(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114731457 B
(45)授权公告日 2024.07.09

(21)申请号 202080079220.1
(22)申请日 2020.09.30
(65)同一申请的已公布的文献号
申请公布号 CN 114731457 A
(43)申请公布日 2022.07.08
(30)优先权数据
16/597,698 2019.10.09 US
(85)PCT国际申请进入国家阶段日
2022.05.13
(86)PCT国际申请的申请数据
PCT/US2020/053523 2020.09.30
(87)PCT国际申请的公布数据
WO2021/071721 EN 2021.04.15

(73)专利权人 索尼互动娱乐股份有限公司
地址 日本东京都
(72)发明人 X.叶 W.贝内德托
(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 张晓明

(51)Int.Cl.
H04N 21/44(2006.01)
(56)对比文件
CN 107580196 A,2018.01.12
CN 108471510 A,2018.08.31
CN 109040781 A,2018.12.18
审查员 周立秋

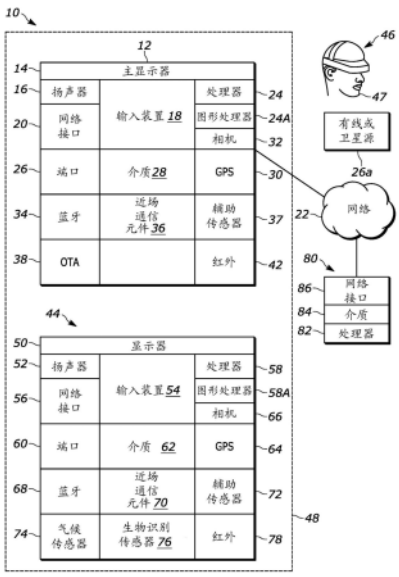
权利要求书2页 说明书11页 附图12页

(54)发明名称

使用区块链的伪造视频检测

(57)摘要

使用区块链检测视频是否是自原始视频(200)衍生并且发生更改的伪造视频(212),所述区块链禁止(912)将已发生更改的原始视频的复本添加到所述区块链,或者在所述区块链中指示经更改视频已发生更改。描述了用于确定(908、1002)试图添加到区块链的视频是否已发生更改的图像指纹技术。



CN 114731457 B

1.一种视频验证系统,其包括:

至少一个处理器,所述至少一个处理器被配置来利用指令访问至少一个计算机存储装置,所述指令能够由所述处理器执行以:

访问存储视频信息的至少一个区块链;

确定试图添加到所述区块链的第一视频是否是所述区块链中表示的原始视频的经更改复本;

响应于确定试图添加到所述区块链的所述第一视频是所述原始视频的经更改复本,拒绝将与所述第一视频有关的信息添加到所述区块链,或者将所述第一视频与所述第一视频已从所述原始视频发生更改的指示一起添加到所述区块链,

其中,确定试图添加到所述区块链的所述第一视频是否是所述经更改复本包括:

访问至少一个频率变换,所述至少一个频率变换被配置用于接收与所述第一视频中的帧序列相关的音频并且被配置用于输出频谱;

访问至少一个神经网络,所述至少一个神经网络被配置用于接收所述频谱并且输出表示所述音频的音频特征向量;以及

访问至少一个分析模块,所述至少一个分析模块被训练来学习自然人类语音特性,被配置用于接收所述音频特征向量并且基于所述音频特征向量输出关于所述音频从原始音频发生更改的指示。

2.如权利要求1所述的系统,其中所述指令能够执行以:至少部分地通过以下方式来确定所述第一视频是否是所述原始视频的经更改复本:

确定在所述第一视频中是否存在所述第一视频中的面部上的或所述第一视频中的所述面部与背景之间的或两者的至少一个纹理不规则性;以及

至少部分地基于所述纹理不规则性确定所述第一视频是否从所述原始视频发生更改。

3.如权利要求1所述的系统,其中所述指令能够执行以:至少部分地通过以下方式来确定所述第一视频是否是所述原始视频的经更改复本:

使用至少一个离散傅里叶变换(DFT)输出表示所述第一视频的至少一部分的频谱;以及

至少部分地使用所述频谱确定所述第一视频是否从所述原始视频发生更改。

4.如权利要求3所述的系统,其中所述指令能够执行以:

检测所述频谱中的至少一个不规则性,其中所述频谱中的所述不规则性包括比所述原始视频中的对应区域更亮的至少一个亮度区域;并且

响应于检测到所述频谱中的所述不规则性而确定所述第一视频已发生更改。

5.如权利要求1所述的系统,其中所述指令能够执行以:

