



(12) 发明专利申请

(10) 申请公布号 CN 115375908 A

(43) 申请公布日 2022. 11. 22

(21) 申请号 202210533429.1

G06N 3/04 (2006.01)

(22) 申请日 2022.05.17

G06N 3/08 (2006.01)

(30) 优先权数据

G06V 10/764 (2022.01)

102021205034.4 2021.05.18 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 I·索斯诺维克 A·斯摩德斯

K·格劳

(74) 专利代理机构 中国专利代理(香港)有限公司

司 72001

专利代理师 刘书航 吕传奇

(51) Int. Cl.

G06V 10/40 (2022.01)

G06V 10/774 (2022.01)

G06V 10/82 (2022.01)

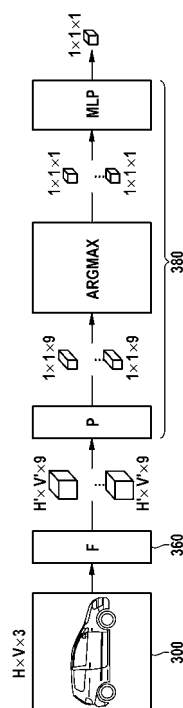
权利要求书3页 说明书15页 附图7页

(54) 发明名称

训练机器可学习模型以估计相对对象尺度

(57) 摘要

提供了一种用于训练机器可学习模型以估计图像中对象的相对尺度的系统和计算机实现的方法。为此目的,可以提供特征提取器(F,360)和包括机器可学习模型部件(G)的尺度估计器(380)。所述特征提取器可以是已知类型的特征提取器,其例如可以被预先训练,而尺度估计器可以由该系统和方法训练,以将特征提取器生成的特征图转换成对象的相对尺度估计。为此,尺度估计器可以以特定但无监督的方式在训练数据上被训练,这可能不需要尺度标签。在推断期间,尺度估计器可以应用于图像的若干个图像补片。所得的补片级尺度估计可以被组合到场景几何图中,所述场景几何图可以指示图像中描绘的场景的几何。



1. 一种训练机器可学习模型以估计图像中对象的相对尺度的计算机实现的方法(200), 包括:

- 提供(210)特征提取器(360), 其被配置为:
- 接收图像的图像补片作为输入;
- 检测所述图像补片中的多个特征以获得多个特征图作为输出, 其中所述多个特征与所述图像中的一个或多个对象相关联, 其中通过对所述图像补片的图像数据应用滤波器来生成相应的特征图, 其中相应的特征图包括沿着尺度维度跨一组不同空间尺度的滤波器响应;
- 提供(220)尺度估计器(380)来处理特征提取器的输出, 其中所述尺度估计器包括机器可学习模型部件, 并且被配置为:
 - 将相应的特征图聚合成特征级尺度估计, 其中所述聚合包括标识跨所述不同空间尺度的最大滤波器响应, 从而获得多个特征级尺度估计, 以及
 - 利用机器可学习模型部件, 从所述多个特征级尺度估计中推断补片级尺度估计, 其中所述方法进一步包括:
 - 访问(230)包括训练图像集的训练数据;
 - 在所述训练数据上训练(240)尺度估计器的机器可学习模型部件, 以从所述多个特征级尺度估计中推断补片级尺度估计, 其中所述训练包括:
 - 通过至少两个已知的尺度因子来在空间上缩放(250)训练图像的图像补片(300)的图像数据, 以获得至少两个另外的图像补片(310, 320);
 - 将特征提取器和尺度估计器应用(260)于所述至少两个另外的图像补片, 以获得至少两个补片级尺度估计;
 - 通过最小化损失函数的误差项来优化(270)机器可学习模型部件的参数, 其中所述误差项表达了实际相对尺度和估计相对尺度之间的失配, 其中实际相对尺度被确定为两个已知的已知尺度因子之间的差, 并且估计相对尺度被确定为至少两个补片级尺度估计之间的差。

2. 根据权利要求1所述的计算机实现的方法(200), 其中相应的特征图包括尺度维度和至少两个空间维度, 其中尺度估计器被配置为通过平均、加权或多数选择来在所述至少两个空间维度上聚合相应的特征图。

3. 根据权利要求1或2所述的计算机实现的方法(200), 其中所述尺度估计器被配置为通过标识在其处滤波器响应最大的空间尺度并且通过使用空间尺度的标识符作为特征级尺度估计或作为特征级尺度估计的一部分, 来在尺度维度上聚合相应的特征图。

4. 根据权利要求1至3中任一项所述的计算机实现的方法(200), 其中所述尺度估计器的机器可学习模型部件包括神经网络。

5. 根据权利要求4所述的计算机实现的方法(200), 其中神经网络是具有一个隐藏层的浅层神经网络。

6. 根据权利要求1至5中任一项所述的计算机实现的方法(200), 其中所述误差项定义了实际相对尺度和估计相对尺度之间的均方误差或均方偏差。

7. 一种估计图像中对象的相对尺度的计算机实现的方法(700), 包括:

- 提供(710)特征提取器(360), 其被配置为:

- 接收图像的图像补片作为输入；
 - 检测所述图像补片中的多个特征以获得多个特征图作为输出，其中所述多个特征与所述图像中的一个或多个对象相关联，其中通过对所述图像补片的图像数据应用滤波器来生成相应的特征图，其中相应的特征图包括沿着尺度维度跨一组不同空间尺度的滤波器响应；
 - 提供(720)尺度估计器(380)以处理所述特征提取器的输出，其中所述尺度估计器包括通过根据权利要求1至6中任一项所述的方法训练的机器可学习模型部件，其中所述尺度估计器被配置为：
 - 将相应的特征图聚合成特征级尺度估计，其中所述聚合包括标识跨所述不同空间尺度的最大滤波器响应，从而获得多个特征级尺度估计，以及
 - 利用机器可学习模型部件，从所述多个特征级尺度估计中推断补片级尺度估计；
 - 将所述特征提取器和所述尺度估计器应用(730)到所述图像的至少一个图像补片，以获得所述至少一个图像补片的补片级尺度估计；和
 - 输出(750)所述补片级尺度估计的数据表示。
8. 根据权利要求7所述的计算机实现的方法(700)，进一步包括通过以下步骤生成指示图像的场景几何的场景几何图(420)：
- 将特征提取器和尺度估计器应用于图像的多个图像补片(410,412)，以获得多个补片级尺度估计；
 - 生成图像的场景几何图，作为与所述多个图像补片相关的多个补片级尺度估计的表示。
9. 根据权利要求8所述的计算机实现的方法(700)，进一步包括将特征提取器和尺度估计器应用于图像的重叠图像补片。
10. 根据权利要求8或9所述的计算机实现的方法(700)，进一步包括以下各项中的至少一个：
- 从场景几何图中的多个补片级尺度估计中减去所述多个补片级尺度估计的最小值；
 - 和
 - 将场景几何图在空间上放大(430)至图像的空间分辨率。
11. 根据权利要求8至10中任一项所述的计算机实现的方法(700)，进一步包括：
- 从传感器(620,622)获得图像，所述传感器被配置为感测计算机控制的实体(610)的环境(600)；
 - 分析图像的场景几何图；
 - 基于所述分析的结果为所述计算机控制的实体生成控制数据，以使对计算机控制的实体的控制适配其环境。
12. 根据权利要求11所述的计算机实现的方法(700)，其中计算机控制的实体是机器人系统或自主车辆(610)。
13. 一种计算机可读介质(800)，包括表示指令的暂时性或非暂时性数据(810)，所述指令被布置成使处理器系统执行根据权利要求1至12中任一项所述的计算机实现的方法。
14. 一种用于训练机器可学习模型以估计图像中对象的相对尺度的系统(100)，包括：
- 输入接口子系统(140)，其被配置为访问：

- 包括训练图像集的训练数据(152)；
- 特征提取器(154,360),其被配置为:
 - 接收图像的图像补片作为输入；
 - 检测图像补片中的多个特征以获得多个特征图作为输出,其中所述多个特征与图像中的一个或多个对象相关联,其中通过对图像补片的图像数据应用滤波器来生成相应的特征图,其中相应的特征图包括沿着尺度维度跨一组不同空间尺度的滤波器响应；
- 尺度估计器(156,380),用于处理所述特征提取器的输出,其中所述尺度估计器包括机器可学习模型部件,并且被配置为:
 - 将相应的特征图聚合成特征级尺度估计,其中所述聚合包括标识跨所述不同空间尺度的最大滤波器响应,从而获得多个特征级尺度估计,以及
 - 利用机器可学习模型部件,从所述多个特征级尺度估计中推断补片级尺度估计；
- 处理器子系统(120),被配置为在训练数据上训练尺度估计器的机器可学习模型部件,以从多个特征级尺度估计中推断补片级尺度估计,其中所述训练包括:
 - 通过至少两个已知的尺度因子来在空间上缩放训练图像的图像补片(300)的图像数据,以获得至少两个另外的图像补片(310,320)；
 - 将特征提取器和尺度估计器应用于所述至少两个另外的图像补片,以获得至少两个补片级尺度估计；
 - 通过最小化损失函数的误差项来优化机器可学习模型部件的参数,其中所述误差项表达了实际相对尺度和估计相对尺度之间的失配,其中所述实际相对尺度被确定为所述两个已知的已知尺度因子之间的差,并且所述估计相对尺度被确定为所述至少两个补片级尺度估计之间的差。

15.一种用于估计图像中对象的相对尺度的系统(500),包括:

- 输入接口子系统(540),被配置为访问:
 - 图像(552)；
 - 特征提取器(554,360),其被配置为:
 - 接收图像的图像补片作为输入；
 - 检测图像补片中的多个特征以获得多个特征图作为输出,其中所述多个特征与图像中的一个或多个对象相关联,其中通过对图像补片的图像数据应用滤波器来生成相应的特征图,其中相应的特征图包括沿着尺度维度跨一组不同空间尺度的滤波器响应；
 - 尺度估计器(556,380),用于处理所述特征提取器的输出,其中所述尺度估计器包括通过根据权利要求1至6中任一项所述的方法训练的机器可学习模型部件,其中所述尺度估计器被配置为:
 - 将相应的特征图聚合成特征级尺度估计,其中所述聚合包括标识跨所述不同空间尺度的最大滤波器响应,从而获得多个特征级尺度估计,以及
 - 利用机器可学习模型部件,从所述多个特征级尺度估计中推断补片级尺度估计；
 - 处理器子系统(520),被配置为:
 - 将所述特征提取器和所述尺度估计器应用于所述图像的至少一个图像补片,以获得所述至少一个图像补片的补片级尺度估计；和
 - 输出所述补片级尺度估计的数据表示。

