



(12)发明专利

(10)授权公告号 CN 105144247 B

(45)授权公告日 2018.10.12

(21)申请号 201380065265.3
 (22)申请日 2013.12.12
 (65)同一申请的已公布的文献号
 申请公布号 CN 105144247 A
 (43)申请公布日 2015.12.09
 (30)优先权数据
 13/711,639 2012.12.12 US
 (85)PCT国际申请进入国家阶段日
 2015.06.12
 (86)PCT国际申请的申请数据
 PCT/US2013/074807 2013.12.12
 (87)PCT国际申请的公布数据
 W02014/093706 EN 2014.06.19
 (73)专利权人 微软技术许可有限责任公司
 地址 美国华盛顿州
 (72)发明人 F·杨 M·龚 Z·刘 X·童
 (74)专利代理机构 上海专利商标事务所有限公
 司 31100
 代理人 顾嘉运

(51)Int.Cl.
 G06T 17/20(2006.01)
 G06K 9/00(2006.01)
 G06K 9/62(2006.01)
 (56)对比文件
 US 2012293635 A1,2012.11.22,
 Jing Tong et al.Scanning 3D Full
 Human Bodies Using Kinects.《IEEE
 TRANSACTIONS ON VISUALIZATION AND
 COMPUTER GRAPHICS》.2012,第18卷(第4期),
 Brett Allen et al.The space of human
 body shapes: reconstruction and
 parameterization from range scans.《ACM
 Transaction on Graphics(TOG)》.2003,第22卷
 (第3期),
 Michael Zollhofer et al.Automatic
 reconstruction of personalized avatars
 from 3D face scans.《COMPUTER ANIMATION
 AND VIRTUAL WORLDS》.2011,
 审查员 王佳楠

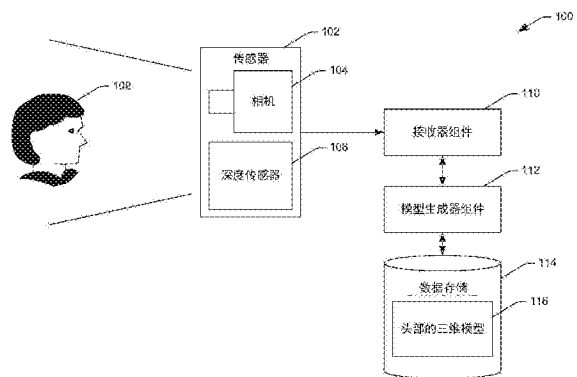
权利要求书2页 说明书10页 附图6页

(54)发明名称

用户的三维表示的生成

(57)摘要

本文描述了涉及生成用户头部/脸部的相对准确的虚拟三维模型的技术。从深度传感器接收深度帧并且从相机接收彩色帧,其中这些帧捕捉用户的头部。基于深度帧和彩色帧,生成用户头部的三维模型。



1. 一种方便构造用户的头部的计算机实现的三维表示的方法,所述方法包括:

从相机接收所述用户的头部的多个RGB帧,所述多个RGB帧是由所述相机在一时间范围内捕捉到的;

从深度传感器接收多个深度帧,所述深度帧指示所述用户的头部的各个部分离所述深度传感器的距离,所述深度帧是由所述深度传感器在所述时间范围内生成的,其中所述用户的头部相对于所述相机和所述深度传感器的定向在所述时间范围内变化;

标识所述多个RGB帧中的所述用户的头部的至少一个特征,所述至少一个特征是以下各项之一:所述用户的眼睛的中心、所述用户的鼻子的中心、所述用户的第一鼻翼或所述用户的第二鼻翼;

至少部分地基于所述多个RGB帧中所标识的所述用户的头部的至少一个特征来使所述多个深度帧中的各个深度帧彼此对齐;

至少部分地基于所述多个深度帧彼此对齐以及对所述RGB帧中的所述用户的头部的至少一个特征的标识来生成三维点云;

基于所述三维点云来生成所述用户的头部的三维网格;以及

至少部分地基于所述多个RGB帧来纹理化所述三维网格。

2. 如权利要求1所述的方法,其特征在于,所述RGB帧的分辨率在长度上小于1280个像素,且在高度上小于960个像素。

3. 如权利要求1所述的方法,其特征在于,所述深度帧的分辨率小于1280个像素长度乘960个像素高度。

4. 如权利要求1所述的方法,其特征在于,所述多个RGB帧以及所述多个深度帧在所述用户的头部旋转时被捕捉到。

5. 如权利要求1所述的方法,其特征在于,进一步包括:

响应于接收到来自所述多个RGB帧的至少一个RGB帧以及来自所述多个深度帧的至少一个深度帧,估计所述用户的头部的三维姿势;以及

至少部分地基于对所述用户的头部的三维姿势的估计来向所述用户输出指令。

6. 如权利要求1所述的方法,其特征在于,还包括在纹理化所述三维网格后自动地动画化所述用户的三维模型。

7. 如权利要求1所述的方法,其特征在于,进一步包括:

在接收到来自所述深度传感器的所述多个深度帧后,选择PCA基础模型;以及

至少部分地基于来自所述深度传感器的所述多个深度帧以及对所述RGB帧中的所述用户的头部的至少一个特征的标识来使得所述PCA基础模型变形。

8. 一种用于生成用户的三维表示的系统,包括:

处理器;以及

包括能由所述处理器执行的多个组件的存储器,所述多个组件包括:

接收以下各项的接收器组件:

由RGB相机在一时间范围内捕捉到的多个RGB帧,所述多个RGB帧中的每一个RGB帧都包括用户的头部的图像;以及

由深度传感器在所述时间范围内捕捉到的多个深度帧,所述多个深度帧中的每一个深度帧都包括用户的头部的深度图,其中所述用户的头部相对于所述RGB相机和所述深度传

传感器的定向在所述时间范围内变化;以及

模型生成器组件,所述模型生成器组件:

标识所述多个RGB帧中的所述用户的脸部特征;

至少部分地基于所述多个RGB帧中所标识的所述用户的脸部特征来使所述多个深度帧中的各个深度帧彼此对齐;以及

至少部分地基于彼此对齐的所述深度帧来生成所述用户的头部的动画化的三维模型。

9. 如权利要求8所述的系统,其特征在于,所述系统包括游戏控制台。

10. 如权利要求8所述的系统,其特征在于,所述模型生成器组件包括:

分析每一个RGB帧并标识其中的所述用户的脸部特征的脸部跟踪器组件,其中所述脸部特征包括以下各项中的至少一个:所述用户的眼睛的中心、所述用户的鼻子的中心、所述用户的第一鼻翼或所述用户的第二鼻翼;以及

网格构造器组件,所述网格构造器组件至少部分地基于所述多个深度帧来选择表示三维头部的模板线网格,并且至少部分地基于所述脸部跟踪器组件标识的所述用户的脸部特征来修改所述三维头部。

11. 一种计算机系统,包括用于执行如权利要求1-7中的任一项所述的方法的装置。

12. 一种具有指令的计算机可读存储介质,所述指令在被执行时使机器执行如权利要求1-7中的任一项所述的方法。

用户的三维表示的生成

[0001] 背景

[0002] 相对新近地,人们已经利用数字化身来在计算机环境中表示他们自己。数字化身的创建通常通过利用在线特征库来完成,其中用户能够从多个不同特征中进行选择以生成一般表示该用户的化身。例如,与生成用户的虚拟表示相关地,用户可以选择特定头形、特定肤色、预定义形状的眉毛等。该过程是相对耗时的,并且经常导致创建不表示个体的实际外貌的化身。

[0003] 已经开发出用于自动生成用户的虚拟化三维表示的系统。然而,这些系统是不准确、复杂和/或昂贵的。例如,已经开发出使用红-蓝-绿(RGB)视频流来生成用户脸部的虚拟三维模型的系统。用户被要求在相机前旋转他或她的头部,并且基于视频来生成该用户脸部的虚拟三维模型。然而,这一系统遭受相对糟糕的质量,因为它无法捕捉到在人脸中找到的有细微差别的深度。

[0004] 另一示例性系统利用若干激光测量单元来获取用于生成用户脸部的虚拟三维表示的相对密集的点云。然而,激光测量单元是相对昂贵的。因此,这一系统不适用于普通消费者。此外,这一系统不容易被运输到另一位置,并且缺乏允许将所得点云变换成可动画化的化身的底层技术。