接收所述第一视频中的帧序列;

输出表示如所述帧序列所示的人的面部的运动是否表现出自然运动的特征向量;并且

基于所述特征向量确定所述帧序列是否从所述原始视频发生更改。

6.如权利要求1所述的系统,其中所述指令能够执行以:

在所述第一视频上执行视频哈希;

将所述第一视频上的所述视频哈希与所述区块链中表示的所述原始视频的视频哈希进行比较;并且

响应于将所述第一视频上的所述视频哈希与所述原始视频的所述视频哈希进行比较，确定所述第一视频是否是所述原始视频的经更改复本。

7.如权利要求1所述的系统，其中所述指令能够执行以：

针对所述原始视频的至少第一帧，在所述第一帧上创建哈希；

将所述第一帧中的所述哈希的元素隐写地嵌入在所述第一帧中的已知位置处；

针对所述原始视频的复本的至少第二帧，仅使用来自所述第二帧中的除所述第二帧中与所述第一帧中的所述已知位置相对应的位置之外的位置的信息来在所述第二帧上创建哈希；

将所述第一帧上的所述哈希与所述第二帧上的所述哈希进行比较；并且

响应于所述比较，指示所述原始视频的所述复本是否已从所述原始视频发生更改。

8.如权利要求7所述的系统，其中所述指令能够执行以：

响应于确定所述原始视频的所述复本尚未从所述原始视频发生更改，将表示所述原始视频的所述复本的信息添加到包含表示所述原始视频的信息的至少一个区块链；并且

响应于确定所述原始视频的所述复本已从所述原始视频发生更改，不将表示所述原始视频的所述复本的信息添加到所述区块链。

9.一种视频验证方法，其包括：

检测自原始视频衍生的第一视频已从所述原始视频发生更改；以及

将所述第一视频连同所述第一视频已发生更改的指示一起添加到包含表示所述原始视频的信息的区块链；或者

拒绝将所述第一视频添加到所述区块链的请求，

其中检测自原始视频衍生的第一视频已从所述原始视频发生更改包括：

访问至少一个频率变换，所述至少一个频率变换被配置用于接收与所述第一视频中的帧序列相关的音频并且被配置用于输出频谱；

访问至少一个神经网络，所述至少一个神经网络被配置用于接收所述频谱并且输出表示所述音频的音频特征向量；以及

访问至少一个分析模块，所述至少一个分析模块被训练来学习自然人类语音特性，被配置用于接收所述音频特征向量并且基于所述音频特征向量输出关于所述音频从原始音频发生更改的指示。

10.如权利要求9所述的方法，其中所述检测至少部分地通过以下方式来执行：

渲染所述第一视频的数字指纹；

将所述第一视频的所述数字指纹与包含在所述区块链中的所述原始视频的数字指纹进行比较。

11.如权利要求10所述的方法，其中所述原始视频的所述数字指纹包括哈希，所述哈希在所述原始视频的至少一个帧上执行并且嵌入所述哈希自其衍生的所述帧中。

12.如权利要求11所述的方法，其中所述哈希以人眼不能检测的方式隐写地嵌入所述帧中。

13.如权利要求12所述的方法，其中所述哈希跨所述帧分布在所述帧中的已知位置中，所述哈希仅由所述帧中的非隐写更改的像素创建。

使用区块链的伪造视频检测

技术领域

[0001] 本申请总体上涉及必须植根于计算机技术并且产生具体技术改进的技术上创新的非常规解决方案。

背景技术

[0002] 如本文所理解,现代数字图像处理,与深度学习算法相结合,呈现如下有趣且令人愉快但潜在地险恶的能力:将一个人的视频图像更改为另一个人的图像,或者通过以此人的声音说出此人实际上从未说过的话的这种方式更改此人的视频。虽然这种能力可以轻松的方式使用,但它也可用于通过使得看起来好像此人说了诋毁的话来诽谤个人。因此,本文提供了用于确定视频是真的还是由机器学习生成的伪造品的技术。

发明内容

[0003] 因此,一种系统包括:至少一个处理器,所述至少一个处理器被配置来利用指令访问至少一个计算机存储装置,所述指令能够由所述处理器执行以访问存储视频信息的至少一个区块链。所述指令还能够执行以:确定试图添加到所述区块链的第一视频是否是所述区块链中表示的原始视频的经更改复本。所述指令进一步能够执行以:响应于确定试图添加到所述区块链的所述第一视频是所述原始视频的经更改复本,拒绝将与所述第一视频有关的信息添加到所述区块链,或者将所述第一视频与所述第一视频已从所述原始视频发生更改的指示一起添加到所述区块链。