训练机器可学习模型以估计相对对象尺度

技术领域

[0001] 本发明涉及一种用于训练机器可学习模型以估计图像中对象的相对尺度的系统和计算机实现的方法。本发明还涉及一种估计图像中对象的相对尺度,例如以确定图像的场景几何的系统和计算机实现的方法。本发明进一步涉及一种计算机可读介质,其包括表示用于处理器系统以执行任何计算机实现的方法的指令的数据。

背景技术

[0002] 计算机视觉领域关注的是使机器能够“看见”并理解它们的环境。计算机视觉中的核心任务是对图像中的对象进行标识和分类,例如由自主车辆获取的相机图像中的行人或车辆,或者要由制造机器人处置的零件等。

[0003] 尺度是图像中示出的对象的自然属性,与位置和外观一样是基本特性。这里,术语“尺度”可以指代图像中对象的表观大小,其可以取决于对象到相机的距离、相机的焦点等。计算机视觉任务典型地需要计及图像中对象的(变化的)尺度。例如,图像分类优选地是尺度不变的,以获得准确的分类结果。在图像分割中,尺度等变性很重要,因为输出图应当与输入在尺度上成比例。在对象检测或对象跟踪中,尺度不变和尺度等变这两者都很重要。

[0004] 在计算机视觉中,通常通过在训练数据中提供各种各样的充足示例,例如不同尺度的对象,来处理尺度不变性或等变性。还可以适配机器可学习模型和/或它们的训练,以提供尺度不变性或等变性。例如,出版物“Scale-Equivariant Steerable Networks”, 2019, <https://arxiv.org/abs/1910.11093v1>描述了将尺度等变性机制并入到CNN中以提高CNN的性能,其中性能可以被理解为CNN正确分类图像的能力。尺度等变性机制基于构建神经网络卷积层的滤波器,使得它们是基滤波器(也称为基函数)的加权和,其中权重可以在CNN的训练期间被训练。

[0005] 虽然已知将用于计算机视觉任务的机器可学习模型适配和/或训练成尺度不变或等变的,或者至少在足够程度上尺度不变或等变的,但是可能合期望的是获得图像中对象的相对尺度的明确指示。这可以例如允许估计图像中描绘的场景的几何。例如,如果场景密集地分布有类似的对象,诸如花田中的花,则场景图像中对象的相对尺度可以指示场景的几何,在这种情况下是包含对象的平面表面,该表面朝着地平线倾斜。这可以是位于更靠近图像底部的对象中更显而易见的,该对象具有更大的表观大小,然后表观大小朝着图像的中间逐渐减小。换句话说,对象的相对尺度可以允许关于场景的几何得出结论,这在许多现实生活应用中可能是有价值的,诸如自动驾驶,其中相机图像示出稠密的车辆区域可能指示交通堵塞。另一个示例是具有许多行人的环境,并且其中场景的几何可以用于标识更靠近自动驾驶汽车的行人,以便能够关于重要性对行人进行分类,其中自动驾驶汽车可以忽略不重要的行人,因为他们离得太远。一般而言,通过尺度获得对象之间的理解可以允许标识对象之间的关系,因为相同或相似尺度的对象可能是相关的。这可以用来生成对象的关系图。

[0006] 已知通过使用有监督学习来训练机器可学习模型以提供图像中对象尺度的明确

指示,例如通过提供训练数据,其中对象被标注,包括它们的尺度,例如以像素为单位或作为相对于图像分辨率的值。不利的是,这要以各种尺度示出对象的大量训练数据,并且需要大量的人工参与,例如,提供标签并仔细构建机器可学习模型以能够明确指示尺度。

[0007] 能够获得更容易训练(例如以非监督的方式)并且具有有限的计算复杂度的尺度估计器将是有益的。

发明内容

[0008] 根据本发明的第一方面,如分别由权利要求1和14定义的,提供了一种计算机实现的方法和对应的系统,用于训练机器可学习模型以估计图像中对象的相对尺度。根据本发明的另一方面,如分别由权利要求7和15定义的,提供了一种计算机实现的方法和对应的系统,用于估计图像中对象的相对尺度。根据本发明的另一方面,如由权利要求13所定义的,提供了一种计算机可读介质,包括用于使处理器系统执行所述计算机实现的方法中的任一项的指令。

[0009] 根据上述措施,提供了尺度估计器,其可以包括机器可学习模型部件,诸如神经网络,并且可以被训练以提供相对尺度估计。尺度估计器可以作为特征提取器的“插件”提供。这样的特征提取器可以是现有的“已知的”特征提取器,其被配置为接收图像补片作为输入,例如 64×64 像素或者具有任何其他合适的空间维度的图像补片,并且从图像补片中提取与图像中的一个或多个对象相关联的多个特征。特征的示例包括不同类型的边缘、纹理、拐角等。这些特征可以是人工定义的,但是也可以是机器学习的,例如通过先前已经在包括对象示例的训练数据上训练过的特征提取器来进行机器学习的。如本身已知的,特征提取器可以提供多个特征图作为输出。例如,如果特征提取器包括卷积神经网络(CNN),则这样的特征图可以由CNN的输出通道构成或表示。更进一步地,特征提取器可以被适配,并且在机器学习的特征提取器的情况下,被训练为至少在适当的程度上是尺度等变的。这可以在具有尺度维度的特征图中显露出来。因此,特征提取器可以沿着尺度维度提供相应的滤波器响应。这样的特征提取器本身是已知的,例如从共同未决的专利申请EP 20195059中已知,该专利申请在涉及其中描述的尺度等变CNN(SE-CNN)的范围内通过引用特此并入,其输入和卷积层可以构成本说明书中描述的特征提取器的示例。

[0010] 尺度估计器可以被配置为聚合从特征提取器的输出获得的每个特征图。例如,这样的聚合可以涉及沿着特征图的各个维度(诸如其空间维度)聚合滤波器响应。特别地,沿着尺度维度,可以标识最大滤波器响应,这可以导致提供特征级尺度估计的特征图的聚合。在特定示例中,如果CNN具有512个输出频道,则该聚合可以产生512个特征级尺度估计,每个都是从沿着特征图的相应尺度维度的最大滤波器响应中导出的。尺度估计器还可以包括机器可学习模型部件,诸如前述的神经网络,其模型部件可以被配置为接收特征级尺度估计作为输入,并且输出表示作为输入提供的图像补片的整体尺度估计的补片级尺度估计。

[0011] 为此,机器可学习模型部件可以在训练数据上训练。然而,不是依赖于有监督训练——在有监督训练中,补片级尺度估计是作为基准真值人工提供的或者至少是外部提供的——而是可以在训练期间生成用于训练的合适目标。也就是说,对于尺度估计器来说,能够学习对象的相对尺度可能就足够了,例如学习一个对象比另一个对象更靠近相机。这样的相对尺度可能不表示尺度的绝对度量,并且从而可能不允许对对象的绝对大小(例如,对

2m宽的对象等)得出结论。然而,这样的相对尺度对于各种目的可能是足够的,包括前述的对场景几何的估计。根据所要求保护的措施,用于估计相对尺度的训练目标可以通过根据至少两个已知尺度因子对训练图像的图像补片的图像数据进行空间缩放来生成。例如,图像补片中的图像数据可以缩小,例如缩小0.75倍,以及被放大,例如放大1.5倍。应当领会,这样的放大可以包括裁剪,而缩小可以包括填充,以便获得相等尺寸的图像补片。另一个示例是,图像补片的图像数据可以用因子1.0(例如,用单位尺度因子)和因子1.5来缩放。同样可以想到这样的尺度因子的各种其他示例。这样的尺度因子可以被称为“已知的”或“实际的”尺度因子。

[0012] 然后,可以将特征提取器和尺度估计器应用于包括经缩放的图像数据的图像补片,从而产生至少两个补片级尺度估计。虽然可能不知道任一图像补片中对象的绝对尺寸是多少,但是现在可以知道两个图像补片之间对象的相对大小,由两个已知尺度因子之间的关系来表示。例如,该关系可以表示为差(例如, $2.0-0.5 = 1.5$)或比率(例如, $2.0/0.5 = 4.0$)等,并且也可以称为实际相对尺度,其中“实际”指的是图像数据实际上根据相应的尺度因子进行了缩放的事实,并且“相对”指的是表示尺度因子之间的关系(例如,差或比率)的数字。可以针对补片级尺度估计计算同样的关系,从而产生估计的相对尺度。损失函数可以被公式化,通过该损失函数,训练努力适配机器可学习模型部件的参数,诸如神经网络的权重,以学习估计实际相对尺度。即,损失函数可以表示实际相对尺度和估计相对尺度之间的失配,并且训练可以寻求最小化失配。因此,尺度估计器可以学习更好地估计相对尺度。这不需要人工提供基准真值,因为实际相对尺度可以是内部生成的。这样,可以促进尺度估计器的训练。此外,尺度估计器可以简单地作为(现有的)特征提取器的插件来提供。这有助于关注点的分离,因为人们可能不需要过分关注特征提取器本身,而是训练尺度估计器使其自身适配特定特征提取器提供的特征图。更进一步地,已经发现,这样的尺度估计器在结构上也可以相对简单,因为它可以“仅仅”需要将特征图聚合成特征级尺度估计,并将特征级尺度估计组合成补片级尺度估计。与特征提取器本身相比,这样的组合可以由相对简单的机器可学习模型部件来完成。例如,在许多应用中,仅具有一个隐藏层的浅层神经网络可能就足够了。因此,如果特征提取器已经可用,例如用于对象检测或分类目的,则可以在计算复杂性和/或训练工作方面以相对较小的成本添加尺度估计器。