[0005] 概述

[0006] 以下是在本文详细描述的主题的简要的概述。本概述不旨在是关于权利要求的范围的限制。

[0007] 本文描述了涉及利用日用RGB相机和深度传感器来以半自治方式生成用户头部/脸部的相对准确的虚拟三维模型的各种技术。在一示例性实施例中,这一三维模型可被用作一个人的虚拟化表示,该虚拟化表示可以在在线/游戏环境中标识这个人。该三维模型能够相对准确地表示用户头部的形状、用户头部的大小、用户的肤色以及用户的脸部特征,包括眼睛、鼻子、耳朵、脸颊线、嘴巴等。

[0008] 为了生成用户头部/脸部的相对准确的模型,在一段时间范围内从彩色相机接收多个RGB帧,并且在该时间范围内从深度传感器接收多个深度帧。相机和深度传感器两者都被配置成捕捉用户头部/脸部的图像。对RGB帧执行脸部跟踪算法以标识在这些RGB帧中捕捉到的脸部中的特征。这些特征可包括但不限于:用户眼睛的中心、用户鼻子的中心和用户的鼻翼。可基于深度帧和RGB帧中标识的特征来生成点云,其中RGB帧中的特征被映射到深度帧以定位这些特征在三维空间中的位置,并且这些特征后续被用来对齐深度帧。点云中的点表示用户头部/脸部在这些点处的离深度传感器的相应距离。

[0009] 随后,可使人类头部形状的预先存在的模型变形以适合三维点云(用脸部特征标记)以表示用户头部/脸部的形状。可以对这一三维模型应用纹理以使该虚拟三维模型准确地表示用户。可以对RGB帧应用降噪和皮肤估计技术,以便为用户的头部/脸部的虚拟三维模型提供审美上令人愉快的纹理。例如,可以利用针对RGB帧的诸如自动曝光调整和自动降噪等技术来纹理化(texturize)用户的头部/脸部的三维模型。从RGB帧中生成的纹理图然后可被映射到三维模型,以使得脸部和颜色特征在三维模型上适当地对齐。

[0010] 可利用日用传感器单元来采用上述技术。例如,这一传感器单元可以是耦合到诸如游戏控制台等计算设备的外围设备。在另一示例中,传感器单元可被包括在诸如台式计算设备、平板计算设备或便携式电话等计算设备中。因此,例如相机和/或深度传感器的分辨率可能相对较低。然而,利用本文描述的技术,能够生成用户头部/脸部的相对准确的虚拟三维表示。

[0011] 在阅读并理解了附图和描述后,可以明白其他方面。

[0012] 附图简述

[0013] 图1是方便生成用户头部/脸部的相对准确的虚拟三维模型的示例性系统的功能框图。

[0014] 图2是用于生成用户头部/脸部的相对准确的虚拟三维模型的模型生成器组件的功能框图。

[0015] 图3是示出用于构造用户头部/脸部的虚拟三维网格的示例性方法的流程图。

[0016] 图4是示出用于生成表示用户脸部纹理的纹理图的示例性方法的流程图。

[0017] 图5是示出用于构造用户头部/脸部的相对准确的虚拟三维模型的示例性方法的流程图。

[0018] 图6是示例性计算系统。

[0019] 详细描述

[0020] 现在将参考附图来描述涉及构造用户的虚拟三维头部/脸部的各种技术,在全部附图中相同的附图标记表示相同的元素。另外,本文出于解释的目的示出并描述了各示例性系统的若干功能框图;然而可以理解,被描述为由特定系统组件执行的功能可以由多个组件来执行。类似地,例如,一组件可被配置成执行被描述为由多个组件执行的功能。另外,如此处所用的,术语“示例性”旨在表示用作某些事物的图示或示例,而不意图指示优选。

[0021] 如本文所使用的,术语“组件”和“系统”旨在涵盖被配置有计算机可执行指令的计算机可读数据存储,这些指令在被处理器执行时使得特定功能得以执行。计算机可执行指令可包括例程、功能等。还应理解,组件或系统可以位于单个设备上或分布在若干设备上。

[0022] 现在参考图1,示出了方便构造用户头部/脸部的相对准确的虚拟三维模型的示例性系统100。系统100包括传感器单元102,该传感器单元包括捕捉彩色(RGB)图像的相机104和捕捉深度图像的深度传感器106。在一示例性实施例中,相机104捕捉到的图像的分辨率可能相对较低;诸如举例而言小于或等于1280个像素乘960个像素。例如,相机104捕捉到的图像的分辨率可以是640个像素乘480个像素。类似地,深度传感器106所生成的深度图像的分辨率可能相对较低;诸如举例而言小于或等于1280个像素乘960个像素。在另一示例中,深度传感器106所生成的深度图像的分辨率可以是640个像素乘480个像素。在图1所示的示例中,相机104和深度传感器106被容纳在传感器单元102中。然而,应当理解,相机104和深度传感器106可被包括在单独的外壳中。此外,传感器单元102可被包括在可以耦合到计算设备的外围设备中,诸如耦合到游戏控制台的传感器。在另一示例性实施例中,传感器单元102可被包括在诸如个人计算设备、膝上型计算设备、移动电话、平板计算设备等计算设备中。因此,例如,相机104可以是web摄像头。另外,深度传感器106可利用任何合适的技术来生成深度图像。例如,深度传感器106可利用基于红外线的深度传感系统、飞行时间深度传感系统或任何合适的深度传感技术。

[0023] 如图1所示,相机104和深度传感器106被对齐以便分别捕捉用户108头部/脸部的RGB图像和深度图像。例如,对应于相机104捕捉到的图像中的用户108头部/脸部的分辨率可低至50个像素乘50个像素。类似地,深度传感器106捕捉到的图像中的用户108头部/脸部的分辨率可低至 100×100 个像素。因此,可以理解,传感器单元102可以是日用传感器单元,以使得相机104和深度传感器106可以是相对不昂贵的、便携的并且可适于用户108的不同环境。

[0024] 系统100包括接收器组件112,该组件在一时间范围内接收相机104捕捉到的多个RGB图像/帧。多个RGB图像/帧中的每一个RGB图像/帧都包括用户108头部/脸部的图像。接收器组件110还接收在该时间范围内深度传感器106捕捉到的多个深度图像/帧。类似于RGB图像/帧,多个深度图像/帧中的每一个深度图像/帧都包括对应于用户108头部/脸部的深度图。在一示例性实施例中,相机104和深度传感器106可以与共同的时钟进行通信,以使得深度相机104和深度传感器106在时间上彼此同步。因此,相机104生成的每一个RGB图像/帧在时间上与深度传感器106生成的相应深度图像/帧对齐。在另一示例性实施例中,接收器组件110可以被配置成在时间上使相机104输出的RGB帧与深度传感器106输出的深度帧同步。

[0025] 模型生成器组件112与接收器组件110通信并接收同步的相机104捕捉到的RGB图像和深度传感器106捕捉到的深度图像。如将在下文中更详细地描述的,模型生成器组件112被配置成至少部分地基于相机104输出的RGB帧与深度传感器106输出的深度帧来生成用户108头部/脸部的相对准确的虚拟三维模型。系统100包括数据存储114,该数据存储114可保存模型生成器组件112生成的用户108头部/脸部的虚拟三维模型116。三维模型116可以在其中期望虚拟化用户108头部/脸部的各种环境中使用。例如,三维模型116可被用来表示诸如在线游戏环境等在线环境中的用户108。即,三维模型116可被用作用户108的化身。另外,三维模型116可被动画化。例如,三维模型116可被用来映射到用户表情,这使用RGB帧中所标识的特征作为指导。在另一个示例中,三维模型116可使用预定义表情来动画化。在这一示例性实施例中,三维模型116中的特征位置的知识可以在动画化该三维模型时使用。简而言之,当用户108的头部/脸部处在相对于传感器单元102的各种不同的可能位置时,模型生成器组件112可基于相机104输出的RGB帧和深度传感器106输出的深度帧来生成三维模型116。在第一示例性实施例中,用户108最初可将她的头部定向在相对于传感器单元102向左偏移大约 20° (从用户的角度)。当用户108的头部被这样定位时,相机104和深度传感器106可以分别捕捉用户108头部/脸部的多个RGB图像和多个深度图像(大约40个)。随后,用户108可以直接看着传感器单元102。当用户108的头部/脸部以这一方式定向时,可由相机104和深度传感器106分别捕捉到大约40个RGB图像和40个深度图像。之后,用户可将她的头部定向在相对于传感器单元102向右偏移大约 20° (从用户的角度),并且可由相机104和深度传感器106分别捕捉到同样的大约40个RGB图像和40个深度图像。接收器组件110可接收对应于用户108的头部/脸部相对于传感器单元102的不同定向的这些图像,并且模型生成器组件112可基于这些RGB图像和深度图像来构造用户108的头部/脸部的三维模型116。