[0004] 在一些示例中,所述指令能够执行以:响应于确定试图添加到所述区块链的所述第一视频是所述原始视频的经更改复本,拒绝将与所述第一视频有关的信息添加到所述区块链。在其他示例中,所述指令能够执行以:响应于确定试图添加到所述区块链的所述第一视频是所述原始视频的经更改复本,将所述第一视频与所述第一视频已从所述原始视频发生更改的指示一起添加到所述区块链。

[0005] 在示例性实现方式中,所述指令可能够执行以:至少部分地通过以下方式来确定所述第一视频是否是所述原始视频的经更改复本:确定在所述第一视频中是否存在所述第一视频中的面部上的或所述第一视频中的所述面部与背景之间的或两者的至少一个纹理不规则性;以及至少部分地基于所述纹理不规则性确定所述第一视频是否从所述原始视频发生更改。

[0006] 在示例性实现方式中,所述指令可能够执行以:至少部分地通过以下方式来确定所述第一视频是否是所述原始视频的经更改复本:使用至少一个离散傅里叶变换(DFT)输出表示所述第一视频的至少一部分的频谱;以及至少部分地使用所述频谱确定所述第一视频是否从所述原始视频发生更改。在此类实现方式中,所述指令可能够执行以:检测所述频谱中的至少一个不规则性,所述频谱中的所述至少一个不规则性可包括比所述原始视频中的对应区域更亮的至少一个亮度区域。所述指令可能够执行以:响应于检测到所述频谱中的所述不规则性而确定所述第一视频已发生更改。

[0007] 在非限制性示例中,所述指令能够执行以:接收所述第一视频中的帧序列,并且输出表示如所述帧序列所示的人的面部的运动是否表现出自然运动的特征向量。所述指令可进一步能够执行以:基于所述特征向量确定所述帧序列是否从所述原始视频发生更改。

[0008] 如果需要,所述系统可包括:至少一个频率变换,所述至少一个频率变换被配置用于接收与所述第一视频中的帧序列相关的音频并且被配置用于输出频谱。所述系统还可包括:至少一个神经网络,所述至少一个神经网络被配置用于接收所述频谱并且输出表示所述音频的音频特征向量。此外,所述系统可包括:至少一个分析模块,所述至少一个分析模块被训练来学习自然人类语音特性,被配置用于接收所述音频特征向量并且基于所述音频特征向量输出关于所述音频从原始音频发生更改的指示。

[0009] 在非限制性的示例性实施方案中,所述指令可能够执行以:在所述第一视频上执行视频哈希。所述指令进一步可能够执行以:将所述第一视频上的所述视频哈希与所述区块链中表示的所述原始视频的视频哈希进行比较,并且响应于将所述第一视频上的所述视频哈希与所述原始视频的所述视频哈希进行比较,确定所述第一视频是否是所述原始视频的经更改复本。

[0010] 在另一个方面,一种方法包括:检测自原始视频衍生的第一视频已从所述原始视频发生更改。所述方法还包括:将所述第一视频连同所述第一视频已发生更改的指示一起添加到包含表示所述原始视频的信息的区块链;或者拒绝将所述第一视频添加到所述区块链的请求。

[0011] 在另一个方面,一种设备包括:至少一个计算机存储介质,所述至少一个计算机存储介质包括指令,所述指令能够由至少一个处理器执行以:针对原始视频的至少第一帧,在所述第一帧上创建哈希。所述指令能够执行以:将所述第一帧中的所述哈希的元素隐写地嵌入在所述第一帧中的已知位置处;并且针对所述原始视频的复本的至少第二帧,仅使用来自所述第二帧中的除所述第二帧中与所述第一帧中的所述已知位置相对应的位置之外的位置的信息来在所述第二帧上创建哈希。所述指令能够执行以:将所述第一帧上的所述哈希与所述第二帧上的所述哈希进行比较;并且响应于所述比较,指示所述原始视频的所述复本是否已从所述原始视频发生更改。

[0012] 参考附图可最好地理解本申请的关于其结构和操作两者的细节,在附图中,相同的附图标记指代相同的部分,并且在附图中:

附图说明

[0013] 图1是包括根据本发明原理的示例的示例性系统的框图;

[0014] 图2是示出真实视频和自真实视频衍生的伪造视频的示意图;

