(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117120963 A
(43)申请公布日 2023.11.24

(21)申请号 202280027110.X

(22)申请日 2022.03.29

(30)优先权数据
21167514.5 2021.04.08 EP

(85)PCT国际申请进入国家阶段日
2023.10.08

(86)PCT国际申请的申请数据
PCT/EP2022/058273 2022.03.29

(87)PCT国际申请的公布数据
WO2022/214357 EN 2022.10.13

(71)申请人 皇家飞利浦有限公司
地址 荷兰艾恩德霍芬

(72)发明人 C·韦雷坎普 J·G·H·科庞

(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 孟杰雄

(51)Int.Cl.
G06F 3/01(2006.01)

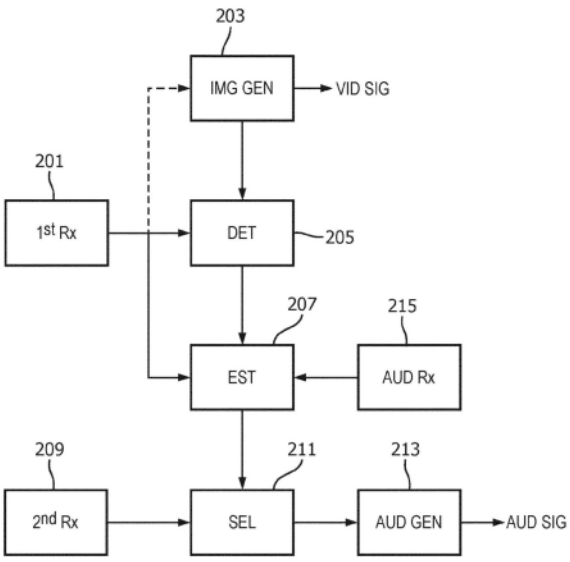
权利要求书2页 说明书18页 附图3页

(54)发明名称

用于生成音频信号的装置和方法

(57)摘要

一种装置包括从图像传感器接收深度和/或视觉数据的实时图像序列的第一接收器(201)。第二接收器(209)接收一组音频对象和指示音频对象与材料特性之间的联系的元数据。图像生成器(203)生成输出图像序列,所述输出图像序列包括与现实世界场景中的虚拟场景对象相对应的图像对象,并且检测器(205)检测所述虚拟场景对象与现实世界场景对象之间的相互作用。估计器(207)响应于图像帧的序列的图像数据而确定所述现实世界场景对象的材料性质,并且选择器(211)响应于所述材料性质和联系到所述一组音频对象的材料特性而选择音频对象。输出电路(213)生成包括所述第一音频对象的所述输出音频信号。



CN 117120963 A

1.一种用于生成输出音频信号的装置,所述装置包括:

第一接收器(201),其被布置为从图像传感器接收现实世界场景的实时图像序列,所述实时图像序列包括图像帧的序列,每个图像帧包括视觉图像数据和深度图像数据中的至少一项;

第二接收器(209),其被布置为接收一组音频对象和针对所述一组音频对象中的音频对象的元数据,所述元数据指示所述一组音频对象中的音频对象与材料特性之间的联系;

图像生成器(203),其被布置为生成输出图像序列,所述输出图像序列包括与所述现实世界场景中的虚拟场景对象相对应的图像对象;

检测器(205),其被布置为响应于对所述虚拟场景对象与所述现实世界场景的现实世界场景对象之间的接近度的检测而检测所述虚拟场景对象与所述现实世界场景对象之间的相互作用;

估计器(207),其被布置为响应于所述图像帧的序列的图像数据而确定针对所述现实世界场景对象的材料性质;

选择器(211),其被布置为响应于所述材料性质和被联系到所述一组音频对象中的音频对象的所述材料特性而选择所述一组音频对象中的第一音频对象;以及

输出电路(213),其被布置为生成包括所述第一音频对象的所述输出音频信号。

2.根据权利要求1所述的装置,其中,所述估计器(207)被布置为:

确定所述图像帧的序列中的至少一个图像帧中的相互作用图像区域,所述相互作用图像区域是所述至少一个图像帧中发生所述相互作用的图像区域;并且

响应于所述相互作用图像区域的图像数据而确定针对所述场景对象的所述材料性质。

3.根据权利要求1或2所述的装置,其中,所述第二接收器(209)被布置为从远程服务器接收所述元数据。

4.根据任一前述权利要求所述的装置,其中,针对至少一些音频对象的所述元数据包括对所述至少一些音频对象与现实世界场景对象的材料特性之间的联系以及所述至少一些音频对象与虚拟场景对象的材料特性之间的联系的指示;并且其中,所述选择器(211)被布置为响应于所述材料性质和被联系到所述一组音频对象的现实世界对象的材料特性并且响应于所述虚拟场景对象的材料性质和被联系到所述一组音频对象的虚拟场景对象的材料特性而选择所述第一音频对象。

5.根据任一前述权利要求所述的装置,其中,所述选择器(211)被布置为响应于所述虚拟场景对象的动态性质而选择所述第一音频对象。

6.根据任一前述权利要求所述的装置,其中,所述检测器(205)被布置为确定所述相互作用的性质,并且所述选择器(211)被布置为响应于所述相互作用的所述性质而选择所述第一音频对象。

7.根据权利要求6所述的装置,其中,所述相互作用的所述性质是从包括以下各项的组中选择的至少一个性质:

所述相互作用的速度;

所述虚拟场景对象与所述现实世界场景对象之间的碰撞的力;

所述虚拟场景对象与所述现实世界场景对象之间的碰撞的弹性;

所述相互作用的持续时间;以及

所述虚拟场景对象相对于所述现实世界场景对象的移动的方向。

8.根据任一前述权利要求所述的装置,其中,所述选择器(211)被布置为响应于所述虚拟对象相对于所述现实世界场景对象的取向而选择所述第一音频对象。

9.根据任一前述权利要求所述的装置,其中,所述估计器(207)被布置为确定所述现实世界场景对象与多个对象类目中的至少第一类目的匹配指示;并且响应于所述匹配指示和被联系到所述对象类目的材料性质而确定所述材料性质。

10.根据权利要求9所述的装置,还包括用于接收在所述现实世界场景中捕获的实时音频的音频信号的音频接收器(215),并且其中,所述估计器被布置为响应于所述音频信号而确定所述匹配指示。

11.根据任一前述权利要求所述的装置,其中,所述选择器(211)被布置为如果没有检测到满足选择准则的音频对象,则选择所述第一音频对象作为默认音频对象。

12.根据任一前述权利要求所述的装置,其中,至少一个图像帧包括深度图像数据,并且其中,所述估计器(207)被布置为响应于检测到表示所述现实世界场景对象的所述至少一个图像帧的图像区域的至少一部分具有不超过阈值的针对深度图像数据的置信度水平而确定针对所述现实世界场景对象的所述材料性质。

13.一种生成输出音频信号的方法,所述方法包括:

从图像传感器接收现实世界场景的实时图像序列,所述实时图像序列包括图像帧的序列,每个图像帧包括视觉图像数据和深度图像数据中的至少一项;

接收一组音频对象和针对所述一组音频对象中的音频对象的元数据,所述元数据指示所述一组音频对象中的音频对象与材料特性之间的联系;

生成输出图像序列,所述输出图像序列包括与所述现实世界场景中的虚拟场景对象相对应的图像对象;

响应于对所述虚拟场景对象与所述现实世界场景的现实世界场景对象之间的接近度的检测而检测所述虚拟场景对象与所述现实世界场景对象之间的相互作用;

响应于所述图像帧的序列的图像数据而确定针对所述现实世界场景对象的材料性质;

响应于所述材料性质和被联系到所述一组音频对象中的音频对象的所述材料特性而选择所述一组音频对象中的第一音频对象;并且

生成包括所述第一音频对象的所述输出音频信号。

14.一种包括计算机程序代码单元的计算机程序产品,当所述程序在计算机上运行时,所述计算机程序代码单元适于执行根据权利要求13所述的所有步骤。

15.一种音频信号,包括一组音频对象和针对所述音频对象的元数据,所述元数据指示音频对象与针对现实世界对象的材料特性之间的联系。

用于生成音频信号的装置和方法

技术领域

[0001] 本发明涉及一种用于生成用于虚拟世界场景对象与现实世界场景对象之间的相互作用的音频信号并且具体地但非排他地用于生成用于增强现实应用的音频信号的装置和方法。

背景技术

[0002] 近年来,基于视听内容的体验的种类和范围已经大幅增加,其中,不断开发和引入利用和消费此类内容的新服务和方式。特别地,正在开发许多空间和交互式服务、应用和体验,以给予用户更复杂且沉浸式的体验。

[0003] 这样的应用的示例是快速成为主流的虚拟现实(VR)、增强现实(AR)和混合现实(MR)应用,其中,许多解决方案针对消费者市场。许多标准化机构也正在开发许多标准。这种标准化活动正在积极开发用于VR/AR/MR系统的各个方面的标准,包括例如流传输、广播、渲染等。

[0004] VR应用往往提供对应于用户处于不同世界/环境/场景中的用户体验,而AR(包括混合现实MR)应用往往提供对应于用户处于现实本地环境中但添加了附加信息或虚拟对象或信息的用户体验。因此,VR应用往往提供完全沉浸式的合成生成的世界/场景,而AR应用往往提供叠加用户物理存在的现实场景的部分合成的世界/场景。然而,术语通常可互换使用并且具有高度重叠。在下文中,术语增强现实/AR将用于表示增强现实和混合现实两者(以及有时表示虚拟现实的一些变型)。

