



(12) 发明专利

(10) 授权公告号 CN 110310175 B

(45) 授权公告日 2020.12.22

(21) 申请号 201910558391.1
(22) 申请日 2019.06.25
(65) 同一申请的已公布的文献号
申请公布号 CN 110310175 A

(43) 申请公布日 2019.10.08

(30) 优先权数据
16/020,920 2018.06.27 US

(73) 专利权人 北京京东尚科信息技术有限公司
地址 100086 北京市海淀区知春路76号8层
专利权人 京东美国科技公司

(72) 发明人 吴宇章 徐毅 苗丹 周辉

(74) 专利代理机构 中科专利商标代理有限责任
公司 11021
代理人 纪雯

(51) Int.Cl.
G06Q 30/06 (2012.01)
G06T 19/00 (2011.01)
G06T 19/20 (2011.01)
G06T 17/05 (2011.01)

(56) 对比文件
US 2015193982 A1, 2015.07.09
CN 107977834 A, 2018.05.01
CN 108009871 A, 2018.05.08
CN 107665508 A, 2018.02.06
审查员 张千

权利要求书3页 说明书22页 附图10页

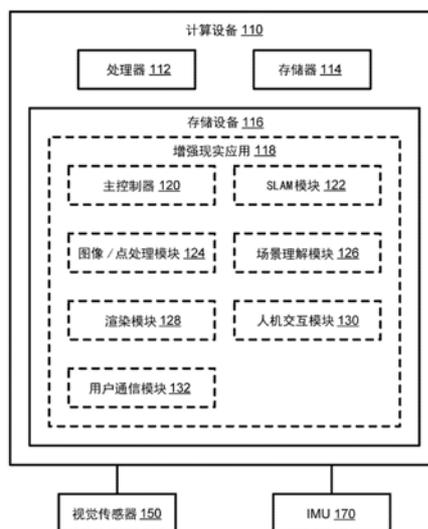
(54) 发明名称

用于移动增强现实的系统和方法

(57) 摘要

一种用于实现增强现实的系统和方法。该系统包括被配置为捕获环境的图像的视觉传感器和与视觉传感器通信的计算设备。计算设备具有处理器和存储计算机可执行代码的存储设备。当在处理器处被执行时,计算机可执行代码被配置为:识别所捕获图像中的二维(2D)对象;使用所捕获图像来构建所述环境的三维(3D)地图;通过将所捕获图像中识别的所述2D对象映射到所述3D地图中的对应点,来在所述3D地图中定义3D对象;以及基于3D地图中定义的3D对象在3D地图上渲染3D模型。

100



1. 一种用于实现增强现实的系统,包括:
 - 视觉传感器,被配置为捕获环境的图像;以及
 - 计算设备,与所述视觉传感器通信,其中所述计算设备包括处理器和存储计算机可执行代码的存储设备,并且所述计算机可执行代码当在所述处理器处被执行时被配置为:
 - 识别所捕获图像中的二维2D对象;
 - 使用所捕获图像来构建所述环境的三维3D地图;
 - 通过将所捕获图像中识别的所述2D对象映射到所述3D地图中的对应特征点,来在所述3D地图中定义3D对象;
 - 测量所述2D对象的尺寸并识别所述2D对象的功能和风格;
 - 将所述2D对象投影到所述3D地图以定义与所述2D对象相对应的3D对象在3D地图中的尺寸、功能和风格;
 - 基于所述3D对象的尺寸、功能和风格来推荐在所述环境中不存在的产品;以及
 - 基于所述3D地图中定义的所述3D对象将所述产品的3D模型放置到所述3D地图并进行渲染,
 - 其中所述计算机可执行代码还被配置为通过以下操作来推荐产品:
 - 基于所述环境的类别来确定多个产品;
 - 基于所述3D对象的风格,从所述多个产品中选择第一子集的产品;
 - 从所述第一子集的产品中选择功能与所述3D对象的功能互补的第二子集的产品;以及
 - 从所述第二子集的产品中确定至少一个产品,其中所述至少一个产品具有适配所述3D对象周围的空置空间的大小。
2. 根据权利要求1所述的系统,其中,所述计算机可执行代码还被配置为通过确定3D对象的平面来定义所述3D对象。
3. 根据权利要求2所述的系统,其中,所述平面包括水平平面、垂直平面和具有不同定向的平面。
4. 根据权利要求2所述的系统,其中,所述计算机可执行代码还被配置为:基于所述3D对象的平面,检测所述3D模型与所述3D地图中定义的所述3D对象之间的碰撞。
5. 根据权利要求1所述的系统,其中,所述计算机可执行代码还被配置为:基于所述3D对象的尺寸、功能和风格将所述3D模型放置在所述3D地图中的位置。
6. 根据权利要求1所述的系统,其中,所述计算机可执行代码被配置为:
 - 使用图像/点处理模块来识别所述2D对象;以及
 - 使用渲染模块在所述3D地图上渲染所述3D模型,
 - 其中,所述图像/点处理模块和所述渲染模块二者是使用C++编码的。
7. 根据权利要求1所述的系统,其中,所述计算机可执行代码被配置为通过以下操作来在所述3D地图中定义3D对象:
 - 确定所捕获图像中与所述2D对象中的一个2D对象相对应的多个2D特征点;
 - 确定所述3D地图中与所确定的2D特征点相对应的多个3D特征点;以及
 - 将对象标签添加到所确定的3D特征点以定义与所述2D对象相对应的3D对象。
8. 根据权利要求1所述的系统,其中,所述计算机可执行代码还被配置为:响应于用户的姿势或语音,在所述3D地图上操作所述3D模型。

9. 根据权利要求1所述的系统,其中,所述计算机可执行代码还被配置为:测量尺寸;以及使用机器学习来识别所述2D对象的功能和风格。

10. 一种用于使用视觉传感器和与所述视觉传感器通信的计算设备来实现增强现实的方法,所述方法包括:

由所述视觉传感器捕获环境的图像;

由所述计算设备识别所捕获图像中的二维2D对象;

由所述计算设备使用所捕获图像构建所述环境的三维3D地图;

由所述计算设备通过将所捕获图像中识别的所述2D对象映射到所述3D地图中的对应特征点,来在所述3D地图中定义3D对象;

测量所述2D对象的尺寸并识别所述2D对象的功能和风格;

将所述2D对象投影到所述3D地图以定义与所述2D对象相对应的3D对象在3D地图中的尺寸、功能和风格;

基于所述3D对象的尺寸、功能和风格来推荐在所述环境中不存在的产品;以及

由所述计算设备基于所述3D地图中定义的所述3D对象将所述产品的3D模型放置到所述3D地图并进行渲染,

其中推荐产品的步骤包括:

基于所述环境的类别来确定多个产品;

基于所述3D对象的风格,从所述多个产品中选择第一子集的产品;

从所述第一子集的产品中选择功能与所述3D对象的功能互补的第二子集的产品;以及

从所述第二子集的产品中确定至少一个产品,其中所述至少一个产品具有适配所述3D对象周围的空置空间的大小。

11. 根据权利要求10所述的方法,还包括:通过确定所述3D对象的平面来定义所述3D地图中的所述3D对象。

12. 根据权利要求11所述的方法,其中,所述平面包括水平平面、垂直平面和具有不同定向的平面。

13. 根据权利要求11所述的方法,还包括基于所述平面检测碰撞。

14. 根据权利要求10所述的方法,还包括:基于所述3D对象的尺寸、功能和风格将所述3D模型放置在所述3D地图中的位置。

15. 根据权利要求10所述的方法,还包括:测量尺寸;以及使用机器学习来识别所述2D对象的功能和风格。

16. 一种存储计算机可执行代码的非暂时性计算机可读介质,其中,所述计算机可执行代码当在计算设备的处理器处被执行时被配置为:

接收视觉传感器捕获的图像;

识别所捕获图像中的二维2D对象;

使用所捕获图像来构建环境的三维3D地图;

通过将所捕获图像中识别的所述2D对象映射到所述3D地图中的对应特征点,来在所述3D地图中定义3D对象;

测量所述2D对象的尺寸并识别所述2D对象的功能和风格;

将所述2D对象投影到所述3D地图以定义与所述2D对象相对应的3D对象在3D地图中的

尺寸、功能和风格；

基于所述3D对象的尺寸、功能和风格来推荐在所述环境中不存在的产品；以及
基于所述3D地图中定义的所述3D对象将所述产品的3D模型放置到所述3D地图并进行渲染，

其中所述计算机可执行代码还被配置为通过以下操作来推荐产品：

基于所述环境的类别来确定多个产品；

基于所述3D对象的风格，从所述多个产品中选择第一子集的产品；

从所述第一子集的产品中选择功能与所述3D对象的功能互补的第二子集的产品；以及

从所述第二子集的产品中确定至少一个产品，其中所述至少一个产品具有适配所述3D对象周围的空置空间的大小。

17. 根据权利要求16所述的非暂时性计算机可读介质，其中，所述计算机可执行代码还被配置为：

通过确定所述3D对象的水平平面和垂直平面来定义3D对象；以及

基于所述3D对象的平面来检测碰撞。

用于移动增强现实的系统和方法

[0001] 相关申请的相交引用

[0002] 在本发明的描述中引用并讨论了一些参考文献,这些参考文献可以包括专利、专利申请和各种出版物。提供对这些参考文献的引用和/或讨论仅仅是为了阐明本发明的描述,而不是承认任何这种参考文献是本文所述发明的“现有技术”。本说明书中引用和讨论的所有参考文献均通过引用整体并入本文,如同每篇参考文献单独地通过引用并入本文。

技术领域

[0003] 本发明总体上涉及增强现实,更具体而言,涉及用于推荐虚拟对象并将虚拟对象放置到现实环境中的跨平台系统和方法。

背景技术

[0004] 本文提供的背景描述是为了总体上呈现本发明的上下文。既不明确地承认也不暗示地承认当前列名的发明人的在背景技术部分中描述的范围内的、以及在提交时可能不是作为现有技术的描述的各方面为与本发明相对的现有技术。

[0005] 增强现实(AR)是物理的真实世界环境的直接或间接实时视图,其元素通过理想地跨多种感觉形态(包括视觉、听觉、触觉、体觉和嗅觉)的计算机生成的感知信息“增强”。用户可以使用AR来将虚拟对象与现实环境组合,并且产生虚拟对象是现实世界的一部分的感觉。随着移动设备和可穿戴设备的广泛采用,消费者可以轻松地访问AR以进行游戏、购物、教育、观光等。然而,困难的是将虚拟对象实时平滑地适配在现实环境中。

[0006] 因此,本领域存在未解决的需求以解决上述缺陷和不足。

发明内容

[0007] 在某些方面,本发明涉及一种用于实现增强现实的系统。在某些实施例中,所述系统包括:视觉传感器,被配置为捕获环境的图像;以及与所述视觉传感器通信的计算设备。所述计算设备具有处理器和存储计算机可执行代码的存储设备。当在所述处理器处被执行时,所述计算机可执行代码被配置为:识别所捕获图像中的二维(2D)对象;使用所捕获图像来构建所述环境的三维(3D)地图;通过将所捕获图像中识别的所述2D对象映射到所述3D地图中的对应特征点,来在所述3D地图中定义3D对象;以及基于所述3D地图中定义的所述3D对象将3D模型放置到所述3D地图并进行渲染。

[0008] 在某些实施例中,所述计算机可执行代码还被配置为通过确定所述3D对象的水平平面和垂直平面来定义3D对象。在某些实施例中,所述计算机可执行代码还被配置为:基于所述3D对象的平面,检测虚拟对象与真实物理对象之间,即所述3D模型与所述3D对象之间的碰撞。在某些实施例中,所述平面包括所述3D地图中的水平平面、垂直平面和/或具有任意定向的平面。

[0009] 在某些实施例中,所述计算机可执行代码还被配置为:测量所述对象的尺寸。在某些实施例中,所述计算机可执行代码还被配置为:将从所述图像识别的所述2D对象投影到

所述3D地图以测量与所述2D对象相对应的3D对象的尺寸,并基于真实3D对象的尺寸来推荐3D虚拟模型。在某些实施例中,所述计算机可执行代码还被配置为:基于真实对象的尺寸将一个或多个3D模型放置在所述3D地图中的位置。

[0010] 在某些实施例中,所述计算机可执行代码还被配置为:使用机器学习来识别所述2D对象的功能和风格。在某些实施例中,所述计算机可执行代码还被配置为:基于所述对象的功能和风格来推荐3D虚拟模型。在某些实施例中,所述计算机可执行代码还被配置为:基于真实对象的功能和风格将所述3D模型放置在所述3D地图中的位置。

[0011] 在某些实施例中,所述计算机可执行代码被配置为识别场景的分类,例如卧室、起居室或浴室。在某些实施例中,所述计算机可执行代码被配置为识别所述2D图像中的所述2D对象的角色或名称。所述2D对象的识别名称包括餐桌、书桌、沙发等。

[0012] 在某些实施例中,所述计算机可执行代码被配置为:使用图像/点处理模块识别所述2D对象;并使用渲染模块来在所述3D地图上渲染所述3D模型。所述图像/点处理模块和所述渲染模块都使用C++被编码。

[0013] 在某些实施例中,所述计算机可执行代码被配置为通过以下操作在所述3D地图中定义3D对象:确定所捕获图像中与所述2D对象中的一个2D对象相对应的多个2D特征点;确定所述3D地图中与所确定的2D特征点相对应的多个3D特征点;以及将对象标签添加到所确定的3D特征点以定义与所述2D对象相对应的3D对象。

[0014] 在某些实施例中,所述计算机可执行代码还被配置为:响应于用户的姿势或语音,在所述3D地图上操作所述3D模型;以及定义所述3D地图与另一计算设备中的另一3D地图的坐标关系。

[0015] 在某些方面,本发明涉及一种使用视觉传感器和与所述视觉传感器通信的计算设备实现增强现实的方法。所述方法包括:所述视觉传感器捕获环境的图像;所述计算设备识别所捕获图像中的二维(2D)对象;所述计算设备使用所捕获图像构建所述环境的三维(3D)地图;所述计算设备通过将所捕获图像中识别的所述2D对象映射到所述3D地图中的对应特征点,来在所述3D地图中定义3D对象;以及所述计算设备基于所述3D地图中定义的所述3D对象将3D模型放置到所述3D地图并进行渲染。

