



(12)发明专利

(10)授权公告号 CN 108885683 B

(45)授权公告日 2020.05.22

(21)申请号 201680083830.2

(22)申请日 2016.03.28

(65)同一申请的已公布的文献号
申请公布号 CN 108885683 A

(43)申请公布日 2018.11.23

(85)PCT国际申请进入国家阶段日
2018.09.20

(86)PCT国际申请的申请数据
PCT/CN2016/077505 2016.03.28

(87)PCT国际申请的公布数据
W02017/166019 EN 2017.10.05

(73)专利权人 北京市商汤科技开发有限公司
地址 100084 北京市海淀区中关村东路1号
院3号楼7层710-712房间

(72)发明人 王晓刚 初晓 欧阳万里 李鸿升

(74)专利代理机构 北京英赛嘉华知识产权代理
有限责任公司 11204
代理人 王达佐 王艳春

(51)Int.Cl.
G06K 9/00(2006.01)

(56)对比文件
CN 1573660 A,2005.02.02,
CN 102509086 A,2012.06.20,
US 2007268295 A1,2007.11.22,
US 2014002338 A1,2014.01.02,
丁肖.针对静态图像的人体姿态估计.《中国
优秀硕士学位论文全文数据库 信息科技辑》
.2016,第2016年卷(第1期),I138-775页.

审查员 张玲

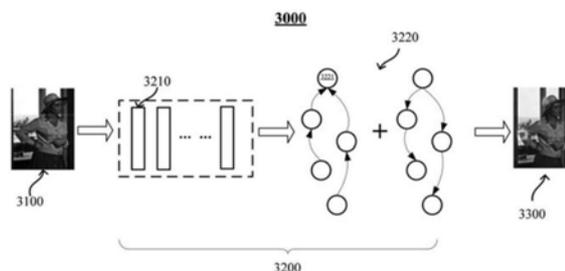
权利要求书2页 说明书5页 附图6页

(54)发明名称

用于位姿估计的方法和系统

(57)摘要

本公开涉及一种用于位姿估计的方法和系统。所述方法包括：从图像中提取多组部位特征图，所提取的每一组部位特征图对特定身体部位的消息进行编码并形成部位特征网络的节点；将所提取的每一组部位特征图的消息传递通过所述部位特征网络，以更新所提取的部位特征图，从而使所提取的每一组部位特征图包含上游节点的消息；基于所更新的部位特征图，估计所述图像内的身体部位。



1. 一种用于位姿估计的方法,包括:

从图像中提取多组部位特征图,所提取的每一组部位特征图表示身体部位并形成部位特征网络的节点;

将所提取的每一组部位特征图的消息在两个相反的方向上各传递通过所述部位特征网络一次,以更新所提取的部位特征图,其中,通过所述更新使得所提取的每一组部位特征图包含上游节点的消息,且在不同方向上执行更新后的每对部位特征图组合成得分图;以及

基于所组合的得分图,估计所述图像内的身体部位。

2. 根据权利要求1所述的方法,其中,对所述部位特征图的提取通过CNN执行。

3. 根据权利要求2所述的方法,其中,所述CNN是VGG网。

4. 根据权利要求3所述的方法,其中,在所述VGG网中采用三个池化层。

5. 根据权利要求1所述的方法,其中,所述消息的传递通过利用几何变换核的卷积操作来执行。

6. 一种用于位姿估计的系统,包括:

存储器,存储可执行组件;以及

处理器,电联接到所述存储器以执行所述可执行组件以进行以下操作:

从图像中提取多组部位特征图,所提取的每一组部位特征图表示身体部位并形成部位特征网络的节点;

将所提取的每一组部位特征图的消息在两个相反的方向上各传递通过所述部位特征网络一次,以更新所提取的部位特征图,其中,通过所述更新使得所提取的每一组部位特征图包含先前传递节点的消息,且在不同方向上执行更新后的每对部位特征图组合成得分图;以及

基于所组合的得分图,估计所述图像内的身体部位。

7. 根据权利要求6所述的系统,其中,对所述部位特征图的提取通过CNN执行。

8. 根据权利要求7所述的系统,其中,所述CNN是VGG网。

9. 根据权利要求8所述的系统,其中,在所述VGG网中采用三个池化层。

10. 根据权利要求6所述的系统,其中,所述消息的传递通过利用几何变换核的卷积操作来执行。

11. 一种计算机可读存储介质,所述计算机可读存储介质存储有可由处理器运行以执行以下操作的计算机可读指令,所述操作包括:

从图像中提取多组部位特征图,所提取的每一组部位特征图表示身体部位并形成部位特征网络的节点;

将所提取的每一组部位特征图的消息在两个相反的方向上各传递通过所述部位特征网络一次,以更新所提取的部位特征图,其中,通过所述更新使得所提取的每一组部位特征图包含上游节点的消息,且在不同方向上执行更新后的每对部位特征图组合成得分图;以及