[0005] 作为示例,使用手持设备的增强现实的服务和应用已经变得越来越流行,并且软件API(应用编程接口)和工具包(诸如ARKit(由Apple Inc.开发)和ARCore(由Google Inc.开发))已经被引入以支持智能电话和平板电脑上的增强现实应用。在这些应用中,设备的内置相机和其他传感器用于生成环境的实时图像,其中,虚拟图形叠加所呈现的图像。应用可以例如生成具有叠加实况视频的图形对象的实况视频馈送。这样的图形对象可以例如用于定位虚拟对象,使得它们被感知为存在于现实世界场景中。

[0006] 作为另一示例,已经开发了头盔和眼镜,其中,现实世界场景可以通过AR眼镜直接观看,但是这些也能够生成用户在通过眼镜观看时看到的图像。这也可用于呈现被感知为所观看的现实世界场景的一部分的虚拟图像。运动传感器用于跟踪头部移动,并且所呈现的虚拟对象可以对应地适于提供虚拟对象是在现实世界中观看的现实世界对象的印象。

[0007] 这些方法分别被称为直通(pass-through)和透视(see-through),并且两者都可以提供新颖且令人兴奋的用户体验。

[0008] 除了视觉渲染之外,一些AR应用可以提供对应的音频体验。此外,对于视觉体验,已经提出要提供可以对应于虚拟对象的音频。例如,如果虚拟对象是将生成噪声的对象,则可以通过生成的对应声音来补充对象的视觉呈现。在一些情况下,还可以生成反映虚拟对象的动作的声音。

[0009] 通常,通过在适当的时间渲染预先录制的音频片段来将声音生成为默认声音。在

一些情况下,声音可以适于反映当前环境,诸如例如通过根据当前环境调整感知混响,或可以例如被处理以被感知为从与现实世界中的虚拟对象的感知位置相对应的位置到达。通常,这种定位可以通过双耳处理来实现,以生成合适的耳机音频输出。

[0010] 然而,尽管在许多实施例中这样的方法可以提供有趣的应用和用户体验,但是常规方法往往是次优的,并且往往难以实施和/或往往提供次优的性能和用户体验。

[0011] 因此,改进的方法将是有利的。特别地,允许改进的操作、增加的灵活性、降低的复杂性、促进的实施、改进的音频体验、改进的音频质量、减少的计算负担、针对混合/增强现实应用的改进的适合性和/或性能、改进的用户沉浸感和/或改进的性能和/或操作的生成音频信号的方法将是有利的。

发明内容

[0012] 因此,本发明试图优选地单独地或以任何组合减轻、缓解或者消除上文所提到的缺点中的一个或多个。

[0013] 根据本发明的一个方面,提供了一种用于生成输出音频信号的装置,所述装置包括:第一接收器,其被布置为从图像传感器接收现实世界场景的实时图像序列,所述实时图像序列包括图像帧的序列,每个图像帧包括视觉图像数据和深度图像数据中的至少一项;第二接收器,其被布置为接收一组音频对象和针对所述一组音频对象中的音频对象的元数据,所述元数据指示所述一组音频对象中的音频对象与材料特性之间的联系;图像生成器,其被布置为生成输出图像序列,所述输出图像序列包括与所述现实世界场景中的虚拟场景对象相对应的图像对象;检测器,其被布置为响应于对所述虚拟场景对象与所述现实世界场景的现实世界场景对象之间的接近度的检测而检测所述虚拟场景对象与所述现实世界场景对象之间的相互作用;估计器,其被布置为响应于所述图像帧的序列的图像数据而确定针对所述现实世界场景对象的材料性质;选择器,其被布置为响应于被联系到所述一组音频对象中的音频对象的所述材料特性和所述材料性质而选择所述一组音频对象中的第一音频对象;以及输出电路,其被布置为生成包括所述第一音频对象的所述输出音频信号。

[0014] 本发明可以在许多实施例中提供改进的用户体验,并且特别地可以在许多实施例中提供改进的且更沉浸式的增强现实体验。在许多实施例中,可以在维持低复杂性和/或实施的同时实现该方法。该方法可以非常适合于AR系统,其中,音频和支持元数据可以由远程服务器提供。该方法可以促进和/或支持其中集中式服务器可以为多个远程客户端提供支持并且可以基本上促进远程客户端的实施的系统。该方法可以支持音频的集中生成和管理以便增强AR应用和体验。

[0015] 音频对象可以是音频剪辑/片段等,并且可以以任何合适的方式表示。在许多实施例中,每个音频对象可以表示时间间隔中的声音。在许多实施例中,对于音频对象中的任何一个,时间间隔可以不超过5秒、10秒或20秒。

[0016] 材料特性可以是现实世界对象的材料特性。

[0017] 输出图像序列可以包括视觉图像。

[0018] 可以在表示现实世界场景的三维坐标系中检测相互作用。相互作用可以是虚拟场景对象与现实世界场景对象之间的接近度/碰撞/接触。

[0019] 该装置可以是增强现实装置。该装置可以是用于为增强现实应用提供输出图像序

列和输出音频信号的装置。增强现实应用可以在现实世界场景中呈现虚拟场景对象。

[0020] 根据本发明的可选特征,所述估计器被布置为:确定所述图像帧的序列中的至少一个图像帧中的相互作用图像区域,所述相互作用图像区域是所述至少一个图像帧的发生所述相互作用的图像区域;以及响应于所述相互作用图像区域的图像数据而确定所述场景对象的所述材料性质。

[0021] 这在许多实施例中可以提供特别有效和/或有利的材料性质估计,并且在许多实施例中可以具体地允许更准确的材料性质估计。因此,该方法可以提供改进的用户体验。

[0022] 根据本发明的可选特征,所述第二接收器被布置为从远程服务器接收所述元数据。

[0023] 该方法可以提供特别有效的应用,其中,音频可以被远程地并且可能集中地生成和管理,同时在本地有效地适应于适当的当前条件。

[0024] 根据本发明的可选特征,针对至少一些音频对象的所述元数据包括对所述至少一些音频对象与现实世界场景对象的材料特性之间的联系以及所述至少一些音频对象与虚拟场景对象的材料特性之间的联系的指示;并且其中,所述选择器被布置为响应于联系到所述一组音频对象的现实世界对象的材料特性和材料性质并且响应于联系到所述一组音频对象的虚拟场景对象的材料特性和虚拟场景对象的材料性质而选择所述第一音频对象。

[0025] 在许多实施例中,这可以允许改进的性能,并且可以具体地允许对特定相互作用的改进的调整。通常可以实现更沉浸式的用户体验。

[0026] 根据本发明的可选特征,所述选择器被布置为响应于所述虚拟场景对象的动态性质而选择所述第一音频对象。

[0027] 在许多实施例中,这可以允许改进的性能,并且可以具体地允许对特定相互作用的改进的调整。通常可以实现更沉浸式的用户体验。

[0028] 根据本发明的可选特征,所述检测器被布置为确定所述相互作用的性质,并且所述选择器被布置为响应于所述相互作用的所述性质而选择所述第一音频对象。

[0029] 在许多实施例中,这可以允许改进的性能,并且可以具体地允许对特定相互作用的改进的调整。通常可以实现更沉浸式的用户体验。

[0030] 根据本发明的可选特征,所述相互作用的所述性质是从包括以下各项的组中选择的至少一个性质:所述相互作用的速度;所述虚拟场景对象与所述现实世界场景对象之间的碰撞的力;所述虚拟场景对象与所述现实世界场景对象之间的碰撞的弹性;所述相互作用的持续时间;以及所述虚拟场景对象相对于所述现实世界场景对象的移动的方向。

[0031] 根据本发明的可选特征,所述选择器被布置为响应于所述虚拟对象相对于所述现实世界场景对象的取向而选择所述第一音频对象。

[0032] 在许多实施例中,这可以允许改进的性能,并且可以具体地允许对特定相互作用的改进的调整。通常可以实现更沉浸式的用户体验。

[0033] 根据本发明的可选特征,所述估计器被布置为确定所述现实世界场景对象与多个对象类目中的至少第一类目的匹配指示;以及响应于所述匹配指示和联系到所述对象类目的材料性质而确定材料性质。

[0034] 这可以提供材料性质的特别有利的且通常更低复杂性的确定,然而,其仍然可以具有高准确性。在许多实施例中,归类/分类可以有利地使用神经网络来实现。

[0035] 根据本发明的可选特征,所述装置还包括用于接收在所述现实世界场景中捕获的实时音频的音频信号的音频接收器,并且其中,所述估计器被布置为响应于所述音频信号而确定所述匹配指示。

[0036] 在许多实施例中,这种方法可以显著改进材料性质估计的准确性,从而导致改进的整体性能。

[0037] 根据本发明的可选特征,所述选择器被布置为如果没有检测到满足选择准则的音频对象,则选择所述第一音频对象作为默认音频对象。

[0038] 根据本发明的可选特征,至少一个图像帧包括深度图像数据,并且其中,所述估计器被布置为响应于检测到表示所述现实世界场景对象的所述至少一个图像帧的图像区域的至少一部分具有不超过阈值的深度图像数据的置信度水平而确定所述现实世界场景对象的所述材料性质。

[0039] 在一些实施例中,所述图像帧包括视觉图像数据和深度图像数据,并且所述估计器被布置为响应于对于所述图像区域的至少一部分检测到所述视觉图像数据的亮度超过阈值并且深度图像数据的置信度水平不超过阈值而将所述现实世界场景对象确定为具有金属分量。