[0015] 图3是用于使用图像处理和频域分析两者来检测伪造视频的示例性逻辑的流程图;

[0016] 图4是用于训练图3中使用的神经网络的示例性逻辑的流程图;

[0017] 图5示出真实视频帧和对应的伪造视频帧的图示,其示出伪造帧中的伪影;

[0018] 图6是用于执行图3的逻辑的示例性神经网络架构的框图;

[0019] 图7是用于使用视频序列分析来检测伪造视频的示例性逻辑的流程图;

[0020] 图8是用于执行图7的逻辑的示例性神经网络架构的框图;

[0021] 图9是用于采用区块链技术处理伪造视频生成的示例性逻辑的流程图;

[0022] 图10是示例性用户界面(UI)的屏幕截图,所述UI用于向互联网服务提供商(ISP)或经销商报告伪造视频,使得ISP/经销商可将视频从公众视野移除;

[0023] 图11是用于记录或上传或下载视频以及视频中的嵌入验证哈希的示例性逻辑的流程图;

[0024] 图12是用于回放图11中所记录或所访问的视频的示例性逻辑的流程图,其中使用哈希来验证真实性;

[0025] 图13是使用先前原理的混合逻辑的示例性逻辑的流程图;

[0026] 图14示出两组真实图像和经更改图像中的示例性照明伪影;

[0027] 图15示出图像中的示例性生成对抗网络(GAN)伪影或不规则性;并且

[0028] 图16示出使用真实图像和经更改图像的另一个GAN相关伪影或不规则性。

具体实施方式

[0029] 本公开总体上涉及计算机生态系统,其包括消费电子(CE)装置网络的各方面,诸如但不限于计算机模拟网络(诸如计算机游戏网络)以及独立计算机模拟系统。本文中的系统可包括通过网络连接使得可在客户端与服务器部件之间交换数据的服务器和客户端部件。客户端部件可包括一个或多个计算装置,所述一个或多个计算装置包括诸如索尼(Sony)PlayStation®的游戏控制台或由微软(Microsoft)或任天堂(Nintendo)制作的游戏控制台、或其他制造商的虚拟现实(VR)头戴式装置、增强现实(AR)头戴式装置、便携式电视机(例如智能TV、支持互联网的TV)、便携式计算机(诸如膝上型计算机和平板计算机)以及其他移动装置(包括智能电话和下文讨论的另外的示例)。这些客户端装置可在多种操作环境下操作。例如,客户端计算机中的一些可采用例如Linux操作系统、来自微软的操作系统、或Unix操作系统或由苹果计算机(Apple Computer)或谷歌(Google)生产的操作系统。这些操作环境可用于执行一个或多个浏览程序(诸如由微软或谷歌制造的浏览器、或Mozilla)或可访问由下文讨论的互联网服务器托管的网站的其他浏览器程序。另外,根据本发明原理的操作环境可用于执行一个或多个计算机游戏程序。

[0030] 服务器和/或网关可包括一个或多个处理器,所述一个或多个处理器执行配置服务器以通过网络诸如互联网来接收和传输数据的指令。或者,客户端和服务器可通过本地内联网或虚拟专用网连接。服务器或控制器可由游戏控制台(诸如索尼PlayStation®)、个人计算机等实例化。

[0031] 信息可通过网络在客户端与服务器之间交换。为此,并且出于安全起见,服务器和/或客户端可包括防火墙、负载平衡器、暂时存储装置和代理以及其他网络基础设施,以实现可靠性和安全性。一个或多个服务器可形成实现向网络成员提供安全社区(诸如在线社交网站)的方法的设备。

[0032] 如本文所用,指令是指用于处理系统中的信息的计算机实现的步骤。指令可在软件、固件或硬件中实现,并且包括由系统部件实现的任何类型的编程步骤。

[0033] 处理器可以是任何常规通用单芯片或多芯片处理器,其可借助于各种线(诸如地址线、数据线和控制线)以及寄存器和移位寄存器执行逻辑。

[0034] 通过本文中的流程图和用户界面的方式描述的软件模块可包括各种子例程、程序

等。在不限制本公开的情况下,陈述为由特定模块执行的逻辑可被再分配到其他软件模块和/或在单个模块中组合在一起和/或在可共享库中可获得。

[0035] 本文描述的本发明原理可实现为硬件、软件、固件或它们的组合;因此,说明性部件、框、模块、电路和步骤是依据它们的功能性加以阐述。

[0036] 除上文已经提到的内容之外,下文描述的逻辑块、模块和电路可用通用处理器、数字信号处理器(DSP)、现场可编程门阵列(FPGA)或被设计成执行本文描述的功能的其他可编程逻辑装置(诸如专用集成电路(ASIC)、离散门或晶体管逻辑、离散硬件部件或它们的任何组合)来实现或执行。处理器可由控制器或状态机或计算装置的组合来实现。