[0013] 可选地,相应的特征图包括至少两个空间维度和尺度维度,其中尺度估计器被配置为通过平均、加权或多数选择来在至少两个空间维度上聚合相应的特征图。特征图的空间维度对于尺度估计可能不是特别相关。这样,通过在空间维度上聚合特征图,例如通过平均、加权、多数选择或类似技术,可以将空间维度减少到例如 1×1 。在特定示例中,全局平均池化层可以用于将 $H \times W \times S$ 的特征图减少为 $1 \times 1 \times S$ 的特征图(用“S”表示尺度维度)。

[0014] 可选地,尺度估计器被配置为通过标识滤波器响应在其处最大的空间尺度并通过使用空间尺度的标识符作为特征级尺度估计的一部分,来在尺度维度上聚合相应的特征图。通过标识滤波器响应在其处最大的空间尺度,并通过使用空间尺度的标识符作为特征级尺度估计器的一部分,可以例如将特征图从 $1 \times 1 \times S$ 聚合到 $1 \times 1 \times 1$ 。例如,如果存在预定义的尺度集 $\{1, \sqrt{2}, 2, 2\sqrt{2}, 4, 4\sqrt{2}, \dots\}$,其例如可以被定义为特征提取器的超参数,则特征级尺度估计可以是与在其处获得最大特征响应的尺度相对应的集合的索引,例如由argmax算子标识。

[0015] 可选地,尺度估计器的机器可学习模型部件包括神经网络。例如,神经网络可以是一个具有一个隐藏层的浅层神经网络。已经发现,这样的神经网络,或者通常的浅层多层感知器(MLP),可以足以学会将特征级尺度估计组合成补片级尺度估计。这样的MLP可能需要较少的资源来实现,并且在给定其相对较少的参数的情况下可以易于训练。

[0016] 可选地,误差项定义实际相对尺度和估计相对尺度之间的均方误差或均方偏差。均方误差(MSE)或均方偏差(MSD)这两者都非常适合作为误差函数,同时在运行时需要很少的资源来实现和评估。

[0017] 可选地,指示图像的场景几何的场景几何图通过以下方式生成:

- 将特征提取器和尺度估计器应用于图像的多个图像补片,以获得多个补片级尺度估计;

- 生成图像的场景几何图,作为与多个图像补片相关的多个补片级尺度估计的表示。

[0018] 如在其他地方所讨论的,例如通过构建表示图像补片的阵列,补片级尺度估计可以被组合到场景几何图中,其中阵列中的每个位置包括相应的补片级尺度估计。这样的阵列可以类似于针对图像的图,并且可以指示场景几何,这也将在本说明书的其他地方进行阐述。

[0019] 可选地,特征提取器和尺度估计器被应用于图像的重叠图像补片。由于尺度估计器可以为每个图像补片产生一个补片级尺度估计,因此如果尺度估计器被应用于非重叠图像补片,则得到的场景几何图与输入图像相比可能具有相对较低的分辨率。为了增加场景几何图的分辨率,可以将特征提取器和尺度估计器应用于重叠的图像补片。这可以为图像提供更详细且准确的场景几何图。

[0020] 可选地,场景几何图通过以下方式生成:

- 从场景几何图中的多个补片级尺度估计中减去多个补片级尺度估计的最小值;

和/或

- 在空间上将场景几何图放大到图像的空间分辨率。

[0021] 可选地,可以从被配置为感测计算机控制的实体的环境的传感器获得图像,分析图像的场景几何图,并且基于所述分析的结果为计算机控制的实体生成控制数据,以使对计算机控制的实体的控制适配其环境。可以基于场景几何图的分析结果来控制计算机控制的实体,诸如机器人系统或自主车辆。例如,据其生成场景几何图的图像可以由机载相机获得,其中场景几何图指示由机载相机获取的场景的几何。例如,场景几何图可以指示自主车辆前方存在交通堵塞,在这种情况下,可能合期望的是以不同方式控制车辆,例如从而进行减速。

[0022] 本领域技术人员应当领会,本发明的上述实施例、实现和/或可选方面中的两个或更多个可以以任何被认为有用的方式进行组合。

[0023] 基于本说明书,本领域技术人员可以执行任何系统、任何计算机实现的方法或任何计算机可读介质的修改和变型,所述修改和变型对应于所述实体中的另一个的所描述的修改和变型。

附图说明

[0024] 本发明的这些和其他方面将通过参考在以下描述中以示例方式描述的实施例以及参考附图而变得显而易见并得到进一步阐述,其中:

图1示出了用于训练机器可学习模型以估计图像中对象的相对尺度的系统;

图2示出了用于训练机器可学习模型以估计图像中对象的相对尺度的计算机实现的方法的步骤;

图3示出了应用于图像补片的图像数据以生成图像补片的补片级尺度估计的特征提取器和尺度估计器;

图4图示了用于尺度估计器的训练的误差项的计算,其包括根据两个不同的尺度因子并且通过使用尺度估计器估计尺度因子来缩放图像补片的图像数据,其中误差函数表示相应尺度因子之间的关系的不匹配;

图5A示出了特征提取器和尺度估计器的输入图像;

图5B示出了被分割成非重叠图像补片的输入图像,其中每个图像补片用作特征提取器和尺度估计器的输入;

图5C示出了由尺度估计器生成的场景几何图,该场景几何图包括每个图像补片的补片级尺度估计,其中场景几何图在空间上被放大到图像分辨率;

图6示出了用于估计图像中对象的相对尺度的系统;

图7示出了包括图5的系统的(半)自主车辆,该系统用于生成和分析场景几何图并基于此控制车辆;

图8示出了用于估计图像中对象的相对尺度的计算机实现的方法的步骤;

图9示出了包括数据的计算机可读介质。

[0025] 应当注意的是,所述图纯粹是示意性的,并没有按比例绘制。在各图中,与已经描述的元件相对应的元件可以具有相同的附图标记。

[0026] 参考编号和缩写列表

提供下面的参考数字列表是为了便于解释附图,并且不应当被解释为限制权利要求:

100 用于训练机器可学习模型以估计图像中对象的相对尺度的系统

120 处理器子系统

140 数据存储接口

150 数据存储装置

152 训练数据

154 特征提取器的数据表示

156 尺度估计器的数据表示

200 训练机器可学习模型以估计图像中对象的相对尺度的方法

210 提供特征提取器

220 提供尺度估计器

230 访问训练数据

240 训练

245 针对下一图像补片进行重复

- 250 在空间上缩放图像补片以获得经缩放的图像补片
- 260 应用特征提取器和尺度估计器
- 270 优化尺度估计器的机器可学习模型部件
- 300 图像补片
- 310 具有缩小的图像数据的图像补片
- 320 具有放大的图像数据的图像补片
- 360 特征提取器
- 380 尺度估计器
- 400 示出花田的图像
- 410 被分割成图像补片的图像
- 412 图像补片
- 420 场景几何图
- 430 放大
- 440 放大到图像分辨率的场景几何图
- 500 用于估计图像中对象的相对尺度的系统
- 520 处理器子系统
- 540 数据存储接口
- 550 数据存储装置
- 552 图像数据
- 554 特征提取器的数据表示
- 556 尺度估计器的数据表示
- 560 传感器数据接口
- 562 传感器数据
- 570 控制接口
- 572 控制数据
- 600 环境
- 610 (半)自主车辆
- 620 传感器
- 622 相机
- 630 致动器
- 632 电马达
- 700 估计图像中对象的相对尺度的方法
- 710 提供特征提取器
- 720 提供尺度估计器
- 730 将特征提取器和尺度估计器应用于图像补片
- 740 针对下一图像补片进行重复
- 750 输出(一个或多个)补片级尺度估计的数据表示
- 800 计算机可读介质
- 810 非暂时性数据。

具体实施方式

[0027] 下面参考图1和2描述了用于训练机器可学习模型以估计图像中对象的相对尺度的系统和计算机实现的方法,参考图3,描述了被应用于图像补片的图像数据以生成图像补片的补片级尺度估计的特征提取器和尺度估计器,参考图4,描述了尺度估计器的机器可学习模型部件的训练,参考图5A-5C,描述了将尺度估计器应用于输入图像以生成场景几何图,参考图6和8,描述了用于估计图像中对象的相对尺度的系统和计算机实现的方法,以及参考图7,描述了合并有图6的系统的自主车辆。图9示出了如所要求保护的本发明的实施例中使用的计算机可读介质。

[0028] 图1示出了用于训练机器可学习模型以估计图像中对象的相对尺度的系统100。系统100可以包括用于访问用于训练的训练数据152的输入接口子系统。例如,如图1图示的那样,输入接口子系统可以包括数据存储接口140或者由数据存储接口140构成,数据存储接口140可以提供对数据存储装置150上的训练数据152的访问。例如,数据存储接口140可以是存储器接口或永久存储接口,例如硬盘或SSD接口,但也可以是个域网、局域网或广域网接口,诸如蓝牙、Zigbee或Wi-Fi接口或以太网或光纤接口。数据存储装置150可以是系统100的内部数据存储装置,诸如存储器、硬盘驱动器或SSD,但是也可以是外部数据存储装置,例如网络可访问的数据存储装置。

[0029] 在一些实施例中,数据存储装置150可以进一步包括特征提取器的数据表示154和尺度估计器的数据表示156,这两者将在下文中详细讨论,并且可以由系统100从数据存储装置150访问。然而,应当领会,训练数据152、特征提取器的数据表示154和尺度估计器的数据表示156也可以每个从不同的数据存储装置中访问,例如,经由不同的数据存储接口访问。每个数据存储接口可以是如上面针对数据存储接口140描述的类型。在其他实施例中,特征提取器和/或尺度估计器的数据表示154、156可以由系统100例如基于设计参数或设计规范内部生成,并且因此可以不明确地存储在数据存储装置150上。