[0016] 在某些实施例中,所述方法还包括:通过确定所述3D对象的平面来定义所述3D地图中的所述3D对象。在某些实施例中,所述方法还包括:基于所述平面检测虚拟对象(产品的3D模型)与真实对象(3D地图中的3D对象)之间的碰撞。在某些实施例中,所述平面包括所述3D地图中的水平平面、垂直平面和/或具有任意定向的平面。

[0017] 在某些实施例中,所述方法还包括测量所述3D对象的尺寸。在某些实施例中,所述方法还包括:将所述2D对象投影到所述3D地图以测量与所述2D对象相对应的3D对象的尺寸,并基于所述3D对象的尺寸推荐所述3D模型。在某些实施例中,所述方法还包括基于所述3D对象的尺寸将所述3D模型放置在所述3D地图中的位置。

[0018] 在某些实施例中,所述方法还包括使用机器学习识别3D对象的功能和风格。在某些实施例中,所述方法还包括基于所述3D对象的功能和风格将所述3D模型放置在所述3D地图中的位置。

[0019] 在某些方面,本发明涉及一种存储计算机可执行代码的非暂时性计算机可读介质。当在计算设备的处理器处被执行时,所述计算机可执行代码被配置为:接收视觉传感器

捕获的图像;识别所捕获图像中的二维(2D)对象;使用所捕获图像来构建所述环境的三维(3D)地图;通过将所捕获图像中识别的所述2D对象映射到所述3D地图中的对应特征点,来在所述3D地图中定义3D对象;以及基于所述3D地图中定义的所述3D对象将一个或多个3D模型放置到所述3D地图并进行渲染。

[0020] 在某些实施例中,所述计算机可执行代码还被配置为:通过确定所述3D对象的平面来定义3D对象;并基于所述3D对象的平面来检测虚拟对象与真实对象(所述3D模型和所述3D对象)之间的碰撞。在某些实施例中,所述平面包括所述3D地图中的水平平面、垂直平面和/或具有任意定向的平面。

[0021] 在某些实施例中,所述计算机可执行代码还被配置为:测量尺寸;将所述2D对象投影到所述3D地图以测量与所述2D对象相对应的3D对象的尺寸;基于所述3D对象的尺寸来推荐所述3D模型;以及基于所述3D对象的尺寸将所述3D模型放置在所述3D地图中的位置。

[0022] 在某些实施例中,所述计算机可执行代码还被配置为:使用机器学习来识别所述2D对象的功能和风格,并将识别的功能和风格投影到所述3D对象;基于真实3D对象的功能和风格来推荐3D模型;以及基于所述3D对象的功能和风格将3D模型放置在所述3D地图中的位置并进行渲染。

[0023] 根据以下结合附图对优选实施例的详细描述,本发明的这些和其他方面将变得更加清楚,尽管可以在不脱离本发明新颖构思的精神和范围的情况下对本文进行多种改变和修改。

附图说明

[0024] 附图示出了本发明的一个或多个实施例,并与书面描述一起用于解释本发明的原理。可能的话,贯穿附图,相同的附图标记用于表示实施例的相同或类似的元件。

[0025] 图1示意性地描绘了根据本发明某些实施例的示例性计算设备。

[0026] 图2A示意性地描绘了根据本发明某些实施例的示例性同时定位和映射(SLAM)模块。

[0027] 图2B示意性地描绘了根据本发明某些实施例的示例性图像/点处理模块。

[0028] 图2C示意性地描绘了根据本发明某些实施例的示例性场景理解模块。

[0029] 图2D示意性地描绘了根据本发明某些实施例的示例性渲染模块。

[0030] 图2E示意性地描绘了根据本发明某些实施例的示例性人机交互模块。

[0031] 图2F示意性地描绘了根据本发明某些实施例的示例性用户通信模块。

[0032] 图3描绘了根据本发明某些实施例的将三维(3D)产品模型渲染到3D环境地图的流程图。

[0033] 图4描绘了根据本发明某些实施例的SLAM的流程图。

[0034] 图5A描绘了根据本发明某些实施例的图像/点处理的流程图。

[0035] 图5B描绘了根据本发明某些实施例的3D地图处理的流程图。

[0036] 图6A描绘了根据本发明某些实施例的训练场景理解模块的流程图。

[0037] 图6B描绘了根据本发明某些实施例的执行场景理解的流程图。

[0038] 图7描绘了根据本发明某些实施例的执行渲染的流程图。

[0039] 图8示意性地描绘了根据本发明某些实施例的多用户增强现实(AR)购物系统。

[0040] 图9描绘了根据本发明某些实施例的执行多用户AR购物的流程图。

具体实施方式

[0041] 在以下示例中更具体地描述了本发明,因为本文中的许多修改和变化对于本领域技术人员而言是显而易见的,这些示例仅用于说明。现在详细描述本发明的各种实施例。参考附图,在所有附图中相同的附图标记表示相同的组件。如本文的描述和所附权利要求中所使用的,“一”、“一个”和“所述”的含义包括复数指示物,除非上下文另有明确说明。此外,如本文的描述和所附权利要求中所使用的,“在...中”的含义包括“在.....中”和“在.....上”,除非上下文另有明确说明。此外,为了方便读者,可以在说明书中使用标题或副标题,这不会影响本发明的范围。另外,在下文中对本说明书中使用的一些术语进行了更具体地定义。

[0042] 本说明书中使用的术语在本发明的上下文中以及在每个术语所使用的特定上下文中通常具有其在本领域中的通常含义。用于描述本发明的某些术语在下面或在说明书中的其他地方进行了讨论,以向实施者提供关于本发明的描述的额外指导。应当理解,可以以不止一种方式对其进行描述。因此,替代语言和同义词可以用于本文所讨论的任何一个或多个术语,并且对于术语是否在本文中详细阐述或讨论也没有任何特殊意义。提供了某些术语的同义词。对一个或多个同义词的叙述不排除使用其他同义词。本说明书中任何地方的示例(包括本文所讨论的任何术语的示例)的使用仅是说明性的,并且决不限制本发明或任何示例性术语的范围和含义。同样,本发明不限于本说明书中给出的各种实施例。

[0043] 将理解,当提及一个元件在另一元件“上”时,该元件可以直接在该另一元件上,或者在它们之间可以存在中间元件。相反,当提及一个元件在另一元件的“直接上面”时,不存在中间元件。如本文所用,术语“和/或”包括关联列出的项目中的一个或多个项目的任意和所有组合。

[0044] 将理解,虽然本文中可以使用术语“第一”、“第二”、“第三”等来描述各种元件、组件、区域、层和/或部分,但是这些元件、组件、区域、层和/或部分不应受这些术语限制。这些术语仅用于将一个元件、组件、区域、层或部分与另一元件、组件、区域、层或部分分开。因此,以下讨论的第一元件、组件、区域、层或部分可被称为第二元件、组件、区域、层或部分,而不脱离本方面的教导。

[0045] 此外,在本文中可以使用诸如“下”或“底”和“上”或“顶”的相对术语来描述一个元件与图示另一元件的关系。应当理解,除了附图中所描绘的定向之外,相对术语旨在包括设备的不同定向。例如,如果一幅图中的设备翻转,则被描述为位于其它元件的“下”侧的元件将定向在所述其它元件的“上”侧。因此,示例性术语“下”可以包含“下”和“上”的定向,这取决于图的特定定向。类似地,如果一幅图中的设备翻转,则被描述为位于其它元件的“下方”或“下侧”的元件将定向在所述其它元件的“上方”。因此,“下方”或“下方”的示例性术语可以包括上下朝向。

[0046] 除非另外定义,否则这里使用的所有术语(包括技术和科学术语)具有本发明所属领域的普通技术人员通常所理解的含义。还将理解的是,诸如在常用词典中定义的术语应被解释为其含义与在相关技术和本公开的上下文中的含义相同,而不应将其解释为理想的或过于正式的含义,除非本文明确如此定义。

[0047] 如本文所用,“大约”、“约”、“基本上”或“近似”通常表示在给定值或范围的20%、优选10%、更优选5%之内。本文给出的数值是近似的,意味着如果没有明确说明,可以推断出术语“大约”、“约”、“基本上”或“近似”。

[0048] 如本文所用,“多个”意指两个或更多个。

[0049] 如本文所用,术语“包含”、“包括”、“携带”、“具有”、“含有”、“涉及”等应理解为开放式的,即意味着包括但不限于。

[0050] 如本文所用,短语A、B和C中的至少一个应该被解释为使用非排他逻辑或(OR)表示逻辑(A或B或C)。应当理解,方法中的一个或多个步骤可以以不同的顺序(或同时)执行而不改变本发明的原理。

[0051] 如本文所用,术语“模块”可以指代以下各项的一部分或包括以下各项:专用集成电路(ASIC);电子电路;组合逻辑电路;现场可编程门阵列(FPGA);执行代码的处理器(共享、专用或组);提供所述功能的其他合适的硬件组件;或者例如在片上系统中的上述部分或全部的组合。术语模块可以包括存储由处理器执行的代码的存储器(共享、专用或组)。

[0052] 本文使用的术语“代码”可以包括软件、固件和/或微代码,并且可以指代程序、例程、功能、类和/或对象。如上所使用的术语“共享”意味着可以使用单个(共享)处理器来执行来自多个模块的一些或所有代码。另外,来自多个模块的一些或所有代码可以由单个(共享)存储器存储。如上所使用的术语“组”意味着可以使用一组处理器来执行来自单个模块的一些或所有代码。另外,可以使用一组存储器来存储来自单个模块的一些或所有代码。

[0053] 本文使用的术语“接口”通常是指在用于在组件之间执行数据通信的组件之间的交互点处的通信工具或装置。通常,接口可以适用于硬件和软件两个级别,并且可以是单向接口或双向接口。物理硬件接口的示例可以包括电连接器、总线、端口、缆线、终端和其他I/O设备或组件。与接口通信的组件可以是例如计算机系统的多个组件或外围设备。

[0054] 本发明涉及计算机系统。如附图中所示,计算机组件可以包括物理硬件组件(其被示为实线框)以及虚拟软件组件(其被示为虚线框)。除非另有说明,否则本领域普通技术人员将理解,这些计算机组件可以以但不限于软件、固件或硬件组件或其组合的形式来实现。

[0055] 本文描述的装置、系统和方法可以由一个或多个处理器执行的一个或多个计算机程序来实现。计算机程序包括存储在非暂时性有形计算机可读介质上的处理器可执行指令。计算机程序还可以包括所存储的数据。非暂时性有形计算机可读介质的非限制性示例是非易失性存储器、磁存储装置和光存储装置。

[0056] 现在将在下文中参照附图更全面地描述本发明,在附图中,示出了本发明的实施例。然而,本发明可以按照多种不同形式来表现,并且不应当被看做是限制在本文中阐述的实施例;相反,提供这些实施例,使得本发明将会全面和完整,并且将会向本领域技术人员完全传达本发明的范围。

[0057] 根据本发明的目的,如本文所体现和宽泛描述的,在某些方面,本发明涉及增强现实系统和方法。在某些实施例中,该系统和方法包括:(1)在跟踪相机相对于环境的姿态期间,同时构建和更新环境的地图;(2)实时地在屏幕上对3D模型或动画进行3D渲染;(3)智能场景理解,允许用于自动且智能地将虚拟对象放置在环境中的全新体验。此外,本发明的某些实施例具有以下有益优点。

[0058] 首先,根据本发明某些实施例的系统和方法是跨平台的,可以在Apple®设备、

Android™设备、Windows®设备等上运行。在某些实施例中，该系统被配置用于移动AR并且易于在电子商务中使用。移动AR系统和平台可以帮助消费者在真实场景中虚拟地“试用”商品并考虑该商品是否适合该环境(例如，一件家具的尺寸和/或风格是否适合起居室)。

[0059] 其次，该系统包括SLAM模块、渲染模块、图像/点处理模块、场景理解模块、人机交互模块、以及可选的用户通信模块。这个完整的系统提供了AR所需的所有功能，尤其适用于AR购物。具体地，(1) SLAM模块使用手机的相机来构建和更新环境的地图，同时跟踪相机相对于环境的姿态。(2) 渲染模块或渲染引擎被配置为在物理环境内使虚拟对象可视化。它可以支持静态渲染和动态动画两者。(3) 图像/点处理模块被配置为从图像获得特征并将这些数据用于后续操作。(4) 场景理解模块被配置为处理获取的图像以及SLAM模块获取的点集，以了解场景的内容和3D几何结构。为了增强用户体验，在场景理解模块中使用了人工智能。通过智能对象检测和场景理解，可以智能地且自动地推荐虚拟对象并将其放置在物理世界中。(5) 交互模块被配置为使用姿势、语音识别或触摸屏交互，以允许用户利用简单的语音命令或友好的手势与虚拟对象交互。

[0060] 第三，通过智能对象检测或/和场景理解，本发明能够推荐与对象和环境的风格和尺寸/大小匹配的一种或多种产品，并将产品放置在环境中的合适位置。例如，该系统能够向现代风格的环境推荐现代风格的产品；根据检测到的TV架或TV桌的大小推荐合适大小的TV；并自动地在墙面上放置一幅绘画。

[0061] 图1示意性地描绘了根据本发明某些实施例的移动增强现实系统或平台。在某些实施例中，系统100包括计算设备110、一个或多个视觉传感器150、以及可选地一个或多个惯性测量单元(IMU) 170。视觉传感器150和IMU 170与计算设备110通信并且受计算设备110的控制。在某些实施例中，计算设备110可以是提供增强现实服务的移动设备、平板电脑、通用计算机、无头计算机、可穿戴设备、专用计算机、服务器计算机、集群或云计算机。在某些实施例中，计算设备110优选地是一个或多个移动设备，诸如智能电话、可穿戴设备或平板电脑。如图1所示，计算设备110可以包括但不限于处理器112、存储器114和存储设备116。在某些实施例中，计算设备110可以包括其他硬件组件和软件组件(未示出)以执行其相应的任务。这些硬件组件和软件组件的示例可以包括但不限于其他所需的存储器、接口、总线、输入/输出(I/O) 模块或设备、网络接口和外围设备。

[0062] 处理器112可以是中央处理单元(CPU)，其被配置为控制计算设备110的操作。处理器112可以执行计算设备110的操作系统(OS)或其他应用。在一些实施例中，计算设备110可以具有多于一个CPU作为处理器，例如两个CPU、四个CPU、八个CPU或任何合适数量的CPU。

[0063] 存储器114可以是易失性存储器，例如随机存取存储器(RAM)，用于存储在计算设备110的操作期间的数据和信息。在某些实施例中，存储器114可以是易失性存储器阵列。在某些实施例中，计算设备110可以在多于一个存储器114上运行。