基于所组合的得分图,估计所述图像内的身体部位。

12. 根据权利要求11所述的计算机可读存储介质,其中,对所述部位特征图的提取通过CNN执行。

13. 根据权利要求12所述的计算机可读存储介质,其中,所述CNN是VGG网。
14. 根据权利要求13所述的计算机可读存储介质,其中,在所述VGG网中采用三个池化层。
15. 根据权利要求11所述的计算机可读存储介质,其中,所述消息的传递通过利用几何变换核的卷积操作来执行。

用于位姿估计的方法和系统

技术领域

[0001] 本公开涉及一种用于位姿估计的方法和系统。

背景技术

[0002] 人体位姿估计旨在估计图像中的身体部位的位置,这可应用于各种视觉任务,例如动作识别、跟踪和人机交互。尽管已经努力了很长时间,但是位姿估计仍然是具有挑战性且未被解决的问题。肢体取向、衣物、视角、背景杂物、截断和遮蔽的较大变化会使得位姿估计更加困难。

[0003] 众所周知,位姿估计可通过优化从输入图像生成的外观得分图来实现。关于得分图,某一位置处的视觉图案的信息被概括成单个概率值,以指示存在对应的身体部位的可能性。例如,如果肘部位置在得分图上具有较大响应,那么可以得出此位置可能属于肘部的结论。然而,肘部的面内和面外旋转、与肘部相连接的上臂和下臂的定向、遮蔽状态等都是不确定的。此类详细信息对估计其它身体部位的位置来说是有价值的,但是在得分图中无法看出,这使得身体部位当中的结构学习的有效性变低。

发明内容

[0004] 下文呈现对本公开的简化概述,以便提供对本公开的一些方面的基本理解。此概述并非本公开的详尽综述。此概述既不标识本公开的重要或关键要素,也不划定本公开的特定实施方式的任何范围或权利要求书的任何范围。其唯一目的是以简化形式呈现本公开的一些概念,以作为稍后呈现的更详细描述的前言。

[0005] 可以观察到,部位间信息在利用卷积神经网络学习分层特征表示的特征级下得到很好的保留。为了至少部分地解决上述问题中的一个问题,在本申请的一个方面中提出一种用于位姿估计的方法。该方法包括:从图像中提取多组部位特征图,所提取的每一组部位特征图表示身体部位并形成部位特征网络的节点;将所提取的每一组部位特征图的消息传递通过部位特征网络,以更新所提取的部位特征图,通过所述更新使所提取的每一组部位特征图包含上游节点的消息;基于所更新的部位特征图,估计图像内的身体部位。因为身体部位之间的部位特征图所含的丰富信息由于特征级中的消息传递过程而可以被大部分保留下来,所以位姿估计准确性将会增强。

[0006] 在本申请的一个实施方式中,在相反方向上传递消息,且在不同方向上执行更新后的每一对部位特征图组合成得分图。使用所生成的得分图以更高的准确性来估计图像内的位姿。

[0007] 在本申请的一个实施方式中,通过CNN提取部位特征图,且优选地通过VGG网提取部位特征图。

[0008] 在本申请的一个实施方式中,仅启用VGG网的3个池化层来保留较高分辨率。

[0009] 在本申请的一个实施方式中,通过利用几何变换核的卷积操作传递消息。

[0010] 另一方面,提供一种用于位姿估计的系统。该系统包括:存储可执行组件的存储

器;以及处理器,其电联接到存储器以执行可执行组件从而进行以下操作:从图像中提取多组部位特征图,所提取的每一组部位特征图表示身体部位并形成部位特征网络的节点;将所提取的每一组部位特征图的消息逐个节点地传递通过部位特征网络,以更新所提取的部位特征图,从而使所提取的每一组部位特征图包含先前传递的节点的消息;基于所更新的部位特征图,估计图像内的身体部位。

[0011] 另一方面,提供了一种计算机可读存储介质,所述计算机可读存储介质存储有可由处理器运行以执行以下操作的计算机可读指令,所述操作包括:从图像中提取多组部位特征图,所提取的每一组部位特征图表示身体部位并形成部位特征网络的节点;将所提取的每一组部位特征图的消息传递通过部位特征网络,以更新所提取的部位特征图,通过所述更新使所提取的每一组部位特征图包含上游节点的消息;基于所更新的部位特征图,估计图像内的身体部位。

附图说明

[0012] 在下文参照附图描述本申请的示例性非限制性实施方式。附图是说明性的,且一般未按确切比例绘制。不同图上的相同或相似元件用相同附图标号标记。

[0013] 图1示出根据本申请的一个实施方式的用于根据输入图像估计位姿的示例性系统;

[0014] 图2是示出根据本申请的一个实施方式的用于根据输入图像估计位姿的过程的流程图;

[0015] 图3示出根据本申请的一个实施方式的用于根据输入图像估计位姿的另一示例性系统;

