



(12) 发明专利申请

(10) 申请公布号 CN 117063201 A

(43) 申请公布日 2023. 11. 14

(21) 申请号 202280024083.0

(72) 发明人 姜制远 俞彩花 朴胜煜

(22) 申请日 2022.03.28

(74) 专利代理机构 北京康信知识产权代理有限公司 11240

(30) 优先权数据

10-2021-0041562 2021.03.31 KR

专利代理师 李静

10-2022-0037097 2022.03.25 KR

(51) Int.Cl.

(85) PCT国际申请进入国家阶段日
2023.09.22

G06T 9/00 (2006.01)

(86) PCT国际申请的申请数据

PCT/KR2022/004312 2022.03.28

(87) PCT国际申请的公布数据

W02022/211409 KO 2022.10.06

(71) 申请人 现代自动车株式会社

地址 韩国首尔

申请人 起亚株式会社

梨花女子大学校产学协力团

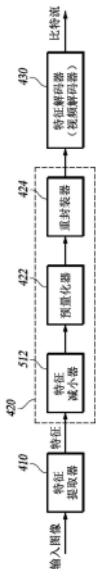
权利要求书2页 说明书12页 附图11页

(54) 发明名称

使用特征图减小来编解码机器视觉数据的方法和装置

(57) 摘要

公开了一种用于使用特征图的减小对机器视觉数据进行编解码的装置和方法。为了减小由机器任务专用深度学习模型提取的特征图的大小,提供一种机器视频编解码(VCM)编解码装置和方法。VCM编解码装置和方法利用减小特征图的空间和通道方面的冗余的稀疏方法,并且VCM编解码装置和方法也利用基于张量分解的特征图分解方法。



1. 一种由机器视觉编码设备执行的用于对特征图进行编码的编码方法,所述编码方法包括:

使用深度学习模型从输入图像提取所述特征图,其中,所述特征图从所述深度学习模型的中间层生成;

通过减小所述特征图的大小来生成减小的特征图;

通过转换所述减小的特征图的数据类型并重新布置所述减小的特征图来生成转换的特征图;以及

通过使用视频编码器编码所述转换的特征图而生成比特流。

2. 根据权利要求1所述的编码方法,其中,所述特征图包括:

与存在的通道同样多的具有相等高度和相等宽度的2D特征图。

3. 根据权利要求1所述的编码方法,其中,生成所述减小的特征图包括:

基于所述特征图的空间或通道方面的特征图稀疏化来减小所述特征图。

4. 根据权利要求3所述的编码方法,其中,生成所述减小的特征图包括:

用所述特征图稀疏化和组合的张量分解来减小所述特征图。

5. 根据权利要求3所述的编码方法,其中,生成所述减小的特征图包括:

当构成所述特征图的2D特征图具有特征值小于预设阈值的区域时,将所述区域的特征值设置为零。

6. 根据权利要求3所述的编码方法,其中,生成所述减小的特征图包括:

计算具有预设通道跨度的两个2D特征图之间的距离;

选择具有小于预设阈值的距离的所有2D特征图对;以及

对于所选的所有对中的每一者,将一个2D特征图的所有值稀疏化为零或删除所述一个2D特征图。

7. 根据权利要求6所述的编码方法,其中,生成所述比特流包括:

在所述一个2D特征图被稀疏化时对稀疏的2D特征图和所述预设通道跨度进行编码;或当所述一个2D特征图被删除时,对所删除的2D特征图的索引和所述预设通道跨度进行编码。

8. 根据权利要求1所述的编码方法,其中,生成所述减小的特征图包括:

基于张量分解来减小所述特征图。

9. 根据权利要求8所述的编码方法,其中,生成所述减小的特征图包括:

通过使用Tucker分解将所述特征图分解成一个核张量和三个因子矩阵。

10. 根据权利要求8所述的编码方法,其中,生成所述减小的特征图包括:

通过使用规范多元分解(CP)将所述特征图分解成P个秩1张量(其中,P是自然数)。

11. 一种机器视觉编码设备,包括:

特征提取器,被配置为使用深度学习模型从输入图像中提取特征图,其中,所述特征图从所述深度学习模型的中间层生成;

特征减小器,被配置为减小所述特征图的大小以生成减小的特征图;

预量化器,被配置为量化所述减小的特征图以转换所述减小的特征图的数据类型;

重封装器,被配置为以视频序列的形式重新布置量化的特征图以生成转换的特征图;以及

特征编码器,被配置为通过使用视频编码器对所述转换的特征图进行编码以生成比特流。

12. 根据权利要求11所述的机器视觉编码设备,其中,所述特征减小器被配置为基于所述特征图的空间或通道方面的特征图稀疏化来减小所述特征图。

13. 根据权利要求11所述的机器视觉编码设备,其中,所述特征减小器被配置成用于基于张量分解来减小所述特征图。

14. 一种由机器视觉解码设备执行的解码方法,所述解码方法包括:

使用视频解码器从比特流解码转换的特征图;

通过重新布置所述转换的特征图并且通过转换重新布置的所述转换的特征图的数据类型来重构减小的特征图;以及

通过扩展所述减小的特征图的大小来生成重构的特征图,

其中,所述重构的特征图对应于从所述机器视觉编码设备中深度学习模型的中间层生成的特征图。

15. 根据权利要求14所述的解码方法,其中,所述重构的特征图包括:

与存在的通道同样多的具有相等高度和相等宽度的2D特征图。

16. 根据权利要求14所述的解码方法,进一步包括:

当基于通道方面的特征图稀疏化来减小所述减小的特征图时,对预设通道跨度和稀疏的2D特征图进行解码,或者对所述预设通道跨度和删除的2D特征图的索引进行解码。

17. 根据权利要求16所述的解码方法,其中,生成所述重构的特征图包括:

对于所述稀疏的2D特征图,通过将所述预设通道跨度之前或之后的重构的2D特征图复制到所述稀疏的2D特征图的位置来生成所述重构的特征图,或者,对于所述删除的2D特征图,参考解码的索引将所述预设通道跨度之前或之后的重构的2D特征图复制到所述删除的2D特征图的位置来生成所述重构的特征图。

18. 根据权利要求14所述的解码方法,其中,生成所述重构的特征图包括:

当所述减小的特征图已通过使用Tucker分解而减小时,通过使用构成所述减小的特征图的核张量和因子矩阵来生成所述重构的特征图。

19. 根据权利要求14所述的解码方法,其中,生成所述重构的特征图包括:

当所述减小的特征图通过使用规范多元分解(CP)已经被减小时,通过使用构成所述减小的特征图的P个秩1张量生成所述重构的特征图(其中P是自然数)。

使用特征图减小来编解码机器视觉数据的方法和设备

技术领域

[0001] 本公开涉及一种用于使用特征图减小来编解码机器视觉数据的设备和方法。

背景技术

[0002] 该部分中的陈述仅提供与本公开相关的背景信息,并且不一定构成现有技术。

[0003] 随着深度学习技术和计算能力的进步,伴随着机器视觉应用的激增,机器逐渐实现为利用大部分视频业务。机器对机器应用预期在未来占据互联网视频流量的最大部分。因此,优化机器所使用的视频数据中的信息的方式可能是视频处理技术的创新和新解决方案的商业化的关键因素。

[0004] 现有视频编解码方案针对人类视觉进行优化,因为现有视频编解码方案针对某些比特率约束下的最佳图片或视频质量。相对照地,对于机器视觉的编解码,重构的图像/视频不一定必须保持高视觉性能。在延时和规模上具有严格限制的技术领域(包括连接的车辆、物联网(IoT)装置、超大型视频监控网络、智能城市、质量检查等)的出现引领了用于机器视觉的新范例,从而严格地需要针对机器视觉的新图像/视频编解码方法。

[0005] 因而,运动图像专家组(MPEG)——标准化组织讨论了机器视觉的标准化的需要,导致提出作为下一代视频编解码设备来提供机器视觉数据的压缩编解码和人机混合视觉的压缩编码的机器视频编解码(VCM)。

[0006] 虽然VCM编解码设备的结构可能存在许多变化,但是在图1中示出了VCM编解码设备的基本结构。当接收到作为传感器的输出的视频时,VCM编码器提取特征作为用于机器视觉的信息,根据需要转换特征,然后执行特征编码。此外,VCM编码器可以参考在编码输入图像或视频时的编码特征。最后,VCM编码器对用于机器视觉的特征和输入图像(或残差图像)进行编码以生成比特流。VCM编码器复用每个通过编码特征和视频而生成的比特流,并将比特流一起发送。

[0007] VCM解码器将传输的比特流解复用为特征比特流和视频比特流,然后分别对特征和视频进行解码。当解码视频时,VCM解码器可以参考重构的特征。在逆转换之后,重构的特征可同时用于机器视觉和人类视觉。

[0008] 此外,在图1的示例中,用于神经网络(NN)的接口可以用于在VCM编码器处应用深度学习模型以提取特征,并且在VCM解码器处应用深度学习模型以执行用于机器视觉的任务等。

[0009] 具体地,机器视觉的代表类型的信息是由机器任务专用深度学习模型提取的特征图。如果VCM编码器发送特征图而不是发送图像/视频,则包括VCM解码器的云服务器上的计算负荷可以显著减小,这具有个人信息保护的附加益处。然而,特征图(多通道的堆叠阵列)固有地大于正常图像/视频。此外,深度学习模型的特定层的输出特征图的大小可以小于输入图像/视频。然而,这一益处被相应的层随着深度学习模型改变而改变的问题所抵消。

[0010] 如果特征图过于大于输入图像/视频,那么计算负荷减小、隐私保护等前述优点可能会衰减。因此,需要设计一种通过反映所提取的特征图的特性来减小所提取的特征图的