[0064] 存储设备116是用于存储计算设备110的OS(未示出)和其他应用的非易失性数据存储介质。存储设备116的示例可以包括非易失性存储器，诸如闪存、存储卡、USB驱动器、硬盘驱动器、软盘、光盘驱动器或任何其他类型的数据存储设备。在某些实施例中，计算设备110可以具有多个存储设备116，其可以是相同的存储设备或不同类型的存储设备，并且计算设备110的应用可以被存储在计算设备110的一个或多个存储设备116中。如图1所示，存储设备116包括增强现实(AR)应用118。AR应用118提供跨平台系统，以提供增强现实，例如

用于在线购物。

[0065] 其中AR应用118包括主控制器120、同时定位和映射 (SLAM) 模块122、图像/点处理模块124、场景理解模块126、渲染模块128、人机交互模块130以及可选地用户通信模块132等。在某些实施例中,AR应用118可以包括模块120-132的操作所需的其他应用或模块。应当注意,模块均由计算机可执行代码或指令,或者数据表或数据库来实现,其中计算机可执行代码或指令,或者数据表或数据库共同形成一个应用。在某些实施例中,每个模块还可以包括子模块。备选地,一些模块可以组合为一个堆栈。在其他实施例中,某些模块可以实现为电路而不是可执行代码。在某些实施例中,AR应用118的一些模块(例如,场景理解模块126)可以位于远程计算设备(例如,服务器计算机)处,并且AR应用118的其他模块经由有线网络或无线网络与场景理解模块126进行通信。

[0066] 主控制器120被配置为接收(或检索)由视觉传感器150捕获的图像和由IMU 170测量的IMU信息,并且与SLAM模块122、图像/点处理模块124、场景理解模块126、渲染模块128、人机交互模块130和用户通信模块132进行通信。主控制器120将相关数据输入到这些模块,指示模块执行它们相应的功能,并接收或重定向从那些模块输出的数据。在某些实施例中,主控制器120对于AR应用118可以不是必需的,并且其他模块直接相互通信而无需通过主控制器120。

[0067] SLAM模块122被配置为在直接从视觉传感器150或经由主控制器120接收到所捕获图像时,计算视觉传感器150的姿态并构建环境的3D地图。通过视觉传感器150持续捕获新的2D图像,SLAM模块122也持续地更新3D地图。在某些实施例中,姿态和3D地图可以由主控制器120和/或其他相关模块访问,例如图像/点处理模块124和场景理解模块126。视觉传感器150可以是一个或多个相机。相机的姿态可以表示为三自由度平移和三自由度旋转,并且3D地图可以是被形成为特征点的点云的点图。每个特征点基于在2D图像中的对应特征点的特征而具有特定特征描述符。如图2A所示,SLAM模块122包括特征提取模块1220、定位和映射模块1222以及可选地光检测模块1224和融合模块1226中的至少一个。在某些实施例中,特征提取模块1220被配置为检测特征且从所捕获图像中提取特征点,并计算特征点的特征描述符。在某些实施例中,特征检测包括从图像中检测角部。角部是具有较高梯度变化的区域,并且在不同图像中容易识别。因此,角部是用于在2D图像中跟踪的优选类型的特征。定位和映射模块1222被配置为执行SLAM模块122的主要功能,即,确定相机的姿态并使用提取的特征点构建环境的3D地图。光检测模块1224被配置为与3D地图中的特征点相关地检测环境中的光。例如,可以基于在3D地图中检测到的面元来确定环境中的多个光源。在某些实施例中,系统100还包括IMU 170,并且融合模块1226被配置为将由IMU 170采集的惯性数据与由视觉传感器150捕获的视觉数据融合,以辅助SLAM过程。在某些实施例中,SLAM模块122可以由ARKit、ARCore或其他定制的SLAM模块提供。

[0068] 图像/点处理模块124被配置为从2D图像提取特征,并从点云检测和适配平面。可以直接从视觉传感器150或经由主控制器120接收2D图像,或者在某些实施例中,图像/点处理模块124使用的2D图像是SLAM模块122使用的相同2D图像。在某些实施例中,从SLAM模块122获得点云数据和相机姿态。在图像/点处理模块124利用平面信息对于3D地图更新之后,SLAM中的3D地图被命名为3D平面地图,其中平面信息被添加作为在3D地图中特征点的关联平面标签。SLAM模块122持续地更新3D平面地图,唯一的区别在于3D平面地图中的特征点不

仅包括特征描述符和光信息,还包括关联平面信息。如图2B所示,图像/点处理模块124包括点云平面适配模块1240、图像特征提取模块1242和图像分割模块1244。点云平面适配模块1240将平面信息添加到3D地图,而图像特征提取模块1242和图像分割模块1244从2D图像获得特征以及分割信息。

[0069] 点云平面适配模块1240被配置为适配水平平面和垂直平面两者以及可选地与水平平面和垂直平面形成不同角度的平面。具体地,点云平面适配模块1240检测属于诸如垂直平面之类的平面的多个3D点,并基于这些3D特征点的分布确定垂直平面的边界。然后,点云平面适配模块1240在3D地图中标记具有边界的垂直平面。在某些实施例中,点云平面适配模块1240可以使用参数化表示来定义平面。在一个示例中,点云平面适配模块1240使用平面的中心点、长度和宽度以及法线方向来定义平面。在一个示例中,点云平面适配模块1240使用平面上一个点和平面定向来定义平面。在一个示例中,用 $Ax+By+Cz+D=0$ 定义平面,其中 x 、 y 、 z 是属于平面的点的3D坐标,且四个参数 A 、 B 、 C 和 D 是常数。在某些实施例中,点云平面适配模块1240可以通过将一组特征点标记为属于该平面来使用平面的非参数化表示,以便定义平面。这些标签现在是3D地图的一部分。在结合水平平面、垂直平面以及沿其他方向的可能平面的信息之后,来自SLAM的3D地图被定义为3D平面地图。SLAM模块122和图像/点处理模块124两者持续地更新该3D平面地图。

[0070] 图像特征提取模块1242被配置为从2D图像提取角部以外的特征。在某些实施例中,与特征提取模块1220相比,图像特征提取模块1242被配置为从图像中提取更多的特征。特征提取模块1220被配置为从图像中检测诸如角部或其他有限数量的特征之类的特征,以便容易地跟踪图像中的特征以执行SLAM。相比之下,图像特征提取模块1242被配置为提取用于场景识别的更复杂的特征,诸如边、斑点、脊、线、圆、平面等。在某些实施例中,图像特征提取模块1242检测并提取通过机器学习方法(特别是深度学习)等学习的特征。这里,2D图像中的检测到的特征然后可以由场景理解模块126使用。当从2D图像中检测或提取角部和其他特征时,图像分割模块1244基于检测到的特征对图像进行分割。在某些实施例中,由图像特征提取模块1242提取的2D特征也可以被重新投影到3D地图。

[0071] 在某些实施例中,图像/点处理模块124使用C++编码,其中C++可以在Android™或iOS™下运行。因此,根据本发明某些实施例的AR应用118可以用作跨平台应用。

[0072] 场景理解模块126被配置为在接收到来自2D图像的特征和分割时,在2D图像中检测和识别对象,并将2D图像中的识别的对象投影到3D平面地图,以更新3D平面地图从而形成3D场景地图。所获得的3D场景地图仍然是SLAM的一部分,并且由SLAM模块122利用来自新2D图像的更新后的特征点来持续更新,由图像/点处理模块124利用更新后的平面信息来更新,并且由场景理解模块126利用对象信息来更新。然后,3D场景地图可以由渲染模块128使用。在某些实施例中,场景理解模块126被配置为基于3D场景地图向用户推荐一个或多个产品。

[0073] 如图2C所示,场景理解模块126包括离线训练模块1260、推理引擎1270和2D-3D投影模块1272。离线训练模块1260被配置为使用标记的图像数据集和可选的未标记的图像数据集来训练不同的机器学习模型。在某些实施例中,当预先训练模型时,推理引擎1270可以直接使用训练好的模型。换言之,场景理解126可以仅包括训练好的模型,而不包括执行训练过程的机制。在通过推理引擎1270运行模型之后,2D-3D投影模块1272被配置为将对从2D

图像获得的场景的理解投影到3D平面地图,以便更新3D平面地图从而形成3D场景地图。

[0074] 如图2C所示,离线训练模块1260包括图像识别模型1262、图像分割模型1264、图像分类模型1266和图像深度模型1268。在某些实施例中,标记的图像数据集包括多个2D图像。在某些实施例中,标记的2D图像的数量超过数千个或超过数百万个,并且每个2D图像具有关联类别的场景(例如,卧室、厨房等)。在某些实施例中,每个2D图像中的对象分别被边界框包围,并用对象的类别或名称(例如,桌子或椅子)标记。在某些实施例中,图像识别模型1262、图像分割模型1264、图像分类模型1266和图像深度模型1268中的每个都是机器学习机制或神经网络,诸如基于区域的卷积神经网络(R-CNN)。在离线训练之后,图像识别模块1262能够确定对象的位置和标签(名称),并且从输入图像将对象识别为桌子、椅子、人等;图像分割模块1264能够将输入图像分割成分割块,以便定位对象和对象的边界;图像分类模块1266能够将输入图像分类为不同的类别,例如厨房、起居室等;且图像深度模块1268被配置为估计对象的深度。可以离线训练那些机器学习机制,然后基于AR应用118的目的使用那些机器学习机制。在某些实施例中,基于项目的目的,AR应用118可以调用模型1262-1268中的一个或多个。此外,模型1262-1268可以独立地起作用,或者是以预定顺序执行的。例如,图像分类模块1266可以在图像识别模块1262之前被训练和使用,使得对诸如起居室的类的空间的分类的理解用于帮助识别空间中的可能对象。备选地,可以在分类模块1266之前训练和执行图像识别模块1262,使得空间中的对象的识别用于帮助识别具有对象的空间的可能分类。在某些实施例中,离线训练模块1260可以不包括所有模型1262-1268,只要可以训练可用模型并将它们用于从图像中识别对象即可。

[0075] 推理引擎1270可以是安装在计算设备110上的AR应用118的一部分,或者是计算设备110可访问的独立云模块。例如,训练的模型和推理引擎1270可以位于服务器计算机处,并且当多个2D图像作为输入从计算设备110被上载到服务器计算机时,推理引擎1270使用那些训练的模型来分析输入的2D图像以生成场景信息(例如图像中的对象的位置、边界和名称),并将信息发送回计算设备110,使得计算设备110中的AR应用118可以使用该信息来更新3D地图。

[0076] 推理引擎1270被配置为使用2D图像和图像特征来执行模型1262-1268中的一个或多个。2D图像可以是直接从SLAM模块122、从图像/点处理模块124、直接从视觉传感器150或经由主控制器120被输入的,并且图像特征是直接从图像/点处理模块124或经由主控制器120被输入的。在某些实施例中,2D图像和图像特征两者是直接从图像/点处理模块124输入的。在某些实施例中,SLAM模块122、图像/点处理模块124和场景理解模块126使用同一组2D图像。

[0077] 推理引擎1270被配置为将提取特征的输入2D图像运行通过模型1262-1268中的一个或多个,并产生推理结果。在某些实施例中,推理引擎1270使用TensorFlow®框架来实现。在某些实施例中,使用CoreML®来实现推理引擎1270以在Apple环境下执行推理。在某些实施例中,当推理引擎1270接收2D图像和2D特征时,推理引擎1270使用模型(在Android™环境下使用TensorFlow®或在Apple®环境下使用CoreML®)进行推理。在某些实施例中,2D图像和推理引擎1270的结果也可以用于重新训练模型1262-1268。推理引擎1270的结果是场景的类别、2D图像中的对象的标签以及可选地2D图像中的对象的定位(位

置)和边界。

[0078] 2D-3D投影模块1272被配置为在推理引擎1270获得场景的类别、对象的位置和边界框以及2D图像中的对象的标签之后,将这些特征投影到3D平面地图,以便将3D平面地图更新为3D场景地图。在SLAM期间,已经建立了2D图像中的2D特征点与3D地图中的3D特征点之间的对应关系。当推理引擎1270确定2D图像中的一组2D特征点属于对象时,用对应的对象对该组2D特征点进行标记。随后,2D特征点的对象标签被投影到对应的3D特征点,使得该组3D特征点也具有该对象的标签。作为示例,推理引擎1270识别属于“桌子”的第一组2D特征点和属于“墙”的第二组2D特征点。然后,推理引擎1270用“桌子”标记第一组2D特征点,并用“墙”标记第二组2D特征点。存在对应于第一组2D特征点的第三组3D特征点以及对应于第二组2D特征点的第四组3D特征点。因此,推理引擎1270用“桌子”标记3D平面地图中的第三组3D特征点,并用“墙”标记3D平面地图中的第四组3D特征点。如上所述,推理引擎1270能够确定对象的名称或类别,例如桌子、椅子或墙。在某些实施例中,推理引擎1270被配置为确定对象的风格,例如家具的经典风格或现代风格、房间的当代风格、沿海风格或现代风格。在某些实施例中,推理引擎1270被配置为确定房间的功能,例如厨房、起居室、卧室;确定家具的功能,例如用餐、办公、娱乐等。

[0079] 在某些实施例中,使用体素标记3D平面地图以形成3D场景地图。具体地,3D平面地图在3D空间中被划分为体素,并且当可用时,基于2D图像的场景理解以及2D特征点与3D特征点之间的对应关系,用对象名称或类别对体素进行标记。在某些实施例中,3D平面地图由平面标记以形成3D场景地图。具体地,使用3D空间中的适配平面来分割3D平面地图,并且基于对2D图像的场景理解以及2D特征点和3D特征点之间的对应关系,用对象名称或类别来标记平面。因此,无论是用体素还是用平面进行标记,都有用一个对象标记的一组3D特征点和用另一对象标记的另一组3D特征点。用对象标签更新3D平面地图会得到3D场景地图。换言之,3D场景地图包括来自SLAM模块122的3D点、由图像/点处理模块124定义的任意定向的平面以及来自场景理解模块126的对象名称的标签和/或场景的类别和/或场景中的对象的分割和/或场景中的每个像素的深度值。在某些实施例中,可以使用JSON(JavaScript对象表示法)来定义附于3D地图的数据。例如,场景可以具有以下数据结构:

```
{ "场景": {  
  "类别": "起居室",  
  "对象": {  
    "ID": 0,  
    "标签": "桌子",  
  }  
}
```