[0030] 系统100可以进一步包括处理器子系统120,该处理器子系统120可以被配置为在系统100的操作期间,以在本说明书中其它地方描述的方式,在训练数据152上训练尺度估计器156,特别是尺度估计器156的机器可学习模型部件。例如,处理器子系统120的训练可以包括执行算法,该算法使用训练目标(例如损失函数)来优化尺度估计器156的参数。在一些实施例中,特征提取器154也可以包括机器可学习模型或由机器可学习模型组成,并且处理器子系统120可以被配置为也在训练数据152上或者不同的或附加的训练数据上来训练特征提取器154。

[0031] 系统100可以进一步包括输出接口,用于输出经训练的尺度估计器的数据表示,该尺度估计器也被称为机器“学习的”尺度估计器,并且该数据也被称为经训练的尺度估计器数据。应当领会,“经训练的”在这里和其他地方指代至少尺度估计器的机器可学习模型部件已经被训练。例如,也如图1所图示的那样,输出接口可以由数据存储接口140构成,在这些实施例中,所述接口是输入/输出(“IO”)接口,经训练的尺度估计器可以经由所述接口存储在数据存储装置150中。例如,定义“未经训练的”尺度估计器的数据表示156可以在训练期间或训练之后至少部分地被经训练的尺度估计器的数据表示所替代,因为尺度估计器156的参数,特别是尺度估计器156的机器可学习模型部件的参数,可以被适配为反映在训练数据152上的训练。在其他实施例中,经训练的尺度估计器的数据表示可以与“未经训练

的”尺度估计器的数据表示156分开存储。在一些实施例中,输出接口可以与数据存储接口140分离,但是通常可以是如上面针对数据存储接口140所描述的类型。

[0032] 图2示出了用于训练机器可学习模型、特别是尺度估计器以估计图像中对象的相对尺度的计算机实现的方法200。方法200可以对应于图1的系统100的操作,但不是必须的,因为它也可以对应于另一种类型的系统、装置、设备或实体的操作,或者因为它可以对应于计算机程序的步骤。

[0033] 方法200被示出为包括:在题为“提供特征提取器”的步骤中,提供210如本说明书中其它地方所述的特征提取器;在题为“提供尺度估计器”的步骤中,提供220如本说明书中其它地方所述的特征提取器;以及在题为“访问训练数据”的步骤中,访问230包括训练图像集的训练数据。方法200进一步被示出为,在题为“训练”的步骤中,包括在训练数据上训练240尺度估计器的机器可学习模型部件,其中训练包括在题为“在空间上缩放图像补片以获得缩放的图像补片”的步骤中,通过至少两个已知的尺度因子在空间上缩放250训练图像的图像补片的图像数据以获得至少两个另外的图像补片,在题为“应用特征提取器和尺度估计器”的步骤中,将特征提取器和尺度估计器应用260到至少两个另外的图像补片,以获得至少两个补片级尺度估计,并且在题为“优化尺度估计器的机器可学习模型部件”的步骤中,通过最小化损失函数的误差项来优化270机器可学习模型部件的参数,其中所述误差项表示实际相对尺度和估计相对尺度之间的失配,其中实际相对尺度被确定为两个已知的已知尺度因子之间的差,并且估计相对尺度被确定为至少两个补片级尺度估计之间的差。训练240可以包括多个迭代循环,例如在训练图像的不同图像补片上迭代,如图2中的箭头245所示,以及在不同的训练图像上迭代(未示出)。

[0034] 继续参考图像中对象的相对尺度的估计,本说明书中描述的测量利用了特征提取器和尺度估计器。特征提取器可以但不必是机器可学习的特征提取器,其例如可以与尺度估计器分开训练,例如通过不同类型的系统、在不同的时刻在不同类型的训练数据上进行训练。例如,用于训练尺度估计器的系统和方法可以利用预先训练的特征提取器,该特征提取器可以以本身已知的方式预先训练。一个非限制性示例是,特征提取器可以是被训练来从图像补片中提取特征的尺度等变卷积神经网络(SE-CNN)。这样的特征提取可以导致每个特征的特征图的输出,该特征图在CNN的示例中也被称为CNN的“通道”。

[0035] 例如,考虑 $F: x \rightarrow y$, 其中 x, y 是输入和输出张量,其中 F 表示特征提取器的函数,例如 SE-CNN。输入张量可以具有以下形状: $batch_size \times 3 \times height \times width$, 而输出张量可以具有以下形状: $batch_size \times num_channels \times num_scales \times height' \times width'$ 。这里,“batch_size”可以指代用作输入的图像补片的数量,而“3”可以表示图像数据的三个颜色分量(例如,RGB或YUV),“height”和“width”可以是每个图像补片的高度和宽度(例如,64×64像素),“num_channels”可以表示作为输出生成的特征图的数量,“num_scales”可以表示在其处检测到特征的尺度的数量,该数量进而可以对应于特征图的尺度维度,并且 $height'$ 和 $width'$ 可以表示特征图的高度和宽度,并且从而表示特征图的空间维度。

[0036] 如本身已知的,特征提取器可以被配置为检测图像补片中的多个特征以获得多个特征图作为输出,其中所述多个特征与图像中的一个或多个对象相关联,其中通过对图像补片的图像数据应用滤波器来生成相应的特征图,并且其中相应的特征图沿着尺度维度包

括跨一组不同空间尺度的滤波器响应。因此,特征提取器可以被配置为通过提供跨不同尺度的滤波器响应来考虑尺度信息。如本身已知的,特征提取器可以被配置有要使用哪些尺度,例如就尺度数量和尺度因子而言。例如,尺度可以被定义为特征提取器的超参数。尺度之间的数量和步长可以取决于特定的应用来选择。例如,如果图像是由很可能示出交通堵塞的车辆的车载相机获得的相机图像,则人们可以预期交通堵塞中其他车辆的相对大小从一辆汽车到下一辆汽车仅略微改变。因此,人们可以在尺度之间使用相对小的步长,诸如1.4。人们还可以预期,非常远的汽车可能最多是附近的汽车的八分之一,并且因此选取9个尺度, $1, \sqrt{2}, 2, 2\sqrt{2}, 4, 4\sqrt{2}, 8, 8\sqrt{2}, 16$, 其中数字指代例如用于不同尺度的滤波器的相对滤波器大小或相对内核大小。应当领会,对于其他类型的应用,可以使用与这里描述的不同数量的尺度和/或不同的尺度集。

[0037] 作为特征提取器的示例,可以使用ImageNet预训练的CNN,诸如EP 20195059中描述的SE-CNN。人们可以进一步假设特征图仅示出一个对象的特征。这样,可以例如使用全局空间平均池化层P在空间上聚合每个特征图。该层可以被提供为特征提取器的最后一层,或者被提供为特征提取器之后的单独一层。在特定示例中,特征提取器可以具有512个输出通道。在例如借助于全局空间平均池化层进行聚合之后,输出张量可以具有 $batch_size \times 512 \times 9$ 的形状,其中“9”指代尺度的数量。从每个输出中,可以提取在其处获得最大滤波器响应的尺度。这可以通过例如使用argmax算子在尺度的维度上进行最大池化来完成。结果,可以为每个图像补片获得512个尺度预测。这些预测在其他地方也被称为特征级尺度估计。然后,可以使用浅层多层感知器将这512个特征级尺度估计回归成一个补片级尺度估计。这里,G可以表示在其他地方被称为尺度估计器的机器可学习模型部件的示例。例如,浅层多层感知器可以是具有一个隐藏层的神经网络,或者深度神经网络,或者线性回归器,或者是通常可以将向量(特征级尺度估计)映射成标量(补片级尺度估计)并且是可微分的任何模型。在这方面,注意,尽管尺度估计器可以包括浅层机器可学习模型部件,但这不是必需的,因为尺度估计器也可以包括深度机器可学习模型部件,例如,具有几个隐藏层。

[0038] 图3示出了上述的一个示例,其中,大小为 $H \times V \times 3$ 的图像补片300(其中 $H \times V$ 例如是 64×64 像素)被输入到特征提取器F 360,从而产生了多个大小为 $H' \times V' \times 9$ 的特征图,其中“9”在该示例中是尺度的数量。全局空间平均池化层P可用于将特征图在空间上聚合成大小为 $1 \times 1 \times 9$ 的特征图,其后可以是argmax算子,argmax算子可以为每个图像补片产生 $1 \times 1 \times 1$ 的单个特征级尺度估计。尺度估计器380的机器可学习模型部件,例如浅层多层感知器G,然后可以用于将所有特征级尺度估计(例如数量为512)组合成单个补片级尺度估计。

[0039] 具有一个隐藏层的尺度估计器的示例可以通过类PyTorch伪代码来描述,如以下代码摘录所示:

```

1 import torch
2 import torch.nn as nn
3 import SE_ResNet101
4
5 class ScaleEstimator(nn.Module):
6     def __init__(self):
7         super().__init__()
8         self.feature_extractor = SE_ResNet101(pretrained = True)
9         self.regressor = nn.Sequential (
10             nn.Linear(512, 256),
11             nn.ReLU() ,
12             nn.Linear(256, 1),
13             nn.ReLU()
14         )
15
16     def forward (self, x):
17         # x.shape = B, 3, 64, 64
18         y = self.feature_extractor(x)
19         # y.shape = B, 512, 9, 1, 1
20         y = y.mean(-1).mean(-1)
21         # y.shape = B, 512, 9
22         y = y.argmax(-1)
23         biased_scale = self.regressor(y)
24         return biased_scale

```

继续参考图3,注意,全局空间平均池化层P被示出为与特征提取器F分离。在一些实施例中,全局空间平均池化层P或类似类型的空间聚合器功能可以是也包括特征提取器F的相同的整个网络的一部分。换句话说,可以提供一种网络,该网络既包括特征提取器F,又包括随后的全局空间平均池化层P或类似功能。特征提取器和尺度估计器的元件的各种其他类型的划分也是可能的。

[0040] 在一些示例中,特征提取器F可以是对象检测器或分类器的一部分。这样的对象检测器或分类器可以包括附加的网络层,其处理特征提取器F的特征图以获得对象分割或分类。在这样的示例中,特征图可以表示对象检测器或分类器的内部数据,该内部数据可以被尺度估计器访问以对尺度进行估计。在这样的示例中,特征图因此可以用于由对象检测器或分类器进行的对象检测或分类,以及尺度估计这两者。