[0037] 下文描述的功能和方法当在软件中实现时可用诸如但不限于Java、C#或C++的适当语言来编写,并且可存储在计算机可读存储介质上或通过计算机可读存储介质传输,所述计算机可读存储介质诸如随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、光盘只读存储器(CD-ROM)或其他光盘存储装置(诸如数字通用光盘(DVD))、磁盘存储装置或包括可移动拇指驱动器的其他磁存储装置等。连接可建立计算机可读介质。此类连接可包括例如硬连线电缆,所述硬连线电缆包括光纤和同轴线以及数字用户线路(DSL)和双绞线。此类连接可包括无线通信连接,所述无线通信连接包括红外线和无线电。

[0038] 一个实施方案中所包括的部件可以任何适当的组合用于其他实施方案中。例如,本文描述和/或在图中描绘的各种部件中的任一者可与其他实施方案组合、互换或从其他实施方案排除。

[0039] “具有A、B和C中的至少一者的系统”(同样地,“具有A、B或C中的至少一者的系统”和“具有A、B、C中的至少一者的系统”)包括以下系统:单独具有A;单独具有B;单独具有C;同时具有A和B;同时具有A和C;同时具有B和C;和/或同时具有A、B和C等。

[0040] 现在具体地参考图1,示出了示例系统10,所述系统10可包括根据本发明原理的上文提及和下文进一步描述的示例性装置中的一者或多者。系统10中所包括的示例性装置中的第一者是诸如音频视频装置(AVD)12的消费电子(CE)装置,诸如但不限于带有TV调谐器(等效地,控制TV的机顶盒)的支持互联网的TV。然而,AVD 12替代地可以是电器或家用物品,例如,计算机化的支持互联网的冰箱、洗涤机或烘干机。AVD 12替代地还可以是计算机化的支持互联网的(“智能”)电话、平板计算机、笔记本计算机、可穿戴的计算机化装置(诸如计算机化的支持互联网的手表、计算机化的支持互联网的手镯)、其他计算机化的支持互联网的装置、计算机化的支持互联网的音乐播放器、计算机化的支持互联网的头戴式耳机、计算机化的支持互联网的可植入装置(诸如可植入皮肤装置)等。无论如何,应理解,AVD 12被配置来实现本发明原理(例如,与其他CE装置通信以实现本发明原理,执行本文描述的逻辑,以及执行本文描述的任何其他功能和/或操作)。

[0041] 因此,为了实现此类原理,AVD 12可由图1所示的部件中的一些或全部建立。例如,AVD 12可包括一个或多个显示器14,所述一个或多个显示器14可由高清晰度或超高清晰度“4K”或更高的平面屏幕实现,并且可以是支持触摸的以用于经由显示器上的触摸接收用户输入信号。AVD 12可包括一个或多个扬声器16,所述一个或多个扬声器16用于根据本发明原理输出音频;以及至少一个另外的输入装置18(诸如例如音频接收器/麦克风),所述至少一个另外的输入装置18用于例如向AVD 12输入可听命令以控制AVD 12。示例性AVD 12还可

包括一个或多个网络接口20,所述一个或多个网络接口20用于在所包括的一个或多个处理器24的控制下通过至少一个网络22(诸如互联网、WAN、LAN等)进行通信。还可包括图形处理器24A。因此,接口20可以是但不限于Wi-Fi收发器,所述Wi-Fi收发器是无线计算机网络接口的示例,诸如但不限于网状网络收发器。应理解,处理器24控制AVD 12以实现本发明原理,包括本文描述的AVD 12的其他元件,诸如例如控制显示器14以在其上呈现图像和从其接收输入。此外,应注意,网络接口20可以是例如有线或无线调制解调器或路由器、或其他适当接口(诸如例如无线电话收发器、或如上文所提及的Wi-Fi收发器等)。