[0026] 在另一示例中,用户108可被指示最初将她的头部/脸部定向在相对于传感器单元102向左大约 20° ,并且然后缓慢地且以连续的方式将她的脸部转向传感器单元102并且继续直到她的脸部被定向在相对于传感器单元102向右偏移 20° 。即,用户108在传感器单元

102前从左到右相对缓慢地旋转将她的头部/脸部,以使得相机104和深度传感器106两者都能够分别捕捉到大约100个RGB图像和深度图像。接收器组件110可接收到这些帧并且模型生成器组件112可基于这些RGB图像和深度图像来构造用户108的脸部/头部的三维模型116。

[0027] 在又一示例性实施例中,用户108可以直接看着传感器单元102,并且RGB图像和深度图像可单独地基于这些图像来捕捉。可以理解,使用用户108的头部/脸部相对于传感器单元102的单个位置来捕捉RGB图像和深度图像可导致用户108头部/脸部的所产生的三维模型116的准确性降低。

[0028] 如将在下文中更详细地描述的,模型生成器组件112可以对传感器单元102生成的RGB图像和深度图像执行各种动作,以生成用户108的头部/脸部的三维模型116。如上所述,因为深度传感器106可能生成噪声相对较多的深度图像(也被称为深度图),所以模型生成器组件112可被配置成对齐将用于生成用户108的头部/脸部的三维模型116的深度图像。另外,可分析相机104生成的RGB图像以获取其中的用户108的脸部/头部的特征。这些特征可包括例如,用户108的眼睛的中心、用户108的鼻子的中心、用户108的鼻子的左翼和右翼等。对于每一个RGB图像/深度图像对,特征位置可以从RGB图像映射到相应的深度图像,并且深度图像中的特征位置可以用于使得深度图像彼此对齐。这可导致生成三维点云,该点云可由模型生成器组件112用来生成用户108头部/脸部的三维模型116。该三维点云进而被用来使得现有的三维网格变形以构造相对准确地反映头部/脸部在三个维度中的形状的网格。

[0029] 另外,模型生成器组件112可至少部分地基于RGB图像和RGB图像中所标识的纹理到用户108的头部/脸部的三维网格的映射来纹理化用户108的头部/脸部的三维网格。例如,在其中RGB图像和深度图像的三个单独集合由相机104和深度传感器106分别生成的示例性实施例中,当用户108的头部/脸部被定向在相对于传感器单元102的三个不同位置时,这三个集合中的RGB图像中的纹理可被混合并映射到三维网格,以使得RGB图像中所标识的特征与三维线网格(wire mesh)适当地对齐并且对该三维线网格应用适当的颜色。在纹理化用户108头部/脸部的三维网格之前,可经由自动曝光调整和降噪来提高纹理质量。此外,可基于RGB图像来估计用户108的肤色,并且可以为未在RGB图像中捕捉到的头部/脸部的各部分合成脸部纹理以便为数据存储114中保存的所得三维模型116提供一致的外观。

[0030] 同样,用户108的头部/脸部的三维模型116可以在其中可期望虚拟地表示用户108的任何环境中使用。这些场景包括在线游戏环境、离线游戏环境、诸如在即时消息收发应用中作为化身以便通信、增强现实环境、虚拟化锻炼或理疗程序等。

[0031] 现在参考图2,示出了对模型生成器组件112的示例性描绘。如上所述,模型生成器组件112从相机104接收RGB帧以及从深度传感器106接收深度帧204。RGB帧202中的每一个RGB帧都可以在时间上与来自深度帧204的相应深度帧同步。模型生成器组件112包括接收RGB帧202和可任选的深度帧204的脸部跟踪器组件206。脸部跟踪器组件206标识每一个RGB帧中的特定特征,其中这些特征包括但不限于:用户108的眼睛的中心、用户108的鼻子的中心、用户108的左右鼻翼等。基于对这些特征的标识,脸部跟踪器组件206可以针对每一个RGB帧估计用户108的姿势(例如,用户108的头部/脸部相对于传感器单元102的位置)。脸部跟踪器组件206可任选地使用深度帧来提高特征识别准确性。脸部跟踪器组件206的输出可以是RGB帧中所标识的特征的位置以及对每一个RGB帧中的用户108的姿势的估计。

[0032] 模型生成器组件112还包括校准器组件207,该校准器组件接收脸部跟踪器组件206输出的RGB帧202中的特征的位置以及RGB帧202和深度帧204,并且经由将RGB帧中标识的特征映射到对应的深度帧(或者将对应的深度帧映射到RGB帧)来标识这些特征在三维坐标系中的位置。

[0033] 在一示例性实施例中,传感器单元102可以本地支持将深度帧映射到RGB帧,但可能不支持将RGB帧中的点映射到深度帧。因此,为了确定每一帧中的每一个特征在三维坐标系中的位置,校准器组件204可以针对每一深度帧基于相应的深度帧的深度坐标来生成三角形。校准器组件207可将每一个三角形从深度帧映射到其对应的RGB帧。校准器组件207然后可以为RGB帧中的每一个所标识的特征从相应的特征所属的深度帧中标识三角形。随后,校准器组件207可计算从所标识的特征点到每一个三角形顶点的距离。基于这些距离,可由校准器组件207经由离各三角形顶点的距离的插值来为每一个特征计算三维坐标点。

[0034] 为了降低计算复杂性,校准器组件207可以在标识所标识的特征所属的三角形时首先定位最靠近所标识的特征的顶点。可以推断出,所标识的特征属于具有这一顶点的三角形。因此,校准器组件207的输出可以是每一个RGB帧中的每一个所标识的特征在三维坐标系中的坐标。

[0035] 模型生成器组件112还包括被用来累积深度数据并生成三维点云的深度平滑器组件208。深度平滑器组件208接收深度帧204以及每一个RGB帧中的由脸部跟踪器组件206标识的特征及其在三维空间中的相应位置。

[0036] 深度平滑器组件208至少部分地基于校准器组件207计算出的三维空间中的坐标来累积深度数据。更具体而言,在一示例性实施例中,深度平滑器组件208可通过对齐深度帧204来累积深度数据。为了执行帧对齐,深度平滑器组件208可利用交互式最近点(ICP)算法。关于每一深度帧中的每一点的法线信息由深度平滑器组件208用来获取相对准确的深度数据。深度平滑器组件208可以在创建两个深度帧之间的对应关系时采用双向映射,以确保丢弃边界点。

[0037] 同样,深度平滑器组件208可利用校准器组件207输出的脸部特征坐标来计算两个深度帧之间的粗略刚性变换。这一刚性变换可被用作上述ICP算法的初始值。用于粗略刚性变换的特征点可以是眼睛中心、鼻子中心和两个鼻翼。

[0038] 更具体而言,深度帧204包括第一深度帧,并且立方体坐标系可基于用户108的头部在该第一深度帧中的位置来建立。该第一帧可以与第二帧(在时间上在第一帧后立即被捕捉到)对齐以创建累积结果。每一后续帧都能够首先与当前累积结果对齐,并且之后经转换的点可被映射到这一立方体累积坐标系并且用距离权重计及附近像素。在对针对特定姿势的所有深度帧采取累积以生成累积的结果之后,深度平滑器组件208可利用另一ICP算法来用初始深度帧映射累积的结果,这能够减少在累积过程期间传播的差错。

[0039] 如上所述,在一示例性实施例中,可以捕捉到对应于用户108头部的三个不同位置的单独的三组帧。来自这三个单独视图的帧可单独由深度平滑器组件208来累积。随后,三个累积的结果可被对齐并且然后通过利用阿尔法混合来组合。阿尔法混合带可以在用户108的眼睛和嘴巴之外,这确保脸部特征全都来自具有最高准确性的前视图。累积的深度数据可被称为三维点云。