大小的先进方法。

发明内容

[0011] 【技术问题】

[0012] 本公开试图提供一种用于机器视频编解码 (VCM) 编解码设备和方法,其用于利用减小特征图的空间和通道方面的冗余的稀疏化方法,并且用于利用基于张量分解的特征图分解方法,以减小由机器任务专用深度学习模型提取的特征图的大小。

[0013] 【技术方案】

[0014] 本公开的至少一个方面提供了一种由机器视觉编码设备执行的用于对特征图进行编码的编码方法。所述编码方法包括使用深度学习模型从输入图像中提取特征图。从深度学习模型的中间层生成特征图。编码方法还包括通过减小特征图的大小来生成减小的特征图。编码方法还包括通过转换减小的特征图的数据类型并重新布置减小的特征图来生成转换的特征图。编码方法还包括通过使用视频编码器编码转换的特征图来生成比特流。

[0015] 本公开的另一方面提供了一种机器视觉编码设备。该设备包括被配置为使用深度学习模型从输入图像提取特征图的特征提取器。从深度学习模型的中间层生成特征图。设备还包括特征减小器,被配置为减小特征图的大小以生成减小的特征图。设备还包括预量化器,被配置为量化减小的特征图以转换减小的特征图的数据类型。设备还包括重封装器,被配置为以视频序列的形式重新布置量化的特征图以生成转换的特征图。设备还包括特征编码器,被配置为通过使用视频编码器对转换的特征图进行编码以生成比特流。

[0016] 本公开的另一方面提供了一种由机器视觉解码设备执行的解码方法。所述解码方法包括使用视频解码器从比特流解码转换的特征图。解码方法还包括通过重新布置转换的特征图且通过转换重新布置的转换的特征图的数据类型来重构减小的特征图。解码方法还包括通过扩展减小的特征图的大小来生成重构的特征图。重构的特征图对应于从机器视觉编码设备中的深度学习模型的中间层生成的特征图。

[0017] 【有益效果】

[0018] 如上所述,本公开提供VCM编解码设备和方法,其对于由专门用于机器任务的深度学习模型提取的特征图利用减小特征图的空间和通道方面的冗余度的稀疏化方法。VCM编解码设备和方法还利用基于张量分解的特征图分解方法。由此,特征图的大小可减小。

[0019] 此外,本公开提供一种VCM编解码设备和方法,对于由专门用于机器任务的深度学习模型提取的特征图,VCM编解码设备和方法利用减小特征图的空间和通道方面的冗余的稀疏方法,作为特征转换的整体减小步骤。VCM编解码设备和方法还利用基于张量分解的特征图分解方法。由此,可独立于在特征转换之前和之后的特征提取步骤和特征编码步骤来使用特征图缩减。

附图说明

[0020] 图1是概念性地示出用于机器视频编解码 (VCM) 编解码设备的框图。

[0021] 图2是示出了根据本公开的至少一个实施方式的深度学习模型的逐层特征图的示图。

[0022] 图3是概念性地示出前端和云服务器的操作的示图。

[0023] 图4是概念性地示出根据本公开的至少一个实施方式的执行中间深度特征压缩的VCM编解码设备的框图。

[0024] 图5是概念性地示出根据本公开的至少一个实施方式的利用中间深度特征图减小的VCM编解码设备的框图。

[0025] 图6是概念性地示出根据本公开的至少一个实施方式的利用中间深度特征图重构的VCM解码设备的框图。

[0026] 图7是示出根据本公开的至少一个实施方式的基于特征图稀疏化的特征图减小方法的示意图。

[0027] 图8是示出根据本公开另一实施方式的基于特征图稀疏化的特征图减小方法的示意图。

[0028] 图9是示出根据本公开的至少一个实施方式的使用Tucker分解 (Tucker decomposition, 塔克分解) 的特征图减小方法的示意图。

[0029] 图10是示出根据本公开的至少一个实施方式的利用规范多元 (CP) 分解的特征图减小方法的示意图。

[0030] 图11是根据本公开的至少一个实施方式的利用中间深度特征图减小的VCM编码方法的流程图。

[0031] 图12是根据本公开的至少一个实施方式的利用中间特征图重构的VCM解码方法的流程图。

具体实施方式

[0032] 在下文中,参照附图详细描述本公开的一些实施方式。在以下描述中,相同的参考标号表示相同的元件,尽管元件在不同的附图中示出。此外,在一些实施方式的以下描述中,为了清楚和简洁的目的,已经省略了被认为会使本公开的主题模糊的已知组件和功能的详细描述。

[0033] 本实施方式公开了一种使用特征图的减小对机器视觉数据进行编解码的设备和方法。更具体地,为了减小通过机器任务专用深度学习模型提取的特征图的大小,提供了用于视频编解码机器 (VCM) 编解码设备和方法。VCM编解码设备和方法利用减小特征图的空间和通道方面的冗余的稀疏方法,并且利用基于张量分解的特征图分解方法。

[0034] 这里,VCM编解码设备包括VCM编码设备和VCM解码设备。

[0035] 同时,编码和解码针对人类视觉而优化的视频信号的传统编解码器由视频编码器和视频解码器表示。

[0036] I. 中间特征图编解码 (中间深度特征编解码)

[0037] 从具有分层结构的深度学习模型的每一层输出的多维阵列通常被称为特征、特征图、深度特征或深度学习特征图。如图2的示例中所示,每一层中的特征图采用空间相关的二维阵列的堆叠形成。在这种情况下,二维阵列的高度H和宽度W被称为对应特征图的高度和宽度,并且堆叠的二维阵列的数量被称为相关特征图的通道数量C。在已知适于图像处理的卷积神经网络 (CNN) 结构中,层越深,由于池化,特征图的高度和宽度通常变得越小,并且通道的大小越大。直观地,单通道二维特征图可以被视为单个帧,并且堆叠的特征图可以被视为连续的视频序列。

[0038] 随着网络基础设施的发展,基于云的应用近来已经出现。具体地,前端装置从用户或物理世界获得信息,然后将获得的信息传输到云服务器(即,数据中心),云服务器然后使用获得的信息进行进一步的处理和分析。如图3所示,部署在物理世界中的前端装置(诸如安全监视摄像机、可穿戴设备等)可获取发送至云服务器以用于视觉分析的大量数据。许多基于深度学习的视觉模型可应用于该基于云的范例,例如,在以下中:i) 用于行人检测、人/车辆识别、和牌照识别的自动车辆;ii) 用于面部识别、地标检测、和对象检测的便携式设备(例如,智能电话、智能眼镜、和其他软件应用);以及iii) 用于监视图像/视频检索和人/车辆识别的监视摄像机。

[0039] 在这种基于云的范例中,前端和云服务器之间的视频的压缩和传输可以依赖于传统的“压缩-然后-分析”方法。前端装置捕捉和压缩视觉数据,前端装置还将经编码的比特流发送到云服务器。云服务器对比特流进行解码并进行形状提取和视觉分析。然而,在例如视频监控和IoT等情境中,大量前端装置可同时生成数百万个比特流。因此,在前端侧,视频信号级的视觉压缩可能是大的传输负担,并且通常是不可控的规模。在云服务器侧,同时运行以用于特征提取的许多深度学习模型的计算负荷可能是重大挑战。

[0040] 另一方面,在“分析-然后-压缩”方法中,其中在前端提取的特征被压缩并发送至云服务器,因为所提取的特征用于任务分析而不是视觉信号,“分析-然后-压缩”方法可以是上述的“压缩-然后-分析”方法的可行的替代。对于基于深度学习的特征,通常将深度学习模型的顶层(即,最深层)发送到云,因为最深层比下层(即,浅层)更紧凑,并且可以立即用于分析。在这种情况下,在云服务器上执行轻量级任务(诸如特征至特征比较),而将重型任务(诸如特征提取)分配到前端。这种特征转移对于隐私保护也是有益的。与可以容易地揭示个人信息的视觉信号的传输相比,特征传输可以避免公开可见信息。

[0041] 另一方面,由于针对特定任务设计和训练深度学习模型,并且顶层特征是非常抽象且任务特定的,因此压缩顶层特征可能难以泛化。即,标准化下的经压缩特征可能难以应用于其他场景,这可能成为用深度学习模型来制定特征压缩的标准的障碍。

[0042] 考虑以上问题,存在中间深度特征压缩方法以减小云服务器上的计算负荷,同时保持特征的可用性。这种方法被认为是如上所述的“分析-然后-压缩”和“压缩-然后-分析”的极端之间的折衷,并且可以在计算负荷、通信成本和泛化能力(generalization ability)之间提供良好的平衡。

[0043] 由于深度学习模型通常具有分层结构,因此深度学习模型可以被认为是堆叠特征提取器的组合,而不是单个特征提取器。顶层中的特征基于大感受野(receptive field)是抽象的和任务特定的,而底层中的特征具有小感受野并在2D特征图中编码大量位置信息,这可适合于更宽的分析。由此,当从中间层而不是最终层压缩和发送特征时,云服务器具有取决于在前端处执行的任务而请求适当特征的灵活性。

[0044] 在VCM标准的流程(pipeline)中,中间深度特征编解码非常重要,VCM标准可在图1中所示的VCM编解码设备的结构中依次应用于特征转换、特征编码和特征解码。在下文中,将参考图4描述中间深度特征编解码。在图4中,未示出执行输入图像的视频编码和视频解码的部件,因为这些部件与本实施方式不特别相关。

[0045] 在以下描述中,特征图指的是作为深度学习模型的中间层的输出的中间深度特征图。

[0046] 图4是概念性地示出根据本公开的至少一个实施方式的执行中间深度特征压缩的VCM编解码设备的框图。

[0047] VCM编码设备包括特征提取器410、特征转换器420和特征编码器430的全部或部分。此外,VCM解码设备包括特征解码器450或特征逆转换器460的全部或部分。