[0041] 可以使用合适的数据集来训练尺度估计器,并且特别是其机器可学习模型部件,例如多层感知器G。例如,可以使用各种类的自然图像的数据集,诸如ImageNet或STL-10。所述训练可以包括例如通过定义损失函数来定义训练目标。图4图示了该损失函数的计算。这里,示出了来自训练数据的图像补片300可以根据两个不同的尺度因子例如通过内插而被缩放,该缩放可以被标识为 $L_{\gamma_1}, L_{\gamma_2}$, 其中 γ_1, γ_2 是相应的尺度因子。例如,这样的尺度因子可以从一个范围(例如,从0.5-2.0)内随机采样。这可以产生包含经缩放的图像数据的两个图

像补片310、320,其可以被馈送到由特征提取器和尺度估计器构成的网络 N_{θ} 中,其中尺度估计器具有机器可学习模型部件,该机器可学习模型部件具有参数 θ 。如参考图3所解释的,网络 N_{θ} 可以通过为经缩放的图像补片310、320中的每一个生成补片级缩放估计来预测缩放因子 $\tilde{\gamma}_1, \tilde{\gamma}_2$ 。这可以产生一组实际尺度因子 γ_1, γ_2 和一组估计尺度因子 $\tilde{\gamma}_1, \tilde{\gamma}_2$ 。损失函数 \mathcal{L}_{scale} 可以定义为:

$$\mathcal{L}_{scale}(N_{\theta}) = \|(\tilde{\gamma}_1 - \tilde{\gamma}_2) - (\gamma_1 - \gamma_2)\|_2。$$

[0042] 该损失函数可以定义由表示两个已知的已知尺度因子 γ_1, γ_2 之差的项 $\gamma_1 - \gamma_2$ 表达的实际相对尺度和由表示补片级尺度估计 $\tilde{\gamma}_1, \tilde{\gamma}_2$ 之差的项 $\tilde{\gamma}_1 - \tilde{\gamma}_2$ 表示的估计相对尺度之间的失配。如果失配最小,则可以认为网络 N_{θ} 精确地估计了相对尺度,因为估计的相对尺度类似于实际的相对尺度。根据该损失函数训练尺度估计器可以提供所谓的尺度对比学习,因为尺度估计器的机器可学习模型部件可以被训练来预测一个图像应该被内插多少以匹配另一个图像。这样的方法不需要任何专用的深度或尺度标签,而是仅通过采样的尺度因子 γ_1, γ_2 之间的差(delta)来监督。在具体的实施例中,可以使用Adam优化器和设置为 $1 \cdot 10^{-3}$ 的学习速率,以128的批大小来执行100轮(epoch)的训练。可以使用下面的类PyTorch伪代码来描述训练过程,其中实际相对尺度被称为“true_scale”,估计相对尺度被称为“pred_scale”,并且其中MSE被用作误差项:

```

1 import random
2 import MSE
3
4 def train_model_one_step(model, optimizer, image ):
5     gamma_1 = random.uniform(0, 8)
6     gamma_2 = random.uniform(0, 8)
7     image_1 = rescale(image, gamma_1)
8     image_2 = rescale(image, gamma_2)
9
10    true_scale = gamma_1 - gamma_2
11    pred_scale = model(image_1) - model(image_2)
12    loss = MSE(true_scale, pred_scale)
13    loss.backward()
14    optimizer.step()
15    optimizer.zero_grad()

```

继续参考训练,注意,也可以使用任何其他合适的损失函数,例如使用与MSE不同的误差项、诸如非平方误差等的函数。此外,替代使用尺度因子的差值,可以使用尺度因子关系的比率或其他类型的表达式。例如, \mathcal{L}_{scale} 可以基于尺度因子比率的差异:

$$\mathcal{L}_{scale}(N_{\theta}) = \|(\tilde{\gamma}_1/\tilde{\gamma}_2) - (\gamma_1/\gamma_2)\|_2$$

在特定示例中,尺度因子的关系可以表达为尺度因子的差的对数,或者表达为尺度的对数。

[0043] 在一些示例中,可以考虑多于两个图像补片例如通过使用三个或更多尺度因子来定义损失函数。在一些示例中,尺度因子中的至少一个是1.0,例如,表示单位尺度因子。

[0044] 已经训练了尺度估计器,并且特别是具有参数 θ 的尺度估计器的机器可学习模型部件,特征提取器和尺度估计器的组合可以用于估计图像中对象的相对尺度。

[0045] 图5A-5C图示了尺度估计器的推断。这里,图5A示出了特征提取器和尺度估计器的示例输入图像400,在该特定示例中示出了花田形式的场景。图5B示出了被分割成不重叠的图像补片412的输入图像410,其中每个图像补片用作特征提取器和尺度估计器的输入。图5C示出了由尺度估计器生成的场景几何图420,该场景几何图420可以被生成成为针对每个图像补片的补片级尺度估计的类图像表示。这里,不同的灰色色调指示不同的尺度估计。可以看出,场景几何图指示场景包含在图像底部处更靠近相机并且在图像中间和顶部附近进一步远离相机的对象。如箭头430所示,场景几何图420可以在空间上被放大到更高的分辨率,诸如输入图像410的分辨率,例如使用双三次插值,产生放大的场景几何图440。在一些示例中,例如,在生成场景几何图420、440之前或在那之后,可以从每个单独的补片级尺度估计中减去最小的所估计的补片级尺度估计 γ_{min} 。推断过程可以由下面类PyTorch伪代码来描述:

```
1 import split_into_patches
2 import upsample
3
4 def estimate_scale(model, image ):
5     # image.shape = 1, 3, 512 , 256
6     patches = split_into_patches (image, patch_size =64)
7     # now the image is split into 8x4=32 patches
8     # each of size 64 x64
9     # patches.shape = 32, 3, 64, 64
10    scales = model(patches)
11    # scales.shape = 32, 1
12    scales = scales.view(8, 4)
13    scales = upsample(scales, factor=64, mode ='bicubic')
14    # scales.shape = 512, 256
15    scales = scales - scales.min()
16    return scales
```

[0046] 可以设想对象的相对尺度估计的各种用途,场景几何图的生成仅仅是示例。然而,通过估计输入图像的图像补片的补片级尺度估计来容易地生成场景几何图的能力在许多应用中可能是有利的。这样的场景几何图对于如在交通堵塞、拥挤的空间、体育场、音乐会、田野等图像的情况下其中以密集的布置出现相同或相似类型的对象(诸如汽车、人、花等)的场景的图像可能特别准确。

[0047] 一个具体的示例是车辆的车载相机的图像。在车辆遇到交通堵塞的情况下,道路本身和道路标记(例如,作为线)可能不可见或仅部分可见。该场景可能非常密集,因为在该车辆前面存在密集布置的其他车辆。在该情况下,道路的几何,例如其曲率,可以从场景几何图中估计,该场景几何图进而可以通过估计场景中汽车的相对位置和尺度来获得。

[0048] 另一个示例是宽的道路上方某处的交通相机,它通常观察跨过道路的行人,或汽车。交通相机可以执行自动对象检测,并且可以被训练来检测实际跨过道路的人。场景几何图可以用于执行健全性检查,因为它可以指示,在其中敞篷双层公共汽车经过的示例中,由相机检测到的人位于地面以上的表面上,因此实际上不可能正在跨过道路。在这样的和类似的情况下,场景几何图因此可以用作对在基于图像的对象检测之后的决策逻辑的附加输入。

[0049] 图6示出了系统500,用于使用如其他地方所述的特征提取器和尺度估计器来估计图像中对象的相对尺度。系统500可以包括输入接口子系统,用于访问特征提取器和尺度估计器的数据表示。例如,也如图6所示,输入接口子系统可以包括数据存储接口540,经由该数据存储接口540可以访问特征提取器的数据表示554和尺度估计器的数据表示556。一般而言,数据存储接口540和数据存储装置550可以与参考图1针对数据存储接口140和数据存储装置150所描述的类型相同。图6进一步示出了包括图像的图像数据552的数据存储装置550,特征提取器和尺度估计器可以应用于该数据存储装置550,以估计其中对象的相对尺度。例如,图像数据552可以由相机的图像传感器获取,或者可以是由另一种类型的空间传感器(诸如激光雷达或雷达)获取的传感器数据,该传感器数据可以被表示为图像。在一些实施例中,替代从数据存储装置550访问,这样的传感器数据也可以例如经由传感器数据接口560或另一种类型的接口直接从传感器620接收。在这样的实施例中,传感器数据562可以由系统500“现场”接收,例如实时或伪实时接收。

[0050] 系统500可以进一步包括处理器子系统520,其可以被配置为在系统500的操作期间,将特征提取器和尺度估计器应用于图像数据552和/或作为图像数据的传感器数据562,以生成至少一个补片级尺度估计,或者在一些示例中,例如以场景几何图的形式,生成图像数据的相应图像补片的多个补片级尺度估计。一般而言,处理器子系统520可以被配置为执行如先前参考图3-5C和其他地方描述的任何功能。应当进一步领会,与图1的处理器子系统120一样,相同的考虑和实现选项适用于图6的处理器子系统520。应当进一步领会,除非另有说明,否则与图1的系统100一样,相同的考虑和实现选项通常可以应用于图6的系统500。

[0051] 图6进一步示出了系统500的各种可选组件。例如,在一些实施例中,系统500可以包括传感器数据接口560,用于直接访问由环境600中的传感器620获取的传感器数据562。传感器620可以但不必须是系统500的一部分。传感器620可以具有任何合适的形式,诸如图像传感器或另一种类型的空间传感器。传感器数据接口560可以具有在类型上对应于(一个或多个)传感器的类型的任何合适形式,包括但不限于低级通信接口、电子总线或如上所述用于数据存储接口540的类型的接口。