[0081] 作为SLAM的一部分的3D场景地图由SLAM模块122持续地更新。然后,渲染模块128可以使用该3D场景地图。

[0082] 在某些实施例中,场景理解模块126还被配置为基于所获得的3D场景地图向用户

提供对产品的推荐。在一个实施例中,场景理解模块126使用一个或多个模型确定环境中的家具(对象)的风格,并且基于风格,推荐风格与家具的风格相匹配的产品。在另一实施例中,确定3D场景地图中用于放置一个或多个物品的空置空间,并且推荐在该环境中不存在并且尺寸适配该空置空间的产品。在又一实施例中,确定3D场景地图中的家具和器具的功能,并且推荐功能与该家具和器具的功能互补的产品。例如,当3D场景地图中存在学习桌和椅子且确定在学习桌上没有灯时,场景理解模块126可以向用户推荐灯,其中灯的风格和尺寸非常适合学习桌。在某些实施例中,通过以下操作来执行推荐:确定与房间类别相对应的一组产品,例如用于起居室的沙发和桌子;基于房间风格和当前家具选择多个产品,例如因为起居室中的当前家居风格是现代风格,所以选择现代风格的沙发和桌子;选择功能与当前家具的功能互补的产品,例如当只有咖啡桌时选择沙发,并基于所识别的桌子周围的空置空间确定沙发的大小。之后,可以将所推荐的沙发的3D模型放置并渲染到3D场景地图中。请注意,推荐可以包括一个或多个上述步骤,并且步骤的顺序可以基于用户的偏好而变化。在某些实施例中,场景理解模块126用C++编码,这使得AR应用118能够在不同平台上可用。

[0083] 渲染模块128被配置为:当3D场景地图和产品的3D模型可用时,将3D模型放置并渲染在3D场景地图上,在3D场景地图中检测3D模型与对象之间(即,虚拟对象和真实对象之间)的碰撞,并为用户提供在3D场景地图中操作3D模型的交互。如图2D所示,渲染模块128包括纹理模块1280、光模块1282、碰撞检测模块1284和粒子模块1286。在某些实施例中,纹理模块1280被配置为定义关于3D模型的高频细节、表面纹理或颜色信息;光模块1282被配置为定义3D模型上的光;碰撞检测模块1284被配置为当在3D场景地图中移动3D模型时检测并避免碰撞;并且粒子模块1286被配置为使用大量非常小的子图形、3D模型或其他图形对象来模拟某种“模糊”现象,例如着火、爆炸、烟雾、流动的水、火花、云、雾、或者如发光小路、魔咒等的抽象视觉效果。在某些实施例中,因为3D场景地图中的对象是由点、平面、体素等的3D坐标定义的,所以可以检测3D模型和环境中的真实对象之间的碰撞。因此,当用户围绕例如沙发对象或墙壁环境移动3D模型时,检测到可能的碰撞并且使3D模型在特定位置停止以避免与沙发或墙重叠。在某些实施例中,当选择多个产品并且将产品的3D模型放置在环境中时,还确定3D模型之间的重叠以避免3D模型之间的碰撞。因此,AR应用118不仅可以检测和避免虚拟对象之间的碰撞,还可以检测和避免虚拟对象和真实对象之间的碰撞。

[0084] 在某些实施例中,使用C++对渲染模块128进行编码,使得渲染模块128是跨平台的。在某些实施例中,渲染模块128是开源渲染引擎。在某些实施例中,渲染模块128是符合本发明目的的经修改的开源渲染引擎。

[0085] 人机交互模块130被配置为在从用户接收到姿势或语言时控制3D模型在3D地图中移动。如图2E所示,人机交互模块130包括姿势识别模块1300、语音识别模块1302和自然语言处理模块1304。姿势识别模块1300被配置为解释源自任何身体运动或状态的姿势,例如来自面部的情绪或用户的手势,并使用该解释来作用于3D模型。语音识别模块1302被配置为识别说出来的语言并将其翻译成文本,自然语言处理模块1304被配置为对语音识别模块1302获得的文本进行语义解释,以使用该解释来作用于3D模型。

[0086] 用户通信模块132被配置为在不同用户之间构建实时通信,或者换言之,在不同计算设备之间构建实时通信。如图2F所示,用户通信模块132包括坐标关系模块1320、时钟同步模块1322、图像登记模块1324和设备通信模块1326。坐标关系模块1320被配置为将计算

设备110中的坐标与另一计算设备中的坐标对准。例如,当两个用户均使用智能电话对同一环境运行AR应用118时,两个智能电话可以查看环境的不同部分,并且这两个不同部分具有一定的重叠部分。两个智能电话能够彼此通信以基于重叠部分将一个智能电话中的AR应用中的坐标与另一智能电话中的AR应用中的坐标对准。时钟同步模块1322被配置为对两个用户拍摄的图像进行同步。图像登记模块1324被配置为将来自一个用户的图像中的特征点登记到来自另一用户的图像。设备通信模块1326被配置为为用户之间的通信提供接口。

[0087] 视觉传感器150被配置为捕获环境的视觉图像。视觉传感器150包括灰度相机、红绿蓝(RGB)相机或RGB深度(RGBD)相机和深度相机中的至少一个。在某些实施例中,视觉传感器150还可以包括一维(单光束)或2D(扫描)激光测距仪、3D高清晰度LiDAR、3D闪光LIDAR、2D或3D声纳传感器。在某些实施例中,一个灰度图像足以执行SLAM。视觉传感器150由主控制器120控制以实时捕获环境图像。图像捕获速率可以由主控制器120定义,或者可以由视觉传感器150本身定义。在某些实施例中,主控制器120控制视觉传感器150以恒定速率持续地捕获环境图像。

[0088] 在某些实施例中,系统100还可以包括一个或多个IMU 170。IMU170可以包括加速度计、陀螺仪和磁力计。主控制器120将指示IMU170实时采集特定的力、角速率和可选的磁场。

[0089] 图3描述了根据本发明某些实施例的将产品渲染到环境中的流程图。在某些实施例中,该方法由图1中所示的系统来实现。应特别注意,除非在本发明中另有说明,否则该方法的步骤可以以不同的顺序设置,因此不限于如图3所示的顺序。

[0090] 如图3所示,在过程302处,计算设备110的主控制器120指示视觉传感器150(例如,一个或多个相机)捕获环境的视频帧或图像。可选地,计算设备110还可以指示IMU 170(例如,加速度计、陀螺仪和磁力计)实时采集特定的力、角速率和可选地磁场。主控制器120从视觉传感器150接收或检索那些图像,并将图像发送到SLAM模块122、图像/点处理模块124和场景理解模块126。在某些实施例中,视觉传感器150可以直接将2D图像发送到SLAM模块122、图像/点处理模块124和场景理解模块126。在某些实施例中,SLAM模块122、图像/点处理模块124和场景理解模块126可以使用2D图像的同副本。

[0091] 在过程304,当从视觉传感器150或经由主控制器120接收到图像时,SLAM模块122执行对相机的同时定位和对环境的映射,并获得相机的姿态和环境的3D地图。在某些实施例中,SLAM过程包括:从2D图像检测诸如角部之类的特征,从2D图像提取特征点,计算每个特征点的特征描述符,对不同2D图像中的特征点进行三角测量以形成3D地图。所形成的3D地图包括特征点的点云,并且点云中的每个特征点具有特定的描述符。

[0092] 在过程304之前、期间或之后,图像/点处理模块124执行过程306。具体地,当从视觉传感器150或经由主控制器120接收到图像时,图像/点处理模块124通过以下方式处理2D图像:特征检测、特征提取和分割。例如,如果在2D图像中存在“桌子”,则图像/点处理模块124能够识别桌子的角部或线条以及桌子的平坦表面的特征点,基于特征点定义桌子的表面,并从其余图像分割桌子。可以通过在2D图像中定义围绕桌子的边界框来执行对桌子的分割。在某些实施例中,图像/点处理模块124的特征检测比SLAM模块122进行的特征检测更全面。例如,SLAM模块122可以仅检测2D图像中的角部特征,并且检测到的角部特征足以执行SLAM;图像/点处理模块124可以检测角部、边、斑点、脊、线、圆、平面等。在某些实施例中,

图像/点处理模块124可以使用来自SLAM模块122的检测到的角部信息,因此不需要具有角部检测功能。在某些实施例中,图像/点处理模块124以诸如C++之类的跨平台语言编写,使得图像/点处理模块124可以在Android™、iOS或任何其他合适的环境下使用。然后可以使用检测到的2D特征来更新3D地图。在某些实施例中,图像/点处理模块124将2D特征发送到场景理解模块126。

[0093] 在过程304之后,当访问来自SLAM模块122的3D地图时,在过程308,图像/点处理模块124从3D地图中检测平面。可以在执行过程306之前、期间或之后执行过程308。在过程308,图像/点处理模块124检测多种形式的平面,并且至少包括对垂直平面的检测。通过检测水平平面、垂直平面和可选地其他倾斜平面,可以定义3D地图中的3D对象。基于对象的所确定的3D平面和对象在3D空间中的相对位置,可以容易地预测和避免虚拟对象与真实对象之间的碰撞。当获得3D地图中的平面信息时,图像/点处理模块124将这些信息添加到3D地图中的特征点,以更新3D地图。具体地,3D地图包括特征点,并且每个点具有描述符。当确定平面时,平面的中心、平面的定向以及平面的长度和宽度被添加到该3D地图,并且标记属于该平面的一组特征点。因此,3D地图现在不仅包括具有特征描述符的特征点,还包括平面的标签和每个平面的一组特征点。更新后的3D地图被称为3D平面地图。在某些实施例中,3D平面地图仍然是SLAM的一部分,并且由SLAM模块122使用持续输入的2D图像来进行更新。

[0094] 在过程310,当接收到2D图像和2D特征时,场景理解模块126处理2D图像和检测到的2D特征以理解场景。在某些实施例中,2D图像是从视觉传感器150接收的、经由主控制器120接收的、从SLAM模块122或从图像/点处理模块124接收的。在某些实施例中,SLAM模块122、图像/点处理模块124和场景理解模块126使用同一组2D图像。从图像/点处理模块124接收2D特征。在某些实施例中,场景理解模块126使用各种模型来识别2D图像中的对象。在某些实施例中,对象检测方法包括快速R-CNN以及仅看一眼(you only look once)(YOLO)。对象检测方法使用图像并产生对象内容和位置。例如,在具有“椅子”、“监视器”和“人”的图像中,对象检测方法能够确定对象的位置,并识别对象的名称、类别和风格。在某些实施例中,每个对象由2D图像中围绕对象的矩形框标识。

[0095] 在过程312,当对象的类型和位置被理解时,场景理解模块126还基于2D特征点和3D特征点之间的对应关系将对象信息投影到3D平面地图,并更新3D平面地图以获得3D场景地图。因此,3D场景地图不仅包括特征点、特征点描述符和平面,还包括对象信息,例如桌子、墙或椅子,以及对象的风格,例如现代风格和经典风格。可以通过标记与对象对应的一组特征点来执行对对象的标记。

[0096] 作为一个示例,在过程304由SLAM模块122构建的3D地图包括以下特征点,每个特征点具有特定描述符:点001,描述符001;点002,描述符002...点099,描述符099;点100,描述符100;点101,描述符101;点102,描述符102...点149,描述符149;点150,描述符150;点151,描述符151...点199,描述符199...场景理解模块126在过程310处理2D特征和2D图像,并确定与3D特征点001-199对应的2D特征点属于“桌子1”。场景理解模块126还可以确定与3D特征点001-100对应的2D特征点属于一个平面,桌子1的顶表面;与3D特征点101-149对应的2D特征点属于另一平面,桌子1的左侧表面;并且与3D特征点150-199对应的2D特征点属于又一平面,桌子1的右侧表面。在某些实施例中,桌子1可以由场景理解模块126确定为具有经典风格,并且经典风格也作为标签被添加到3D地图中的对应特征点。通过图像/点处理

模块124组合上述信息和可选的平面适配信息,场景理解模块126更新3D平面地图以获得3D场景地图。结果,3D场景地图中的每个对象被定义为包括一组特征点,具有适配平面,并且标记有对象的名称、类别和风格。利用3D场景地图中的对象的3D信息,可以在将产品的3D模型渲染到3D地图中时检测并避免碰撞。渲染模块128可访问3D场景地图。

[0097] 在某些实施例中,场景理解模块126还被配置为:当3D场景地图可用时,为用户提供产品推荐。3D场景地图包括环境中的家具的风格,因此场景理解模块126被配置为推荐与家具的风格匹配的产品。3D场景地图包括环境中的家具的大小,因此场景理解模块126被配置为推荐与一个或多个家具的大小匹配的产品。

[0098] 在过程314,当3D场景地图可用时,渲染模块128将产品的3D模型渲染到3D场景地图中。在某些实施例中,产品是在步骤312由场景理解模块126推荐的产品。当场景理解模块126推荐多于一个产品时,AR应用118还可以为用户提供从推荐的产品中选择一个或多个产品的界面。在某些实施例中,用户可以选择产品,而不需要推荐。当选择或推荐产品时,产品的3D模型将渲染到3D场景地图中。3D场景地图包括对场景的理解,诸如对象的功能和尺寸,并且渲染模块128使用该信息来渲染3D模型。首先,渲染模块128可以通过将3D模型放置在3D场景地图中与产品的功能匹配的区域或表面中,来执行区域选择。例如,3D场景地图包括环境中的墙的位置和尺寸,并且当选择绘画作为产品时,渲染模块128基于场景理解将绘画渲染在一个墙上。当选择TV作为产品时,渲染模块128基于场景理解将TV放置在TV桌或TV架上。其次,渲染模块128可以通过基于场景理解将3D模型放置在所选区域中的合适位置来执行位置确定。例如,渲染模块128基于对墙和TV架的尺寸的场景理解,将绘画放置在墙上的某个高度处,或者将TV直接放置在TV架的顶表面上。第三,在渲染产品期间,渲染模块128基于3D场景地图中的不同对象的位置和尺寸来检测并避免碰撞。例如,当渲染模块128确定墙和桌子之间的空间用于放置沙发时,渲染模块128将基于墙和桌子的位置和尺寸在该过程期间避免碰撞。