[0048] 首先,特征编码器430和特征解码器450可使用传统的视频编解码设备来实现。

[0049] VCM编码设备中的特征提取器410基于深度学习模型从输入视频提取用于机器视觉的任务的特征图。此外,特征提取器410中的深度学习模型可以执行机器视觉任务,诸如生成机器视觉任务的分析结果。

[0050] 特征提取器410中的深度学习模型可以是能够生成适合于在机器视觉任务中分析的特征图的任何神经网络模型。例如,深度学习模型可以是利用多个编码层实现的基于CNN的模型。这里,一个编码层可包括卷积层和池化层。

[0051] 特征转换器420包括预量化器422和重封装器 (repacker) 424以转换用于特征图的数据类型并重新布置特征图通道等。

[0052] 由深度学习模型提取的特征图可具有与常规视频编码设备的输入不兼容的数字类型。因此,为了确保与特征编码器430的输入兼容性,预量化器422首先执行预量化。例如,典型的深度学习模型的特征图可以是float32类型,而传统的视频编码设备利用8或更多位的整数输入。此外,取决于中间深度特征图的分布分析,可以使用不同的预量化方法。

[0053] 重封装器424将预量化的N特征图 $R^{N \times H \times W \times C}$ (其中,N是自然数) 重新封装成视频序列兼容格式 $R^{H' \times W' \times NC}$,以适合视频编码设备的输入格式。这里,H和W是2D特征图的高度和宽度,并且C是通道的数量。 $H' \times W'$ 是从通过填充扩展的原始特征图的大小 $H \times W$ 获得的,以适合视频编码设备的输入帧大小。 R 表示实数。此外,由于 $H \times W \times C$ 是由深度学习模型从输入视频的一帧中提取的特征图,因此 $N \times H \times W \times C$ 等于从N帧中提取的特征图。

[0054] 在以下描述中,特征图由 $F \in R^{H \times W \times C}$ 表示,并且构成特征图的2D特征图由 $f \in R^{H \times W}$ 表示。

[0055] 在另一实施方式中,重封装器424可将预量化的N特征图 $R^{N \times H \times W \times C}$ (其中N是自然数) 重新封装到一个帧中。例如,在特征编码器430是传统的视频编解码设备的情况下,如上所述,通道特定特征图可以被重新封装成单个超帧以仅使用帧内预测模式而不使用帧间预测模式。对于这样的重新封装的超帧,可以仅使用视频编解码设备的帧内预测来实现更好的压缩效率。在下文中,该重新封装处理被称为帧封装。

[0056] 特征编码器430对转换的特征图进行编码,从而生成比特流。

[0057] VCM解码设备中的特征解码器450从比特流解码转换的特征图。

[0058] 特征逆转换器460包括去重封装器462和后去量化器464,以将特征图逆转换应用于比特流,这是VCM编码设备的特征图转换的逆过程。

[0059] 去重封装器462将已经重新封装的特征图转换为视频序列兼容格式,回到原始特征图大小。

[0060] 后去量化器464将具有整数值的特征图去量化为浮点类型。如上所述,这些重构的特征图可以用于分析机器视觉任务。

[0061] 存在这样的担忧,中间深度特征图在压缩和传输而不是视觉信号时会大于视频/图像流。然而,已知的是,深度学习模型的更高层中的特征通常具有比更低层中的特征更小

的压缩大小,并且具体地,离顶层的距离越近,对应特征的压缩比率越优于原始视觉信息。换句话说,一些种类的特征可使用比图像/视频比特流少的带宽。因此,在这样的情况下,传输中间深度特征图在传输带宽效率方面是有意义的。

[0062] 另一方面,即使中间深度特征图不能被压缩到小于图像/视频比特流的大小,传输特征图仍然是很重要的。在许多应用中,计算资源比带宽更昂贵,所以使用中间深度特征图传输来将计算负荷从云服务器分散到前端装置可能更重要。

[0063] II. 中间深度特征图减小(中间深度特征减小)

[0064] 以下参照图5和图6描述在特征转换步骤中减小中间深度特征图的方法和在特征逆转换步骤中重构中间深度特征图的方法。

[0065] 图5是概念性地示出根据本公开的至少一个实施方式的利用中间深度特征图减小的VCM编码设备的框图。

[0066] 除了特征减小器512(其对于图5的图示是新的),剩余的部件与图4中示出的VCM编码设备的部件相同。如上所述,由于特征转换步骤是基于利用现有的视频编解码设备进行特征编码来预测的,因此在对特征图进行预量化之前,特征减小器512还可以执行各种特征图减小方法来减小原始特征图的大小,以提高特征图的编码效率。

[0067] 对于特征图减小,特征减小器512可以使用两种方法,诸如基于特征图稀疏化的特征图减小和基于张量分解的特征图减小。通过单独或同时使用这两种方法,特征减小器512可以最大化特征图减小。此外,当两种方法一起使用时,对使用顺序没有限制。根据本实施方式的用于特征图减小的方法还可以与单独的预处理方法(诸如如上所述的特征图的帧封装)结合使用。

[0068] 图6是概念性地示出根据本公开的至少一个实施方式的利用中间深度特征图重构的VCM解码设备的框图。

[0069] 除了特征重构器612(其对应图6的图示是新的),其余部件与图4中所示的VCM解码设备的部件相同。如上所述,由于特征逆转换步骤是基于利用用于特征解码的现有视频编解码设备来预测的,所以与特征减小器512相关的特征重构器612从后去量化特征图(即,减小的特征图)重构原始特征图。这里,特定特征图重构方法可以依赖于如由VCM编码设备使用的这种特征图减小方法。

[0070] 以下描述由特征减小器512和特征重构器612执行的特征图减小方法。虽然图5和图6说明用于特征编码/解码的常规视频编解码设备的使用,但本文描述的特征图减小方法不必限于使用所述方法。例如,根据本实施方式的特征图减小方法可用于减小所提取的特征图的大小,而不管用于特征提取的深度学习模型或编解码方法如何。

[0071] 首先,在特征图减小方法中,基于特征图稀疏化的特征图减小利用在特征图中存在的空间或通道方面的稀疏性。

[0072] 当深度学习模型生成中间深度特征图时,取决于输入图像/帧的空间特性或被训练的滤波器的特性,一些区域和一些通道可以被高度激活。另一方面,低激活区域,即使其被输入到下一层,也引起连续的低激活,并且最终对深度学习模型的最后一层的判断具有小的影响。因此,即使将输出这种小的激活的区域的值归零稀疏化,特征图自身的大小也能够减小,而不会显著地劣化深度学习模型的判断性能。换言之,特征图的编解码效率可以基于特征图稀疏化而提高,该特征图稀疏化对被激活的低于某个阈值的特征图区域进行稀疏

化。

[0073] 下面使用图7的示例描述由特征减小器512执行的用于基于特征图在空间方面的稀疏度减小特征图的方法。

[0074] 图7是示出根据本公开的至少一个实施方式的基于特征图稀疏化的特征图减小方法的示意图。

[0075] 特征减小器512在构成中间深度特征图 $F \in \mathbb{R}^{H \times W \times C}$ 的单个2D特征图 $f \in \mathbb{R}^{H \times W}$ 内基于输出低激活的区域的激活程度而对所述区域进行稀疏化。例如,当基于特征图的特定通道的2D特征图 f 的区域(例如,像素)的特征值而数字表示所述区域时,特征减小器512将区域的低于特定阈值的特征值设置为完全为零。在图7的示例中,如果阈值设置为1.5,标记为1的区域可以被设置为0。通过设定该适当的阈值,通过将低于该阈值的区域的特征值设定为零,不会显著影响深度学习模型的性能。特征减小器512可以在用于共计为所有特征图通道的C2D映射 $\{f_1, f_2, \dots, f_C\}$ 的空间方面或用于一些2D映射 $\{f_1, f_2, \dots, f_{C'}\}$ ($C' < C$)的空间方面执行该特征图稀疏化。

[0076] 下面使用图8的示例描述特征减小器512如何基于其在通道方面的稀疏度减小特征图。

[0077] 图8是示出了根据本公开另一实施方式的基于特征图稀疏化的特征图减小方法的示意图。

[0078] 特征减小器512可以通过某个通道跨度S响应于两个相邻通道的2D特征图之间的相似性,即,两个特征图的距离小于某个阈值,用于通过将一个2D特征图设置为零或完全删除一个通道来减小两个2D特征图之间的冗余。因此,可以减小原始特征图中的通道数目,并且因此可以减小整个特征图的大小。

[0079] 对于原始特征图,特征减小器512计算具有特定通道跨度S的两个2D特征图之间的距离 $\|f_{i+S} - f_i\|$ ($1 \leq i \leq C-S$)。该距离可以是范数,诸如L1、L2等。例如,在图8的示例中,通道跨度S被设置为1。特征减小器512将具有小于某一阈值d的距离的两个2D特征图视为非常类似的特征图。在选择具有相似2D特征图的所有通道对之后,对于每个对,特征减小器512可以将两个2D特征图的前(或后)通道的2D特征图的所有值稀疏化为零。可选地,如图8所示,特征减小器512可以为每对完全删除一个通道。特征减小器512可以仅使用非相似通道的2D特征图以生成具有减小的通道总数的中间深度特征图 F' 。

[0080] VCM编码设备检查是否存在稀疏化通道。如果是,则VCM编码设备解码每个稀疏化通道的通道跨度S值和2D特征图,然后将解码的每个稀疏化通道的通道跨度S值和解码的2D特征图发送给VCM解码设备。由于每个稀疏化通道的2D特征图的所有值为零,VCM编码设备可以容易地对每个稀疏化通道的2D特征图进行编码,从而可以大大提高编解码效率。可选地,VCM编码设备可以检查是否存在删除的通道。如果是,则VCM编码设备可对通道跨度S值和每个删除通道的索引进行编码,然后可将编码的通道跨度S值和每个删除通道的编码的索引发送到VCM解码设备。VCM编码设备可仅编码对应于索引的整数值,而不是编码用于删除的通道的大小 $H \times W$ 的2D特征图,这可极大地提高编码效率。