[0016] 图4示出根据本申请的一个实施方式的几何变换过程;

[0017] 图5示出根据本申请的一个实施方式的特征图更新操作;以及

[0018] 图6示出根据本申请的一个实施方式的双向消息传递的过程。

具体实施方式

[0019] 现将详细参考发明人为实施本申请而提出的本申请的一些特定实施方式。附图中示出这些特定实施方式的示例。尽管结合这些特定实施方式描述了本申请,但本领域技术人员应了解,这并不意欲将本申请限于所描述的实施方式。在以下描述中,阐述众多具体细节以便提供对本申请的透彻理解。本申请可以在没有这些具体细节中的一些细节或全部细节的情况下实践。在其它情况下,并未详细描述众所周知的过程操作,以免不必要地混淆本申请。

[0020] 本文中使用的术语仅用于描述具体实施方式的目的,而并不意欲限制本申请。如本文所使用,除非上下文另外清楚地指示,否则单数形式“一”和“所述”旨在还包含复数形式。还应理解,术语“包括(comprise/comprising)”在本说明书中使用时指代所陈述的特征、整体、步骤、操作、元件和/或组件的存在,但并不排除一个或多个其它特征、整体、步骤、操作、元件、组件和/或其群组的存在或添加。

[0021] 现将参照图1描述用于根据输入图像估计位姿的示例性系统1000。利用卷积神经网络(Convolutional Neural Network,CNN)1200从输入图像1100生成语义上有意义的表

示。应了解,CNN 1200可采用多个层1210来学习输入图像的分层特征表示。较低层中的特征采集低级信息,而较高层中的那些特征可表示更抽象的概念,例如属性和对象类别。在此实施方式中,CNN 1200采用例如7个层。例如AlexNet、Clarifai、Overfeat、GoogleNet和视觉几何组(Visual Geometry Group,VGG)等广泛使用的CNN在卷积层之后采用全连接(fc)层,以获得全局特征向量。在全卷积网(FullyConvolutional Net,FCN)中,使用 1×1 卷积框架来替换fc层。在此实施方式中,全卷积VGG网优选地被选为基本模型。在优选实施方式中,仅启用5个池化层中的3个池化层以保留较高分辨率。在层fcn6中的位置(x,y)处获得的特征向量表示为 $h_{fcn6}(x,y)$ 。从层fcn6中提取的位置(x,y)处的身体部位k的特征向量计算为

$$[0022] \quad h_{fcn7}^k(x,y) = f(h_{fcn6}(x,y) \otimes w_{fcn7}^k) \quad (1)$$

[0023] 其中, \otimes 表示卷积操作,f表示非线性函数,且 w_{fcn7}^k 表示部位k的过滤器组。应注意, h_{fcn7}^k 含有从不同通道提取的一组部位特征图。身体部位的部位特征图含有人体位姿和外观的丰富信息和详细描述。

[0024] 因为空间分布和在不同部位处获得的部位特征图的共同出现率高度相关,所以传递部位之间的部位特征图中所含的丰富信息可有效改进在每个部位处习得的特征。在现有技术中,在得分图级下实施传递过程,这会导致重要的部位间信息的丢失。出乎意料地,当在特征级下传递消息时,部位之间的部位特征图中所含的丰富信息很大程度上被保留。

[0025] 在本申请中,可通过朝向相邻部位偏移一个身体部位的部位特征图来强化身体部位当中的几何约束。几何变换核对来自相邻部位的每对部位特征图之间的关系进行建模。为了优化在某一部位处获得的特征,预期利用全连接图形从所有其它部位接收信息。然而,为了直接对各部位的部位特征图之间在距离方面的关系进行建模,必须引入难以训练的较大变换核。第二,一些部位(例如头部和足部)之间的关系是不稳定的。在这些部位之间通过设计好的图形上的中间部位传递消息是有利的,因为两个邻近部位之间的相对空间分布稳定,且易于训练对应的核。图形上的邻近部位在距离上很接近,且在图形中具有相对稳定的关系。所提取的部位特征图组构成由结构化特征学习层1220处理的部位特征网络,其中每一组部位特征图在部位特征网络中占用一个节点1221。在示例性实施例中,所提取的每一组部位特征图的消息沿着单一方向传递通过部位特征网络。传递操作将参照图6详细说明。在消息传递之后,每一组部位特征图通过部位特征网络中传递的消息进行更新。因此,除占用叶节点的部位特征图以外,每一组部位特征图包含先前传递的节点的消息。在优选实施方式中,所更新的部位特征图变换(例如,线性组合)成得分图,可根据该得分图估计最终图像1300中的部位位置。利用所更新的部位特征图,可实现具有较高准确性的位姿估计。