[0040] 根据本发明的另一方面,提供了一种生成输出音频信号的方法,所述方法包括:从图像传感器接收现实世界场景的实时图像序列,所述实时图像序列包括图像帧的序列,每个图像帧包括视觉图像数据和深度图像数据中的至少一项;接收一组音频对象和针对所述一组音频对象中的音频对象的元数据,所述元数据指示所述一组音频对象中的音频对象与材料特性之间的联系;生成输出图像序列,所述输出图像序列包括与所述现实世界场景中的虚拟场景对象相对应的图像对象;响应于所述虚拟场景对象与所述现实世界场景的现实世界场景对象之间的接近度的检测而检测所述虚拟场景对象与所述现实世界场景对象之间的相互作用;响应于所述图像帧的序列的图像数据而确定所述现实世界场景对象的材料性质;响应于联系到所述一组音频对象中的音频对象的所述材料特性和所述材料性质而选择所述一组音频对象中的第一音频对象;以及生成包括所述第一音频对象的所述输出音频信号。

[0041] 所述方法可以包括显示所述输出图像序列和/或渲染所述输出音频信号。

[0042] 参考下文描述的(一个或多个)实施例,本发明的这些和其他方面、特征和优点将是显而易见的并且得以阐明。

附图说明

[0043] 将参考附图仅通过示例的方式描述本发明的实施例,在附图中:

[0044] 图1图示了增强现实系统的元件的示例;

[0045] 图2图示了根据本发明的一些实施例的用于生成输出音频信号的音频装置的示例;

[0046] 图3图示了具有虚拟场景对象的现实世界场景的图像的示例;并且

[0047] 图4图示了根据本发明的一些实施例的用于生成输出音频信号的方法的示例。

具体实施方式

[0048] 以下说明书将集中于音频信号的生成,以补充作为增强现实应用的一部分的现实世界场景中的虚拟对象的图像生成。然而,应当理解,所描述的原理和概念可以用于许多其他应用和实施例中。

[0049] 允许呈现虚拟信息和对象以补充现实世界环境的增强现实体验变得越来越流行,并且正在开发服务以满足这种需求。

[0050] 在许多方法中,AR应用可以通过例如不使用任何远程AR服务器或甚至不具有对任何远程AR服务器的任何访问的独立设备在本地提供给观看者。然而,在其他应用中,AR应用可以基于从远程或中央服务器接收的数据。例如,音频或图形数据可以从远程中央服务器提供给AR设备,并且可以被本地处理以生成期望的AR体验。

[0051] 图1图示了AR系统的这种示例,其中远程AR客户端装置101例如经由网络105(诸如互联网)与AR服务器103联系。服务器103可以被布置为同时支持潜在大量的客户端设备101。

[0052] AR服务器103可以例如通过将定义虚拟环境和对象的元素的数据发送到客户端设备101来支持增强体验。数据可以具体地描述可以由客户端设备101使用以生成可以呈现给用户的叠加图形的多个虚拟对象的视觉特征和几何性质。在一些实施例中,数据还可以包括可以呈现给用户的各种信息。此外,服务器103可以向客户端设备103提供音频数据,该音频数据可以用于在本地生成可以进一步增强用户体验并且具体地沉浸感的虚拟声音/音频。

[0053] 图2图示了根据本发明的一些实施例的设备。该设备可以具体地是图1的客户端设备101,并且将参考这样的实施例进行描述。

[0054] 该装置包括第一接收器201,该第一接收器201被布置为从一个或多个图像传感器接收图像数据。来自图像传感器的图像数据具体地包括现实世界场景的帧/图像的实时图像序列。

[0055] 在许多实施例中,帧可以包括来自视觉图像相机的视觉图像数据。帧/图像数据可以包括视觉图像的实时序列,其中每个像素表示从像素的观看方向接收的光强度。每个像素可以例如包括可见光谱的(可能加权的)区间的一组光亮度/强度值。例如,视觉图像的像素值可以表示一个或多个亮度水平,诸如例如颜色通道亮度/强度值的集合。例如,图像可以是RGB图像。在许多实施例中,视觉图像可以是彩色图像和/或RGB图像(并且在一些实施例中,对视觉图像的引用可以由对彩色图像或RGB图像的引用代替)。

[0056] 替代地或另外地,图像传感器可以是提供深度图像/帧的实时序列的深度传感器。对于这样的深度图像,每个像素值可以表示在像素的观看方向上到对象的深度/距离。例如,每个像素值可以是视差或深度值。深度图像也可以被称为深度图。

[0057] 因此,在不同的实施例中,从图像传感器接收的帧/图像可以是例如视觉/彩色图像、红外图像、深度图像、雷达图像、多光谱图像、声纳图像、相位图像、相位差图像、强度图像、相干幅度图像和/或置信度图像。因此,图像/帧的序列可以是像素的二维结构,其包括表示现实世界在像素的观看方向上的性质的值。

[0058] 在许多实施例中,该装置将被布置为处理视觉图像和深度图像两者,并且第一接收器201可以被布置为接收现实世界场景的实时视觉图像序列和现实世界场景的实时深度

图像序列两者。在许多实施例中,图像可以是组合的视觉深度图像,并且第一接收器201可以被布置为接收现实世界场景的实时深度和视觉图像的序列。

[0059] 应当理解,用于生成视觉和/或深度图像的许多不同的传感器和方法是已知的,并且可以使用任何合适的方法和图像传感器。例如,传统的摄像机可以用作用于生成视觉图像的图像传感器。深度图像可以例如使用专用深度相机(诸如红外测距相机)来生成,或可以例如通过基于例如具有已知物理偏移的两个视觉相机的视差估计来生成。

[0060] (一个或多个)图像传感器正在捕获现实世界场景,并且因此所接收的图像包括对该现实世界场景的捕获。例如,对于许多AR应用,所接收的图像可以是与用户的环境相对应的现实世界场景。作为具体示例,用户可以佩戴AR头盔或眼镜,该AR头盔或眼镜还包括用于在观看者正在观看的方向上捕获现实世界场景的一个或多个图像传感器。

[0061] 该装置还包括图像生成器203,该图像生成器203被布置为生成并输出包括表示虚拟场景对象的图像对象的视觉图像序列。可以使用合适的显示器(诸如AR头盔或AR眼镜的显示器(或多个显示器))将输出视觉图像序列呈现给用户。因此,当被呈现输出视觉图像序列时,用户将感知图像对象对应于现实世界场景中存在的虚拟对象。

[0062] 图像生成器203可以响应于例如由从合适的移动传感器接收的数据提供的用户姿势(位置和/或取向)的变化而调整图像对象(例如,图像中的取向、图像中的位置、观看方向等)。因此,图像对象的呈现可以使得它提供存在于正在观看的现实世界场景中的虚拟对象的印象。

[0063] 在一些实施例中,所生成的输出视觉图像序列可以仅包括虚拟对象(或在包括多于一个的情况下,多个虚拟对象),并且可以例如使用透视眼镜来呈现,从而允许通过眼镜看到现实世界场景。这种方法被称为AR透视应用。

[0064] 在其他实施例中,输出视觉图像序列可以被生成为还包括如由图像传感器捕获的现实世界场景的呈现。因此,在这种方法中,还可以(并且仅)经由所生成的图像观看现实世界场景。这种方法被称为AR直通应用。

[0065] 应当理解,用于生成这样的AR图像以及用于调整和修改所生成的图像对象以提供现实世界场景中存在的虚拟对象的印象的许多不同方法是已知的,并且为了简洁起见,将不再进一步详细描述这样的特征。应当理解,在不脱离本发明的情况下,可以使用任何合适的方法。例如,可以使用来自开发的AR工具包的算法和方法,诸如例如来自ARKit或ARCore的过程。

[0066] 因此,图像生成器203可以生成图像对象,使得感知到虚拟场景对象被添加到现实世界场景。在许多实施例中,虚拟场景对象可以相对于至少一个现实世界场景对象可移动,并且通常虚拟场景对象可以相对于整个现实世界场景可移动。例如,虚拟场景对象可以根据用户动作、现实世界场景改变的性质、预定移动被移动以创建预定效果等。

[0067] 该装置还包括检测器205,该检测器205被布置为检测虚拟场景对象与现实世界场景的现实世界场景对象之间的相互作用。检测器205可以具体地检测在虚拟场景对象和现实世界场景对象之间发生碰撞或接触。例如,其他可能的相互作用可以是虚拟场景对象和现实世界场景对象的接近度、或可以是我们确信虚拟和现实世界对象已经相互触碰的(一个或多个)时间间隔的接触的持续时间的检测。相互作用的另一示例是当在现实世界中下雨并且雨滴落到虚拟对象上时。另一示例是当在现实世界中生成气流并且空气朝向虚拟对

象流动时。

[0068] 因此,检测器205可以响应于检测到虚拟场景对象与现实世界场景对象之间的接近度而检测虚拟场景对象与现实世界场景的现实世界场景对象之间的相互作用。接近度的检测可以根据任何合适的接近度准则。

[0069] 虚拟场景对象与现实世界场景对象之间的接近度的检测可以是现实世界场景对象的位置与满足接近度准则的虚拟场景对象的位置之间的距离的检测。虚拟场景对象与现实世界场景对象之间的接近度的检测可以是现实世界场景对象的位置与虚拟场景对象的位置之间的距离小于阈值的检测。可以在现实世界场景的场景坐标系中确定位置和距离。虚拟场景对象可以是呈现/显示为好像它是存在于现实世界场景中的现实世界场景对象的对象。