[0052] 在一些实施例中,系统500可以包括输出接口,诸如用于向例如环境600中的致动器630提供控制数据572的控制接口570。这样的控制数据572可以由处理器子系统520生成,以基于对尺度估计器的输出的分析来控制致动器630。例如,致动器630可以是电、液压、气动、热、磁和/或机械致动器。具体但非限制性的示例包括电马达、电活性聚合物、液压缸、压电致动器、气动致动器、伺服机构、螺线管、步进马达等。由此,系统500可以响应于图像数据中的(一个或多个)对象的相对尺度的估计而行动,以例如控制制造过程,控制机器人系统或自主车辆等。

[0053] 在其他实施例中(图6中未示出),系统500可以包括到呈现设备的输出接口,呈现

设备诸如显示器、光源、扬声器、振动马达等,其可以用于生成感官可感知的输出信号,该输出信号可以基于尺度估计器的输出来生成。感官可感知输出信号可以直接指示补片级尺度估计或场景几何图,但是也可以表示导出的感官可感知输出信号。使用呈现设备,系统500可以向用户提供感官可感知的反馈。

[0054] 一般而言,本说明书中描述的每个系统,包括但不限于图1的系统100和图6的系统500,可以体现为或体现在单个设备或装置中,诸如工作站或服务器。所述设备可以是嵌入式设备。所述设备或装置可以包括执行适当软件的一个或多个微处理器。例如,相应系统的处理器子系统可以由单个中央处理单元(CPU)来实现,但是也可以由这样的CPU和/或其他类型的处理单元的组合或系统来实现。软件可能已经被下载和/或存储在对应的存储器中,例如诸如RAM的易失性存储器或者诸如闪存的非易失性存储器。可替代地,相应系统的处理器子系统可以以可编程逻辑的形式在设备或装置中实现,例如作为现场可编程门阵列(FPGA)。通常,相应系统的每个功能单元可以以电路的形式实现。相应的系统也可以以分布式方式实现,例如,涉及不同的设备或装置,诸如分布式本地或基于云的服务器。在一些实施例中,系统500可以是配置为控制物理实体或制造过程的控制系统的一部分,或者可以是数据分析系统的一部分。在一些实施例中,系统500可以是车辆、机器人或类似的计算机控制实体的一部分,和/或可以表示被配置为控制所述实体的控制系统。

[0055] 图7示出了上述的示例,其中系统500被示出为在环境600中操作的(半)自主车辆610的控制系统。自主车辆600可以合并有系统500,以基于从集成到车辆600中的相机622获得的传感器数据来控制诸如自主车辆的转向和制动之类的各方面。例如,在自主车辆600预期会遇到交通堵塞的情况下,系统500可以控制电马达632执行(再生)制动,如可以从场景几何图中检测到的。

[0056] 图8示出了估计图像中对象的相对尺度的计算机实现的方法700。方法700可以对应于图6的系统500的操作,但是也可以使用或由任何其他系统、机器、装置或设备来执行。方法700被示出为包括,在题为“提供特征提取器”的步骤中,提供710如本说明书其他地方所述的特征提取器,以及在题为“提供尺度估计器”的步骤中,提供720如本说明书其他地方所述的尺度估计器提取器。方法700进一步被示出为包括,在题为“将特征提取器和尺度估计器应用于图像补片”的步骤中,将特征提取器和尺度估计器应用于730图像的至少一个图像补片,以获得该至少一个图像补片的补片级尺度估计,可选地对其他图像补片重复740所述步骤730,并且在题为“输出(一个或多个)补片级尺度估计的数据表示”的步骤中,输出750(一个或多个)补片级尺度估计的数据表示,例如以如本文其他地方所述的场景几何图的形式。

[0057] 在一些实施例中,图2的计算机实现的方法200和图8的计算机实现的方法700可以由相同的计算机程序来体现,或者可以由相同的系统来执行。在其他实施例中,图2的计算机实现的方法200和图8的计算机实现的方法700可以由不同的计算机程序来体现,或者可以由不同的系统来执行。

[0058] 应当领会,一般而言,分别在图2和图8中的计算机实现的方法200和700的操作或步骤可以以任何合适的顺序(例如,连续地、同时地或其组合)来执行,在适用的情况下,受制于例如通过输入/输出关系,特定的顺序是必要的。

[0059] 本说明书中描述的每个方法、算法或伪代码可以在计算机上实现为计算机实现的

方法、专用硬件或这两者的组合。同样如图9所示,用于计算机的指令,例如可执行代码,可以例如以机器可读物理标记的系列810的形式和/或作为具有不同电(例如,磁)或光学特性或值的元件系列存储在计算机可读介质800上。可执行代码可以以暂时性或非暂时性的方式存储。计算机可读介质的示例包括存储器设备、光存储设备、集成电路、服务器、在线软件等。图9示出了光盘800。在替代实施例中,计算机可读介质800可以包括特征提取器的数据表示和/或尺度估计器的数据表示,如本说明书中其他地方所述的那样。

[0060] 示例、实施例或可选特征,无论是否被指示为非限制性的,都不应被理解为限制所要求保护的本发明。

[0061] 提供数学符号和记号是为了便于解释本发明,而不应被解释为限制权利要求。

[0062] 应当注意,上述实施例说明而非限制了本发明,并且本领域技术人员应当能够在不脱离所附权利要求的范围的情况下设计许多替代实施例。在权利要求中,放置在括号之间的任何附图标记都不应被解释为限制权利要求。动词“包括”及其变形的使用不排除除权利要求中所述的元素或阶段之外的元素或阶段的存在。元素前面的冠词“一”或者“一个”不排除多个这样的元素的存在。诸如“至少一个”的表达当在元素列表或元素组之前时,表示从该列表或组中选择所有元素或元素的任何子集。例如,表述“A、B和C中的至少一个”应该理解为仅包括A、仅包括B、仅包括C、包括A和B这两者、包括A和C这两者、包括B和C这两者、或者包括全部的A、B和C。本发明可以通过包括若干个不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干个部件的设备权利要求中,这些部件中的若干个可以由同一个硬件项目来体现。在相互不同的从属权利要求中引用某些措施的这一事实并不指示不能有利地使用这些措施的组合。