[0026] 图2中示意性地示出了流程图,该流程图示出根据输入图像估计位姿的过程。在步骤S2020,通过CNN从输入图像中提取多组部位特征图。将提取的部位特征图分别分类到表示人体的各个部位的多个组中。在示例性实施方式中,所提取的部位特征图组构成人形网络。在步骤2040,所提取的每一组部位特征图的消息传递通过网络。在优选实施方式中,在步骤S2060,所更新的部位特征图以线性方式组合成得分图。在步骤S2080,基于得分图,检测出图像内的位姿。

[0027] 参照图3,提供位姿估计系统3000的替代实施方式。除CNN 3200还包括结构化特征

学习层3220以外,系统3000类似于上述系统1000。例如,CNN 3200可采用多个层3210来学习输入图像的分层特征表示,并且其中每一组部位特征图在部位特征网络中占用一个节点3221。在此实施方式中,消息双向传递通过部位特征网络,从而使得所提取的部位特征图能够包含在两个方向上邻近的节点的消息。在相反方向上的所更新的多对部位特征图组合得分图,从而进一步提高位姿估计的准确性。因此,最终图像3300中的位姿估计的准确性可得以提高。

[0028] 图4和图5示出可用于消息传递的详细的几何变换操作。如图4中所示,提供高斯分布(Gaussian distribution)下的特征图4100。三个不同的几何变换核4200分别与相同特征图4100卷积,从而产生以不同方式变换的特征图4300。如从图4中可见,所变换的特征图已分别朝左上、右上和左下偏移。在位姿估计中,可通过朝向邻近身体部位偏移一个身体部位的特征图来强化身体部位之间的几何约束。几何变换核对每对邻近部位的特征图之间的关系进行建模。为了示意性地示出上文所描述的过程,图5中示出示例。从输入图像5100中提取分别表示右侧肘部和右侧下臂的特征图5200和5300。下臂特征图5300中的一个特征图具有高响应,这是因为该下臂特征图5300是从描述没有衣服遮盖的朝向下方的下臂的通道提取的。与下臂特征图5300正相关的肘部特征图5200也具有高响应。预期使用下臂特征图5300来减少误报且增强右侧肘部上的响应。然而,直接将肘部特征图5200添加到下臂特征图5300并不适合,因为这两个关节之间存在空间错配。实际上,通过利用如上文所描述的几何变换核5400的卷积操作来将下臂特征图5300朝向右侧肘部偏移。然后,所变换的特征图5500与肘部特征图5200组合,以生成更新的特征图5600,从而改进位姿估计结果。

[0029] 图6中示出示例性双向消息传递过程。尽管在部位特征网络6100中,消息从底部传递到顶部,但是在部位特征网络6200中,消息沿相反方向传递。部位特征网络6100包括叶节点6105、6106、6109和6110以及其它节点6104、6108、6103、6102、6107和6101。部位特征网络6200包括叶节点6201以及其它节点6203、6202、6207、6204、6208、6205、6206、6210和6209。在这两个方向上,消息以类似于上文参照图5所描述的那样的方式传递。网络6100或6200中除叶节点6105、6106、6109、6110和6201以外的每个节点都可利用从上游节点传递来的消息更新。以节点6104为例,此过程可表示为下式,

$$[0030] \quad \mathbf{A}'_4 = f(\mathbf{A}_4 + \mathbf{A}'_5 \otimes \mathbf{w}^{a_5, a_4}) \quad (2)$$

[0031] 其中, \mathbf{A}'_4 表示在消息传递之后更新的部位特征图, \mathbf{A}_4 表示在消息传递之前的部位特征图,且 \mathbf{w}^{a_5, a_4} 表示节点6105和节点6104之间的变换核的组合。因为节点6103可从节点6104和节点6106接收消息,所以节点6103处的部位特征图可通过下式更新,

$$[0032] \quad \mathbf{A}'_3 = f(\mathbf{A}_3 + \mathbf{A}'_4 \otimes \mathbf{w}^{a_4, a_3} + \mathbf{A}'_6 \otimes \mathbf{w}^{a_6, a_3})。 \quad (3)$$

[0033] 网络6200中的部位特征图可以类似方式但在相反方向上更新,并因此不再详细论述。最后,两组更新的部位特征图 $[\mathbf{A}'_k, \mathbf{B}'_k]$ 可以线性方式组合成一组得分图,以指示存在对应的身体部位的可能性。

[0034] 本申请在一方面提供了一种用于位姿估计的方法,其包括:从图像中提取多组部位特征图,所提取的每一组部位特征图表示身体部位并形成部位特征网络的节点;将所提取的每一组部位特征图的消息传递通过所述部位特征网络,以更新所提取的部位特征图,通过所述更新使得所提取的每一组部位特征图包含上游节点的消息;以及基于所更新的部

位特征图,估计所述图像内的身体部位。

[0035] 所述消息的传递在相反方向上执行两次,且在不同方向上执行更新后的每对部位特征图组合成得分图。该方法还包括:基于所组合的得分图,执行对所述身体部位的估计。