[0099] 图4示意性地示出了根据本发明某些实施例的SLAM的示例。在某些实施例中,该示例由计算设备(诸如,图1中所示的计算设备110)执行。在视觉传感器150捕获环境的图像之后,在过程402,SLAM模块122从视觉传感器150或主控制器120接收那些图像。在过程404,特征提取模块1220从图像中提取特征点并计算所提取的特征点的特征描述符。在某些实施例中,特征提取模块1220还从特征点中检测角部。在过程406,定位和映射模块1222使用那些提取的特征点执行SLAM以获得视觉传感器150的姿态并构建3D地图。在过程408,光检测模块1224检测场景中的光方向和强度。当IMU 170可用时,在过程406之前,融合模块1226可以从IMU 170接收信息并将这些信息与捕获的图像融合以获得融合数据,并将融合数据发送到定位和映射模块1222以执行SLAM。在某些实施方案中,使用ARKit或ARCore执行图4中所示的方法,该ARKit或ARCore通过引用整体并入本文。在其他实施例中,SLAM模块122是与不同操作系统(例如,Android™、iOS™和Windows™)兼容的定制模块。

[0100] 图5A示意性地示出了根据本发明某些实施例的2D图像/点处理的示例。在某些实施例中,该示例由计算设备(诸如,图1中所示的计算设备110)执行。如图5A所示,在过程502,图像/点处理模块124从SLAM模块122接收2D图像和检测到的角部信息。在某些实施例中,图像/点处理模块124还可以从视觉传感器150接收图像或者经由主控制器120接收图像和角部特征。然后在过程504,图像特征提取模块1246提取除角部之外的特征,例如边、斑

点、脊、线、圆等。可选地,当图像/点处理模块124从视觉传感器150接收原始图像时,它还被配置为从原始图像中检测角部。在检测并提取这些特征之后,在过程506,图像分割模块1248对主要对象进行分割,例如图像中的桌子、监视器、椅子、门、窗户或墙。尽管这些对象以它们可以被彼此区分开的程度被识别,但是这里它们还未被识别为桌子、监视器、椅子、门、窗户、墙。然后,2D图像中的那些分割对象可以用于场景理解。在某些实施例中,2D特征可以被投影到3D地图。在某些实施例中,过程504和506可以以不同的顺序布置。在某些实施例中,并非所有过程504和506都是项目所需的。

[0101] 图5B示意性地示出了根据本发明某些实施例的对3D地图的图像/点处理的示例。在某些实施例中,该示例由计算设备(诸如,图1中所示的计算设备110)执行。如图5B所示,在过程510处,图像/点处理模块124直接或经由主控制器120接收由SLAM模块122构建的3D地图。然后在过程512处,点云平面适配模块1240检测并适配3D地图中的水平平面、垂直平面或具有任意定向的平面。通过检测各个方向的平面,可以通过这些平面中的一个或多个平面来标识3D地图中的对象。例如,优选地具有所确定的边的垂直平面可以用于定义墙;水平平面和几个垂直平面可用于定义文件柜。

[0102] 图6A示意性地示出了根据本发明某些实施例的场景理解模块的离线训练600A的示例。在某些实施例中,离线训练在除计算设备110之外的强大计算设备上执行,并且在该强大计算设备完成训练之后计算设备110可以使用训练好的模型。在某些实施例中,在过程602,离线训练模块1260接收标记的2D图像数据集。在接收到标记的2D图像数据集时,离线训练模块1260在过程604训练图像识别模块1262以确定标记图像中的对象的位置和标签(名称);在过程606训练图像分割模块1264以将标记图像分割成分割块,以便定位对象和对象的边界;在过程608训练图像分类模块1266以将图像分类到不同的类别,例如厨房、卧室、椅子、沙发、人等;并且在过程612训练图像深度模块1268以估计对象的深度。在某些实施例中,过程604-612可以以不同的顺序或并行布置。在某些实施例中,并非所有过程604-612都是项目所需的。

[0103] 图6B示意性地示出了根据本发明某些实施例的运行场景理解模块的示例600B。如图6B所示,在过程614,推理引擎1270从图像/点处理模块124接收2D图像和2D特征。如本申请前面所述,还可以从视觉传感器150、主控制器120或SLAM模块122接收2D图像。在某些实施例中,SLAM模块122、图像/点处理模块124和场景理解模块126可以使用同一组2D图像。当接收到2D图像和2D特征时,在过程616,推理引擎1270针对2D图像和2D特征运行所选择的算法。在某些实施例中,推理引擎1270可以使用TensorFlow®或CoreML®来实现。推理引擎1270可以在该过程期间调用一个或多个模型1262-1268。在过程616之后,获得场景理解,其包括2D图像中的对象的位置和边界、尺寸、名称和/或类别以及风格。然后在过程618,2D-3D投影模块1272基于2D图像中的2D特征点与3D平面地图中的3D特征点之间的对应关系将信息投影到3D平面地图,使得3D平面地图被更新为3D场景地图。

[0104] 图7示意性地示出了根据本发明某些实施例的渲染3D模型和3D地图的示例。如图7所示,在过程702处,场景理解模块126推荐产品,或者用户针对环境选择产品,并且AR应用118例如从电子商务服务器计算机中检索该产品的3D模型。然后在过程704,纹理模块1280定义3D模型的纹理,并且光模块1282定义3D模型上的光;在过程706,当在3D场景地图中移动3D模型时,碰撞检测模块1284实时检测并避免碰撞;并且在过程708,当3D模型将要与3D

场景地图中的对象接触或与之接触时,粒子模块1286提供特殊的视觉效果或动画。在某些实施例中,过程704-708可以以不同的顺序布置。在某些实施例中,并非所有过程704-708都是项目所需的。

[0105] 在某些示例中,图像/点处理如下所示地执行。图像/点处理模块的输入包括图像和点云,而输出是3D信息和2D图像中的特征。在处理点云时,获得整个环境地图并对特定方向的某些平面进行适配。在处理图像时,可以使用传统的图像处理方法来获得关于图像的特征,这些特征将是深度学习或场景理解部分的输入。提取特征点或线,并执行分割,以帮助对2D图像的场景理解。在示例中,当垂直平面的方向被定义,并且垂直平面中的点根据其坐标 (x_p, y_p, z_p) 而被确定时,可以将绘画放置在垂直平面上,使得绘画被悬挂在该平面上,并且绘画的中心放置在所确定的点处。在另一示例中,确定诸如地板之类的水平平面和诸如墙之类的垂直平面,并且将桌子放置在地板上并靠着墙。当诸如桌子之类的产品被放置在地板上时,AR应用实时计算该产品与墙之间的距离,以避免虚拟产品与真实墙之间的碰撞。

[0106] 在某些示例中,通过使用由相机捕获的2D图像进行深度学习来执行场景理解。AR应用使用诸如快速R-CNN和YOLO之类的神经网络来进行对象检测,尤其是在室内场景中。AR应用实时使用输入图像以通过场景理解模块126知道家具是什么及其位置。然后,AR应用通过SLAM结果将对应于这些对象的2D点映射到3D点。因此,AR应用能够基于对这些对象的识别来了解3D空间中的家具或其他对象是什么及其位置,或者换言之,AR应用标记3D空间中对应于对象的点云集。在示例中,深度学习模型从2D图像中检测椅子、人和监视器。基于2D图像中的2D特征点与3D地图中的3D特征点之间的对应关系,分别标记对应于椅子、人和监视器的3D特征点集,使得3D空间中的椅子、人和监视器的位置和占用可以从3D地图直接获得。

[0107] 在某些示例中,根据本发明某些实施例的AR应用用于碰撞检测。如上所述,3D空间中的真实对象的位置和占用已经由AR应用的图像/点处理模块标记。当AR应用的渲染模块移动虚拟对象(产品的3D模型)时,可以使动画显示碰撞并且可以避免碰撞。

[0108] 在某些示例中,根据本发明某些实施例的AR应用用于智能放置。如上所述,3D空间中的真实对象的位置和占用已经由AR应用的图像/点处理模块标记。当一些虚拟商品被放置在3D地图中时,AR应用能够自动将虚拟商品放置在合适的位置,例如自动将虚拟椅子放置在真实桌子附近。

[0109] 在某些示例中,根据本发明某些实施例的AR应用用于商品推荐。如上所述,AR应用的场景理解模块理解用户周围的环境。根据理解,AR应用还可以被配置为更准确地为用户推荐商品。例如,如果AR应用在用户的房间中检测到桌子,则AR应用被配置为向顾客推荐例如灯或椅子。

[0110] 在某些方面,本发明涉及多用户通信。在某些实施例中,通信涉及增强现实(AR)购物。图8示意性地描绘了根据本发明某些实施例的AR购物系统,并且图9描绘了根据本发明某些实施例的执行多用户AR购物的流程图。

[0111] 图8示意性地描绘了根据本发明某些实施例的多用户增强现实购物系统。如图8所示,系统800包括第一计算设备810和第二计算设备860。第一计算设备810和第二计算设备860经由网络850彼此通信。第一计算设备810和第二计算设备860中的每个可以与计算设备

110相同或不同。第一计算设备810和第二计算设备860也可以分别命名为主计算设备和辅助计算设备。系统800还可以包括与第一计算设备810或/和第二计算设备860通信的其他计算设备。系统800中的计算设备可以在配置方面是相同的或不同的,但是它们都具有相应的购物应用并且能够彼此通信并使用购物应用共享数据。在某些实施例中,网络850可以是有线网络或无线网络,并且可以具有各种形式。网络850的示例可以包括但不限于局域网(LAN)、包括因特网的广域网(WAN)或任何其他类型的网络。最著名的计算机网络是因特网。

[0112] 如图8所示,第一计算设备810可以包括但不限于第一处理器812、第一存储器814和第一存储设备816。第一计算设备810可以与一个或多个第一视觉传感器840集成,并且可选地与一个或多个惯性测量单元(IMU) 845集成,以例如形成诸如智能电话之类的移动设备。如图8所示,第一存储设备816包括第一购物应用818。在某些实施例中,第一购物应用818提供用于在线购物的增强现实(AR)功能。备选地,第一购物应用818也可以离线执行。

[0113] 其中第一购物应用818包括第一产品选择模块820和第一AR模块822。第一AR模块822包括第一同步模块824、第一3D地图模块826、第一坐标对准模块828和第一渲染模块830。第一产品选择模块820被配置为向第一用户提供界面,以浏览电子商务服务器上的产品,选择他感兴趣的一个或多个产品,并将选择发送给第一同步模块824。第一同步模块824被配置为当从第一产品选择模块820接收到选择时,从电子商务网站检索或下载产品信息,并将产品信息发送到第二同步模块874。下载的产品信息可以包括产品的标识、产品的类别和名称/标题以及产品的3D模型。在某些实施例中,如上所述,第一产品选择模块820还被配置为将产品信息或产品信息的一部分发送到第二计算设备860。产品信息的发送可以是推送通知的形式。响应于接收到具有产品信息的推送通知,第二计算设备860启动第二购物应用868。在第二购物应用868运行之后,第二同步模块870被配置为从该通知中提取产品信息。在某些实施例中,第一同步模块824可以仅将产品的标识或URL发送到第二计算设备860,并且第二同步模块874在从通知中提取出标识或URL时,从电子商务服务器下载所有产品信息。进一步地,第二同步模块874被配置为向第一同步模块824发送请求,并与第一同步模块824通信以在第一计算设备810与第二计算设备860之间建立主辅关系,其中第一计算810被确定为主设备,且第二计算设备860被确定为辅助设备。在某些实施例中,第一购物应用818和第二购物应用868还可以为用户提供任意确定主辅关系的选项。一旦建立了主辅关系,第一同步模块824和第二同步模块874被配置为使两个计算设备的时钟同步。在某些实施例中,如果第一视觉传感器840和第二视觉传感器890以不同的帧速率捕获环境图像,则时钟的同步使得两个设备之间的通信处于同一速度并且基于第一图像和第二图像的捕获时间而对应第一图像和第二图像。在某些实施例中,当第一购物应用818将3D模型放置在第一3D地图中时,第一同步模块824被配置为与第二同步模块874通信该3D模型在第一3D地图中的位置和定向,并且第二同步模块874被配置为基于3D模型在第一3D地图中的位置和定向以及第二3D地图与第一3D地图之间的变换将3D模型放置到第二3D地图中。第一同步模块824和第二同步模块874之间的通信是双向通信。一旦一个用户在其计算设备上移动或旋转3D模型,3D模型的变换就被实时传送到另一计算设备,使得另一设备在对应的3D地图中同步该3D模型的相同移动或旋转。第一3D地图模块826被配置为,当建立了第一计算设备810和第二计算设备860之间的主辅关系时,指示第一视觉传感器840和第一IMU 845采集环境的第一图像和第一计算设备810的IMU信息,并且基于采集的第一图像和可选地采集的IMU

信息来构建环境的第一3D地图。在某些实施例中,第一3D环境地图模块826被配置为使用同时定位和映射(SLAM)来构建第一3D地图,其中获得环境的第一3D地图和第一视觉传感器840的姿态(或者第一计算设备810的姿态)两者。类似地,第二3D地图模块876被配置为,当建立了第一计算设备810和第二计算设备860之间的主辅关系时,指示第二视觉传感器890和第二IMU 895采集环境的图像和第二计算设备860的IMU信息,并构建环境的第二3D地图。在某些实施例中,第二3D地图模块876被配置为使用SLAM构建第二3D地图,其中获得环境的第二3D地图和第二计算设备860的姿态两者。通过第一视觉传感器840和第二视觉传感器890持续捕获新图像,第一3D地图模块826和第二3D地图模块876被配置为更新第一3D地图和第二3D地图。第一3D地图可以使用第一相机设备810的起始位置和定向来建立第一3D地图的坐标系,并且第二3D地图可以使用第二相机设备860的起始位置和定向来建立第二3D地图的坐标系。因此,第一3D地图和第二3D地图具有不同的坐标系。第一3D地图模块826和第二3D地图模块876还被配置为实时更新第一3D地图和第二3D地图。在某些实施例中,当确定第一计算设备和第二计算设备具有主辅关系时,第一3D地图也可以被命名为主3D地图,并且第二3D地图也可以被命名为辅助3D地图。第一坐标对准模块828被配置为将构建的第一3D地图发送给第二坐标对准模块878。当接收到第一3D地图时,第二坐标对准模块878被配置为将第二3D地图与第一3D地图对准,并将第二3D地图的坐标系变换为第一3D地图的坐标系。在某些实施例中,第二坐标对准模块878被配置为将第二3D地图中的特征点与第一3D地图中的特征点相匹配,并且基于匹配的特征点将第二3D地图中的坐标变换为第一3D地图中的坐标。在某些实施例中,在构建第一3D地图期间存在第一关键帧,并且在构建第二3D地图期间存在第二关键帧。第一坐标对准模块828被配置为将第一关键帧或来自第一关键帧的特征点发送至第二坐标对准模块878,该第二坐标对准模块878被配置为将第二关键帧中的特征点与来自第一关键帧的特征点进行匹配。基于第一3D地图与第一关键帧之间的对应关系和第二3D地图与第二关键帧之间的对应关系,第二坐标对准模块878被配置为将第一3D地图的坐标系变换为第二3D地图的坐标系。在某些实施例中,第一3D地图和第二3D地图中的特征点之间的匹配或两个图像(或关键帧)中的特征点与第一3D地图之间的匹配也称为图像登记。第一渲染模块830被配置为在执行第一3D地图和第二3D地图之间的对准之后,将3D模型放置并渲染到第一3D地图。