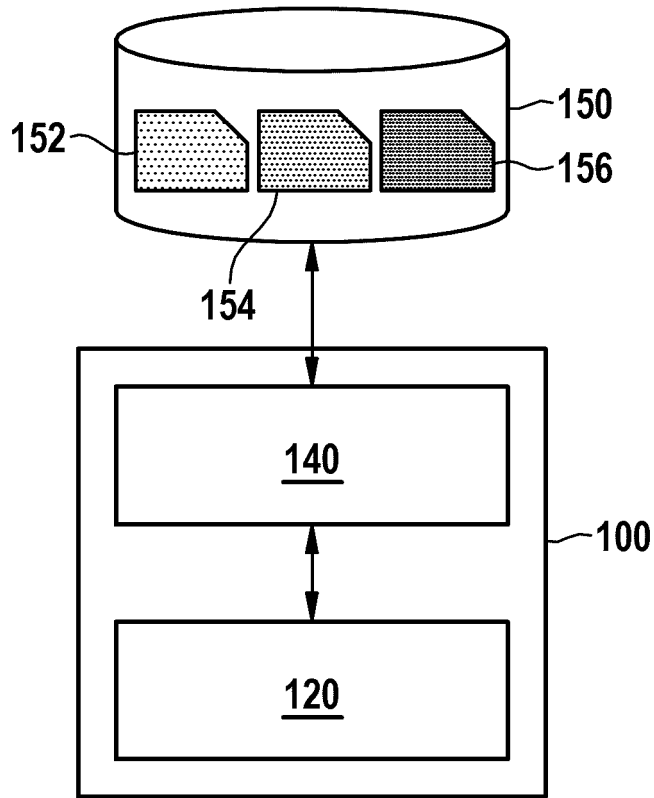


图 1

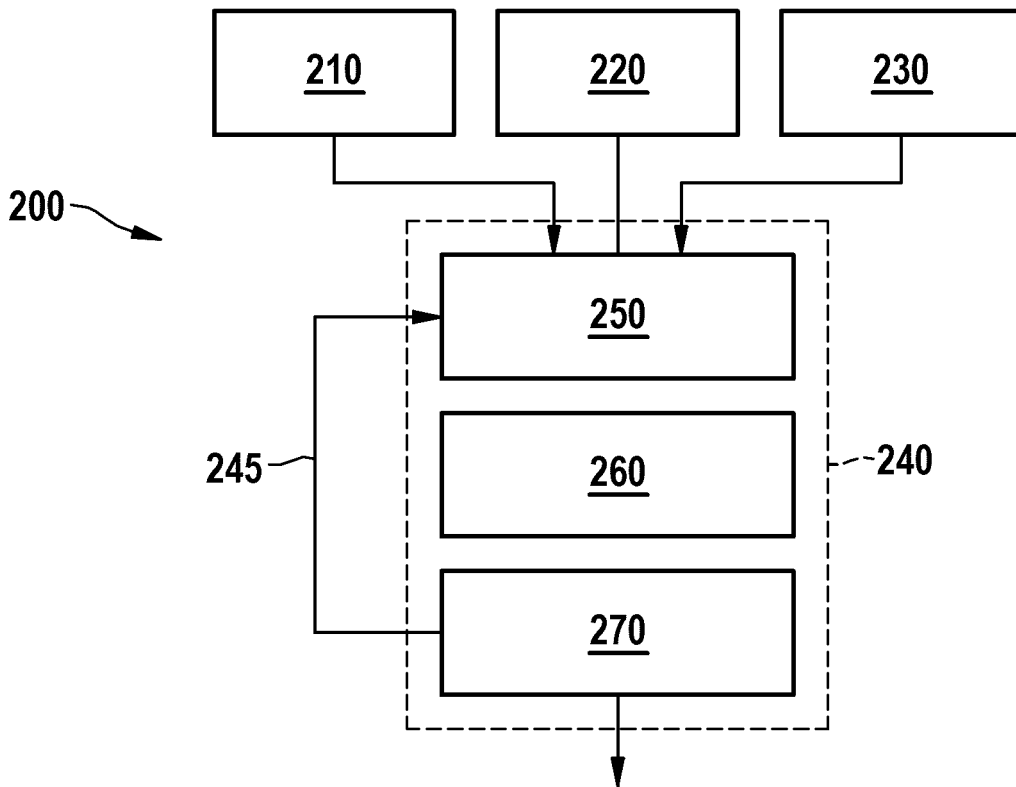


图 2

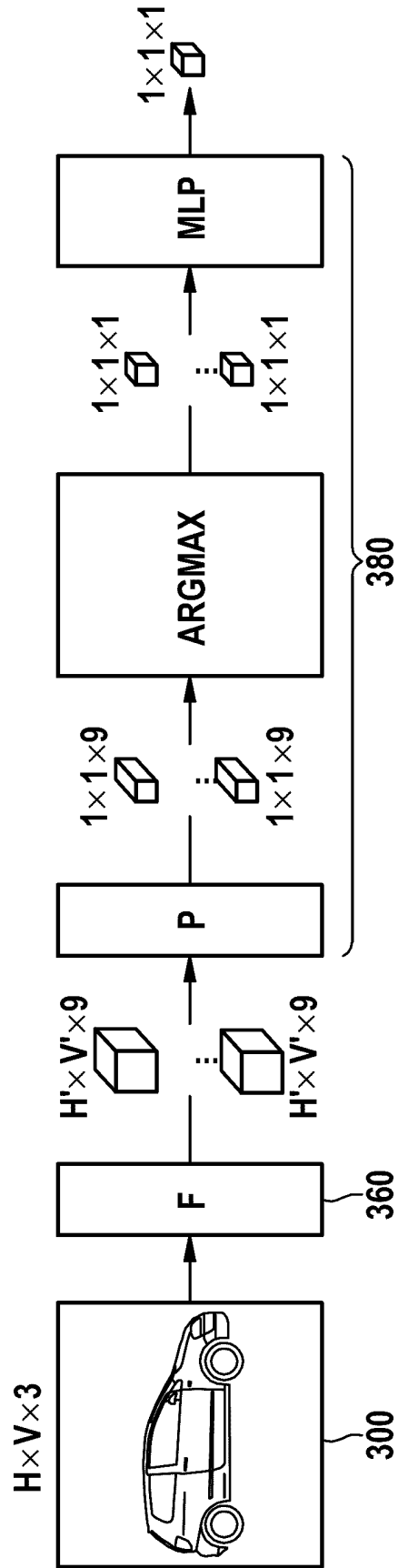


图 3

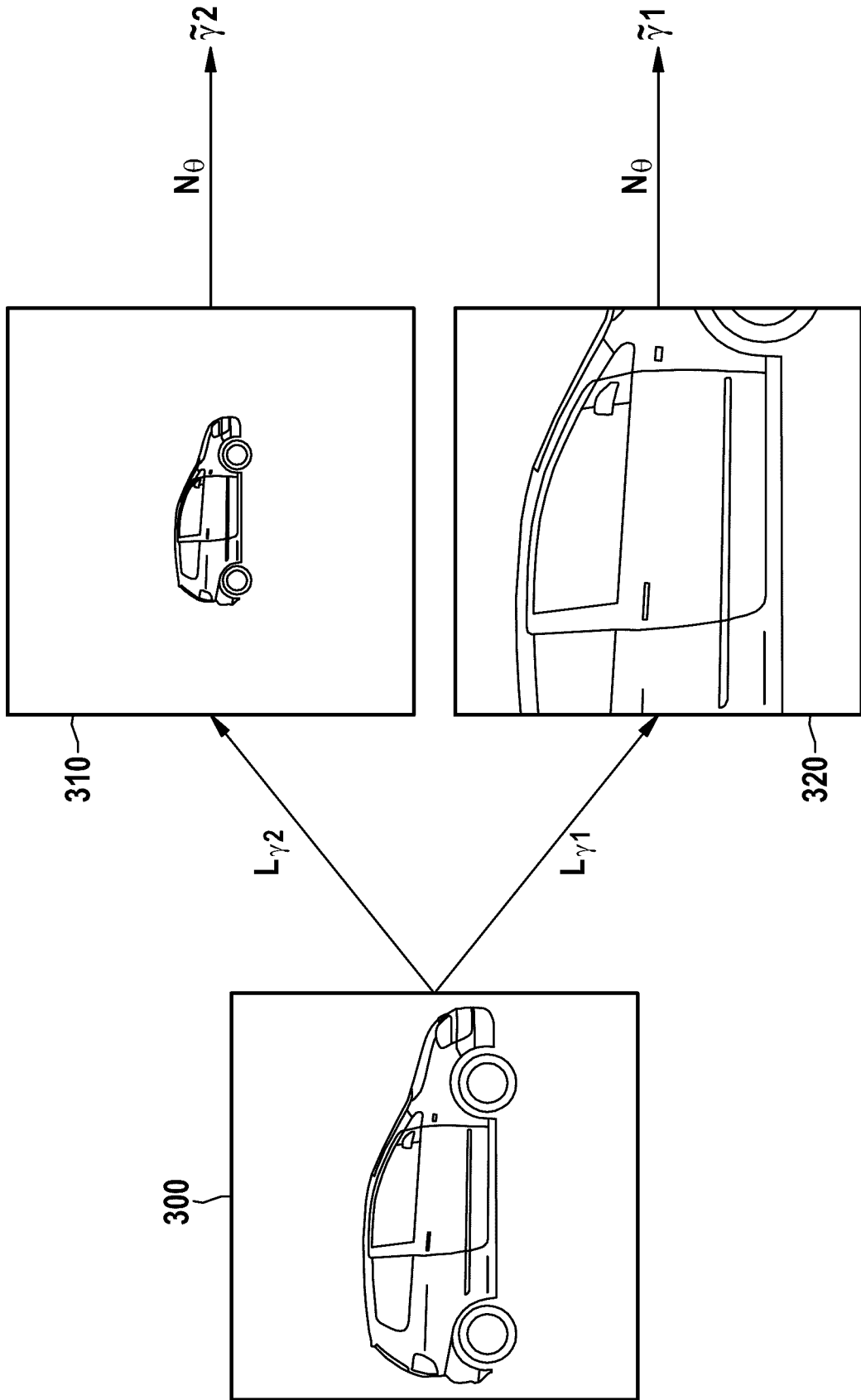


图 4



图 5A

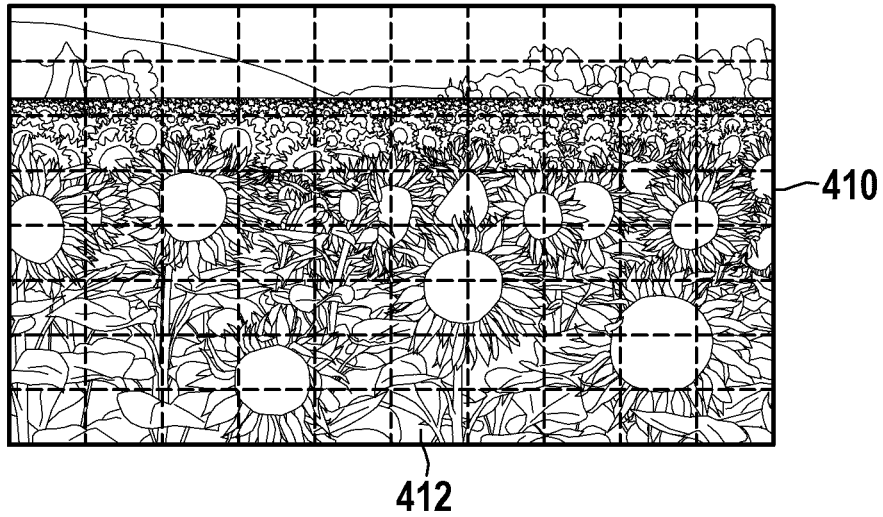