[0036] 根据本申请的实施方式,对所述部位特征图的提取通过CNN执行。所述CNN是VGG网。当所述CNN是VGG网时,所述VGG网中采用三个池化层。

[0037] 所述消息的传递通过利用几何变换核的卷积操作来执行。

[0038] 本申请在另一方面提供了一种用于位姿估计的系统,其包括:存储器,存储可执行组件;以及处理器,电联接到所述存储器以执行所述可执行组件以进行以下操作:从图像中提取多组部位特征图,所提取的每一组部位特征图表示身体部位并形成部位特征网络的节点;将所提取的每一组部位特征图的消息传递通过所述部位特征网络,以更新所提取的部位特征图,通过所述更新使得所提取的每一组部位特征图包含先前传递节点的消息;以及基于所更新的部位特征图,估计所述图像内的身体部位。

[0039] 本申请在又一方面提供了一种计算机可读存储介质,所述计算机可读存储介质存储有可由处理器运行以执行以下操作的计算机可读指令,所述操作包括:从图像中提取多组部位特征图,所提取的每一组部位特征图表示身体部位并形成部位特征网络的节点;将所提取的每一组部位特征图的消息传递通过部位特征网络,以更新所提取的部位特征图,通过所述更新使所提取的每一组部位特征图包含上游节点的消息;基于所更新的部位特征图,估计图像内的身体部位。

[0040] 如本领域技术人员所能够理解的那样,本申请可实现为系统、方法或计算机程序产品。因此,本申请可采取完全为硬件的实施方式和方面,而在本文中,硬件通常被称为“单元”、“电路”、“模块”或“系统”。许多功能和许多原理在实施时最好由集成电路(integrated circuit, IC)支持,例如数字信号处理器和其软件或者专用IC。可以预期的是,本领域普通技术人员根据本文公开的概念和原理的教导能够通过最少的实验而容易地生成IC,而不必考虑例如由可用时间、当前技术和经济考量等驱使的可能繁重的工作量和许多其它设计选项。因此,为了简化和最小化混淆根据本申请的原理和概念的任何风险,对此类软件和IC(如果存在的话)的进一步论述将受限于对于优选实施方式所使用的原理和概念而言必要的部分。另外,本申请可采取完全为软件的实施方式(包含固件、驻存软件、微码等)或可采取组合了软件的实施方式。例如,系统可包括存储可执行组件的存储器以及处理器,所述处理器电联接到存储器以执行可执行组件来执行系统的、如参照图1到6所论述的操作。另外,本申请可采用体现在任何有形表达介质中的计算机程序产品的形式,所述有形表达介质具有体现于介质中的计算机可用程序代码。