[0040] 模型生成器组件112还包括网格构造器组件210,该网格构造器组件接收深度平滑

器组件208输出的三维点云以及校准器组件207为每一个RGB帧输出的三维特征位置。另外，网格构造器组件210接收可以是三个脸部形状样本的集合的模型库。网格构造器组件210利用这一输入来生成用户108头部/脸部的相对准确的三维线网格。网格构造器组件210可构造用户108头部/脸部的分量分析 (PCA) 模型，其中这一模型指的是被表达成正交形状向量加上均值模型的线性组合的目标模型。这些形状向量和均值模型可以从三维脸部形状样本的集合离线计算。具体而言，具有相同的几何拓扑结构的预定义的三维脸部形状样本可被用来计算PCA基础和均值模型。

[0041] 网格构造器组件210可将PCA均值模型与深度平滑器组件208输出的累积深度数据对齐。网格构造器组件210可进行PCA均值模型和累积深度数据之间的粗略刚性变换，该变换可被用来将PCA基础转换成深度空间。利用从校准器组件207接收到的特征点，网格构造器组件210可执行变换矩阵计算。PCA均值模型可能已经在其中计算出这些特征点。

[0042] 关于网格构造器组件210输出的三维线网格，更具体而言，这一网格可以是呈三角形的三维网格。形状向量被表示为 $S = (v_1^t, \dots, v_n^t)$ ，其中 $v_i = (x_i, y_i, z_i)^t$ 是第 i 个顶点的三维坐标。三维脸部的变形可由如下的线性PCA来描述：

$$[0043] \quad S = \mu + \sum_{i=1}^k \alpha_i \Phi_i \quad (1)$$

[0044] 其中 μ 表示来自训练样本 (上述三维脸部形状样本) 的平均形状， Φ_i 是正交形状向量， α_i 是指示来自第 i 个形状向量的形状变形的贡献的标量值，而 k 是形状向量的总数。形状向量可以按照如由它们的对应本征值 σ_i 确定的重要性来排序。

[0045] 对于相对简单的三维可变形模型重构，网格构造器组件210可估计 α_i 以实现以下：

$$[0046] \quad E(\alpha) = \min_{\alpha} \left(\sum \left(\left\| \mu + \sum_{i=1}^k \alpha_i \Phi_i - S' \right\|^2 \right) \right) \quad (2)$$

[0047] 对于该方程， S' 是目标PCA模型需要适应的深度累积数据 (三维点云)。具体而言，因为 S' 可能不具有从用户108的耳朵到用户108头部的背面的信息，并且前额上的数据可能由于用户108头发的存在而具有干扰，所以能够手动生成蒙板以选择要执行重构的顶点。以下呈现的蒙板范围包括主要脸部特征以及大多数人的脸颊。

[0048] $v_j = (x_j, y_j, z_j)^t$ 定义用于蒙板约束的所选顶点，其中 $j \in (1, 2, \dots, n)$ ，且对应的形状向量是 $S_j = (v_1^t, \dots, v_n^t)$ ，这是 v_j 的集合。给定这些约束，PCA适应目标可被修改为如下：

$$[0049] \quad E(\alpha) = \min_{\alpha} \left(\sum (\|S_j - S'\|^2) + \rho \sum_{i=1}^k \alpha_i / \sigma_i \right) \quad (3)$$

[0050] 方程 (4) 的第二部分是正则化约束。如上所述， σ_i 是本征值，该本征值用于确保其他向量 (被认为是较不重要的) 具有与较重要的向量相比相对较小的 α_i ，以避免创建奇怪的形状。

[0051] 网格构造器组件210可以采用各种技术来标识 S' 中的对于每一 v_j 的对应点，其中 $j \in (1, 2, \dots, n)$ 。网格构造器组件210可将约束蒙板上的顶点群集成若干区域，其中这些区域基于所标识的脸部特征定义 (嘴巴、眼睛、鼻子、脸颊等)。同样，可使用不同的技术来为不同的区域找到输入的三维点云上的对应点。例如，对于PCA模型上的顶点，可定位累积点云的平面上的最近点，并且可以将源/目的地点的法线值与该平面上的这一点进行比较，以拒绝其在源/目的地法线之间的角度大于预定义阈值的点对。这能够特别有助于找到关于下

巴/前额的对应点。

[0052] 在另一示例中,网格构造器组件210可以从PCA模型上的法线方向上的顶点建立一条线,并且可标识相关于三维点云的对应关系。随后,可分析顶点法线和点云上的对应点的法线之间的角度。如果该角度大于预定义阈值,则可忽略该顶点;否则,可使用该顶点来建立对应点。该办法可以特别有助于找到对于鼻子、嘴唇和脸颊的对应点。

[0053] 在又一示例中,网格构造器组件210可选择PCA模型上的顶点并在三维点云中找到最近点。这一办法可以特别有助于找到关于头部/脸部中的除了上述那些部分之外的各部分的对应点。

[0054] 网格构造器组件210还可以在PCA模型构造期间使用二维脸部特征信息。由于二维特征点在大多数情况下提供准确的脸部特征位置,因此可使用这些特征点来约束PCA模型构造,如下:

$$[0055] \quad E(\alpha) = \min_{\alpha} \left(\sum (\|S_p - S'\|^2) + \rho \sum_{i=1}^n \alpha_i / \sigma_i + \sum_{i=1}^m \beta_i (\|W(S_p) - F'\|^2) \right), \quad (4)$$

[0056] 其中 S_p 是选自所生成模型的表示用户108的脸部特征位置的一组顶点, $W()$ 是将模型映射到二维彩色图像的映射函数, F' 是二维坐标中的目标特征点、 β_i 是第 i 个特征点的权重,其中可以为不同的特征提供不同的权重。例如,鼻子和眼睛中心可被提供较高的权重,因为不管用户表情如何它们是相对准确的并且对大多数用户而言通常是刚性的,而嘴巴特征将被提供较低的权重,因为即使对该特征的标识是准确的,其位置也可随着用户表情的变化而变化。类似地,脸颊轮廓可被提供还更低的权重(或甚至被忽略),因为难以足够准确地获取脸颊轮廓。网格构造器组件210可多次执行PCA变形以获取合适的结果。

[0057] 在一示例性实施例中,除了ICP算法之外还可采用迭代方法来估计累积的深度数据和最后一步PCA结果之间的刚性变换。PCA基础可以用这一更新后的信息来转换,且PCA模型可同样使用这一更新后的基础来分析。该迭代方法可提高PCA结果准确性,并且可以在 $E(\alpha)$ 小于预定义阈值时停止。

[0058] 在另一示例中,可以与PCA适应一起估计刚性信息,如下:

$$[0059] \quad E(\alpha, R) = \min_{\alpha, R} \left(\sum (\|R(S_p) - S'\|^2) + \rho \sum_{i=1}^n \alpha_i / \sigma_i + \sum_{i=1}^m \beta_i (\|W(R(S_p)) - F'\|^2) \right) \quad (5)$$

[0060] 其中 $R()$ 是刚性变换矩阵。

[0061] 网格构造器组件210还可以对PCA生成的模型采用拉普拉斯(Laplacian)变形以生成更准确的三维网格。例如,上述关于PCA模型的线性组合办法可以准确地描述诸如脸颊形状、脸部中的肥胖等级等大多数总体形状,以及诸如脸颊骨高度和嘴巴形状等中型特征。然而,对于诸如嘴唇沟等详细特征,线性模型可能无法准确地表示这些特征。由此,网格构造器组件210可以对PCA生成的模型采用拉普拉斯变形以使得目标适应输入的累积深度数据。

[0062] 网格构造器组件210可使用正面脸部上的顶点作为对变形的三维约束,并且可使用加权重的二维特征点映射作为对变形的二维约束。这些约束可以一起用来获取最终变形结果。同时,可使用脸部特征点信息来将头部分成若干区域,并且可将区域性权重给予三维顶点作为位置约束。对于人们更刚性且具有更少累积差错的区域将被提供更高的权重,而较不刚性且具有更多累积差错的区域(诸如嘴巴)将被提供较低的权重。网格构造器组件210的输出是用户108的头部/脸部的相对准确的呈三角形的三维网格。