[0114] 在某些实施例中,第一购物应用818和第二购物应用868还可以包括设备通信功能,其为用户之间的通信提供接口。在某些实施例中,第一渲染模块830和第二渲染模块880中的每个被配置为:当3D地图和产品的3D模型可用时,将3D模型放置并渲染在3D地图上,并为用户提供用于在3D地图中操作3D模型的界面。在某些实施例中,渲染模块包括:纹理功能,用于定义3D模型上的高频细节、表面纹理或颜色信息;光功能,用于定义3D模型上的光;碰撞功能,用于当3D地图中移动3D模型时检测并避免碰撞;以及粒子功能,用于使用大量非常小的子图形、3D模型或其他图形对象来模拟某些类型的“模糊”现象,如着火、爆炸、烟雾、流水、火花、云、雾或如发光小路、魔咒等的抽象视觉效果。在某些实施例中,第一购物应用818和第二购物应用868还可以包括人机交互模块,其被配置为在从用户接收到姿势或语言时控制3D模型在3D地图中移动。

[0115] 图9描绘了根据本发明某些实施例的多用户购物的流程图。在某些实施例中,该方法900由图8中所示的系统来实现。应特别注意,除非在本发明中另有说明,否则该方法的步

骤可以以不同的顺序设置,因此不限于如图9所示的顺序。

[0116] 如图9所示,在过程902,第一用户使用第一计算设备810来启动第一购物应用818。第一产品选择模块820在操作时提供界面,使得第一用户可以浏览与购物应用818相对应的一个或多个电子商务平台,并通过该界面选择一个或多个感兴趣的产品。响应于对产品的选择,产品选择模块820将选择发送到第一AR模块822的第一同步模块824。在某些实施例中,产品选择820可以不是购物应用818的一部分。例如,第一用户可以使用网络浏览器来浏览电子商务平台的网站,通过网络浏览器选择产品,并将选择的产品发送到购物应用818,以便触发对购物应用818的初始化。

[0117] 在选择感兴趣的产品之后,在过程904,第一同步模块824从电子商务平台下载产品信息,并将产品信息或产品信息的一部分发送到第二计算设备860。可以以推送通知的形式发送产品信息。推送通知可以包括产品的登记信息的完整条目或者仅包括产品的标识或URL。由卖方和/或电子商务平台提供的登记信息可以包括产品的标识、类别、名称、描述和3D模型。在某些实施例中,推送通知可以包括其他功能,诸如触发被安装在第二计算设备860上的AR购物应用的初始化。

[0118] 在过程906,响应于从第一计算设备810接收到通知,第二计算设备860启动第二购物应用868。在初始化第二购物应用868之后,第二同步模块874从通知中提取产品信息。如果提取的信息是产品信息的一部分,例如仅是产品的标识或URL,则第二同步模块874还从电子商务服务器下载其他必要的产品信息。必要的信息可以包括产品的类别、名称、3D模型和材料。

[0119] 在过程908,当第二购物应用868被初始化并且下载了产品信息时,第二同步模块874向同步模块824发送请求,以建立第一计算设备810和第二计算设备860之间的主辅关系,其中第一计算设备810被确定为主设备,且第二计算设备860被确定为辅助设备。在某些实施例中,当两个以上的计算设备被连接时,那些计算设备的同步模块可以将一个计算设备定义为主设备而将所有其他计算设备定义为辅助设备。在某些实施例中,还可以使用由用户生成的指令来切换主辅关系。第一同步模块824和第二同步模块874分别将主辅关系发送到第一3D地图模块826和第二3D地图模块876。在某些实施例中,过程908还可以包括两个计算设备之间的时钟同步。

[0120] 然后在过程910,响应于从第一同步模块824接收到主辅关系,第一3D地图模块826指示第一视觉传感器840捕获环境的图像,指示第一IMU 845采集第一计算设备810的IMU信息,并基于捕获的图像和采集的IMU信息构建环境的第一3D地图和获得第一计算设备810的姿态。在某些实施例中,第一3D地图模块826使用SLAM构建第一3D地图。类似地,响应于从第二同步模块874接收到主辅关系,第二3D地图模块876指示第二视觉传感器890捕获环境的图像,指示第二IMU 895采集第二计算设备860的IMU信息,并基于捕获的图像和采集的IMU信息构建环境的第二3D地图和获得第二计算设备860的姿态。在某些实施例中,第二3D地图模块876使用SLAM构建第二3D地图。第一3D地图可以使用第一相机设备810的起始位置和定向来建立第一3D地图的坐标系,并且第二3D地图可以使用第二相机设备860的起始位置和定向来建立第二3D地图的坐标系。因此,第一3D地图和第二3D地图具有不同的坐标系。在某些实施例中,视觉传感器840和890具有重叠的视场。换言之,第一视觉传感器840和第二视觉传感器890至少捕获环境的一些公共区域的图像。相应地,第一3D地图和第二3D地图包括

与环境的公共区域对应的特征点。在构建第一3D地图和第二3D地图之后,第一3D地图模块826和第二3D地图模块876分别通知第一坐标对准模块828和第二坐标对准模块878。

[0121] 响应于接收到构建了第一3D地图和第二3D地图的通知,在过程912,第一坐标对准模块828和第二坐标对准模块878将第一3D地图和第二3D地图对准。在某些实施例中,第一坐标对准模块828将第一3D地图发送到第二坐标对准模块878,并且在接收到第一3D地图时,坐标对准模块878将第一3D地图和第二3D地图中的特征点进行比较,以定位匹配的特征点,然后通过将第二3D地图的坐标变换为第一3D地图中的坐标来将第二3D地图与第一3D地图对准。在某些实施例中,在构建第一3D地图期间存在第一关键帧,并且在构建第二3D地图期间存在第二关键帧。第一坐标对准模块828将第一关键帧或来自关键帧的特征点发送至第二对准模块878,该第二坐标对准模块878将第二关键帧中的特征点与来自第一关键帧的特征点进行匹配。基于第一3D地图与第一关键帧之间的对应关系和第二3D地图与第二关键帧之间的对应关系,第二坐标对准模块878将第一3D地图的坐标系变换为第二3D地图的坐标系。

[0122] 在某些实施例中,第一3D地图和第二3D地图之间的对准包括确定刚性变换,使用特征点作出表面表示并匹配表面表示,以及对匹配的点和可选地不匹配的点进行优化。

[0123] 在某些实施例中,用于将第二(辅助)3D地图与第一(主)3D地图对准的变换是3D空间中的刚性变换,其包括旋转和平移。旋转具有三个自由度(DOF),并且平移具有三个DOF。在某些实施例中,应用视觉惯性测距(VIO)方法,其可以使用IMU估计地平面。第二3D地图的坐标系的Y轴和第一3D地图的坐标系的Y轴指向重力的相反方向,并且刚性变换的旋转部分仅具有一个DOF。因此,更易于将第二3D地图与第一3D地图对准。在某些实施例中,通过最小化误差项来计算对准。在某些实施例中,误差项可以被计算为对应特征点之间的平方距离的平均值。在另一实施例中,误差项可以被定义为采样点与第一3D地图和第二3D地图的参数表面模型之间的平方距离的平均值。

[0124] 在某些实施例中,结构表示和相似性度量主要包括基于特征的表面表示、基于点的表面表示和基于模型的表面表示。在某些实施例中,对于基于特征的方法,特征是指表面形状的紧凑描述。当通过比较标量值来匹配特征时,可以容易地直接计算变换,例如通过计算最小二乘方。在某些实施方案中,应用了D.W.Eggert和A Lorusso,R.B.Fisher的“Estimating 3-D rigid body transformations:a comparison of four major algorithms”中描述的不同变换方法,其通过引用整体并入本文。

[0125] 在某些实施例中,对于基于点的方法,使用全部点或点的子集。相似性标准是在表面点对之间要最小化的平均距离。在某些实施例中,对应关系是预先不知道的,但是可以在处理期间通过算法恢复。在一个示例中,应用迭代最近点(ICP)以最小化两个3D地图中包含的两个点云之间的差异。

[0126] 在某些实施例中,结构表示和相似性度量还可以包括基于全局相似性的方法。在某些实施例中,该方法基于全局表面几何法来登记表面,全局表面几何法不依赖于变换的粗略先验估计并且能够处理相对无特征的区块。

[0127] 关于自动3D表面登记的进一步讨论,请参考Michel Audette等人的“An algorithmic overview of surface registration techniques for medical imaging”,其通过引用整体并入本文。

[0128] 不同表示的选择导致不同的匹配和优化方法。对于基于特征的方法,特征匹配和直接计算就足够了。对于基于点的方法,匹配和优化可以涉及迭代地找到最近点和使表面之间的距离最小化的优化。对于基于模型的方法,可以使用表面演化模型,或者可以应用有限元建模来对表面进行建模。

[0129] 在一轮对准之后,第二坐标对准模块878计算对准误差并确定误差是否大于预定阈值。在某些实施例中,对准误差被定义为来自两个点集的对应点对的平方距离的平均值。如果误差大于阈值,则第二坐标对准模块878向第二用户发送通知,并经由坐标对准模块828向第一用户发送通知,使得第一用户和第二用户可以捕获的更多图像,优选地公共区域的图像。基于新添加的图像,第一3D地图模块826更新第一3D地图,第二3D地图模块876更新第二3D地图。在某些实施例中,坐标对准模块828和878可以基于对准结果提供捕获图像的引导。在某些实施例中,坐标对准模块828和878通过将第一用户和第二用户引导到环境中的公共区域来提供引导。

[0130] 当对准误差小于预定阈值时,在过程914,坐标对准模块878经由坐标对准模块828向第一用户发送通知或直接向第二用户发送通知,并请求第一用户将感兴趣的产品的3D模型放置到第一3D地图。在某些实施例中,第二坐标对准模块878可以允许第二用户而非第一用户将3D模型放置在第二3D地图中。在某些实施例中,购物应用818还可以基于第一3D地图和第二3D地图中的至少一个将3D模型自动地放置在环境中。第一同步模块824和第二同步模块874彼此通信,使得当3D模型被放置在第一3D地图和第二3D地图之一中并且定义了3D模型的位置和定向时,位置和定向可以用于将3D模型放置在第一3D地图和第二3D地图中的另一地图中。因为第一3D地图和第二3D地图在对准之后共享相同的坐标系,所以将3D模型放置在两个3D地图中是直截了当的。

[0131] 然后,用户可以在环境中移动第一计算设备810和第二计算设备860,第一视觉传感器840和第二视觉传感器890继续捕获新图像,并且用户可以在环境中操纵3D模型。现在在过程916,第一3D地图模块826和第二3D地图模块876跟踪相对于全局坐标系的位置/定向和用户操纵。

[0132] 此外,在过程918,第一渲染模块830和第二渲染模块880还在3D地图中渲染3D模型。

[0133] 此外,本发明的某些实施例具有以下有益优点。

[0134] (1) 该系统和方法是跨平台的,可以在Apple®设备、Android™设备、Windows®设备等下运行。图像/点处理模块、场景理解模块和渲染模块用C++编码,C++本身可以在不同平台上被识别和执行。因此,如果SLAM系统不依赖于移动设备的IMU,则系统可以使用C++编码模块是跨平台的。在某些实施例中,系统还可以针对Apple®设备和Android™设备单独设计,以利用移动设备上的惯性传感器。

[0135] (2) 该系统包括用于执行SLAM、图像/点处理、场景理解、渲染以及可选地人机交互的功能模块。通过这些模块的功能联合在一起,本发明为AR应用(例如,AR购物)提供了一站式服务。

[0136] (3) 通过结合场景理解模块,本发明的系统了解环境中对象的风格和环境的风格。根据对风格的场景理解,系统可以推荐与该风格匹配的产品。

[0137] (4) 通过结合SLAM模块和场景理解模块,本发明的系统理解对象的尺寸,例如TV架

的大小、桌子的尺寸以及环境的尺寸,例如房间的大小。因此,当具有特定风格的诸如TV的产品被选择时,系统可以确定与TV架的大小和房间的大小相匹配的TV的合适大小;当具有特定风格的椅子被选择时,系统可以确定与桌子的大小相匹配的椅子的合适大小;当餐桌被选择时,系统可以确定与餐厅的大小相匹配的餐桌的大小。

[0138] (5) 通过结合场景理解模块,本发明的系统理解放置所选产品的位置和表面。例如,当绘画被选择时,该绘画的3D模型将有可能竖直地放置在墙上,当灯被选择时,灯的3D模型将有可能被放置在桌子的顶表面上,并且当沙发被选择时,灯的3D模型将被放置在地板上的在咖啡桌和墙之间的较大空置空间中。

[0139] (6) 系统识别2D图像中的对象的特征,并将这些2D特征投影到3D地图。一方面,2D图像中的识别快速而准确。另一方面,被投影到3D地图中的特征有助于对产品的选择,以及自动将3D产品模型放置在适当的位置。

[0140] (7) 多用户可以运行同一在线购物应用并相互通信以对准他们的3D空间并实时共享产品的3D模型的位置和定向。

[0141] 对本发明的示例性实施例的前述描述是为了说明和描述的目的而给出的,而非意在穷举本发明或将本发明限制于所公开的具体形式。鉴于上述教导,许多修改和变化是可能的。

[0142] 选择和描述实施例以便解释本发明的原理及其实际应用,从而使得本领域的其他技术人员能够利用具有适合于预期特定用途的各种修改的本发明和各种实施例。在不脱离本发明的精神和范围的情况下,备选实施例对于本发明所属领域的技术人员来说将变得显而易见。因此,本发明的范围由随附权利要求,而不是上述描述和本文描述的示例性实施例来限定。

[0143] 参考文献:

[0144] 1.Ren,Shaoqing等.Faster R-CNN:Towards real-time object detection with region proposal networks.Advances in neural information processing systems.2015;

[0145] 2.Dong,Jianfeng,Xirong Li和Cees GM Snoek.Predicting Visual Features from Text for Image and Video Caption Retrieval.arXiv preprint arXiv:1709.01362 (2017);以及