[0081] 同时,对于稀疏化或删除的通道,VCM解码设备中的特征重构器612可参考重构的减小特征图、重构的通道跨度S值或删除的通道索引信息。对于所有值被重构为零的稀疏化通道,VCM解码设备通过将跨度S之前(或之后)的通道的2D特征图复制到稀疏化通道的位置

来重构减小之前的原始特征图。可选地,对于删除的通道,VCM解码设备参考删除的通道索引信息,并且因此将跨度S之前(或之后)的通道的2D特征图复制到删除的通道的位置,以重构减小之前的原始特征图。

[0082] 所得的重构特征图是具有与原始特征图相同大小的阵列,并且VCM解码设备然后可以使用重构特征图用于在机器视觉任务中的分析。

[0083] 在另一个实施方式中,VCM编码设备可以将包括在表示每个通道的权重的 $1 \times C$ 矢量中的元素值 c_i 乘以对应的特征图 f_i ,然后可以压缩和传输特征图。当解码特征图时,VCM解码设备可以通过将接收到的特征图除以 c_i 来生成重构的特征图。 c_i 可以基于深度学习从关注模型获得或基于深度学习从压缩和激励(squeeze and excitation, SE)模型的SE模块获得。

[0084] 在下文中,描述使用张量分解的特征图减小方法的示例。

[0085] 如上所述的多维阵列形式的特征图可通过使用张量分解被分解成多个低维简单张量。在这种情况下,分解的张量的总大小可从原始特征图的大小减小。

[0086] 另一方面,特征图减小不限于特定张量分解方法,并且除了以下描述的示例之外,可以应用不同张量分解方法,诸如矩阵分解(例如,奇异值分解SVD、CUR因式分解、QR因式分解等)、高阶SVD(HOSVD)。在应用张量分解方法之后,VCM编码设备可利用分解的低维张量用于代替原始特征图的特征编码。

[0087] 作为张量分解的示例,图9描述了由特征减小器512执行的用于基于Tucker分解减小特征图的方法。

[0088] 图9是示出根据本公开的至少一个实施方式的使用Tucker分解的特征图减小方法的示意图。

[0089] 如图9中所示,特征减小器512可以通过使用Tucker分解将单个图像的中间深度特征图 $F \in R^{n_1 \times n_2 \times n_3}$ 分解成一个核张量 $K \in R^{r_1 \times r_2 \times r_3}$ 和三个因子矩阵 $F^{(i)} \in R^{r_i \times n_i}$ ($i \in \{1, 2, 3\}$)。这里, r_i 由 i -秩表示,并且当特征图被生成为 $F^* \in R^{n_i \times n_1 \dots n_{i-1} n_{i+1} \dots n_3}$ 整形矩阵时, r_i 表示该矩阵的秩。利用一个核张量和三个因子矩阵,特征图 F 可通过对张量的操作来近似,如等式1所示。

[0090] [等式1]

$$[0091] \quad F \approx K \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$$

[0092] 这里,运算符 $\times_1$ 、 $\times_2$ 和 $\times_3$ 可被表示为等式2。

[0093] [等式2]

$$[0094] \quad (K \times_1 F^{(1)})(n_1, r_2, r_3) = \sum_{i_1=1}^{r_1} K(i_1, r_2, r_3) F^{(1)}(i_1, n_1)$$

$$[0095] \quad (K \times_2 F^{(2)})(r_1, n_2, r_3) = \sum_{i_2=1}^{r_2} K(r_1, i_2, r_3) F^{(2)}(i_2, n_2)$$

$$[0096] \quad (K \times_3 F^{(3)})(r_1, r_2, n_3) = \sum_{i_3=1}^{r_3} K(r_1, r_2, i_3) F^{(3)}(i_3, n_3)$$

[0097] 在Tucker分解之后,原始特征图的数据大小 $n_1 n_2 n_3$ 减小到“ $r_1 r_2 r_3 + r_1 n_1 + r_2 n_2 + r_3 n_3$ ”。

[0098] VCM编码设备可使用视频编码器来编码每个分解的张量 K 、 $F^{(1)}$ 、 $F^{(2)}$ 和 $F^{(3)}$,然后将生成的比特流复用并发送到VCM解码设备。

[0099] 另一方面,VCM解码设备使用视频解码器来解码分解的张量的比特流,并且然后将解码的张量应用于去重封装器462和后去量化器464,如图4中所示。VCM解码设备可根据等式1使用去量化的分解张量来重构原始特征图。

[0100] 作为张量分解的另一个实施方式,图10的示例描述了特征减小器512如何基于规范多元(Canonical Polyadic, CP)分解来减小特征图。

[0101] 图10是示出根据本公开的至少一个实施方式的利用CP分解的特征图减小方法的示意图。

[0102] 如图10所示,特征减小器512可以使用CP分解将中间深度特征图 $F \in R^{n_1 \times n_2 \times n_3}$ 分解成秩1张量的线性组合。例如,如等式3所示,可基于P个秩1张量(其中,P是自然数)近似特征图F。

[0103] [等式3]

$$[0104] \quad F \approx \sum_{p=1}^P a^p \circ b^p \circ c^p$$

[0105] 这里,运算符 $\circ$ 表示外积(outer product)。

[0106] VCM编码设备通过级联相同大小的秩1的张量生成张量A、B和C,如等式4所示,然后编码张量A、B和C。

[0107] [等式4]

$$[0108] \quad A = [a_1, a_2, \dots, a_P] \in R^{n_1 \times P}$$

$$[0109] \quad B = [b_1, b_2, \dots, b_P] \in R^{n_2 \times P}$$

$$[0110] \quad C = [c_1, c_2, \dots, c_P] \in R^{n_3 \times P}$$

[0111] 基于CP分解,将原始特征图的数据大小 $n_1 n_2 n_3$ 减小到“ $(n_1 + n_2 + n_3) P$ ”。

[0112] 在另一个实施方式中,VCM编码设备可编码P个分解的秩1张量。

[0113] VCM编码设备将通过分解的张量进行编码而获得的比特流发送到VCM解码设备。

[0114] 另一方面,VCM解码设备使用视频解码器来解码分解的张量的比特流,并且将解码的张量应用于去重封装器462和后去量化器464,如图4中所示。VCM解码设备可根据等式3使用去量化的分解张量来重构原始特征图。

[0115] 在另一个实施方式中,VCM编码设备可编码仅组合P' (<P) 个秩1张量的张量,并且可将编码的张量传输至VCM解码设备。VCM解码设备可通过使用解码的P' 个秩1张量重构原始特征图。

[0116] 现在参考图11和图12,描述了利用中间深度特征图减小的VCM编码方法和利用中

间深度特征图重构的VCM解码方法。

[0117] 图11是根据本公开内容的至少一个实施方式的利用中间深度特征图减小的VCM编码方法的流程图。

[0118] VCM编码设备使用深度学习模型从输入图像提取特征图(S1100)。这里,特征图表示从深度学习模型的中间层生成的中间深度特征图。特征图包括与存在的通道同样多的具有相同高度和宽度的2D特征图。

[0119] VCM编码设备减小特征图的大小以生成减小的特征图(S1102)。

[0120] VCM编码设备基于特征图的空间或通道方面的特征图稀疏化来减小特征图的大小。可替代地,VCM编码设备可以基于张量分解来减小特征图。通过单独或同时使用这两种方法,VCM编码设备可以最大化特征图减小。此外,当两个方法一起使用时,对它们使用的顺序没有限制。

[0121] 为了利用特征图的空间方面的稀疏性,VCM编码设备将构成特征图的2D特征图内的区域的特征值与预设阈值进行比较。如果特征值小于预设阈值,则可以将相应区域的特征值设置为零。

[0122] 在另一实施方式中,为了利用特征图的通道方面的稀疏性,VCM编码设备计算具有预设通道跨度的两个2D特征图之间的距离。在选择具有小于预设阈值的计算距离的所有2D特征图对之后,VCM编码设备对于所选择的对中的每一对,将一个2D特征图的所有值稀疏化为零或者删除一个2D特征图。

[0123] 同时,使用其他张量分解方法中的Tucker分解,VCM编码设备可以将特征图分解成一个核张量和三个因子矩阵。在这种情况下,特征图可以由一个核张量和三个因子矩阵来近似,如等式1所示。

[0124] 在另一实施方式中,使用张量分解方法中的CP分解,VCM编码设备可将特征图分解成P个秩1张量。在这种情况下,如等式3所示,可基于P个秩1张量近似特征图。

[0125] VCM编码设备量化减小的特征图以转换减小的特征图的数据类型(S1104)。

[0126] VCM编码设备以视频序列的形式重新布置量化的特征图以生成转换的特征图(S1106)。

[0127] VCM编码设备使用视频编码器对转换的特征图进行编码以生成比特流(S1108)。

[0128] 同时,当使用特征图的通道方面的稀疏度时,VCM编码设备如下生成比特流。对于稀疏的特征图,VCM编码设备可对2D特征图和预设通道跨度进行编码,并可将编码的2D特征图和编码的预设通道跨度发送到VCM解码设备。可选地,对于删除的特征图,VCM编码设备可以对删除的2D特征图的索引和预设通道跨度进行编码,并且可以将删除的2D特征图的编码的索引和编码的预设通道跨度发送到VCM解码设备。