[0070] 在一些实施例中,检测器205可以确定现实世界场景对象在(用于场景的)(场景)坐标系中的(第一)位置和虚拟场景对象在坐标系中的(第二)位置。检测器205可以响应于(并且具体地,如果)(第一和第二)位置满足接近度准则而检测相互作用。具体地,检测器205可以响应于并且具体地在(第一和第二)位置之间的距离(根据任何合适的距离量度)满足距离接近度准则的情况下并且具体地在距离小于阈值的情况下检测相互作用。

[0071] 图像生成器可以被布置为生成图像对象以将虚拟场景对象表示为在现实世界场景/场景坐标系中具有给定姿势/空间存在的对象。可以响应于指示现实世界场景/场景坐标系中的该姿势/空间存在的位置以及现实世界场景/场景坐标系中的现实世界场景对象的位置而确定接近度。可以响应于该虚拟场景对象/空间过程在现实世界场景/场景坐标系中的位置和现实世界场景对象在现实世界场景/场景坐标系中的位置而确定接近度。

[0072] 取决于特定实施例的期望操作和性能,可以在不同的实施例中使用用于检测相互作用的不同方法,并且稍后将描述各种方法。

[0073] 检测器205耦合到估计器207,该估计器207被布置为确定已经检测到虚拟场景对象与其相互作用的场景对象的材料性质。材料性质可以是实时对象由其制成的材料/物质/填充物的指示(包括在对象包括不同材料的情况下的材料成分)。材料性质例如可以被生成为指示现实世界对象是否由多个类别中的一个类别制成,诸如例如它是否由木材、布料、金属、塑料等制成。

[0074] 估计器207被布置为响应于从图像传感器接收的图像数据而确定材料性质,即,它响应于从传感器接收的图像帧的序列的数据而被确定。在许多实施例中,图像帧的序列可以包括深度图像数据和视觉图像数据两者,并且估计器207可以被布置为响应于深度图像数据和视觉图像数据两者而估计材料性质。在其他实施例中,估计器207可以仅考虑视觉数据或深度数据。

[0075] 应当理解,在不同的实施例中,不同的算法和方法可以用于估计材料性质。还应当理解,在许多实施例中,即使材料性质的非常不准确且不可靠的估计也可以是有用的,并且可以提供改进的用户体验。

[0076] 作为低复杂性示例,估计器207可以例如被布置为通过将基于颜色的分割应用于所接收的视觉图像来确定材料性质,并且确定被认为与实时图像对象匹配的图像对象。然后,它可以将平均颜色与一组预定义类别进行比较,以找到具有最接近匹配的类别。实时图像对象的材料性质然后可以被设置为与最接近的类别相关联的材料性质。作为简单的示

例,如果现实世界图像对象被确定为主要是棕色,则可以估计它由木材制成,如果它主要是银,则可以估计它是金属,如果它主要是明亮的原色,则可以估计它是塑料等。

[0077] 应当理解,在大多数实施例中,可以使用用于估计材料性质的更复杂且更准确的方法,并且稍后将描述进一步的示例。

[0078] 该装置还包括第二接收器209,该第二接收器209被布置为接收多个音频对象以及音频对象的元数据。音频对象可以是声音/音频片段/音频片段,并且具体地可以是有限持续时间的音频信号。应当理解,音频对象可以使用任何表示和格式来表示声音。因此,音频对象可以对应于可以由装置选择并渲染以在各种动作发生时提供声音输出的声音。

[0079] 具体地,该装置包括选择器211,该选择器211被布置为当检测到虚拟场景对象与现实世界音频对象之间的相互作用时从所接收的音频对象中选择音频对象。所选择的音频对象被馈送到输出电路213,输出电路213被布置为生成包括第一音频对象的输出音频信号。因此,输出电路213可以通过生成包括音频对象的渲染的输出音频信号来渲染所选择的音频对象。

[0080] 应当理解,取决于音频对象的格式,输出电路213可以在一些情况下简单地生成输出信号作为音频对象的音频,在一些实施例中可以包括音频对象的解码,可以将输出音频信号生成为包括各种音频分量(例如,环境声音、叙述者音频等)之间的混合以与音频对象组合等。还应当理解,在输出电路213可以生成模拟输出音频信号例如以直接驱动扬声器的实施例中,输出电路213可以例如包括数模转换和模拟信号放大。

[0081] 输出电路213可以包括通过例如利用头部相关脉冲响应(HRIR)、头部相关传递函数(HRTF)或双耳房间脉冲响应(BRIR)的双耳处理来渲染音频对象的处理,或通过例如基于向量的幅度平移(VBAP)和/或(例如)遮挡、衍射、反射、混响、源程度等的进一步声学模拟来渲染到扬声器配置。具体地,渲染处理可以被配置为使得音频对象的声音被感知为源自检测到的相互作用的位置。

[0082] 在一些实施例中,输出电路213可以包括时序调整以控制渲染所选择的音频对象的时序,诸如具体地可能可变延迟。例如,可以包括“视觉”接触/相互作用的时间与渲染声音的时间之间的延迟。例如,可以基于声速和虚拟对象与观看者之间的距离来调整延迟,例如以确保所生成的声音与视觉相互作用同时被感知。在一些实施例中,这种时间调整可以在装置的除了输出电路213之外的其他部分处执行。

[0083] 由第二接收器209接收的元数据包括指示至少一些音频对象与材料特性之间的联系的元数据。元数据可以例如针对每个音频对象指示与其联系的现实世界对象的材料。例如,元数据可以指示一个音频对象联系到由木材制成的现实世界对象,另一音频对象联系到由金属制成的现实世界对象,另一音频对象联系到由塑料制成的现实世界对象等。

[0084] 如稍后将更详细描述的,在许多实施例中,元数据可以包括其他另外的联系,诸如例如音频对象到虚拟场景对象、到虚拟场景对象的材料特性、相互作用的类型等的联系。

[0085] 选择器211响应于检测器205检测到虚拟场景对象与现实世界场景对象之间的相互作用,被布置为基于估计的现实世界场景对象的材料性质和指示音频对象与材料特性之间的联系的元数据来选择音频对象。

[0086] 在低复杂性实施例中,选择器211可以简单地选择估计的材料性质和联系的材料特性匹配的音频对象,例如,如果估计的性质是“木材”,则选择联系到“木材”的音频对象,

并且然后生成对应的输出音频信号。

[0087] 在许多实施例中,该装置可以生成改进的用户体验,并且特别地可以提供更沉浸式且自然发声的体验,其中所感知的音频可以更接近地反映虚拟场景对象与现实世界对象的相互作用。例如,该装置可以提供与用户的视觉感知更紧密匹配的音频输出,并且可以提供由用户感知的视觉输入和音频输入之间的更紧密的对应关系,而不仅仅是伴随虚拟场景对象和具有标准化音频的现实世界对象之间的相互作用(诸如碰撞)。这可以通过装置调整并生成更紧密匹配相互作用的音频来实现。

[0088] 系统可以例如使用为虚拟场景对象提供的记录或合成的多个音频对象/剪辑/片段。每个音频片段可以对可能在场景中发生的与特定类别的现实世界对象的相互作用进行建模。在运行时,该装置可以使用内置相机来对例如虚拟对象所在的表面进行分类,并且基于分类结果,该装置可以播放用于特定相互作用的正确音频片段。

[0089] 在一些实施例中,第二接收器209可以被布置为从内部源接收音频对象和/或元数据。然而,在许多实施例中,第二接收器209可以被布置为从远程服务器(诸如具体地图1的服务器103)接收元数据并且通常音频对象。

[0090] 例如,远程服务器可以维护与对象之间的各种可能的相互作用相对应的大的音频对象库,并且可以进一步维护包括例如定义相互作用中涉及的对象的材料性质、相互作用的性质等的数据的元数据。

[0091] 在应用程序初始化时,或以重复率,第二接收器209可以检索所有音频对象的元数据。当检测器205检测到相互作用时,估计器207可以估计现实世界场景对象的材料性质,并且选择器211可评估元数据以找到匹配的音频对象。然后可以继续控制第二接收器209从服务器103检索所选择的音频对象,并且当被接收时,音频装置可以将音频信号生成为包括所接收的音频对象。

[0092] 在一些实施例中,这种方法对于实时操作可能太慢,并且在一些实施例中,第二接收器209可以从远程服务器103检索子集或可能所有音频对象并将它们本地存储以用于快速检索。例如,在一些实施例中,可以在启动时和/或以规则的间隔进行现实世界场景的分析,其中装置检测尽可能多的对象。然后,装置可以继续检索与在现实世界场景中检测到的对象或材料相对应的对象或材料和/或与活动的或位于用户位置附近的虚拟场景对象相对应的对象或材料的所有音频对象。在其他实施例中,用户可以例如在初始化时提供现实世界场景(例如,起居室、体育场等)的输入,并且装置可以继续检索作为这种环境中的可能对象存储在装置中的一组音频对象。

[0093] 因此,该方法可以提供用于允许中央服务器与远程设备装置相互作用以提供对AR应用的音频支持的有效方法。应当理解,如上所述的中央服务器也可以由相对本地的服务器(诸如5G边缘服务器)表示。

[0094] 在该方法中,虚拟场景对象被呈现给用户,使得它可以被感知为存在于现实世界场景中,并且因此它与现实世界场景中的空间(通常三维)几何结构/形状相关联。检测器205可以被布置为将相互作用检测为虚拟场景对象和三维现实世界空间中的现实世界场景对象之间的距离小于给定阈值或具体地它们彼此接触的检测。在一些情况下,可以考虑进一步的考虑,诸如对象朝向彼此移动的速度等。