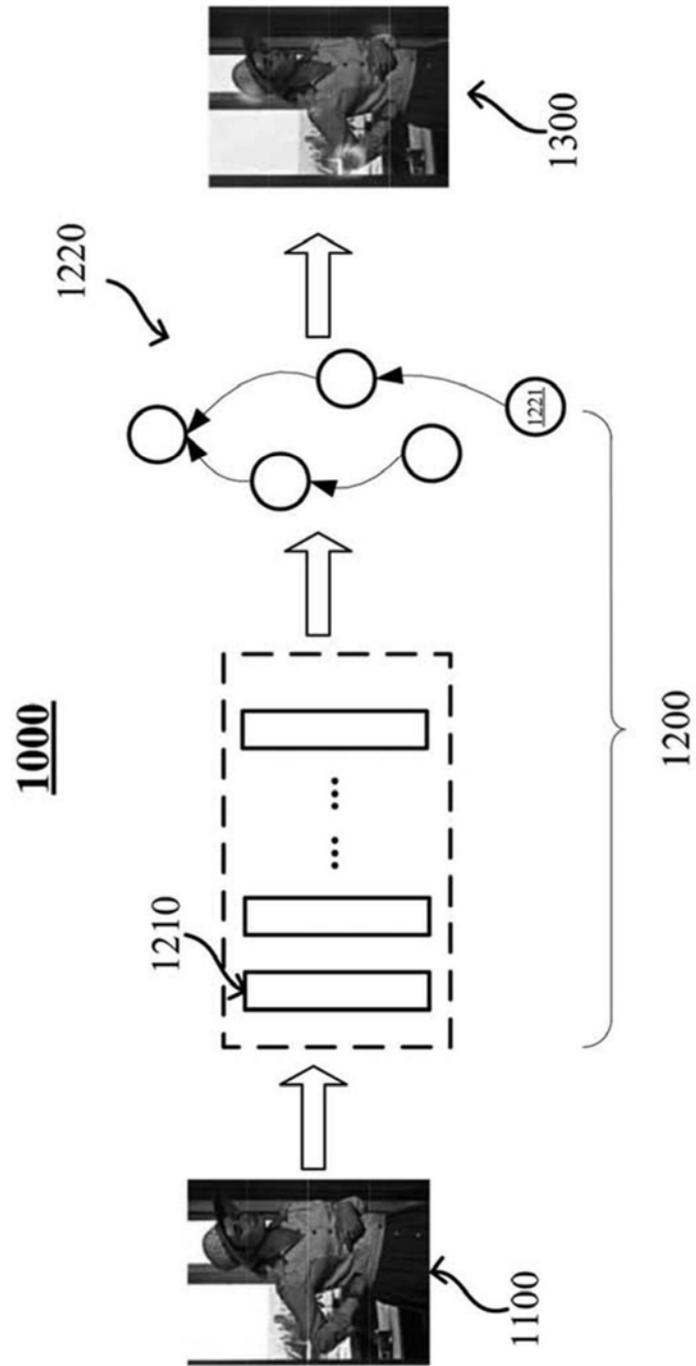


图1

2000

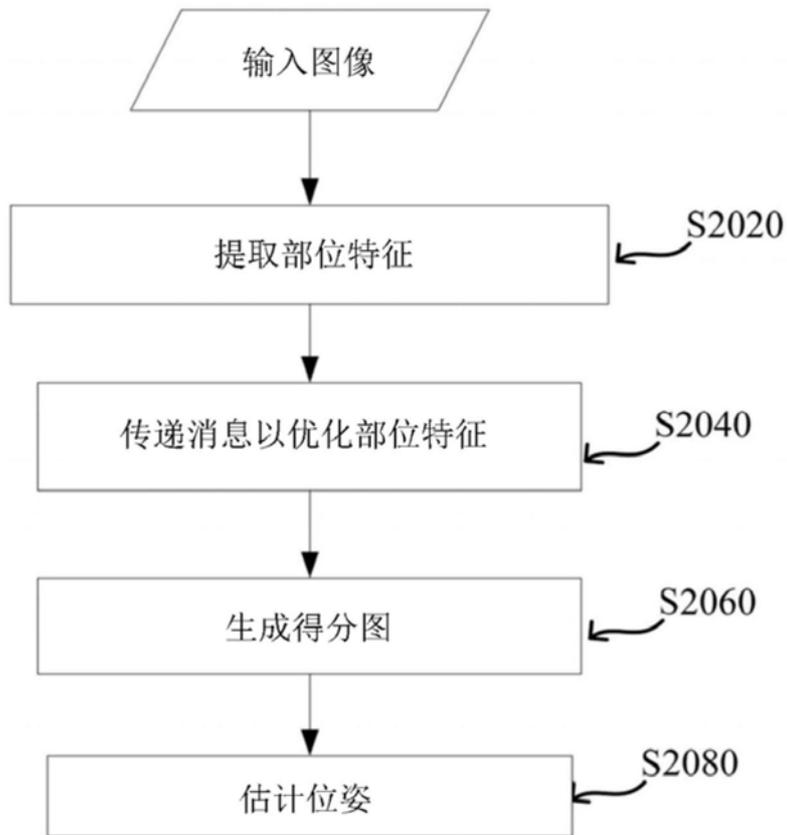