[0129] 当使用Tucker分解时,VCM编码设备可对核张量和因子矩阵中的每一个进行编码以生成相应的比特流,然后可将编码的核张量和编码的因子矩阵发送到VCM解码设备。

[0130] 当使用CP分解时,VCM编码设备可编码P个秩1张量中的每一个以生成相应的比特流,然后可将它们传输至VCM解码设备。

[0131] 在利用CP分解的另一个实施方式中,VCM编码设备可生成组合相同大小的秩1张量的张量。VCM编码设备可编码这些张量以生成对应的比特流,然后可将编码的张量传输至VCM解码设备。

[0132] 图12是根据本公开内容的至少一个实施方式的利用中间深度特征图重构的VCM解码方法的流程图。

[0133] 这里,中间深度特征图表示从机器视觉编码设备中的深度学习模型的中间层生成的特征图。特征图包括与通道的数目同样多的具有相同高度和宽度的2D特征图。

[0134] VCM解码设备使用视频解码器从比特流解码转换的特征图(S1200)。

[0135] 另一方面,当基于通道方面的特征图稀疏化减小减小的特征图时,VCM解码设备如下解码比特流。对于稀疏的2D特征图,VCM解码设备对预设通道跨度和稀疏的2D特征图进行解码。可选地,对于删除的2D特征图,VCM解码设备对预设通道跨度和删除的2D特征图的索引进行解码。

[0136] VCM解码设备以视频序列的形式重新布置转换的特征图(S1202)。

[0137] VCM解码设备去量化重新布置的转换的特征图以转换重新布置的转换的特征图的数据类型,并且从而重构减小的特征图(S1204)。

[0138] VCM解码设备扩展减小的特征图的大小以生成重构的特征图(S1206)。这里,重构的特征图对应于上文描述的中间深度特征图。

[0139] 当基于通道方面的特征图稀疏化减小减小的特征图时,VCM解码设备如下生成重构的特征图。对于稀疏的2D特征图,VCM解码设备可以通过将预设通道跨度之前或之后的重构的2D特征图复制到稀疏的2D特征图的位置来生成重构的特征图。可选地,对于删除的2D特征图,VCM解码设备可以通过参考解码的索引将预设通道跨度之前或之后的重构的2D特征图复制到删除的2D特征图的位置,来生成重构的特征图。

[0140] 在另一实施方式中,当通过使用Tucker分解来减小减小的特征图时,VCM解码设备可以通过使用构成减小的特征图的核张量和因子矩阵来近似重构的特征图,如等式1所示。

[0141] 在又一实施方式中,当减小的特征图通过使用CP分解减小时,VCM解码设备可通过使用组成减小的特征图的P个秩1张量来近似重构的特征图,如等式3所示。

[0142] 尽管各个流程图中的步骤被描述为被顺序地执行,但是这些步骤仅仅例示本公开的一些实施方式的技术构思。因此,本公开所属领域的普通技术人员可通过改变在各个附图中描述的顺序或者通过并行执行两个以上的步骤来执行步骤。所以,各个流程图中的步骤不限于所示的时间序列顺序。

[0143] 应当理解,以上描述呈现了可以以不同其他方式实现的说明性实施方式。在一些实施方式中描述的功能可以通过硬件、软件、固件和/或其组合来实现。还应当理解,本说明书中描述的功能部件用“…单元”标记以强烈强调它们独立实现的可能性。

[0144] 同时,在一些实施方式中描述的各种方法或功能可以被实现为存储在能够由一个或多个处理器读取和执行的非暂时性记录介质中的指令。例如,非暂时性记录介质可包括各种类型的记录装置,其中,数据以计算机系统可读的形式存储。例如,非暂时性记录介质可包括存储介质,诸如可擦除可编程只读存储器(EPROM)、闪存驱动器、光盘驱动器、磁性硬盘驱动器和固态驱动器(SSD)等。

[0145] 尽管出于说明性的目的已经描述了本公开的实施方式,然而,本公开所属领域的普通技术人员应当认识到,在不背离本公开的构思和范围的情况下,各种修改、添加、以及替换是可能的。因此,为了简洁和清晰起见,已经描述了本公开的实施方式。本公开的实施方式的技术构思的范围不受图示的限制。因此,本公开所属领域的普通技术人员应当理解,

本公开的范围不限于以上明确描述的实施方式,而是由权利要求及其等同物限制。

[0146] (参考标号)

[0147] 410 特征提取器

[0148] 420 特征转换器

[0149] 430 特征编码器

[0150] 450 特征解码器

[0151] 460 特征逆转换器

[0152] 512 特征减小器

[0153] 612 特征重构器。

[0154] 相关申请

[0155] 本申请要求于2021年3月31日提交的韩国专利申请号10-2021-0041562和2022年3月25日提交的韩国专利申请号10-2022-0037097的优先权,通过引用将其每一个的全部公开内容结合于此。