[0063] 一旦生成用户108头部/脸部的三维网格,就对这一三维网格应用纹理。一般而言,纹理基于相机104捕捉到的RGB帧202,其中可采用合成来确定用户108头部/脸部中的未在图像/帧202中捕捉到的各部分的纹理。为了将输入的RGB图像用作纹理源,正确地将输入的彩色帧与网格构造器组件210生成的三维网格对齐。如上所述,三维网格与第一深度帧对齐,并且生成从每一帧到该第一帧的刚性变换,这意味着可将头部模型映射到对应于已被捕捉到并用于将颜色映射到网格的每一帧的正确姿势。

[0064] 关于纹理化三维网格,更详细地,模型生成器组件112可包括图像预处理器组件212,该预处理器组件可采用各种预处理技术来提高RGB帧202的总体质量。例如,相机104可以是输出相对较低分辨率图像的日用相机。另外,相机104捕捉到的图像的质量可能受到光照条件的影响,并且一般而言可具有相对较多的噪声。此外,在某些情况下,用户108头部/脸部上的光线可能不足,并且涉及皮肤和/或用户的其他特征的颜色噪声可存在于RGB帧202中。图像预处理器组件212可利用自适应曝光调整技术以及降噪来提高图像质量。

[0065] 结合曝光调整,图像预处理器组件212可分析输入的彩色图像并估计当前曝光分布。随后,图像预处理器组件212可应用非线性变换来使得RGB帧内的主要彩色像素被转换成适当的曝光级别。结合降噪,图像预处理器组件212可单独地分析输入的RGB帧在亮度空间和色彩空间中的边缘。这可允许计算噪声级别,并且图像预处理器组件212可单独地在两个空间中采用基于区域的平滑。

[0066] 模型生成器组件112还包括从图像预处理器组件212接收经处理的RGB帧的纹理化器组件214。纹理化器组件214可通过纹理合成来估计用户108的肤色。更具体而言,纹理化器组件214可选择RGB帧202中的与脸部跟踪器组件206所标识的特征(用户108的脸部正面上的除了眼睛、眉毛、嘴巴之外的特征)分开的各部分。在脸部的除了对应于所标识特征的部分之外的各部分然后可由纹理化器组件214通过K均值办法来群集成五个单独的类别,并且可以为这些类别计算平均颜色。纹理化器组件214对群集の利用可方便移除由过多光线以及图像的镜面和阴影部分导致的干扰。

[0067] 纹理化器组件214还可接收网格构造器组件210输出的三维网格以便对这一网格应用纹理。更详细地,纹理化器组件214可使得围绕该三维网格建立圆柱坐标系。头部的中心点可以是圆柱坐标系的原点,其中柱体的下侧与三维网格的Y平面平行。纹理化器组件214可将三维网格的每一三角形映射到柱面上。因为从三维网格三角形到原始输入彩色帧的映射已经存在,所以可将彩色图像帧转换成柱面纹理。

[0068] 此外,当用户108具有在三个不同位置(面向左、朝前看、面向右)捕捉到的深度和RGB图像时,可选择具有不同视图的三个帧并将其映射到相同的柱体纹理坐标,并且可应用阿尔法混合来组合这三个纹理。左视图和右视图可由纹理化器组件214用来提供关于脸颊的纹理信息。此外,纹理化器组件214可以在纹理化过程期间采用两个蒙板。假定超出边界的顶点在当前视图上不可见,一个蒙板可用于设置纹理映射边界。另一蒙板可用于其中混合纹理的三视图实施例并且可定义正视图可见边界。这两个蒙板都可以在三维线网格拓扑结构中预定义。

[0069] 可由纹理化器组件214在对三维网格应用纹理时考虑的变量是前额位置。由于一些用户在前额上有头发(这导致在最终化的纹理中出现噪声),因此可展示可变前额位置以使得用户能够调整前额位置以移除头发干扰。此外,纹理化器组件214可以从用户108接收

在需要时提纯颜色的输入。使用已知映射,纹理化器组件214可将适当的颜色映射到网格构造器组件210生成的三维网格中的各个位置。另外,纹理化器组件214可以对用户108头部/脸部中的未在RGB图像中捕捉到的各部分利用合成技术。这可通过以下动作来完成:利用肤色估计来填充不具有纹理信息的各部分,然后执行泊松(Poisson)滤波来平滑纹理的边界部分。

[0070] 模型生成器组件112可任选地包括可降低所得三维模型中的逼真度的纹理平滑器组件216。这可以是合乎需要的,因为定义头发的纹理是困难的并且应用于三维模型的头发的外观经常是卡通式的。因此,为了向所得三维模型提供一致的外观和感觉,纹理平滑器组件216可以选择性地平滑用户108的头部/脸部上的纹理,以产生审美上一致的三维模型。

[0071] 模型生成器组件112还可包括头发生成器组件217,该头发生成器组件生成头发以应用至网格构造器组件210生成的三维网格。头发生成器组件217从图像预处理器组件212接收经处理的图像以及从纹理化器组件214接收纹理信息,并且可选择要应用于三维网格的头发模型。例如,头发生成器组件217可以从多个模板中选择发型,其中该发型可基于用户108头部的大致形状以及用户头发的所确定的颜色(以及用户108头发的尺寸)来选择。

[0072] 渲染器组件218可任选地被包括在模型生成器组件112中,其中该渲染器组件218从网格构造器组件210接收三维网格、从纹理平滑器组件216接收要应用于三维网格的平滑纹理,以及接收头发生成器组件214选择的头发,并且输出用户108头部/脸部的三维模型的渲染。或者,渲染器组件218可使得三维模型以准备好渲染的格式被存储在数据储存库中。

[0073] 现在参考图3-5,示出并描述了各种示例性方法。尽管各方法被描述为顺序地执行的一系列动作,但可以理解,这些方法不受该顺序的次序的限制。例如,一些动作能以与本文描述的不同的次序发生。另外,动作可以与另一动作同时发生。此外,在一些情况下,实现本文描述的方法并不需要所有动作。

[0074] 此外,本文描述的动作可以是可由一个或多个处理器实现的和/或存储在一个或多个计算机可读介质上的计算机可执行指令。计算机可执行指令可包括例程、子例程、程序、执行的线程等。另外,这些方法的动作的结果可以存储在计算机可读介质中,显示在显示设备上,等等。计算机可读介质可以是任何合适的计算机可读存储设备,如存储器、硬盘驱动器、CD、DVD、闪存驱动器等。本文中所使用的术语“计算机可读介质”并不旨在涵盖传播信号。

[0075] 现在参考图3,示出了一流程图,该流程图示出用于生成用户108头部/脸部的三维网格的示例性方法。方法300开始于302,并且在304累积深度数据。如上所述,累积包括深度帧之间的帧对齐,其中采用ICP算法来进行帧对齐。另外,可使用脸部跟踪算法标识的脸部特征点来计算深度帧之间的粗略刚性变换,其中这一刚性变换可用作ICP算法的初始值。在306,将累积的深度帧与第一帧对齐。同样,为了对上述三个视图模式利用ICP算法,在304可以单独地累积这三个视图上的帧并且可对齐三个累积结果并将其与阿尔法混合组合。

[0076] 在308,调整PCA模型基础。可计算预先计算出的PCA均值模型和累积的深度数据之间的粗略刚性变换并使用该变换来将PCA基础转换成深度空间。

[0077] 在310,可生成用户108头部/脸部的PCA模型。上文已经描述PCA模型构造。

[0078] 在312,可以对所生成的PCA模型进行拉普拉斯变形。拉普拉斯变形可用于更准确地表示用户108头部/脸部的各部分。方法300在314完成。

[0079] 现在参考图4,示出了方便对用户脸部/头部的三维网格应用纹理的示例性方法400。方法400开始于402,并且在404对RGB图像执行自动曝光纠正过程。在406,对RGB图像执行降噪,并且在408分析经处理的图像以便进行肤色计算。如上所述,该肤色计算可基于用户108脸部中的不对应于所标识的特定特征的部分。

