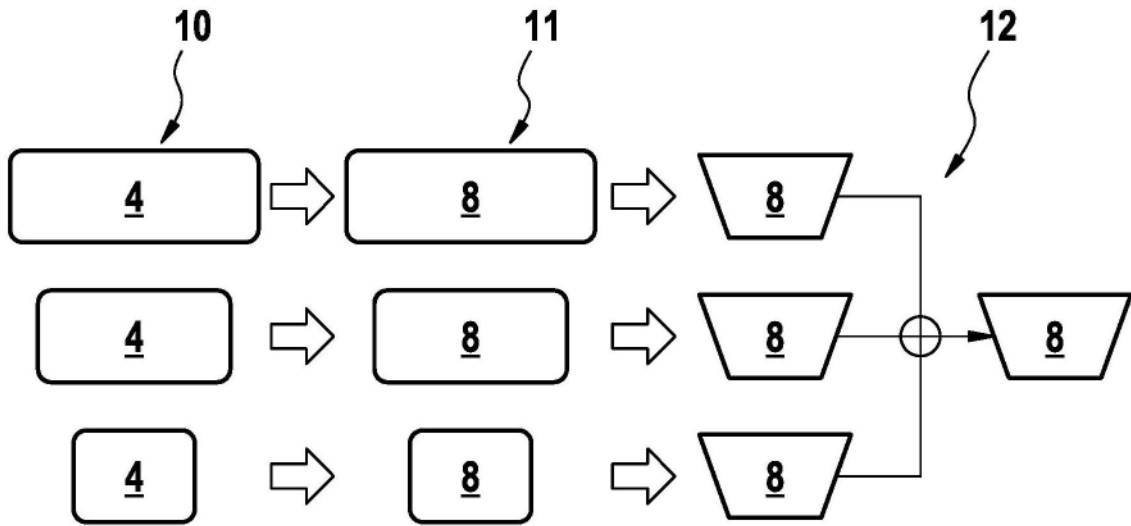


图5

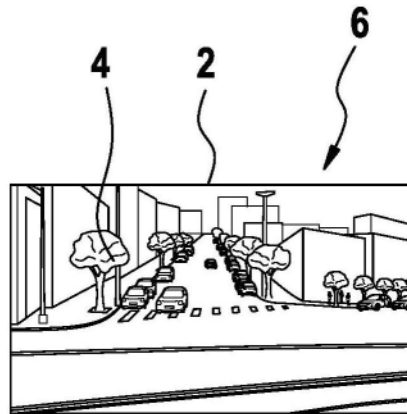


图6a

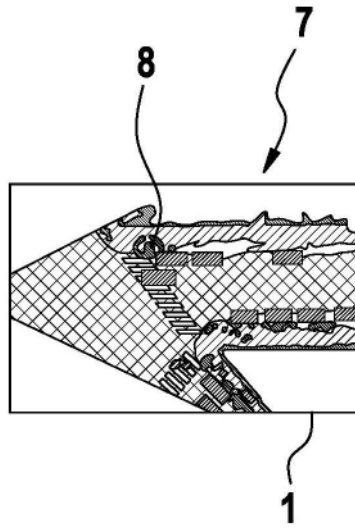


图6b

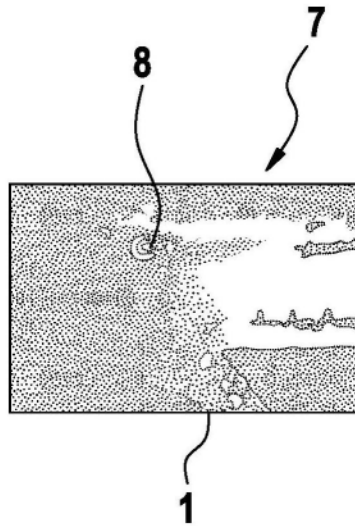