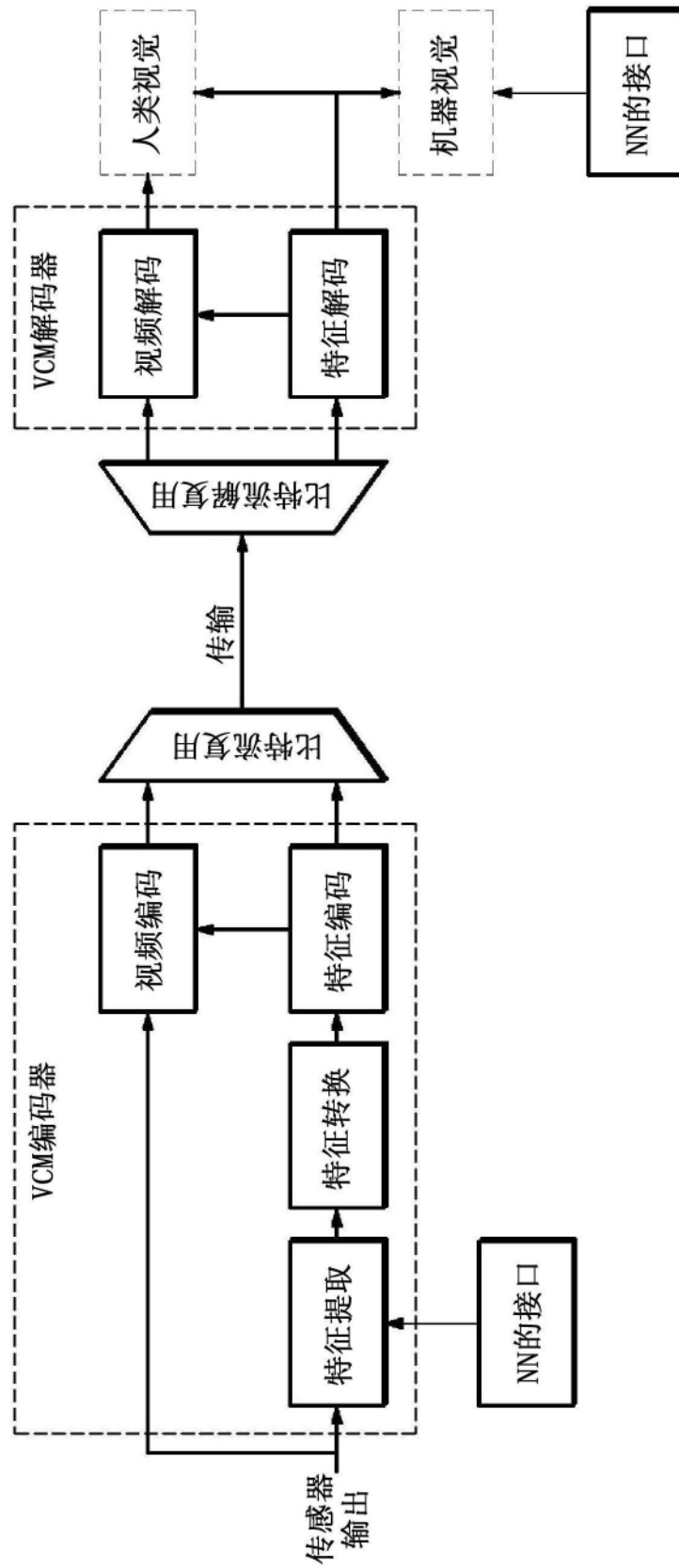


图1

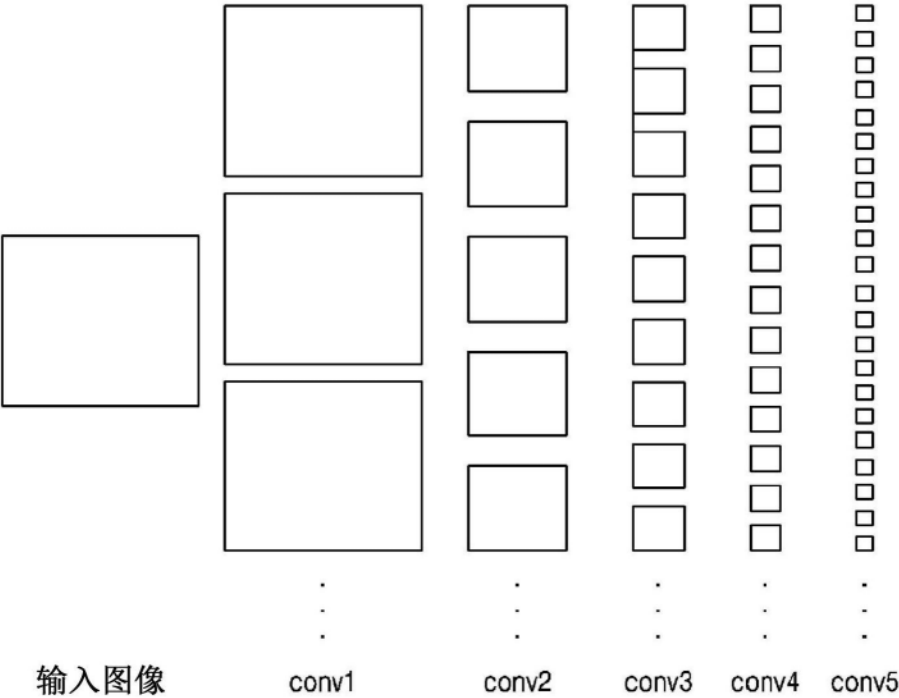


图2

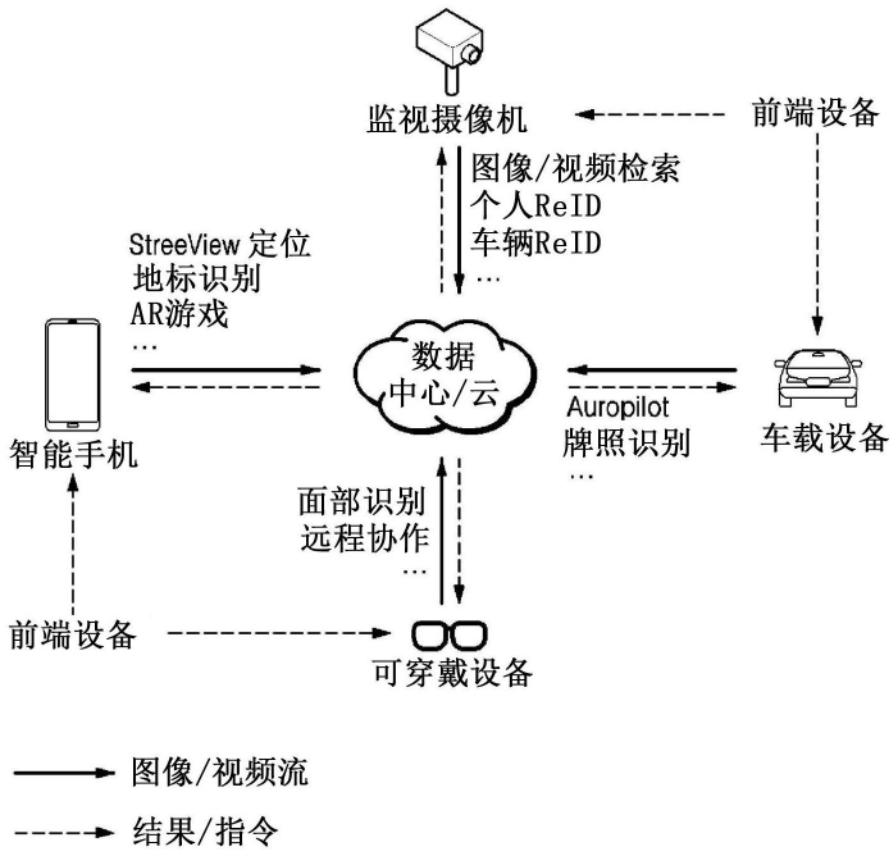


图3

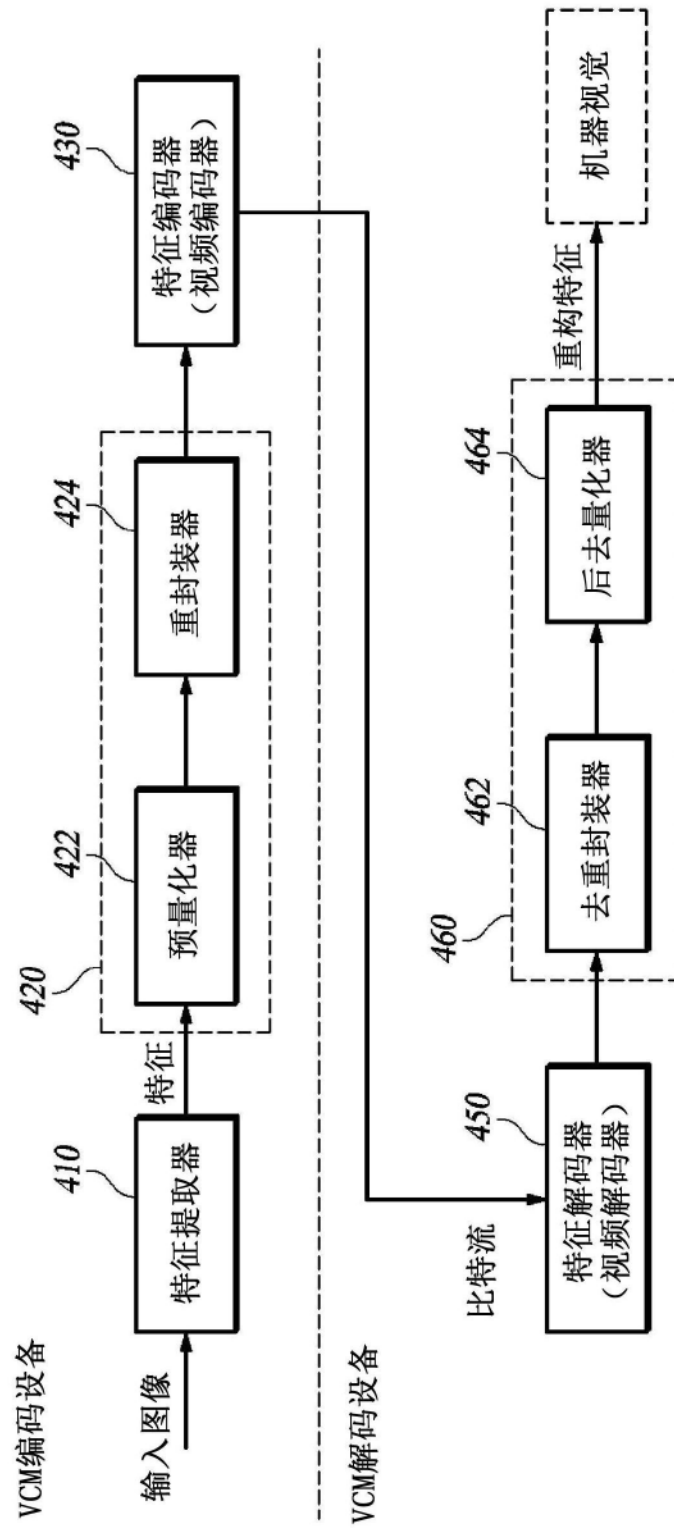


图4