[0080] 在410,出于柱体纹理生成的目的,围绕头部的三维网格建立圆柱坐标系。在412,可以在用户输入的帮助下采取可任选的颜色调整步骤,并且在414,可进行合成以便跨用户108的头部/脸部中的未在RGB帧中捕捉到的各部分统一地应用纹理。在416,可任选地进行滤波以便向用户头部/脸部的所得三维模型提供例如与用户的头发相比一致的外观。方法400在418完成。

[0081] 现在参考图5,示出了方便生成用户脸部/头部的相对准确的三维模型的示例性方法500。方法500开始于502,并且在504从相机接收用户头部的多个RGB图像,其中该多个RGB图像由相机在某一时间范围内捕捉。在506,从深度传感器接收多个深度帧,其中这些深度帧指示用户头部的各个部分与深度传感器的距离。深度帧由深度传感器在该时间范围内生成。另外,可以在时间上同步多个RGB图像和多个深度帧。在508,标识多个RGB图像中的用户的头部的至少一个特征,其中该至少一个特征是以下各项中的一个:用户眼睛的中心、用户鼻子的中心、用户的第一鼻翼或用户的第二鼻翼。

[0082] 在510,至少部分地基于多个深度图以及对RGB图像中用户头部的至少一个特征的标识来生成用户的头部的三维网格。在512,至少部分地基于多个RGB图像来纹理化三维网格。方法500在514完成。

[0083] 现在参考图6,示出了可以根据本文公开的系统和方法使用的示例性计算设备600的高级图示。例如,计算设备600可以在支持生成用户头部/脸部的虚拟三维模型的系统中使用。计算设备600包括执行存储在存储器604中的指令的至少一个处理器602。存储器604可以是或可以包括RAM、ROM、EEPROM、闪存、或其他合适的存储器。这些指令可以是例如用于实现被描述为由上述一个或多个组件执行的功能的指令或用于实现上述方法中的一个或多个的指令。处理器602可以通过系统总线606访问存储器604。除了存储可执行指令之外,存储器604还可存储特征点、变换矩阵、RGB帧、深度帧等。

[0084] 计算设备600还包括可由处理器602通过系统总线606访问的数据存储608。数据存储可以是或可以包括任何合适的计算机可读存储,包括硬盘、存储器等。数据存储608可包括可执行指令、RGB帧、深度帧、所标识的特征点等。计算设备600还包括允许外部设备与计算设备600进行通信的输入接口610。例如,可以使用输入接口610来从外部计算机设备、用户等接收指令。计算设备600还包括将计算设备600与一个或多个外部设备进行接口的输出接口612。例如,计算设备600可以通过输出接口612显示文本、图像等。

[0085] 另外,尽管被示为单个系统,但可以理解,计算设备600可以是分布式系统。因此,例如,若干设备可以通过网络连接进行通信并且可共同执行被描述为由计算设备600执行的任务。