[0095] 例如,开发用于支持AR应用的一些工具包提供3D平面表面检测,其中经由随时间

的特征点跟踪,经由特征点的集群随时间跟随的特定运动来检测平面区域。

[0096] 在一些实施例中,检测器205可以被布置为通过跟踪虚拟图形场景对象的边界网格中的任何顶点与所检测到的平面中的任何一个之间的最短距离来检测相互作用:

[0097] $$d(t) \equiv \min_{i \in V} \left(\min_{j \in P} \left(\text{dist}\left(\mathbf{x}_j(t), \mathbf{x}_i(t) \right) \right) \right)$$

[0098] 其中$i \in V$表示来自虚拟场景对象的边界网格中存在的所有顶点V的集合的顶点i,$j \in P$是针对其计算最短距离的平面,并且函数$\text{dist}(x_j(t), x_i(t))$评估该距离。应当注意,虚拟场景对象边界网格和图像传感器两者通常都在运动中,这意味着网格中的顶点位置和检测到的平面上的位置并且因此还有最终的最短距离d(t)都随时间变化。

[0099] 如果距离小于阈值,则可以检测到相互作用已经发生。具体地,如果距离小于小阈值,或例如它达到零,则可以检测到碰撞已经发生。

[0100] 在一些实施例中,当顶点到平面的最小距离减小到最小阈值距离$\Delta_{\text{碰撞}}$内并且瞬时对象速度v(t)超过给定的最小阈值v_{min}时,可以检测到已经发生相互作用并且具体地碰撞。

[0101] $$(d(t) \leqslant \Delta_{\text{碰撞}}) \wedge (v(t) > v_{min})$$

[0102] 速度条件背后的原因是,在实践中往往难以确定当对象速度低时是否发生实际碰撞,因为虚拟对象可能刚好在检测到的表面的前面或旁边停止。有用的参数值例如是$\Delta_{\text{碰撞}} = 5\text{mm}$并且$v_{min} = 0.05\text{m/s}$。

[0103] 估计器207然后可以继续确定已检测到与其发生相互作用的现实世界场景对象的材料性质。在许多实施例中,检测可以基于估计器207确定输入图像帧中的至少一个输入图像帧中的相互作用图像区域,其中相互作用图像区域是其中发生相互作用的图像帧的图像区域。因此,估计器207可以确定图像帧中的包括由检测器205检测到的对象之间的接触点的二维区域。例如,可以确定接触点在二维图像帧上的投影,并且可以识别其周围的区域。

[0104] 在一些实施例中,图像区域可以例如是以投影的接触点为中心的预定形状。在其他实施例中,图像区域可以适应于图像性质。例如,图像区域可以被识别为所投影的接触点周围的满足类似准则的区域,诸如视觉性质和/或深度不偏离所投影的接触点的性质多于预定量。

[0105] 然后,估计器207可以响应于相互作用图像区域的图像数据而确定场景对象的材料性质。例如,可以将颜色和纹理变化与预定类别集合的对应性质进行比较,并且可以将材料性质设置为最接近的匹配类别的性质。

[0106] 在许多实施例中,可以响应于针对图像数据确定的图像性质与多个图像性质参考的比较而确定材料性质,每个图像性质参考与材料性质值联系。图像性质可以响应于与多个图像性质参考中的匹配图像性质参考联系的材料性质值而被确定/被确定为与多个图像性质参考中的匹配图像性质参考联系的材料性质值。匹配图像性质参考可以被确定为图像性质和匹配图像性质参考满足匹配准则的图像性质参考。图像性质可以是亮度和/或颜色和/或纹理性质。

[0107] 具体地,一旦检测到碰撞,就可以确定图像相互作用区域。这可以通过首先使用相机视图矩阵和相机投影矩阵将最近的网格点x_i投影到图像中来完成。得到的2D图像点(u_i, v_i)然后用于从相机图像确定以(u_i, v_i)为中心的固定尺寸的子图像。图3图示了图像帧的示例,其中球形式的虚拟场景对象301可以与现实世界桌子形式的现实世界场景对象303相互

作用,从而导致图像区域305的检测。

[0108] 在许多实施例中,估计器207可以被布置为确定现实世界场景对象与多个对象类目/类别中的至少第一类目/类别的匹配指示。估计器207可以例如具有为对应于特定材料的图像区域存储的多个性质,诸如颜色、纹理变化、深度变化等。估计器207可以确定图像区域的对应性质,并将这些性质与存储的所有类目的性质进行比较。可以确定匹配指示以反映性质匹配的程度,并且类目可以被选择为匹配指示指示最高匹配的类目。

[0109] 然后可以选择现实世界场景对象的材料性质作为所选择的类目或类别的存储的性质。因此,类目/类别中的每一个与一个或多个材料性质相关联,诸如简单地对象由其制成的材料的指示,并且现实世界场景对象的材料性质被设置为针对所选择的类别存储的材料性质。

[0110] 应当理解,在许多实施例中,估计器207可以使用基本上更复杂的方法来确定现实世界场景对象的材料性质,并且具体地识别被认为最接近地匹配与图像区域相对应的图像区域的性质的类别或类目。

[0111] 实际上,在许多实施例中,估计器207可以包括被布置为基于相互作用图像区域的图像数据来估计现实世界场景对象的材质性质的神经网络。神经网络可以具体地用于识别现实世界场景对象的类目或类别。

[0112] 在许多实施例中,提取的(图像)相互作用区域可以具有比完整图像小得多的尺寸。在完整图像通常具有2K或4K的尺寸的情况下,相互作用子图像可以具有例如256×256像素的恒定的更小尺寸。

[0113] 例如256×256像素的彩色图像可以被表示为尺寸为3×256×256的张量,并且被馈送到经训练的分类器神经网络中。神经网络的输出可以是针对每个预定义的材料类别提供图像属于给定类别的概率的向量。

[0114] 在许多实施例中,以下对象(表面)类别可能是特别有利的:

[0115] -地面地毯(材料的短高度变化)

[0116] -地面地毯(材料高度的变化将会抑制碰撞声)

[0117] -地面石材

[0118] -地面木材

[0119] -台面玻璃

[0120] -台面木材

[0121] -墙面

[0122] -沙发

[0123] -垫子

[0124] -家用物品(炊具、键盘、膝上型计算机)的塑料表面

[0125] 这些类别中的每一个可以具有特定的颜色和纹理统计,或可以由于其边界几何形状(例如桌子或垫子)而被区分。

[0126] 作为训练数据集,可以捕获来自不同对象表面的图像并用地面实况类别标签手动注释。

[0127] 对于神经网络架构,一个选项可以是使用使通道数量加倍的2D卷积层作为基本块,随后是2D平均池化层(步幅=2)和线性整流作为激活。每个块以因子2来按比例缩小张

量的空间分辨率,但是使通道的数量加倍。该方法可以继续连接这些块,直到张量具有分辨率N通道x1x1。然后可以添加两个全连接层,随后是S形激活函数。

[0128] 如果所有类别概率的最大值低于给定阈值(在许多实施例中,0.5似乎是适当的值),则可以播放默认(中性)相互作用声音,即,如果估计器207不能足够准确地确定现实世界场景对象的材料,则可以选择默认音频对象。在所有其他情况下,音频对象/声音文件可以由选择器211选择,诸如具体地与由如通过分类确定的具有最高概率的材料相对应的音频对象。

[0129] 在一些实施例中,装置还可以包括用于接收在现实世界场景中捕获的实时音频的音频信号的音频接收器215,并且估计器207可以被布置为响应于音频信号而确定匹配指示。

[0130] 因此,可以通过基于音频的分类来辅助材料分类,基于音频的分类利用麦克风记录物理空间中的声音。如果在物理对象之间发生相互作用,则所得到的声音可以为材料的分类或更具体的性质提供有价值的信息。它可以用于提供更好的材料分类,或提供材料的进一步属性(实心/中空、薄/厚、张紧/松弛、大/小)。

[0131] 例如,当前和最近过去的环境声音信号S_{env}可以在短时间窗口(例如10秒)内捕获,并且可以馈送到材料分类器中。例如,持有智能手机或佩戴AR头盔的用户或其他人可能正在房间中行走,并且其鞋子在地面上的声音可以给出关于地面材料的信息。可以通过收集不同材料的图像和声音来产生材料分类器。然后,如上所述的更高级的神经网络将多条音频片段与捕获的图像连接,并将这些片段用作训练网络的输入。

[0132] 作为具体示例,在一些实施例中,材料性质的确定可以响应于深度图像数据,并且具体地,估计器可以被布置为响应于检测到表示现实世界场景对象的至少一个图像帧的图像区域的至少一部分具有不超过阈值的深度图像数据的置信水平而确定现实世界场景对象的材料性质。

[0133] 深度估计通常利用指示所生成的深度值/估计的可靠性的置信度数据来生成。例如,对于视差估计,可以确定置信度水平以反映形成视差估计的基础的图像区域在图像中匹配的紧密程度。作为另一示例,对于深度相机,例如基于红外光的反射,可以生成置信度水平以反映针对给定像素接收的红外光的量。如果仅接收到少量光,则距离估计/测距过程可能不如接收到大量光那样准确。在一些情况下,深度图像/图的每个像素可以包括估计的深度值以及指示像素的深度值被认为是多么可靠的置信度值/水平。