图6c

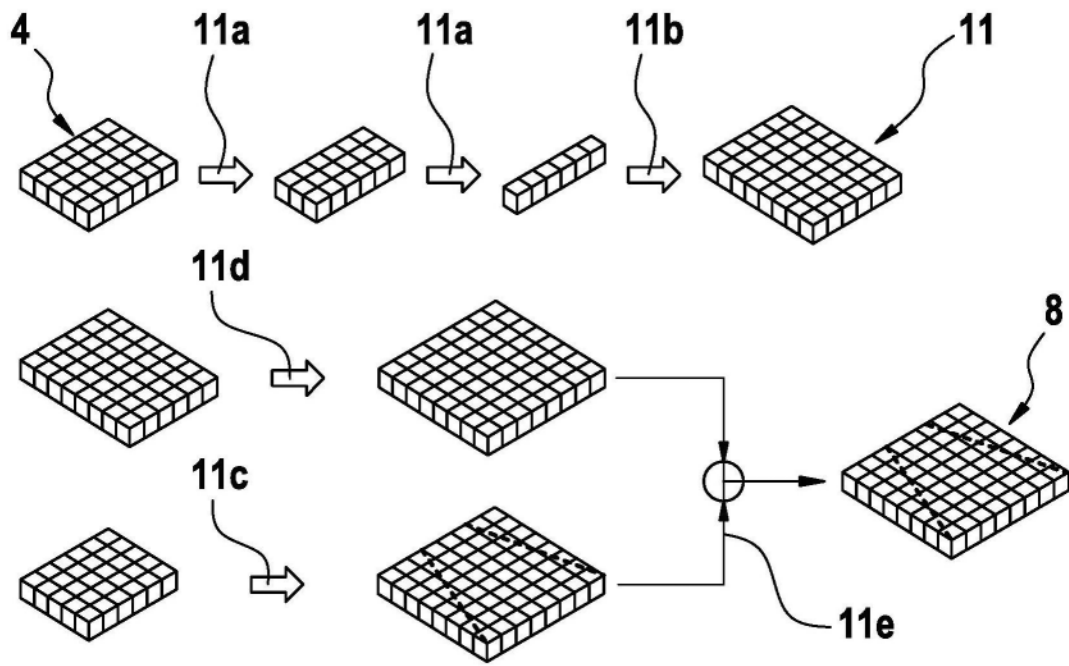


图7

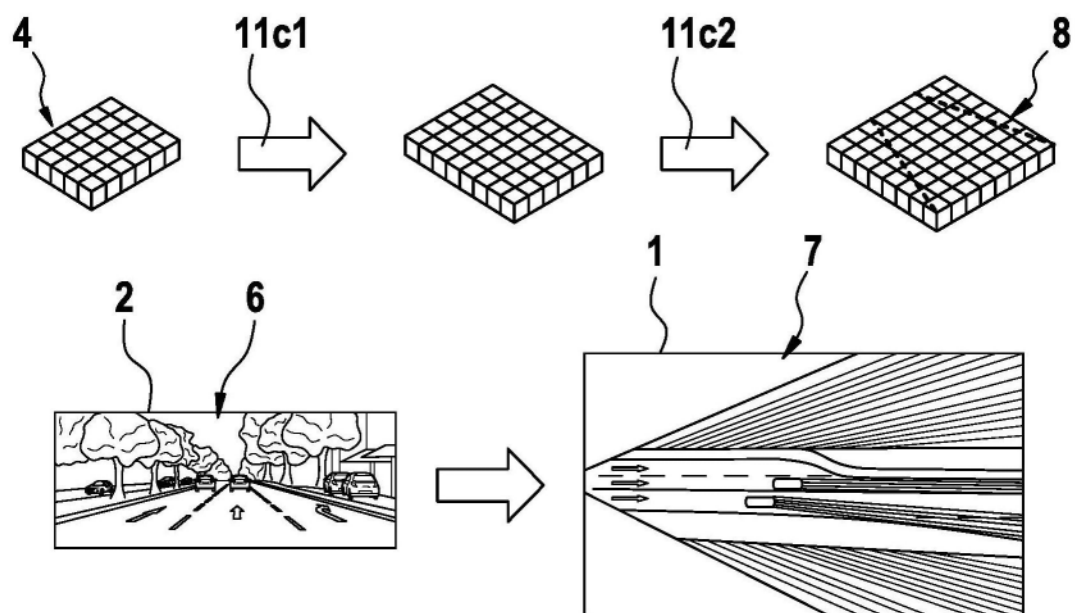


图8

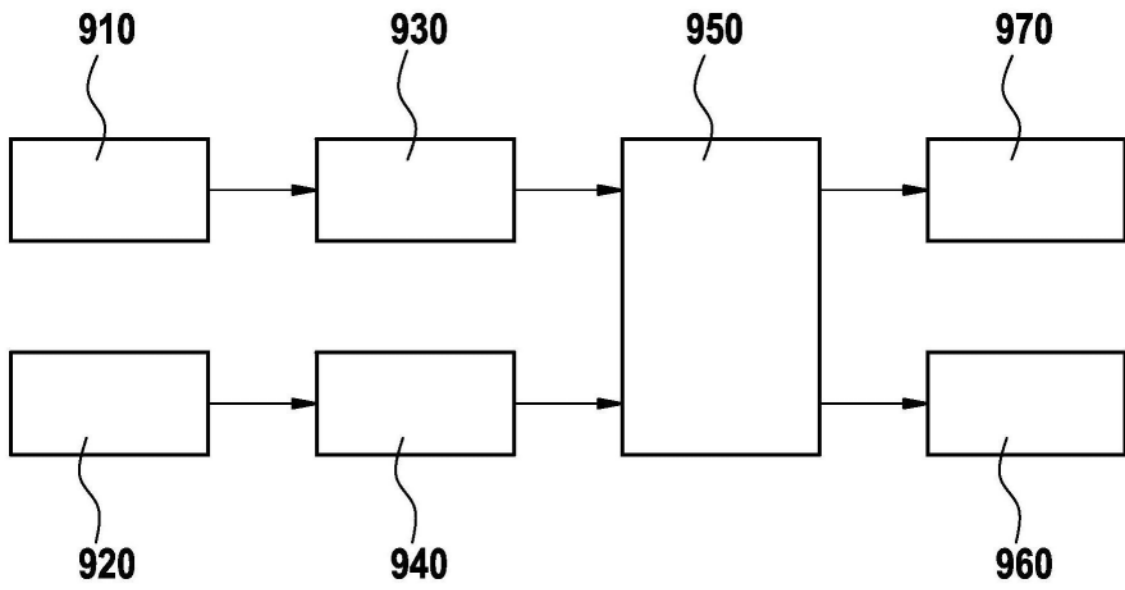


图9