图2

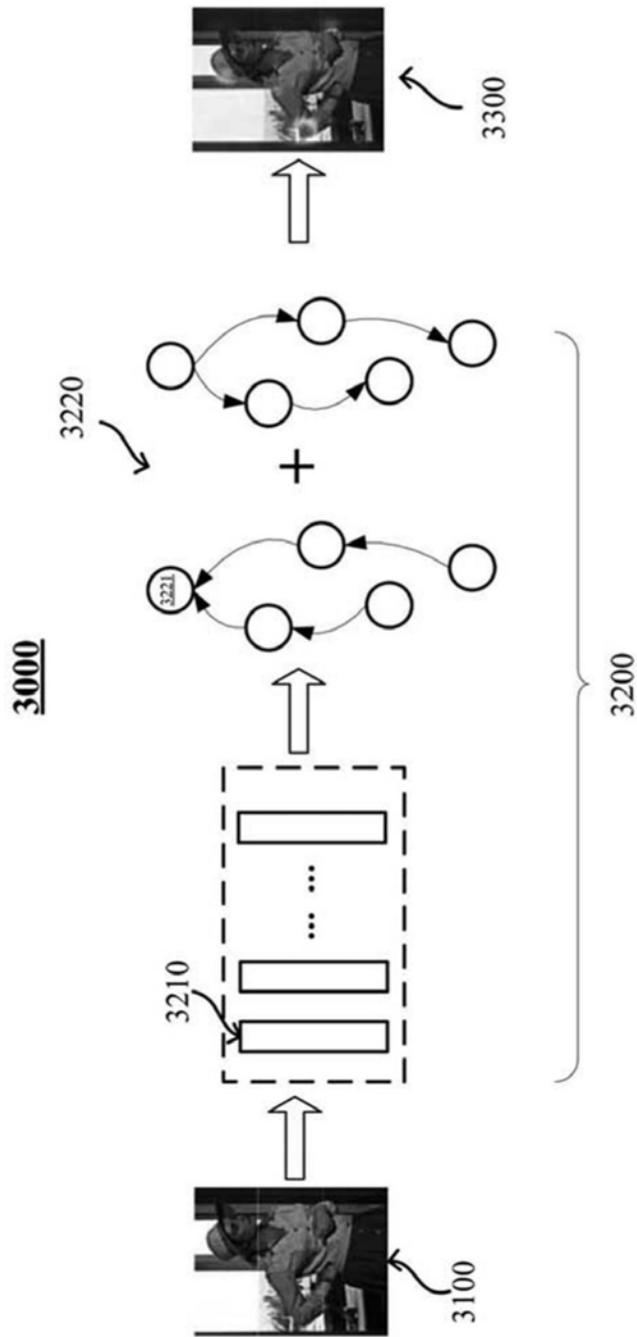


图3

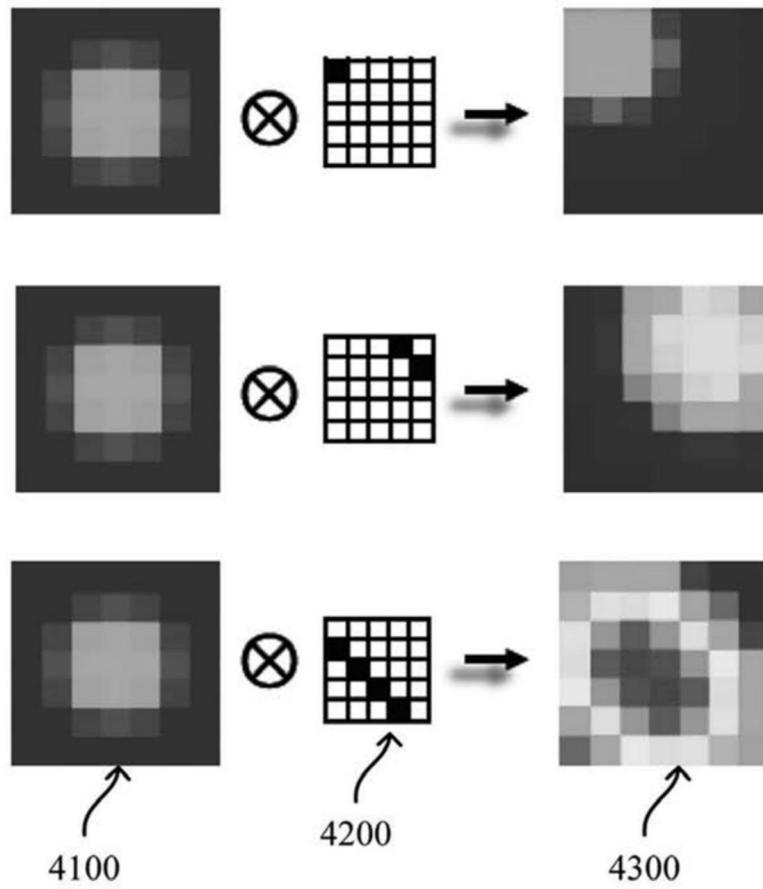