[0134] 在一些实施例中,当确定材料性质时,可以考虑该深度置信度数据。例如,与其他材料相比,可以已知一些材料提供降低的置信度水平。例如,一些材料可能在局部视觉外观上变化很大,这可能使视差估计不太可靠。

[0135] 作为另一示例,对于基于从传感器发射的反射光并且具体地反射红外光的深度估计,由于反射回到传感器的光量显著减少,一些材料可能导致置信度大大降低。这可以是例如不反射红外光但对红外光具有高吸收的材料的情况。对于表现出镜面反射的材料(例如金属对象)也可能是特别的情况,在这种情况下,非常少的红外光被反射回发射器和传感器。注意,对于基于飞行时间或结构化光的有源深度传感器,红外光传感器(例如,CMOS)通常与光发送器/发射器共同定位。同时,由于表面对环境光的反射,从金属表面接收的可见光可能相当大。在许多实施例中,估计器207可以例如检测到相互作用图像区域的区域具有

低深度置信度但是非常亮。这样的区域可以指示高反射但色散的表面,诸如金属表面。

[0136] 在一些实施例中,图像帧包括视觉图像数据和深度图像数据,并且估计器207可以被布置为响应于对于图像区域的至少一部分检测到视觉图像数据的亮度超过阈值并且深度图像数据的置信度水平不超过阈值而将现实世界场景对象确定为具有金属分量。

[0137] 如前所述,元数据可以从远程源接收,并且可以包括音频对象与对象的材料性质之间的联系,该对象具体地可以是现实世界场景对象。然而,在一些实施例中,元数据可以包括到两个对象的材料性质的联系,并且实际上可以在虚拟场景对象和现实世界场景对象之间进行区分。在这种情况下,选择器211可以基于现实世界场景对象的估计的材料性质并且还基于虚拟场景对象的材料性质来选择音频对象。由于虚拟场景对象可以是由装置生成的虚拟对象,因此该对象旨在被感知为由其制成的材料通常是已知的。在这种情况下,音频对象因此可以被选择为尽可能接近地匹配两种性质。

[0138] 因此,在一些实施例中,至少一些音频对象的元数据包括至少一些音频对象与现实世界场景对象的材料特性之间的联系以及至少一些音频对象与虚拟场景对象的材料特性之间的联系的指示。在这样的系统中,选择器211可以被布置为响应于联系到该组音频对象的现实世界对象的材料性质和材料特性质并且响应于联系到该组音频对象的虚拟场景对象的材料性质和虚拟场景对象的材料特性而选择音频对象。

[0139] 作为示例,远程服务器103可以提供将音频对象联系到材料性质的二维结构。这种结构的示例是以下二维查找表:

[0140]

	检测到的现实世界材料			
虚拟对象材料	木材	金属	混凝土	未知
木材	木材在木材上	木材在金属上	木材在混凝土上	木材在未知上
金属	金属在木材上	金属在金属上	金属在混凝土上	金属在未知上
橡胶	橡胶在木材上	橡胶在金属上	橡胶在混凝土上	橡胶在未知上

[0141] 在其他实施例中,元数据可以提供到其他特征的联系,并且音频对象的选择可以考虑这样的附加特征。

[0142] 例如,在许多实施例中,可以进一步考虑虚拟场景对象的动态性质。在此类实施例中,元数据可以指示不同音频对象联系到虚拟场景对象的不同状态/性质。虚拟场景对象的状态/性质可动态地改变,并且选择可以在寻找匹配音频对象时进一步考虑当前状态/性质。

[0143] 例如,放气的虚拟球生成与充气的虚拟球不同的声音。在一些实施例中,球的虚拟场景对象可以在查找表中包括两次,一次放气并且一次充气,其中每个联系到不同的音频对象/声音效果。

[0144] 虚拟对象的其他属性可以影响适当声音效果的选择或生成。例如,虚拟对象的取向可以引起不同的声音效果。

[0145] 虚拟场景对象可以例如具有不同类型的表面,诸如例如对于不同侧面具有不同表面性质的立方体。在这种情况下,虚拟场景对象的取向可以确定哪个表面与现实世界场景对象碰撞,并且音频对象的选择可以考虑该取向。

[0146] 作为具体示例,查找表可以包括用于不同取向的不同条目,因此对音频对象的选择可以包括选择与当前取向相对应的条目。虚拟场景对象可以由网格表示,其中每个面或面组与对象取向范围、虚拟材料性质或音频对象的子集相关联。

[0147] 最后,现实世界场景对象的其他属性可以被检测并在选择或生成声音效果时使用。打开的门可能听起来与关闭的门不同。或当玻璃容纳液体时,它听起来不同。

[0148] 在一些实施例中,音频对象的选择可以进一步响应于相互作用的性质,诸如具体地取决于相互作用的类型。因此,在一些实施例中,在一些实施例中,检测器205可以被布置为确定相互作用的性质,并且选择器211可以被布置为响应于相互作用的性质而选择第一音频对象。

[0149] 因此,在一些实施例中,相互作用本身的属性可以影响声音效果的选择或生成。相互作用的速度或强度可能导致不同的声音,因为物理碰撞可能与碰撞的强度非线性(例如,弹跳球、玻璃可能在特定碰撞强度下破裂等)。其他示例是碰撞的方向或非碰撞相互作用,诸如虚拟对象跨物理表面的滑动/摩擦。

[0150] 作为具体示例,被进一步考虑用于选择音频对象的相互作用性质可以是以下中的一个或多个:

[0151] 相互作用的速度:检测器205可以例如确定虚拟场景对象和现实世界场景对象之间的相对速度,并且可以基于此选择音频对象。例如,撞击地板的慢球可能具有与以高得多的速度撞击地板的球不同的声音。查找表可以例如具有用于不同速度范围的不同条目,其中每个条目联系到表示以不同速度撞击合适地板材料的球的录音的音频对象。作为示例,相互作用检测可以包括速度向量,该速度向量与物理表面取向进行比较,并且可选地,其在相互作用之后不久的行为允许检测相互作用类型。例如,如果对象速度矢量垂直并朝向物理表面,则相互作用可以被认为是"冲击"或"附接"相互作用类型。对象表面旁边的速度矢量将导致"滑动"相互作用类型。这种不同的相互作用类型可以联系到不同的音频对象。

[0152] 虚拟场景对象与现实世界场景对象之间的碰撞力。力可以例如根据当前速度以及虚拟场景对象和现实世界场景对象中的至少一个的估计或假定重量来确定。不同的力冲击可能导致不同的声音,并且可以基于力提供和选择不同的音频对象。

[0153] 虚拟场景对象与现实世界场景对象之间的碰撞的弹性。例如,当撞击地板时,被泵送到高压的球可能听起来与部分放气的球不同。球的弹性可能不同并且导致非常不同的声音。在一些实施例中,具体地,可以为虚拟场景对象的不同弹性提供不同的音频对象,并且元数据可以包括允许基于这种弹性来区分音频对象的数据。

[0154] 相互作用的持续时间:例如,装置可以包括物理引擎或与物理引擎通信,所述物理引擎预测在所检测到的与现实世界对象的相互作用之后的虚拟场景对象的更新轨迹。由于相互作用的动量、惯性、方向或弹性或相互作用中涉及的虚拟和现实世界材料,预测轨迹可以产生相互作用的持续时间。相互作用的长持续时间可以导致音频对象具有长或标称持续时间,而更短持续时间可以导致音频对象具有更短持续时间或标称音频对象的更短渲染。

[0155] 在其他示例中,相互作用可以由用户与虚拟内容相互作用来驱动,诸如将虚拟场景对象拖动通过现实世界环境。只要用户具有接近现实世界对象或与现实世界对象(虚拟)接触的虚拟场景对象,相互作用就可以持续。

[0156] 虚拟场景对象相对于现实世界场景对象的移动方向。例如,在沿着现实世界木质表面触碰的同时滑动虚拟对象通常将产生比在平行于表面法线的方向上局部撞击现实世界木质表面时更高音调的声音。

[0157] 在许多实施例中,元数据可以以查找表的形式提供,该查找表可以是多维的,这取

决于在选择音频对象时可以/应当考虑多少性质。例如,可以提供具有以下维度的查找表:

[0158] • 维度1:虚拟材料

[0159] • 维度2:检测到的现实世界材料

[0160] • 维度3:相互作用的类型(例如冲击/滑动/弹跳/附接/分离)

[0161] • 维度4:相互作用的速度

[0162] 在一些实施例中,可以提供基本上将两种材料类型联系到每个声音效果的元数据表。然而,对于相同的材料甚至相互作用类型(材料1、材料2和相互作用类型)可能存在许多声音效果。基于该信息,它们都可以适合于表示某个相互作用。一种方法是总是选择第一种,但是如果随机选择声音效果,则体验可能更引人注目。对于重复性相互作用,音频的小差异使它听起来不那么人为。声音剪辑之间的选择甚至可以基于库的元数据中未涵盖的相互作用的属性(例如,球的速度、或冲击的方向、或第n次弹跳)而变化。

[0163] 图2的装置的具体方法由图4的流程图表示。在该示例中,分析当前视频图像I以预测可能发生相互作用的图像区域。该分析可以基于虚拟场景对象/图形对象G的3D位置信度息。来自所得到的空间图像区域R的其中虚拟场景对象可能与真实环境的真实世界对象碰撞的图像部分然后被馈送到确定材料类别M的材料分类器(估计器207)。最后,根据材料类别M和虚拟场景对象/图形对象G产生声音。