[0146] 3.Dong,Hao等.I2T2I:Learning Text to Image Synthesis with Textual Data Augmentation."arXiv preprint arXiv:1703.06676 (2017)。

100

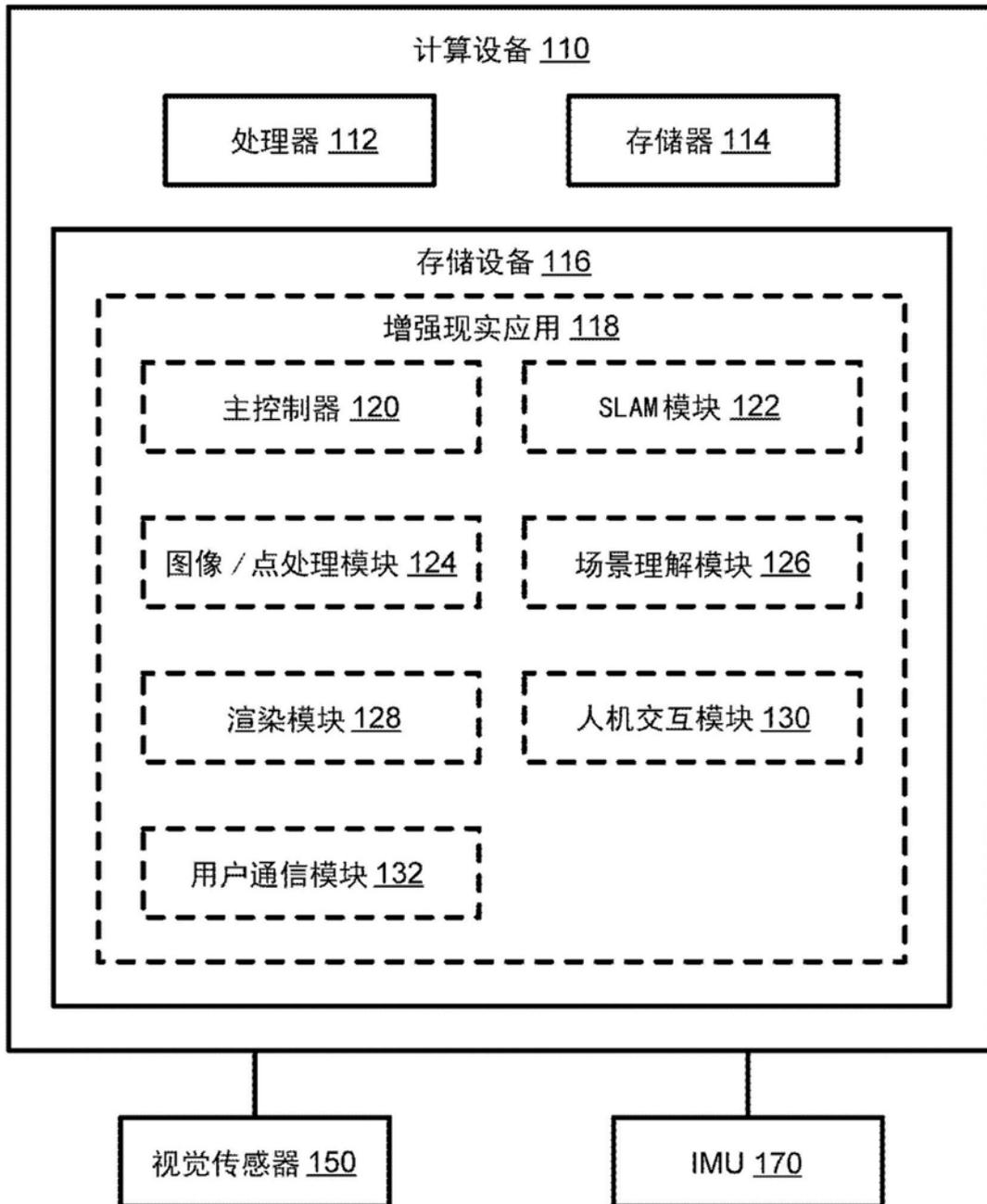


图1

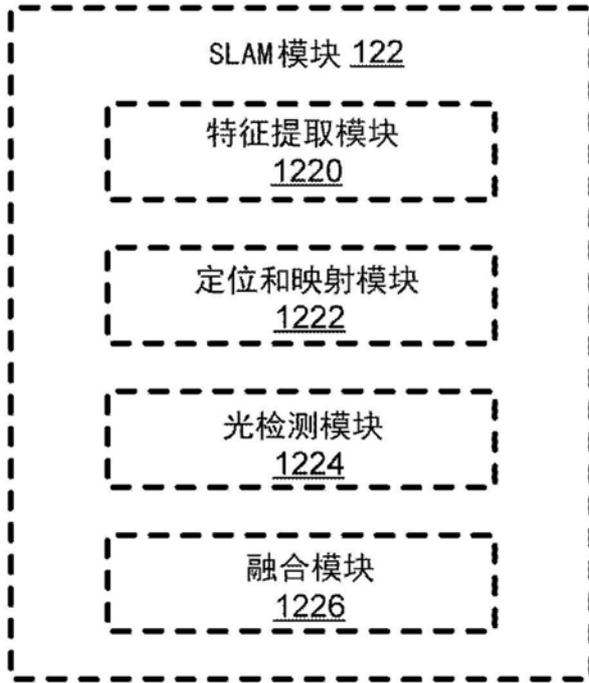


图2A

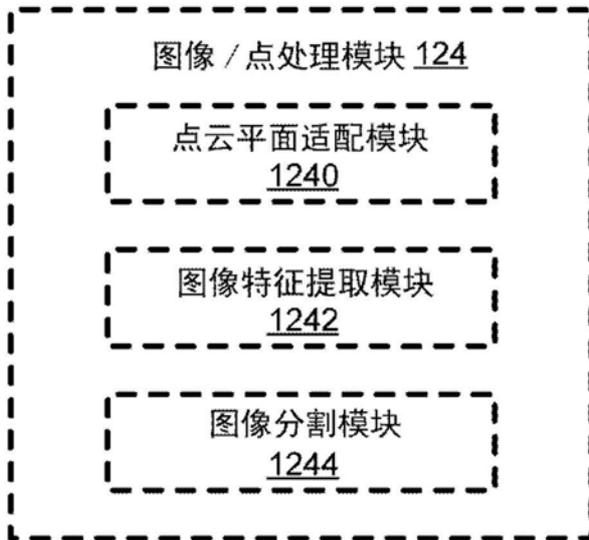


图2B

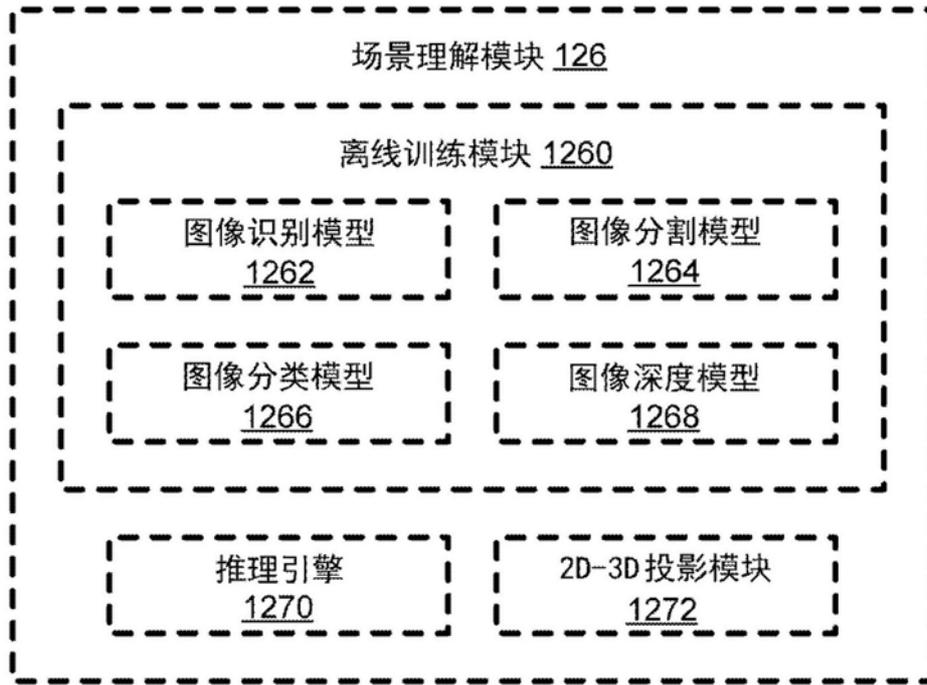


图2C



图2D



图2E

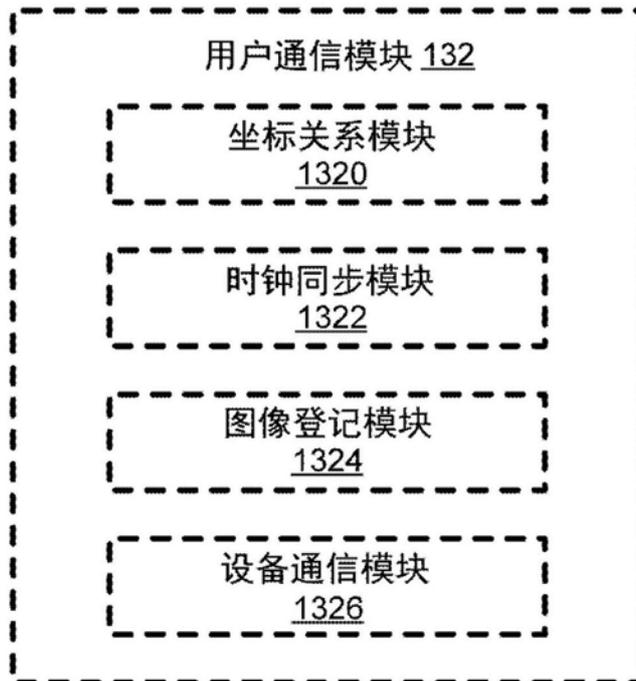


图2F

300