[0086] 注意,出于解释目的提供了若干示例。这些示例不应被解释为限制所附权利要求书。另外,可以认识到,本文提供的示例可被改变而仍然落入权利要求的范围内。

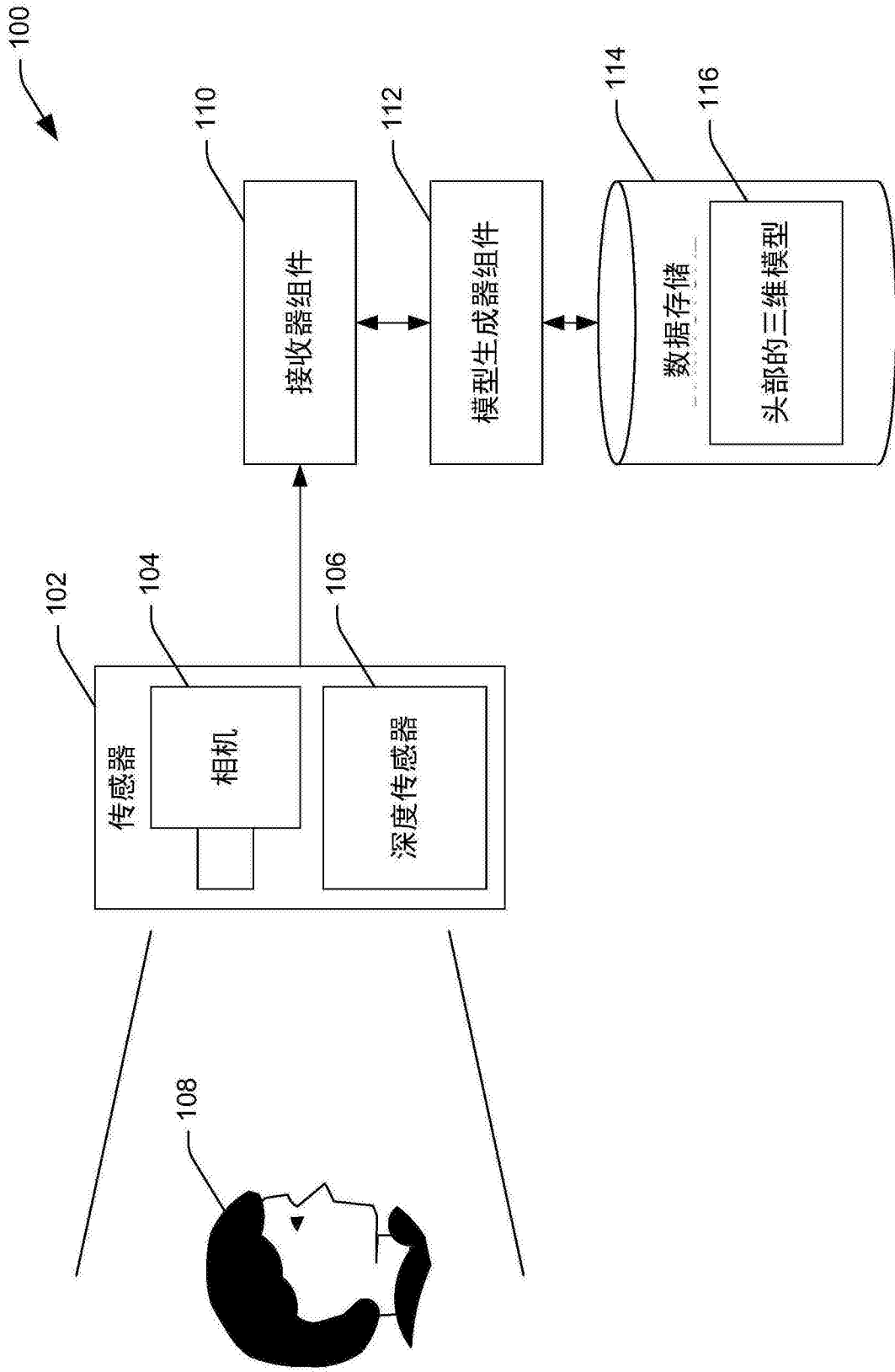


图1

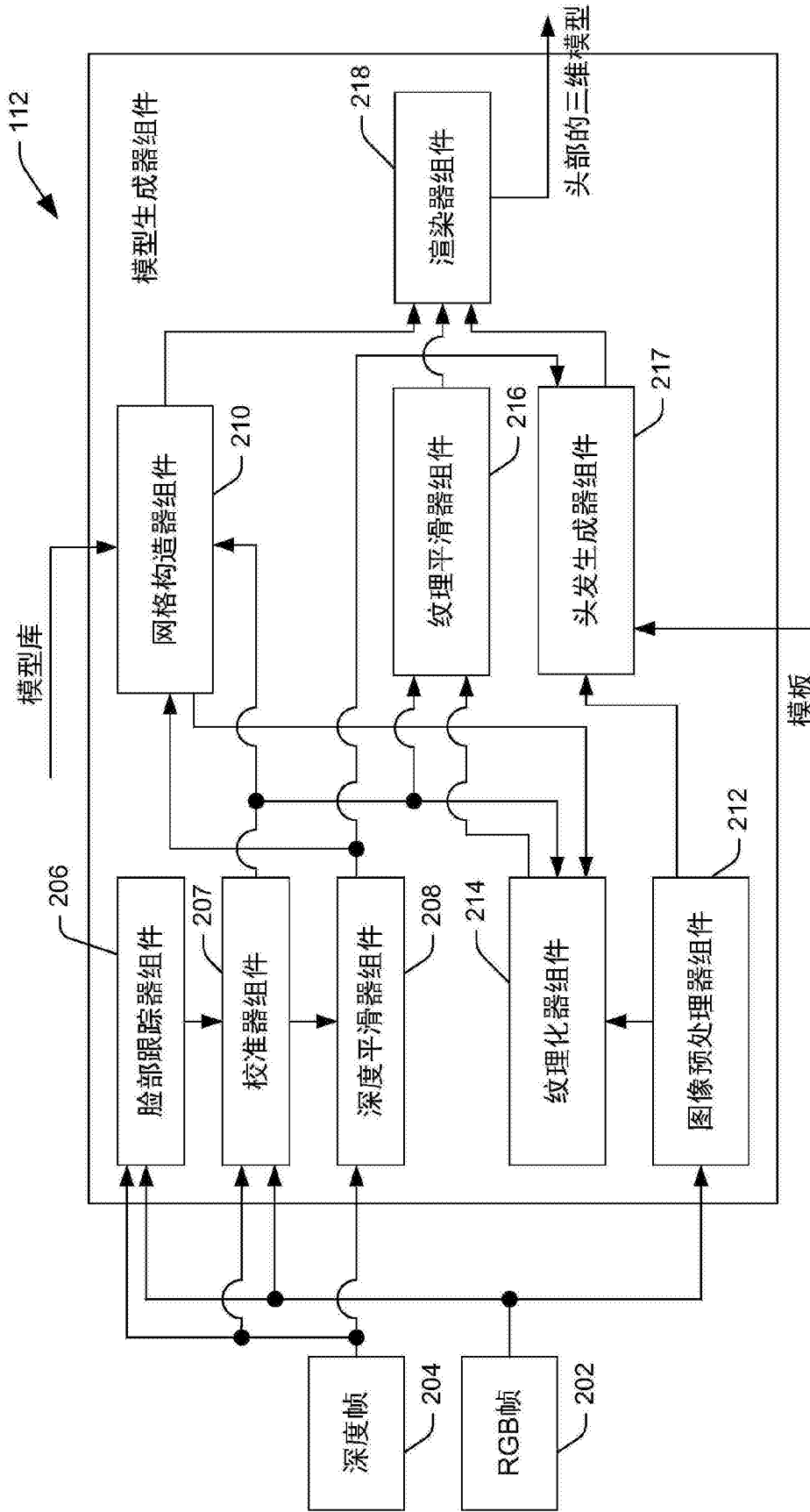


图2