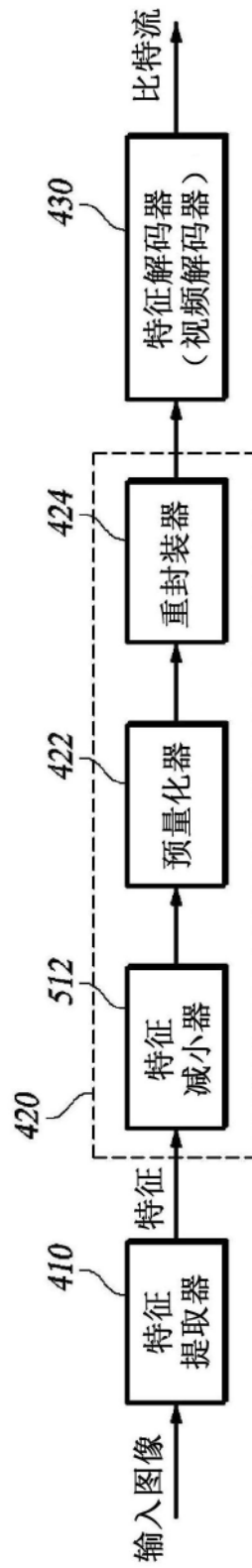


图5

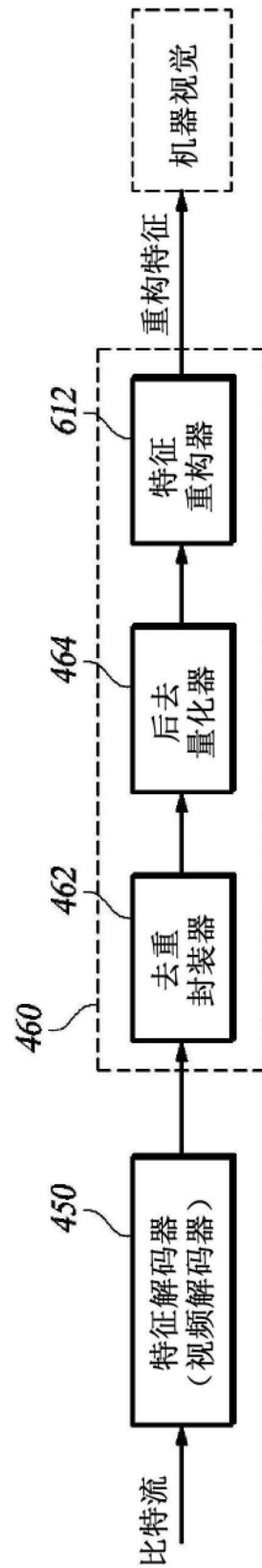


图6

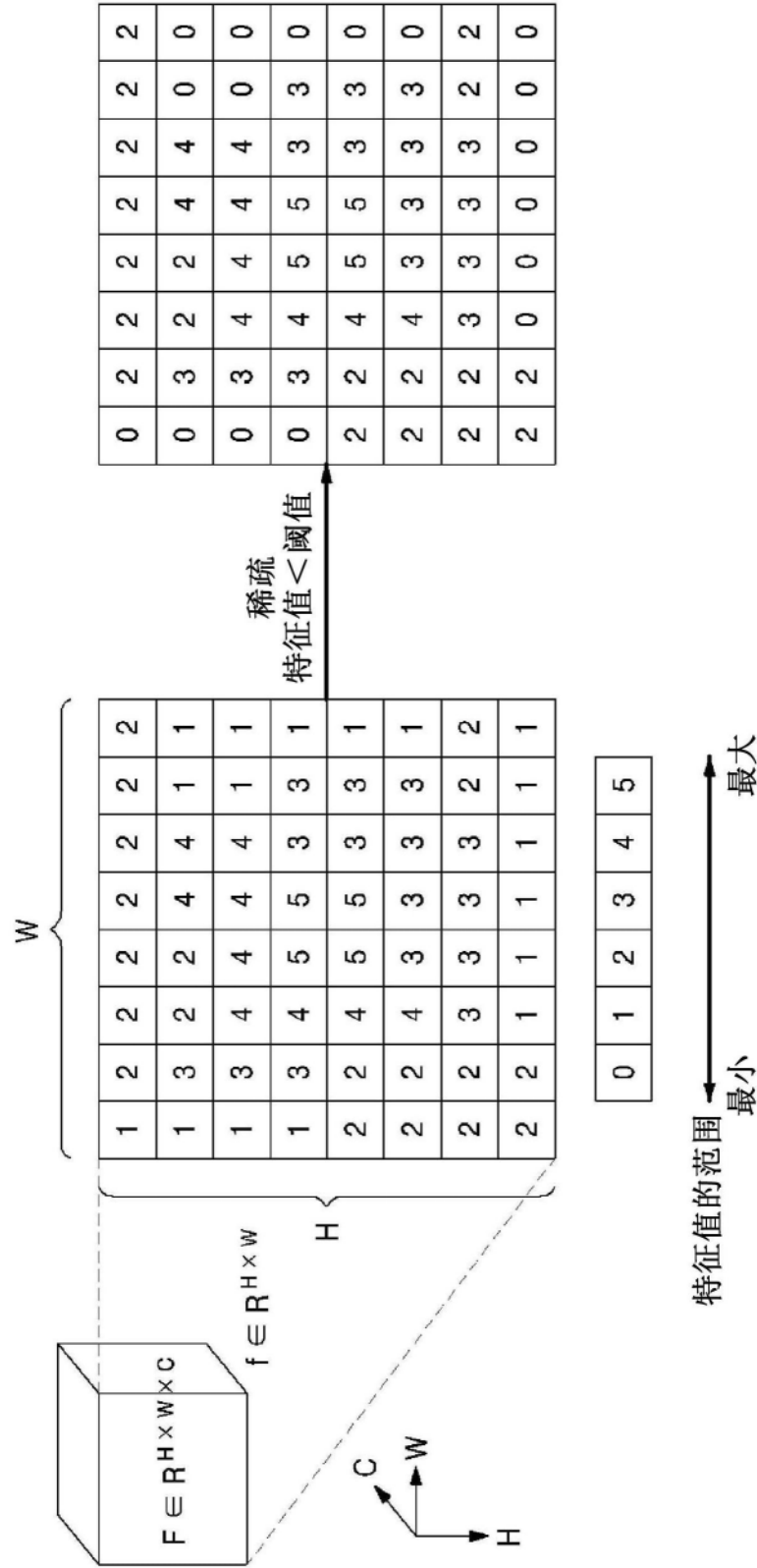


图7

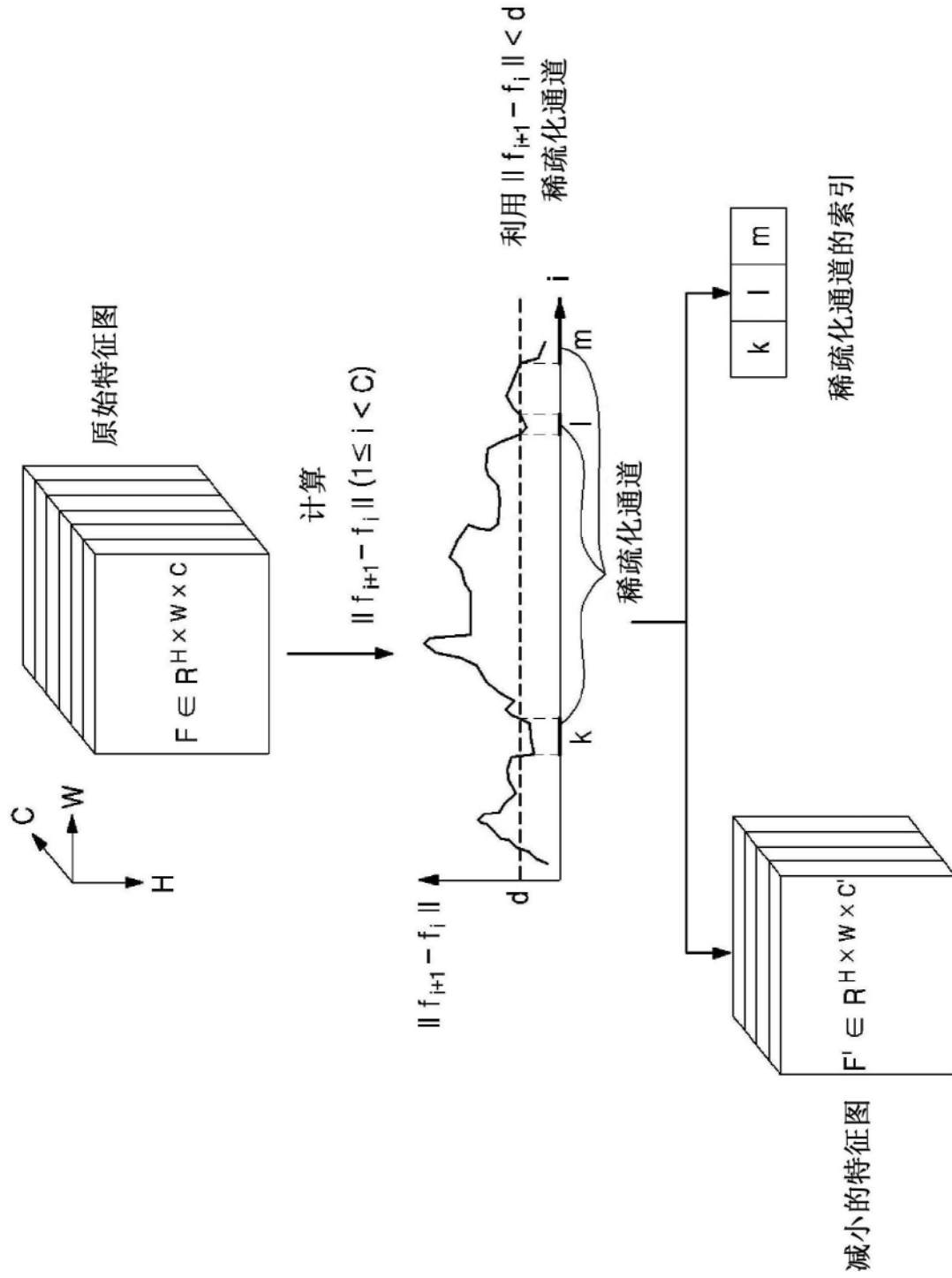


图8