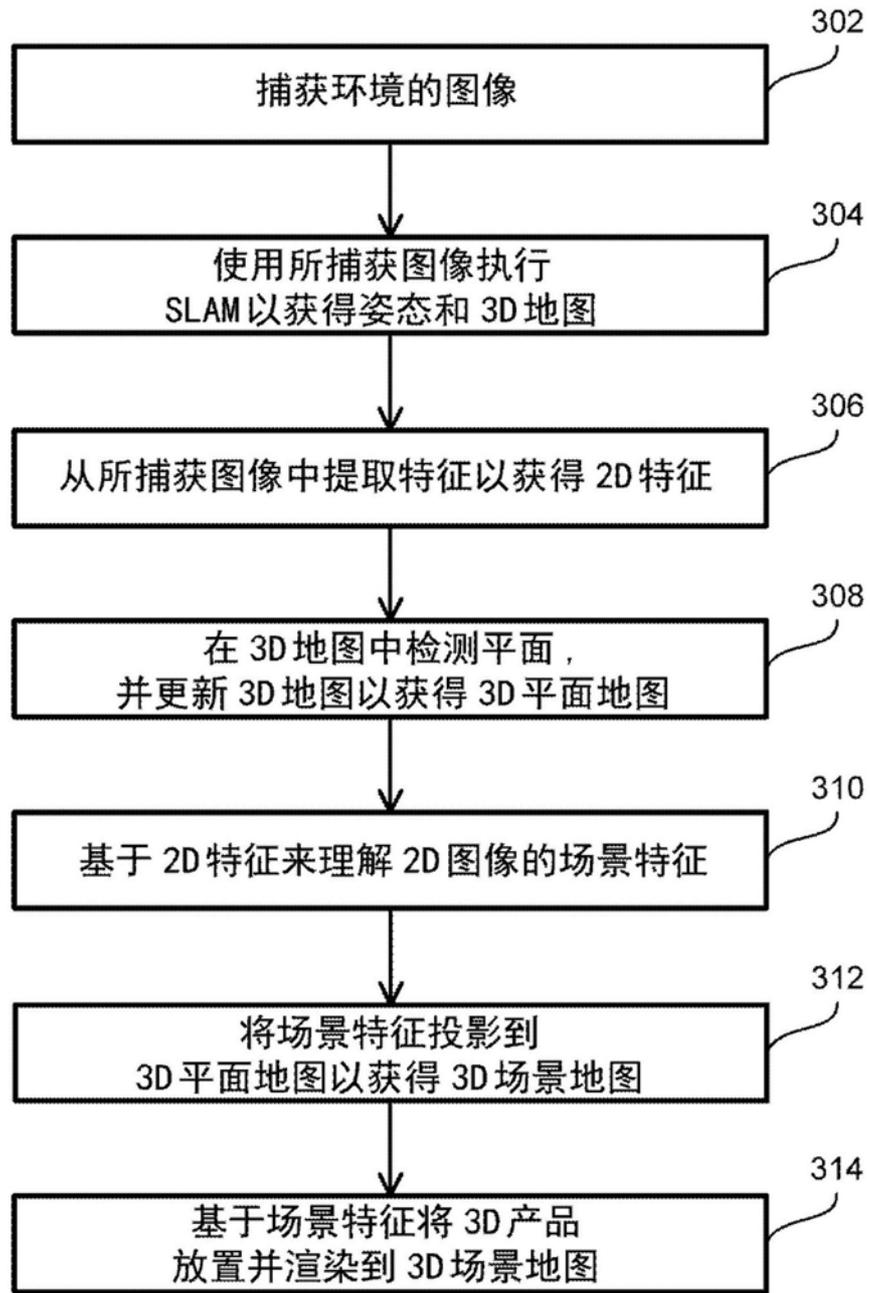


图3

400

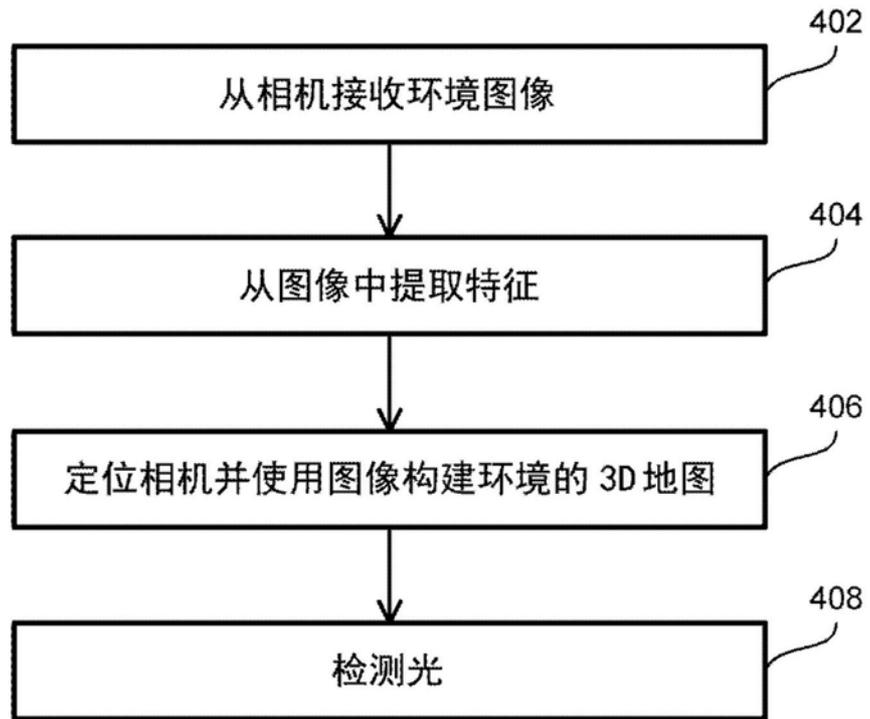


图4

500A

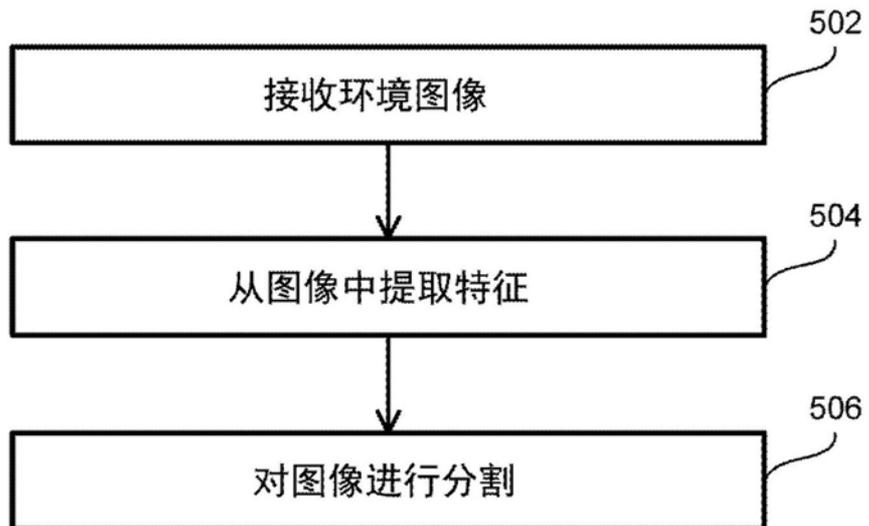


图5A

500B

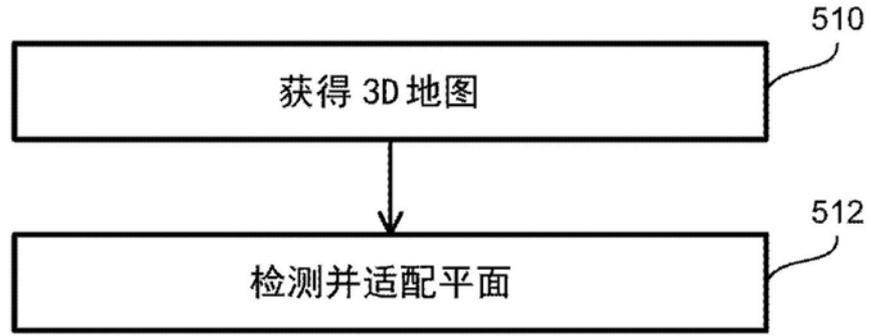


图5B

600A



图6A

600B

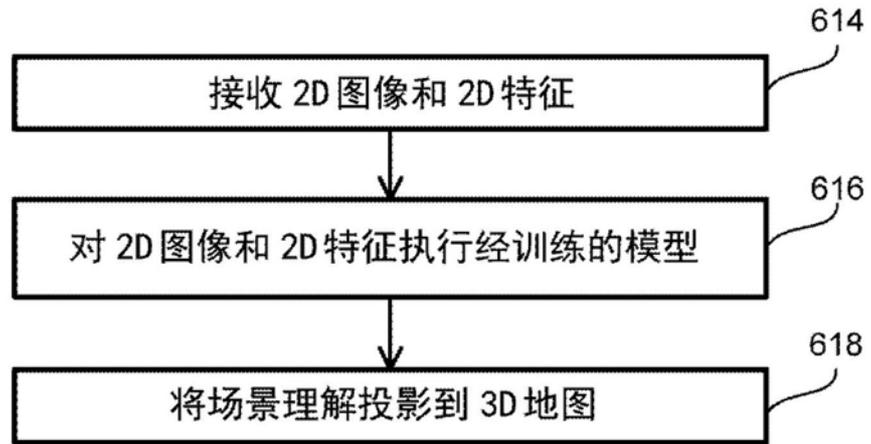


图6B

700

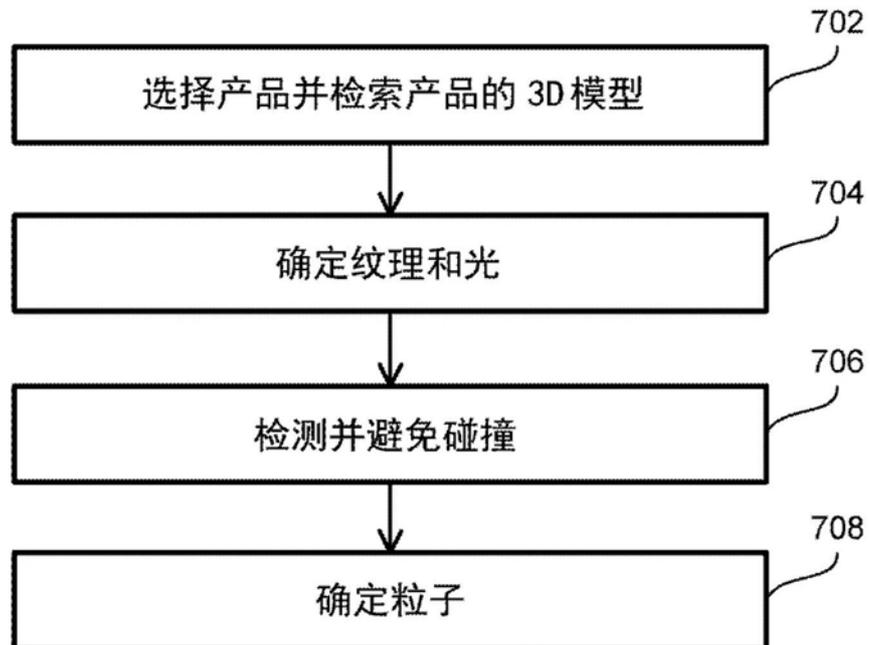


图7

800

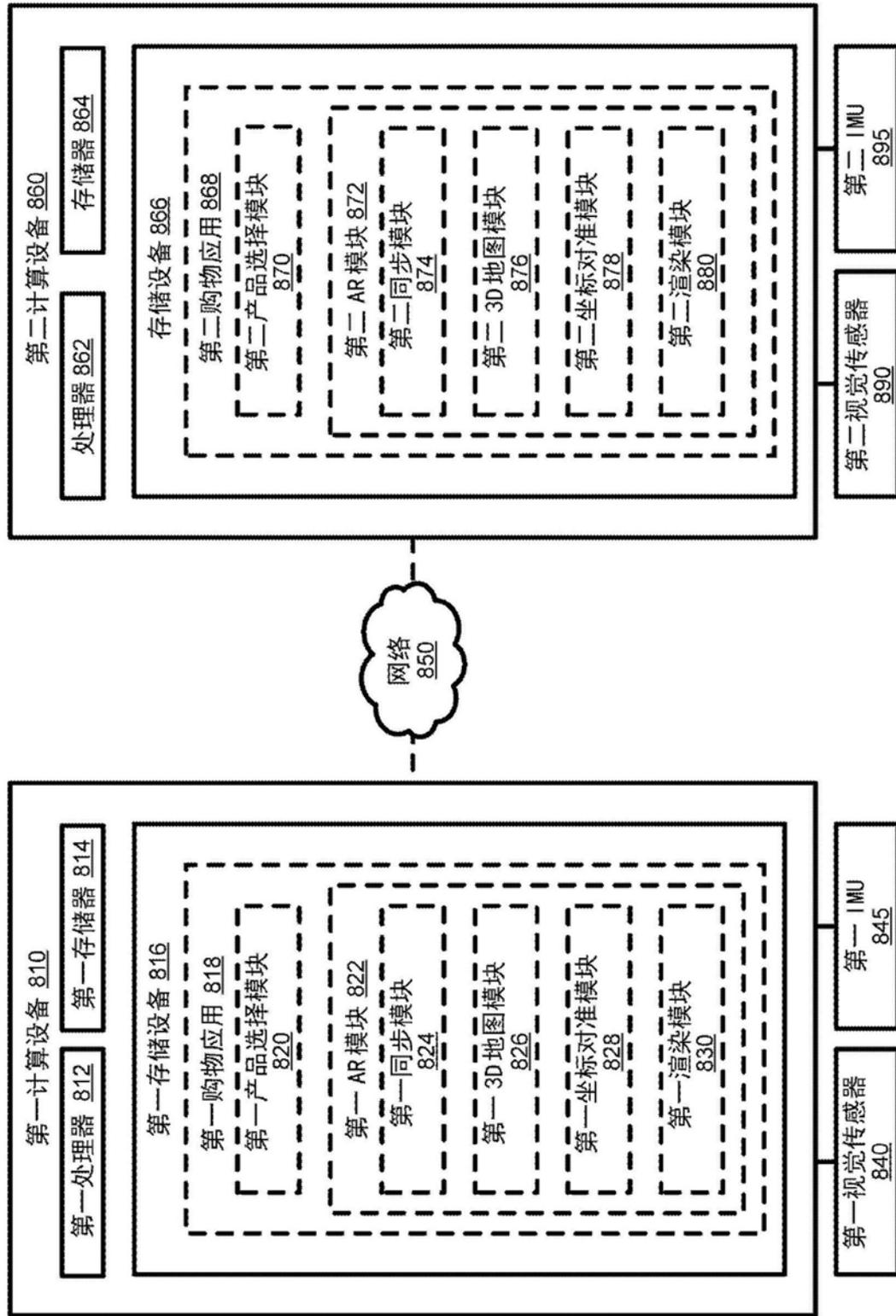


图8

900

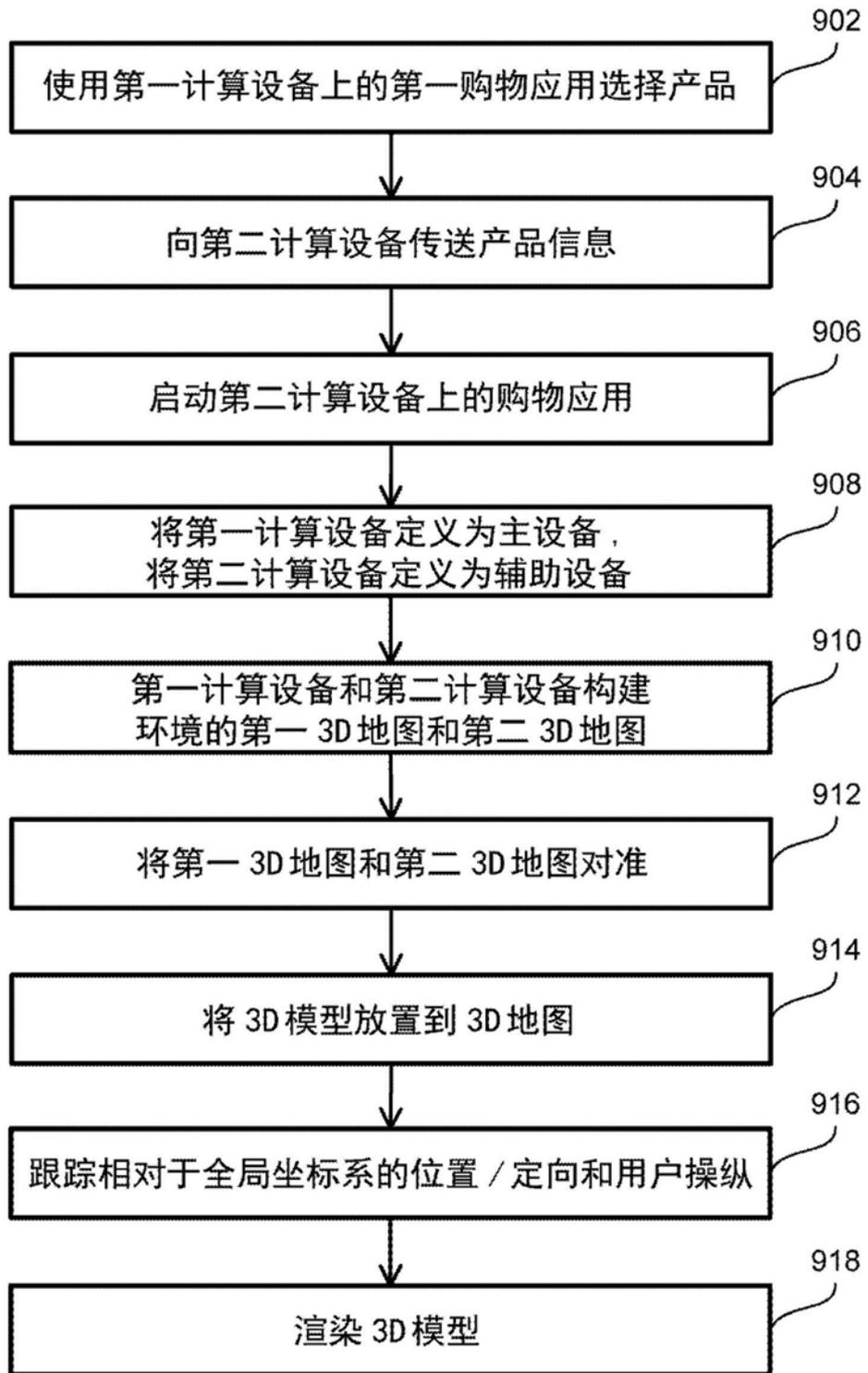


图9