图 5B

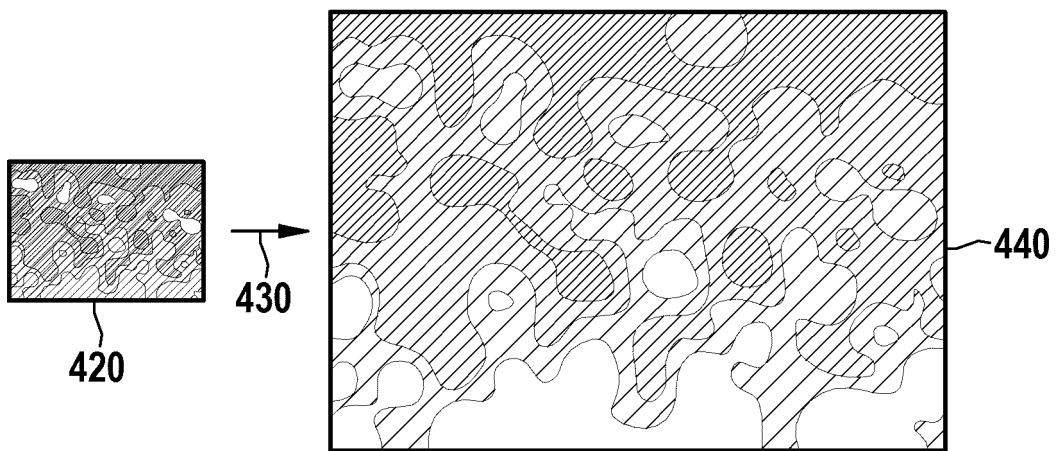


图 5C

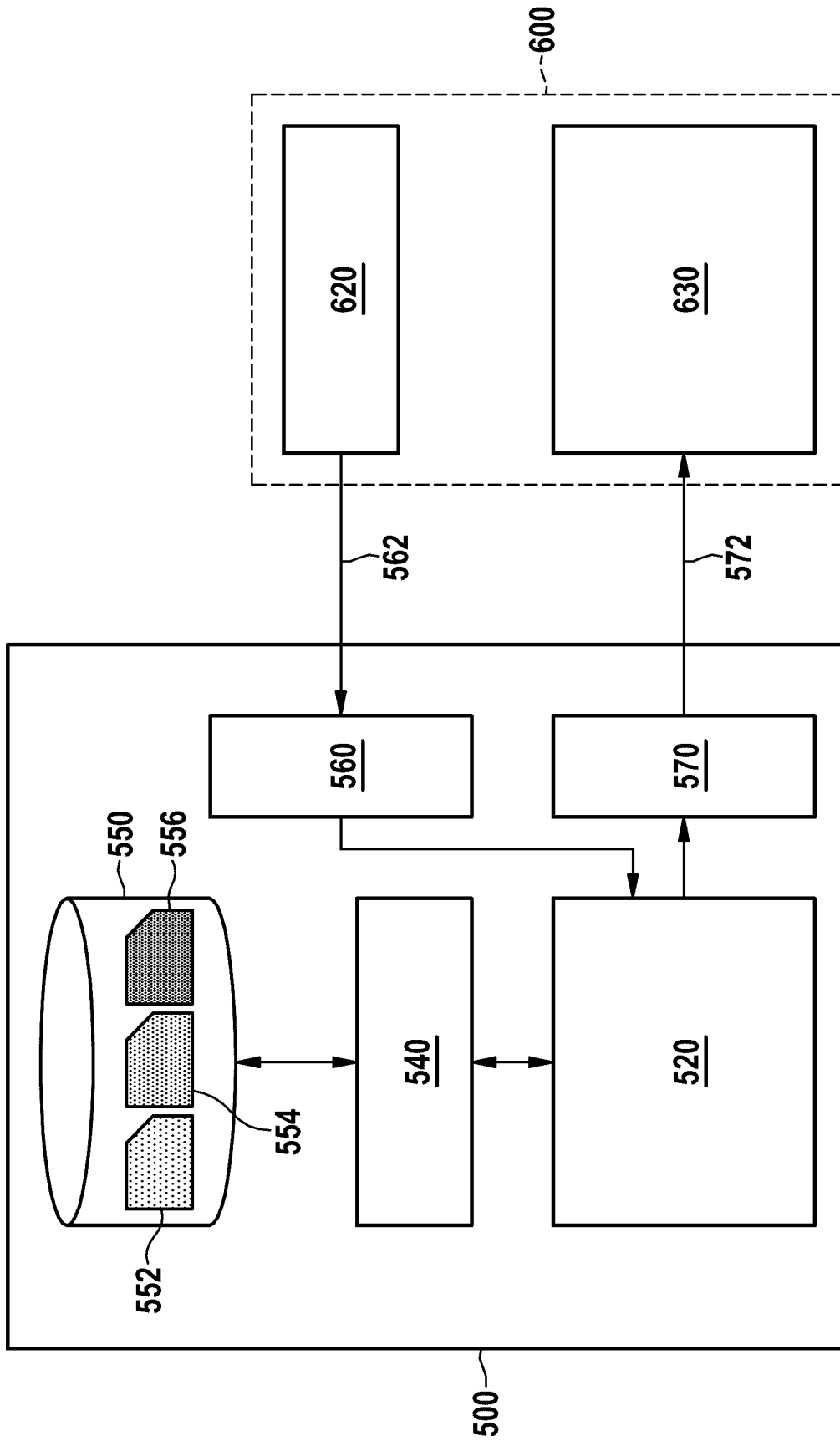


图 6

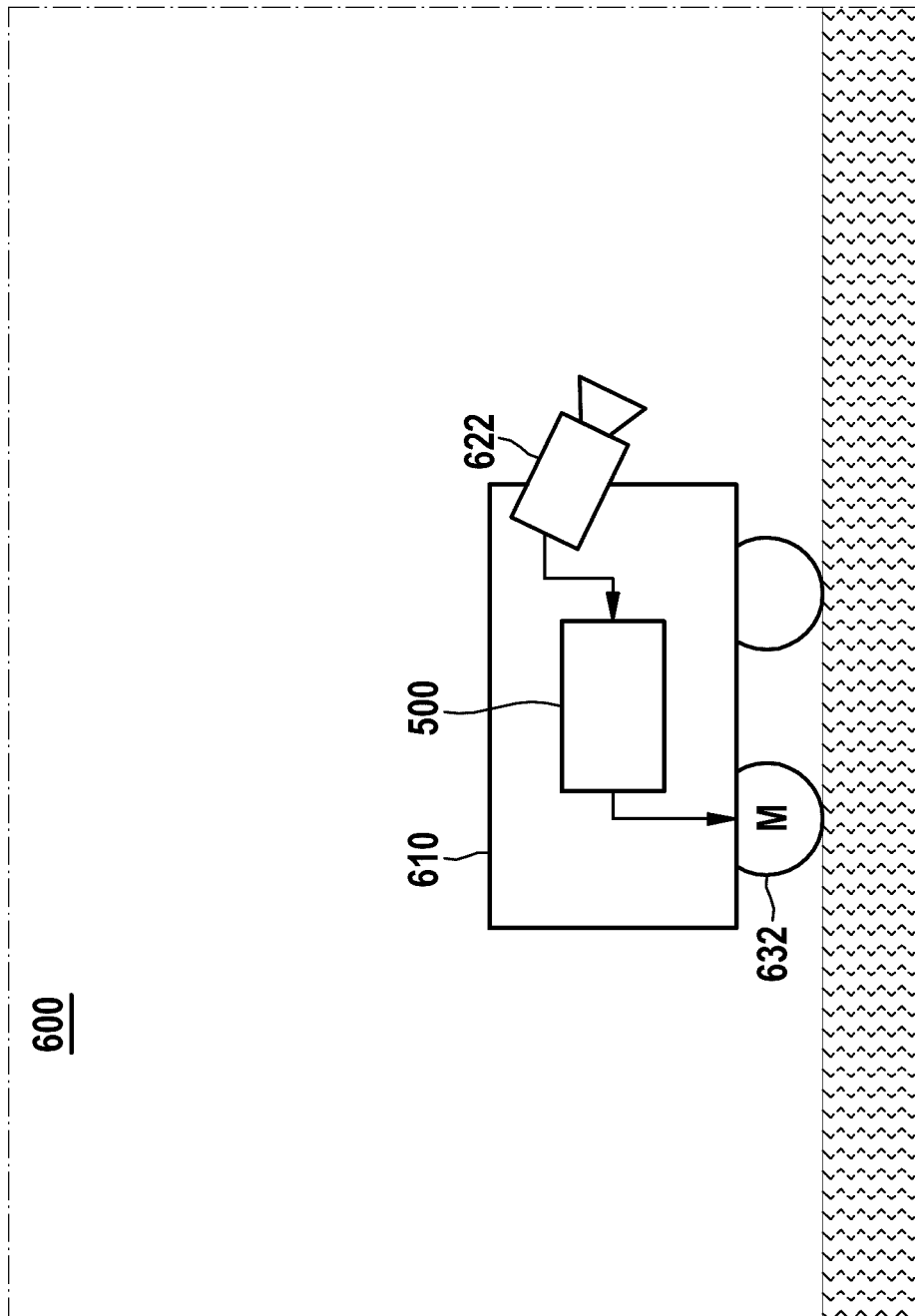


图 7

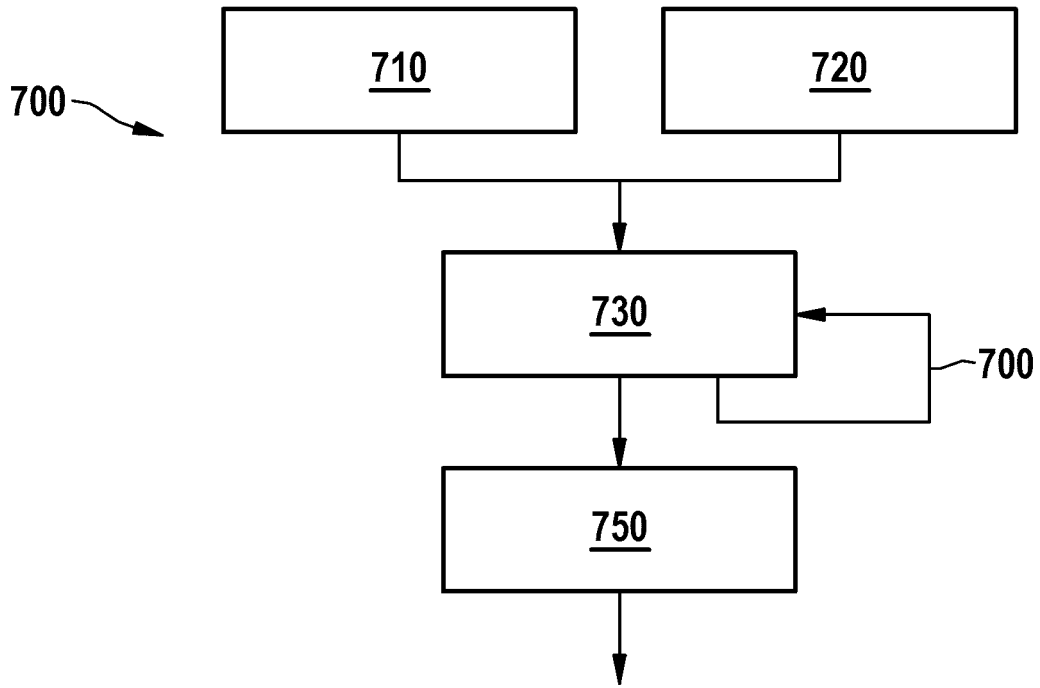


图 8

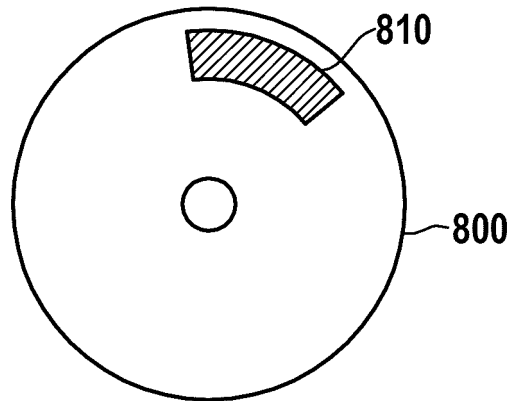


图 9