[0164] 仅基于视觉信息的表面材料分类对于许多实际实施方式和用途中的应用可能是足够的。此外,如果没有识别出匹配的类目或声音,则可以使用虚拟场景对象G的默认相互作用声音,而不是特定于现实世界对象的材料的相互作用声音。可选地,当前和最近过去的环境声音信号S_{env}可以被馈送到材料分类器中。

[0165] 该方法可以例如允许其中一旦发生虚拟场景对象与现实世界场景对象之间的碰撞就开始播放预先记录的声音的应用。代替为每个虚拟对象指定与所有可能的现实世界材料的相互作用声音,每个虚拟对象可以被给予材料性质,并且单个查找表可以用作选择适当声音的基础。当材料分类器对于任何预定义的类别没有产生足够高的概率时,“未知”类别可以用于仍然产生近似正确的声音。例如,当撞击不同的现实世界表面时,虚拟橡胶球将产生差不多相同的声音。

[0166] AR标准可以提供具有用于虚拟元素的音频、视觉和元数据的比特流的语法,以增强特定的物理区域。该标准可以允许用户与物理环境中的这些虚拟元素相互作用。为了具有这些相互作用的音频增强,该标准可以提供用于将许多声音效果剪辑发送到解码器的手段,其中除非它们被触发,否则这些声音效果剪辑不被渲染给用户。

[0167] 所提供的比特流可以包括指示多维查找表的元数据,该元数据将每个声音效果剪辑分配给该查找表中的条目。表的每个维度可以与虚拟元素和物理元素之间的相互作用的方面相对应。当相互作用具有存在于该查找表中的联合标识某个表条目的方面(例如,通过分类和/或逻辑算法来检测到)时,可以在物理元素和虚拟元素之间的相互作用的位置处播放与该表条目相关联的声音效果剪辑。

[0168] 应当理解,为了清楚起见,以上描述已经参考不同的功能电路、单元和处理器描述了本发明的实施例。然而,显而易见的是,可以在不背离本发明的情况下使用不同功能电路、单元或处理器之间的任何合适的功能分布。例如,被示为由分别的处理器或控制器执行的功能可以由相同的处理器执行。因此,对特定功能单元或电路的引用仅被视为对用于提

供所描述的功能的合适设备的引用,而不是指示严格的逻辑或物理结构或组织。

[0169] 本发明可以以任何合适的形式实现,包括硬件、软件、固件或这些的任何组合。本发明可以任选地至少部分地实现为在一个或多个数据处理器和/或数字信号处理器上运行的计算机软件。本发明的实施例的元件和部件可以以任何合适的方式来物理地、功能地和逻辑地实现。实际上,功能可以在单个单元中、在多个单元中或作为其他功能单元的一部分来实现。这样,本发明可以在单个单元中实现,或者可以在不同的单元、电路和处理器之间物理地和功能地分布。

[0170] 尽管已经结合一些实施例描述了本发明,但是并不旨在将本发明限于这里阐述的特定形式。相反,本发明的范围仅受所附权利要求的限制。另外,尽管可能看起来结合特定实施例描述了特征,但是本领域技术人员将认识到,可以根据本发明组合所描述的实施例的各种特征。在权利要求中,术语“包括”不排除存在其他元件或步骤。

[0171] 此外,尽管单独列出,但是多个设备、元件、电路或方法步骤可以通过例如单个电路、单元或处理器来实现。另外,尽管各个特征可以包括在不同的权利要求中,但是这些特征可以有利地组合,并且包含在不同的权利要求中并不意味着特征的组合是不可行和/或不利的。在一类权利要求中包含特征并不意味着对该类别的限制,而是指示该特征在合适时同样适用于其他权利要求类别。此外,权利要求中的特征的顺序并不意味着特征必须工作的任何特定顺序,并且特别地,方法权利要求中的各个步骤的顺序并不意味着必须以该顺序执行这些步骤。而是,可以以任何合适的顺序来执行这些步骤。另外,单数引用不排除多个。因此,对“一”、“一个”、“第一”、“第二”等的引用不排除多个。权利要求中的附图标记仅被提供用于地使示例清楚,不应被解释为以任何方式限制权利要求的范围。

[0172] 可以提供一种包括指令的计算机程序产品,当所述程序由计算机执行时,所述指令引起所述计算机执行根据权利要求13所述的方法[的步骤]。

[0173] 实施例:

[0174] 1、一种用于生成输出音频信号的装置,所述装置包括:

[0175] 第一接收器(201),其被布置为从图像传感器接收现实世界场景的实时图像序列,所述实时图像序列包括图像帧的序列,每个图像帧包括视觉图像数据和深度图像数据中的至少一个;

[0176] 第二接收器(209),其被布置为接收一组音频对象和针对所述一组音频对象中的音频对象的元数据,所述元数据指示所述一组音频对象中的音频对象与材料特性之间的联系;

[0177] 图像生成器(203),其被布置为生成输出图像序列,所述输出图像序列包括与所述现实世界场景中的虚拟场景对象相对应的图像对象;

[0178] 检测器(205),其被布置为检测所述虚拟场景对象与所述现实世界场景对象之间的相互作用;

[0179] 估计器(207),其被布置为响应于所述图像帧的序列的图像数据而确定所述现实世界场景对象的材料性质;

[0180] 选择器(211),其被布置为响应于联系到所述一组音频对象中的音频对象的所述材料特性和所述材料性质而选择所述一组音频对象中的第一音频对象;以及

[0181] 输出电路(213),其被布置为生成包括所述第一音频对象的所述输出音频信号。

[0182] 2、根据权利要求1所述的装置,其中,所述估计器(207)被布置为

[0183] 确定所述图像帧的序列的至少一个图像帧中的相互作用图像区域,所述相互作用图像区域是所述至少一个图像帧的发生所述相互作用的图像区域;以及

[0184] 响应于所述相互作用图像区域的图像数据而确定所述场景对象的所述材料性质。

[0185] 3、根据权利要求1或2所述的装置,其中,所述第二接收器(209)被布置为从远程服务器接收所述元数据。

[0186] 4、根据任一前述权利要求所述的装置,其中,针对至少一些音频对象的所述元数据包括对所述至少一些音频对象与现实世界场景对象的材料特性之间的联系以及所述至少一些音频对象与虚拟场景对象的材料特性之间的联系的指示;并且其中,所述选择器(211)被布置为响应于联系到所述一组音频对象的现实世界对象的材料特性和材料性质并且响应于联系到所述一组音频对象的虚拟场景对象的材料特性和虚拟场景对象的材料性质而选择所述第一音频对象。

[0187] 5、根据任一前述权利要求所述的装置,其中,所述选择器(211)被布置为响应于所述虚拟场景对象的动态性质而选择所述第一音频对象。

[0188] 6、根据任一前述权利要求所述的装置,其中,所述检测器(205)被布置为确定所述相互作用的性质,并且所述选择器(211)被布置为响应于所述相互作用的所述性质而选择所述第一音频对象。

[0189] 7、根据权利要求6所述的装置,其中,所述相互作用的所述性质是从包括以下各项的组中选择的至少一个性质:

[0190] 所述相互作用的速度;

[0191] 所述虚拟场景对象与所述现实世界场景对象之间的碰撞的力;

[0192] 所述虚拟场景对象与所述现实世界场景对象之间的碰撞的弹性;

[0193] 所述相互作用的持续时间;以及

[0194] 所述虚拟场景对象相对于所述现实世界场景对象的移动的方向。

[0195] 8、根据任一前述权利要求所述的装置,其中,所述选择器(211)被布置为响应于所述虚拟对象相对于所述现实世界场景对象的取向而选择所述第一音频对象。

[0196] 9、根据任一前述权利要求所述的装置,其中,所述估计器(207)被布置为确定所述现实世界场景对象与多个对象类目中的至少第一类目的匹配指示;以及响应于所述匹配指示和联系到所述对象类目的材料性质而确定材料性质。

[0197] 10、根据权利要求9所述的装置,还包括用于接收在所述现实世界场景中捕获的实时音频的音频信号的音频接收器(215),并且其中,所述估计器被布置为响应于所述音频信号而确定所述匹配指示。

[0198] 11、根据任一前述权利要求所述的装置,其中,所述选择器(211)被布置为如果没有检测到满足选择准则的音频对象,则选择所述第一音频对象作为默认音频对象。

[0199] 12、根据任一前述权利要求所述的装置,其中,至少一个图像帧包括深度图像数据,并且其中,所述估计器(207)被布置为响应于检测到表示所述现实世界场景对象的所述至少一个图像帧的图像区域的至少一部分具有不超过阈值的深度图像数据的置信度水平而确定所述现实世界场景对象的所述材料性质。

[0200] 13、一种生成输出音频信号的方法,所述方法包括:

[0201] 从图像传感器接收现实世界场景的实时图像序列,所述实时图像序列包括图像帧的序列,每个图像帧包括视觉图像数据和深度图像数据中的至少一个;

[0202] 接收一组音频对象和针对所述一组音频对象中的音频对象的元数据,所述元数据指示所述一组音频对象中的音频对象与材料特性之间的联系;

[0203] 生成输出图像序列,所述输出图像序列包括与所述现实世界场景中的虚拟场景对象相对应的图像对象;

[0204] 检测所述虚拟场景对象与所述现实世界场景对象之间的相互作用;

[0205] 响应于所述图像帧的序列的图像数据而确定所述现实世界场景对象的材料性质;