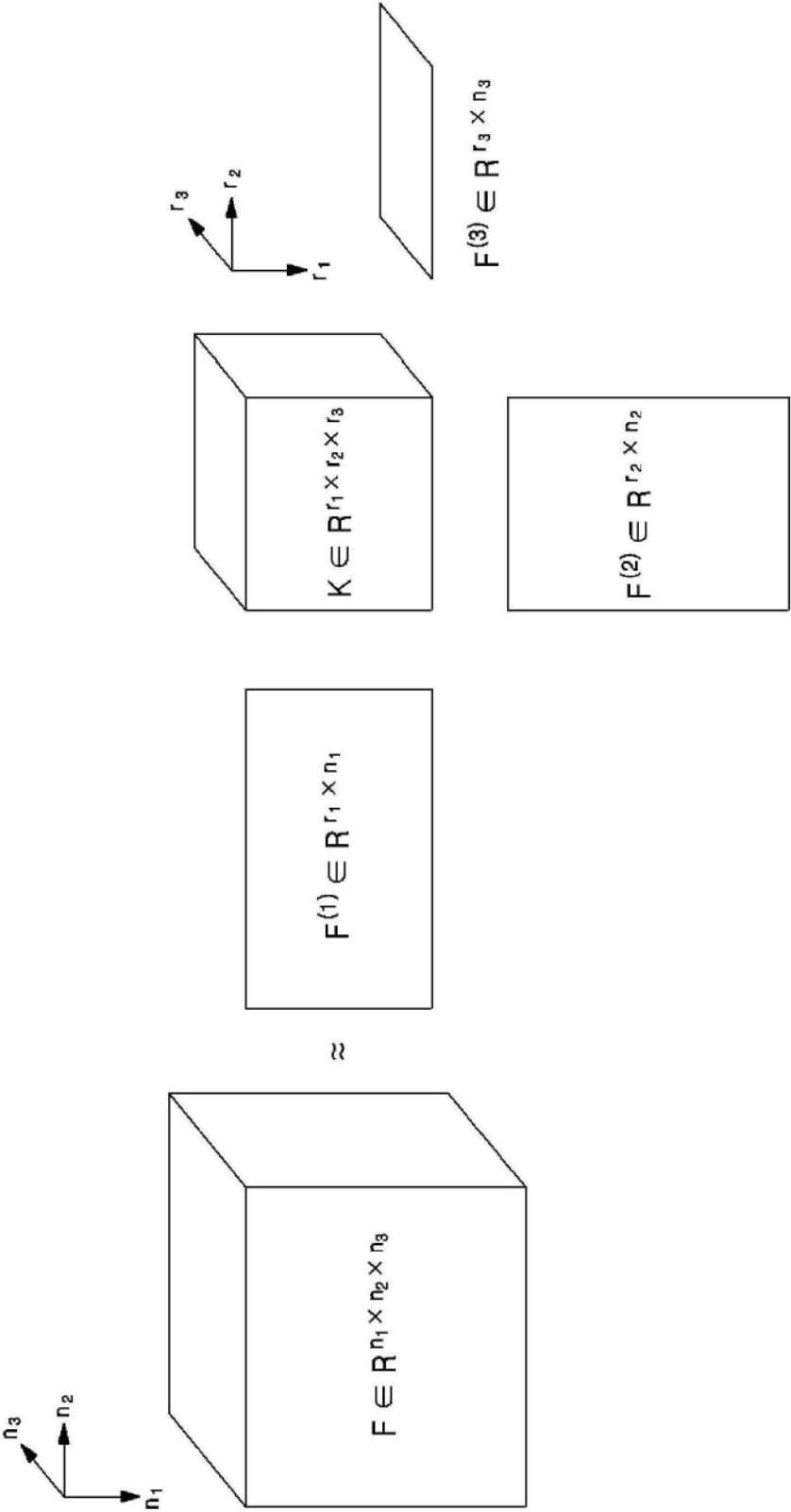


图9

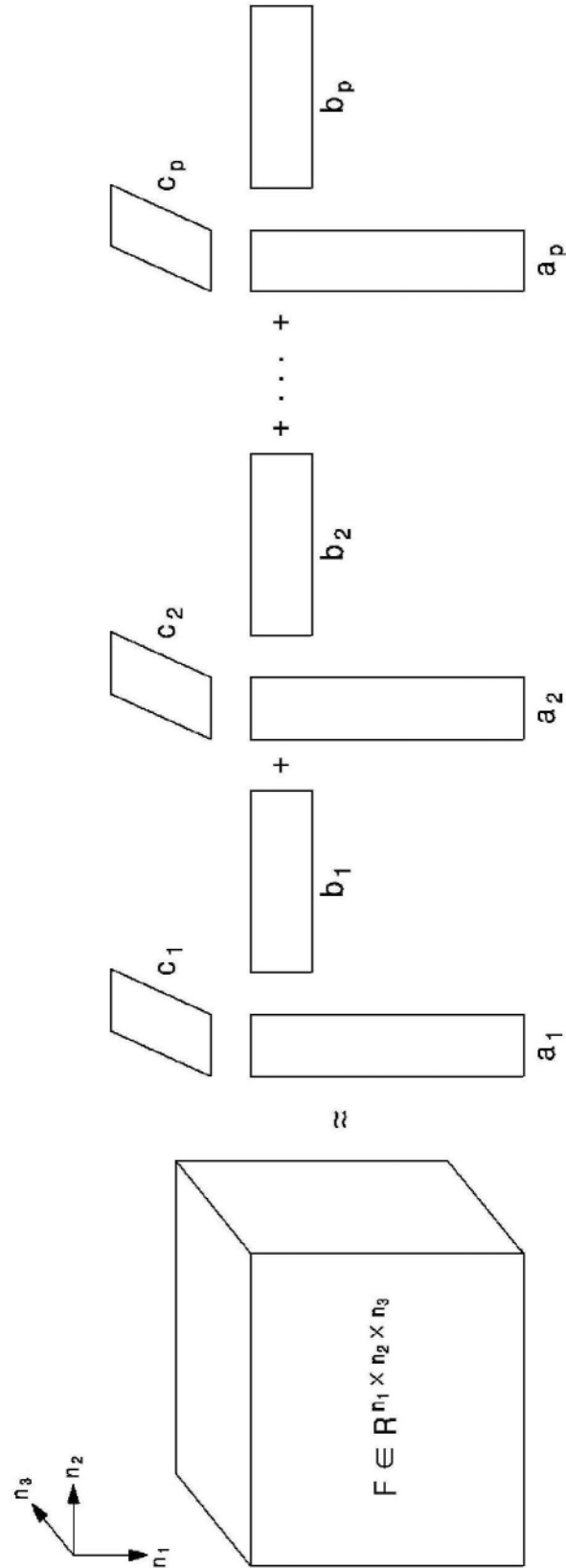


图10

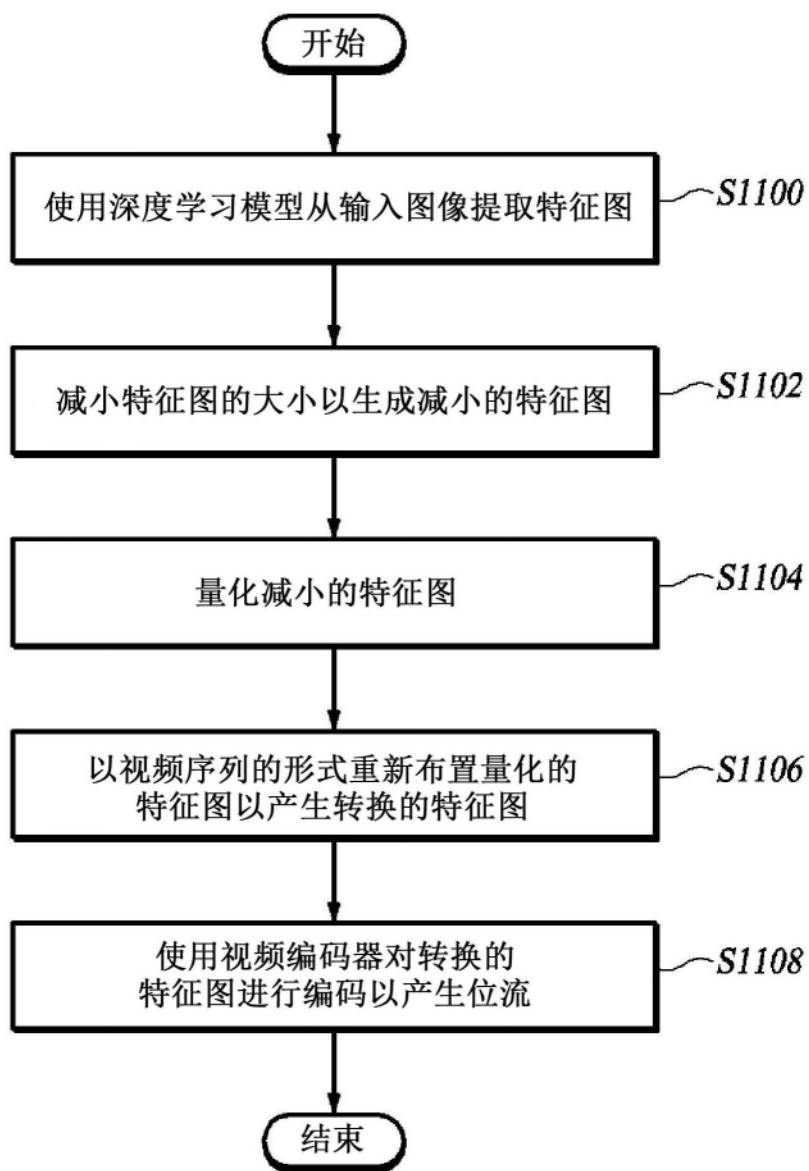


图11

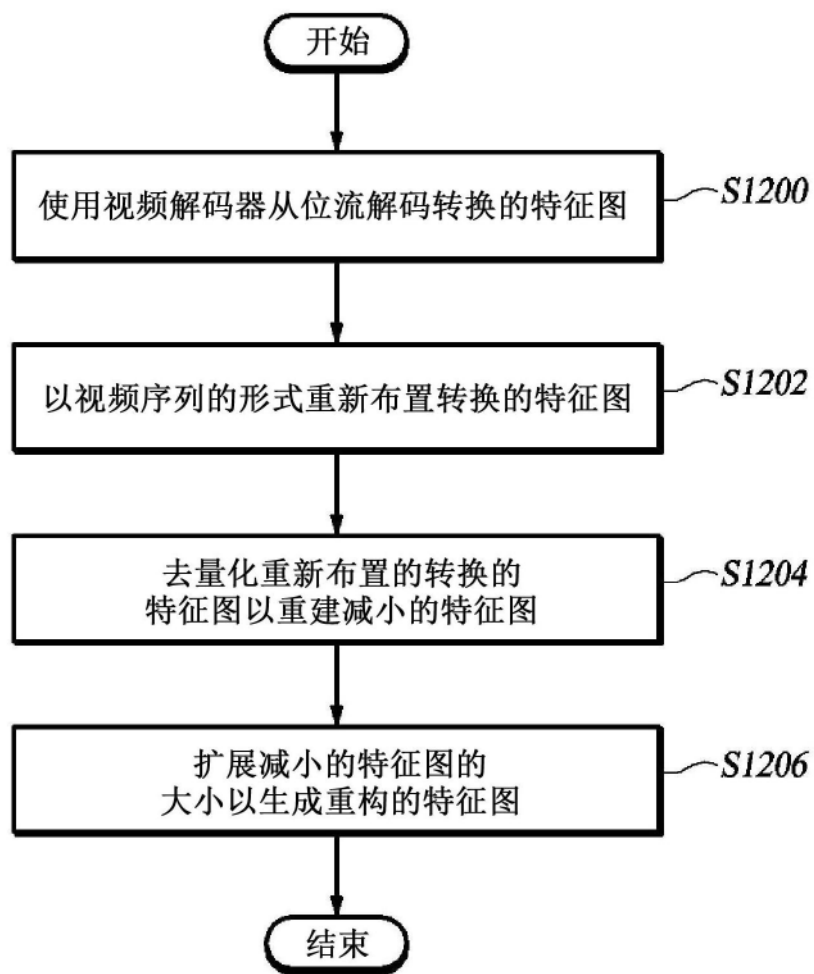


图12