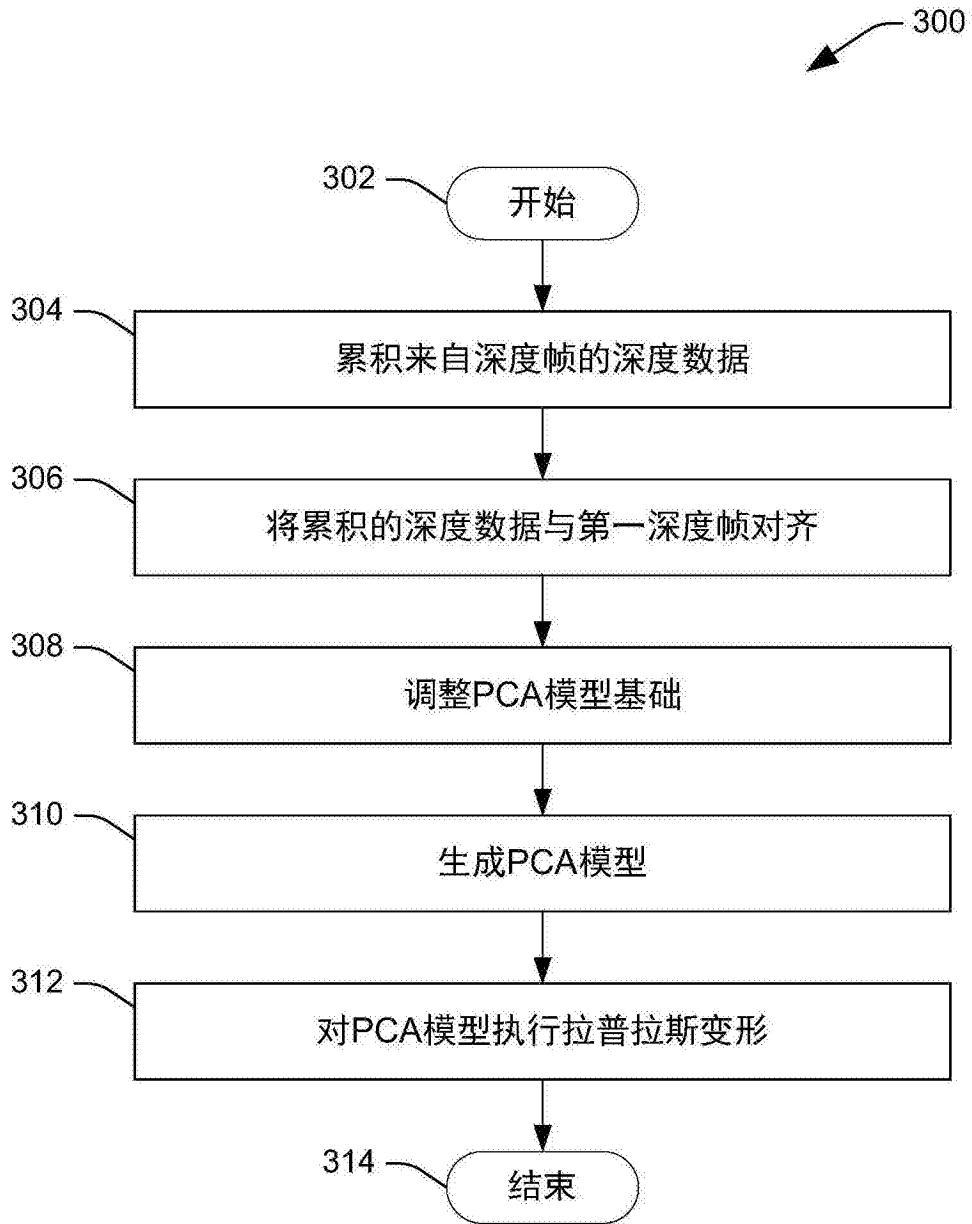


图3

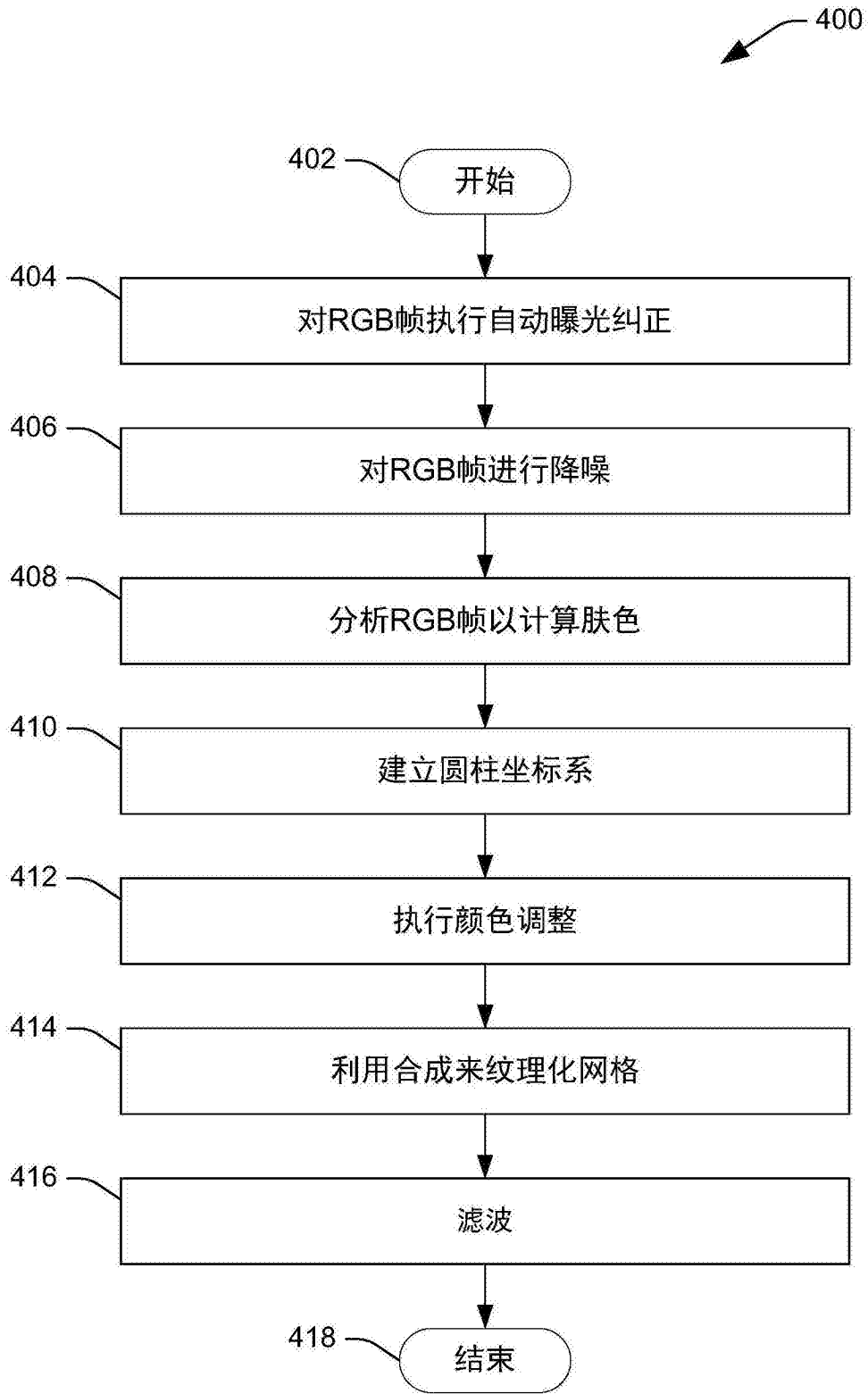


图4

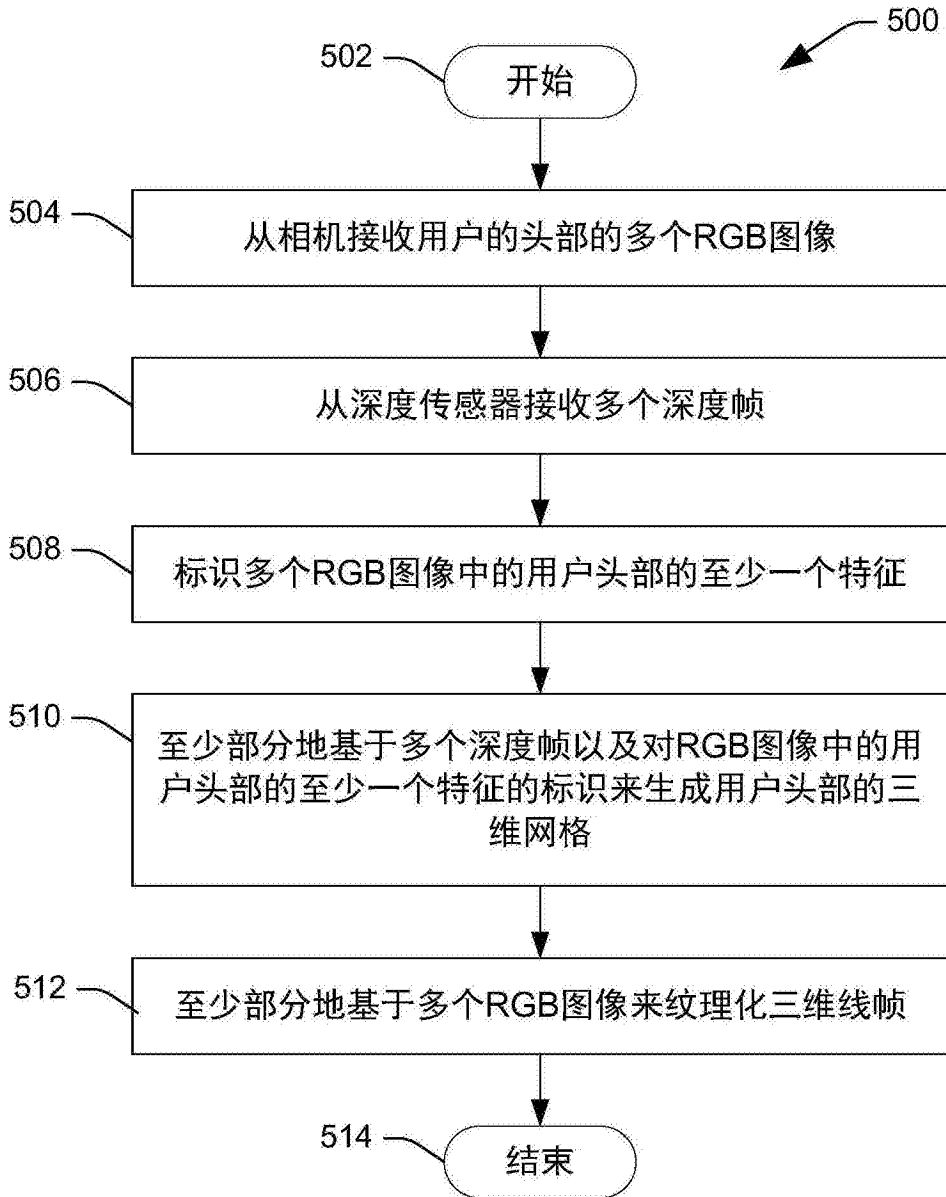


图5

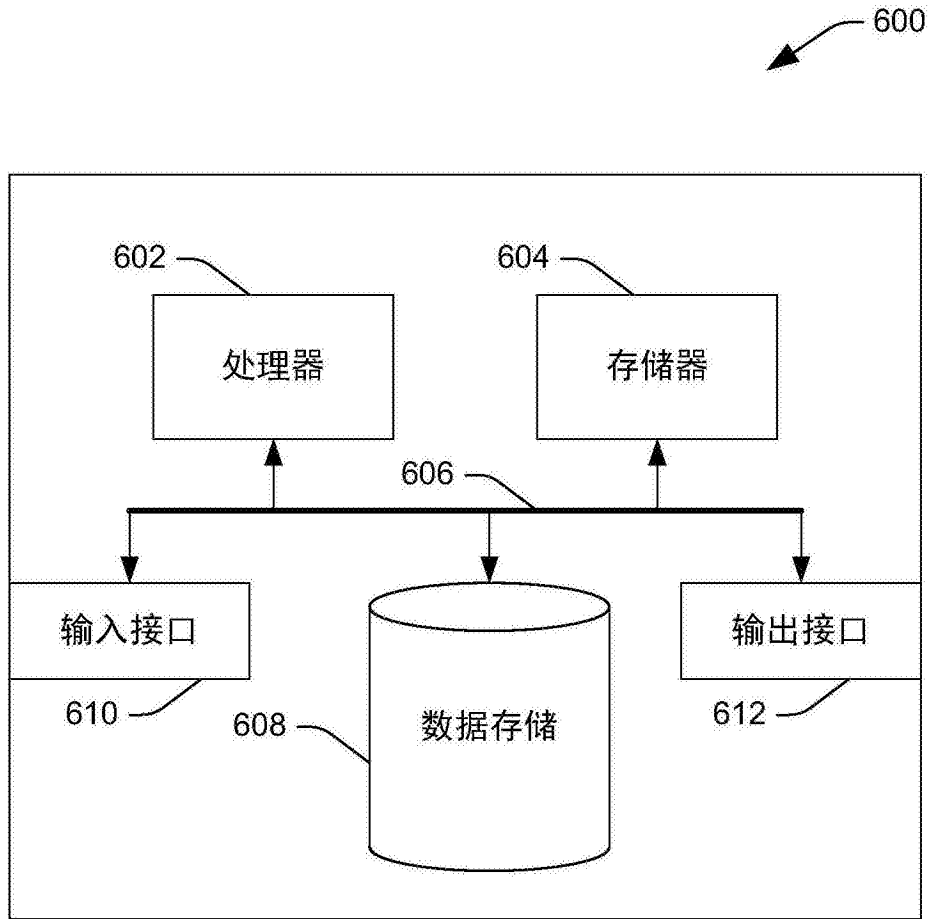


图6