[0206] 响应于联系到所述一组音频对象中的音频对象的所述材料特性和所述材料性质而选择所述一组音频对象中的第一音频对象;以及

[0207] 生成包括所述第一音频对象的所述输出音频信号。

[0208] 14、一种计算机程序产品,包括计算机程序代码单元,当所述程序在计算机上运行时,所述计算机程序代码单元适于执行根据权利要求13所述的所有步骤。

[0209] 15、一种音频信号,包括一组音频对象和用于所述音频对象的元数据,所述元数据指示音频对象与现实世界对象的材料特性之间的联系。

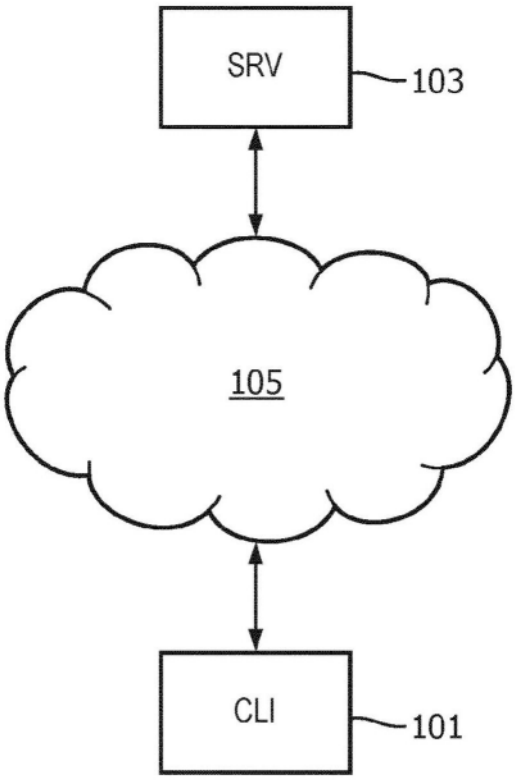


图1

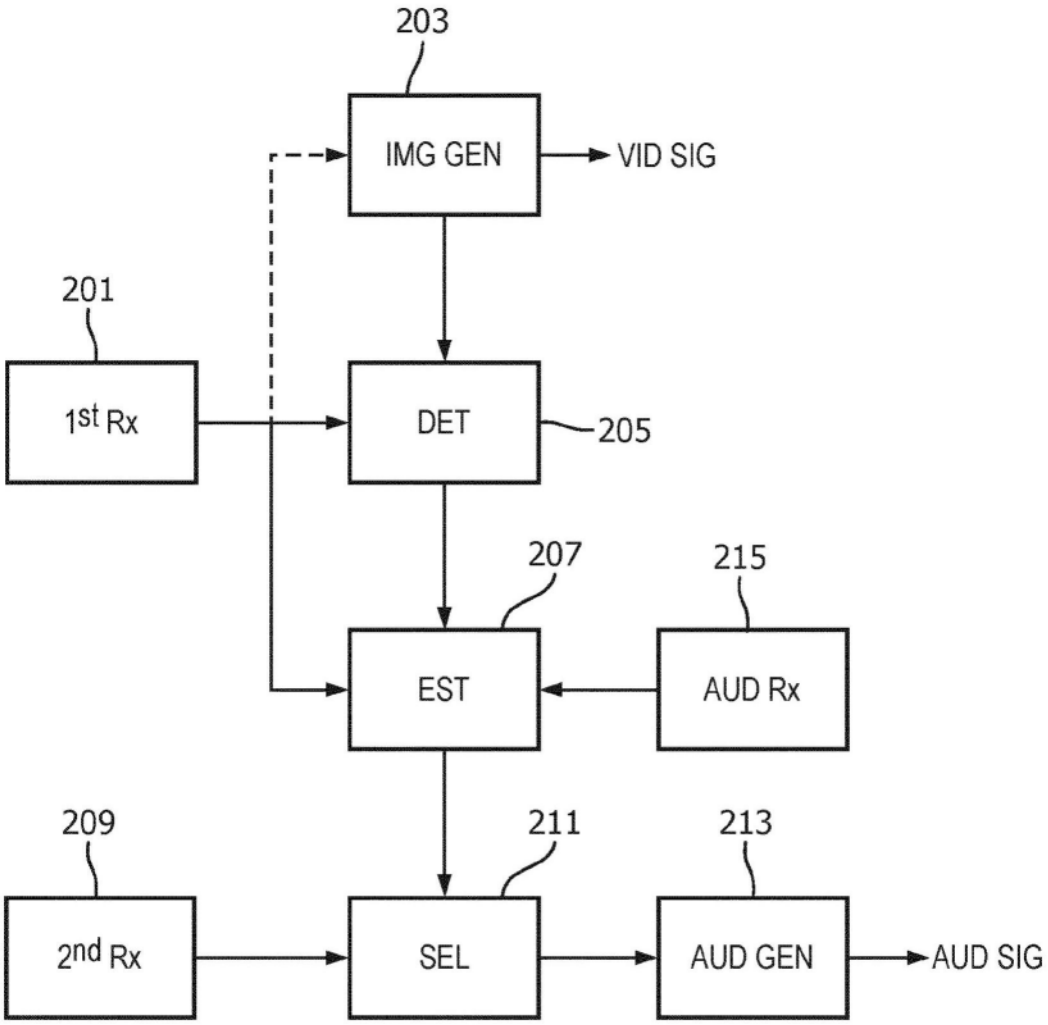

203
IMG GEN
VID SIG
201
1st Rx
DET
205
207
215
EST
AUD Rx
209
211
213
2nd Rx
SEL
AUD GEN
AUD SIG

图2

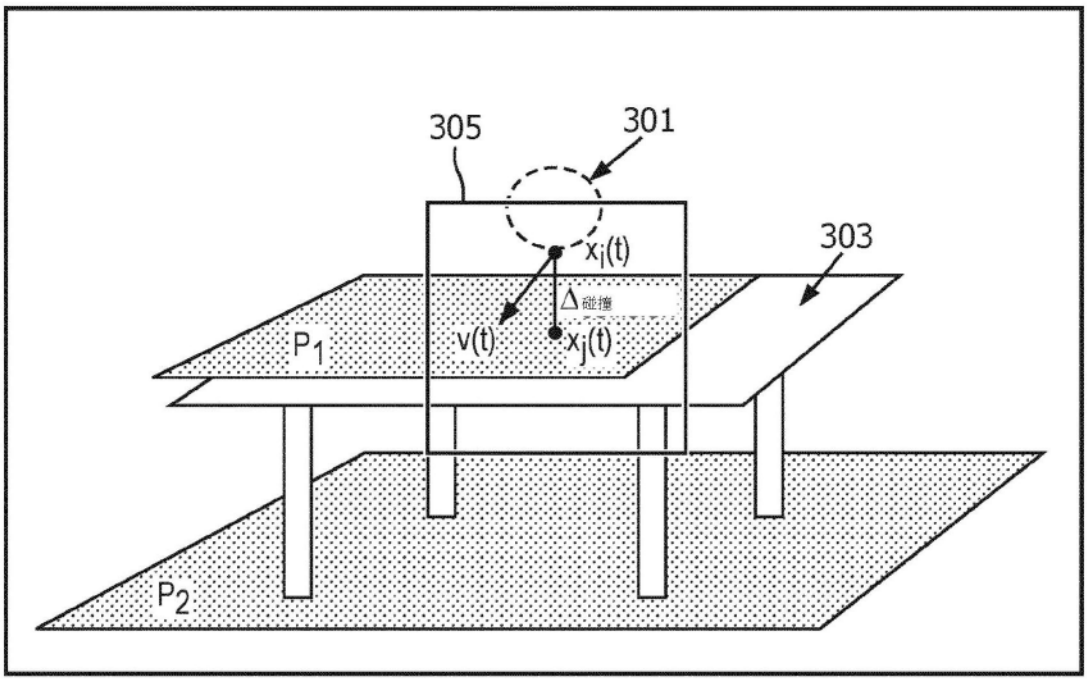

305
301
303
$x_i(t)$
$\Delta_{碰撞}$
v(t)
$x_j(t)$
P_1
P_2

图3

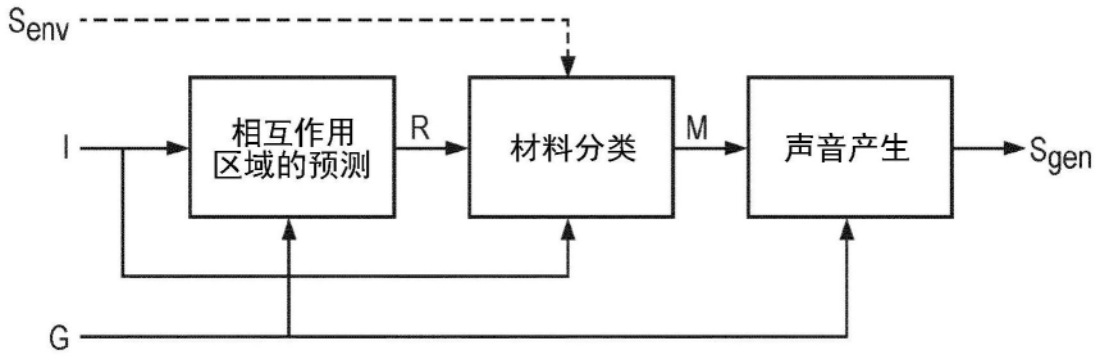

S_{env}
I
相互作用
区域的预测
R
材料分类
M
声音产生
S_{gen}
G

图4