图4

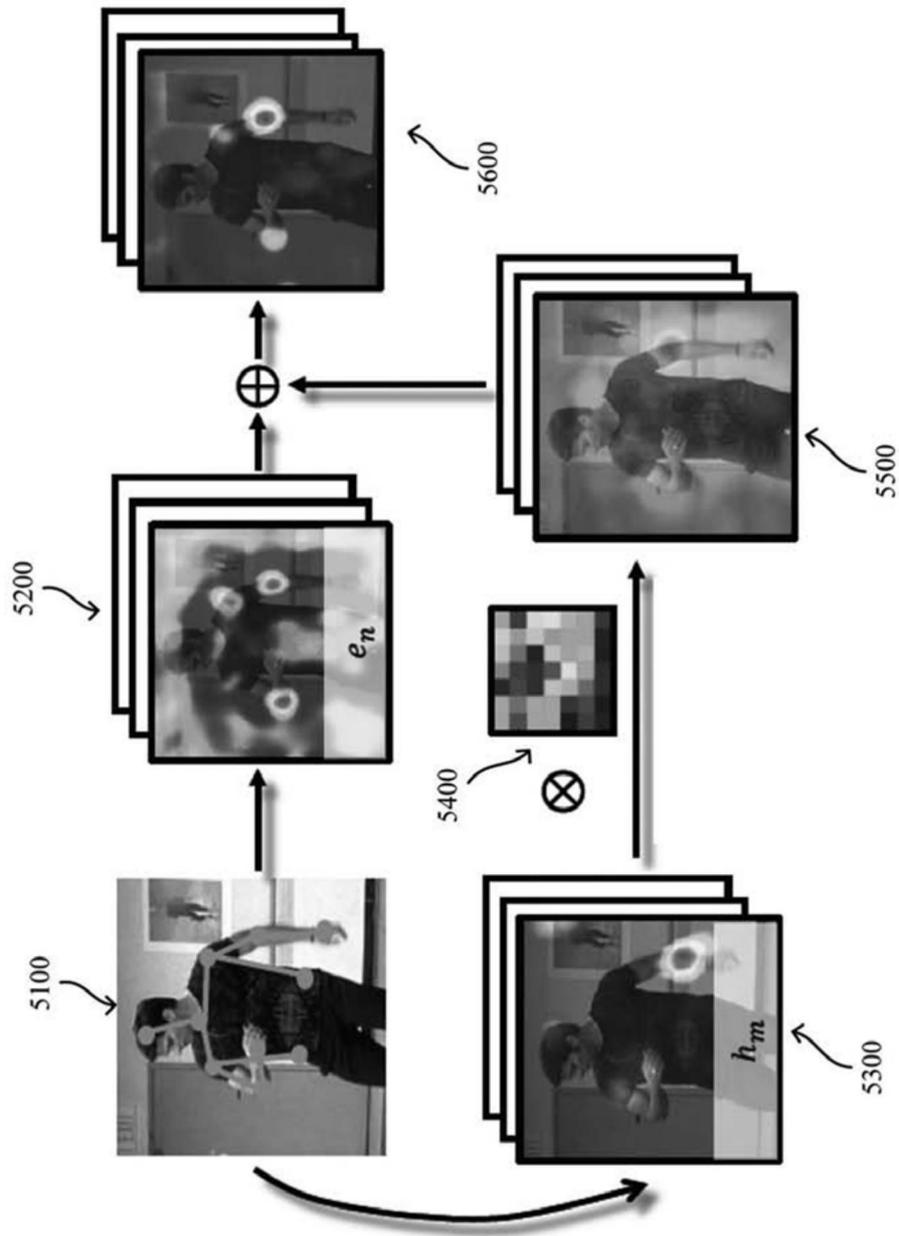


图5

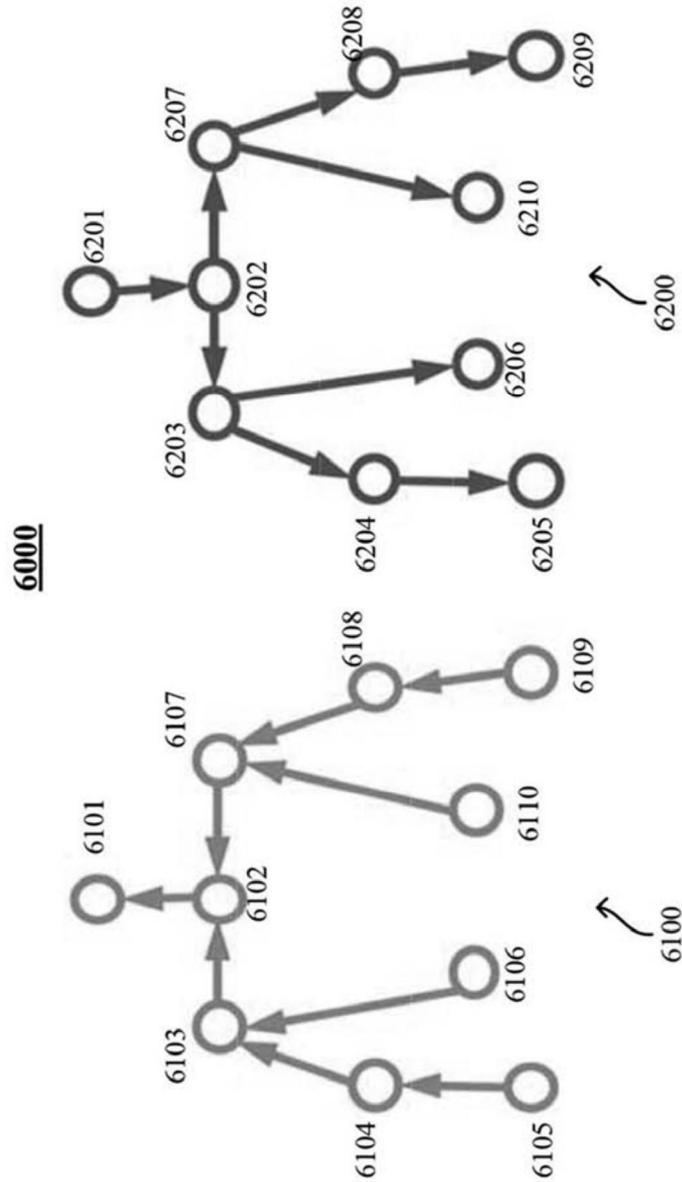


图6