[0042] 除前述内容之外,AVD 12还可包括一个或多个输入端口26,诸如例如用于物理地连接(例如,使用有线连接)到另一个CE装置的高清晰度多媒体接口(HDMI)端口或USB端口和/或用于将头戴式耳机连接到AVD 12以通过头戴式耳机将来自AVD 12的音频呈现给用户的头戴式耳机端口。例如,输入端口26可经由导线或无线地连接到音频视频内容的有线或卫星源26a。因此,源26a可以是例如单独的或集成的机顶盒、或卫星接收器。或者,源26a可以是包含内容的游戏控制台或磁盘播放器,所述内容可能被用户视为收藏以用于下文进一步描述的信道指派目的。源26a在被实现为游戏控制台时可包括下文关于CE装置44描述的部件中的一些或全部。

[0043] AVD 12还可包括不是暂态信号的一个或多个计算机存储器28,诸如基于磁盘的存储装置或固态存储装置,所述一个或多个计算机存储器28在一些情况下在AVD的机壳中体现为独立装置,或者体现为在AVD的机壳内部或外部的用于回放AV节目的个人视频记录装置(PVR)或视频磁盘播放器,或者体现为可移动存储器介质。此外,在一些实施方案中,AVD 12可包括位置或定位接收器,诸如但不限于手机接收器、GPS接收器和/或高度计30,所述位置或定位接收器被配置来例如从至少一个卫星或手机塔接收地理位置信息并且向处理器24提供所述信息和/或结合处理器24来确定AVD 12被设置处于的海拔。然而,应理解,根据本发明原理,可使用除手机接收器、GPS接收器和/或高度计之外的另一种合适的位置接收器来例如确定AVD 12在例如所有三个维度上的定位。

[0044] 继续对AVD 12的描述,在一些实施方案中,根据本发明原理,AVD 12可包括一个或多个相机32,所述一个或多个相机32可以是例如热成像相机、数字相机(诸如网络摄像头)和/或集成到AVD 12中并且可由处理器24控制以收集图片/图像和/或视频的相机。AVD 12上还可包括蓝牙收发器34和另一近场通信(NFC)元件36,所述蓝牙收发器34和另一近场通信(NFC)元件36用于分别使用蓝牙和/或NFC技术与其他装置进行通信。示例性NFC元件可以是射频识别(RFID)元件。

[0045] 更进一步地,AVD 12可包括向处理器24提供输入的一个或多个辅助传感器37(例如,运动传感器(诸如加速度计、陀螺仪、回转计)、或磁传感器、红外(IR)传感器、光学传感器、速度和/或节奏传感器、手势传感器(例如,用于感测手势命令等)。AVD 12可包括空中TV广播端口38,所述空中TV广播端口38用于接收向处理器24提供输入的OTA TV广播。除前述内容之外,应注意,AVD 12还可包括红外(IR)发射器和/或IR接收器和/或IR收发器42,诸如IR数据关联(IRDA)装置。可提供电池(未示出)以为AVD 12供电。

[0046] 仍参考图1,除AVD 12之外,系统10可包括一个或多个其他CE装置类型。在一个示例中,第一CE装置44可用于经由直接发送到AVD 12的命令和/或通过下文描述的服务器向AVD 12发送计算机游戏音频和视频,而第二CE装置46可包括与第一CE装置44相似的部件。

在所示的示例中,第二CE装置46可被配置为玩家47所穿戴的VR头戴式装置,如图所示。在所示的示例中,仅示出两个CE装置44、46,但应理解,可使用更少或更多的装置。例如,下文的原理讨论在玩由游戏控制台向一个或多个AVD 12提供的计算机游戏期间,多个玩家47通过各自的头戴式装置彼此通信。

[0047] 在所示的示例中,为了说明本发明原理,假设所有三个装置12、44、46都是例如家庭中的娱乐网络的成员,或至少彼此接近地存在于诸如房屋的某一位置中。然而,除非另有明确要求,否则本发明原理不限于虚线48所示的特定位置。

[0048] 示例性的非限制性第一CE装置44可由上文提及的装置(例如,便携式无线膝上型计算机或笔记本计算机或游戏控制器)中的任何一者来建立,并因此可具有下文描述的部件中的一者或多者。第一CE装置44可以是用于例如向AVD 12发布AV播放和暂停命令的遥控器(RC),或者它可以是更复杂的装置,诸如平板计算机、经由有线或无线链路与AVD 12通信的游戏控制器和/或游戏控制台、个人计算机、无线电话等。

[0049] 因此,第一CE装置44可包括一个或多个显示器50,所述一个或多个显示器50可以是支持触摸的以用于经由显示器上的触摸来接收用户输入信号。第一CE装置44可包括一个或多个扬声器52,所述一个或多个扬声器52用于根据本发明原理输出音频;以及至少一个另外的输入装置54(诸如例如音频接收器/传声器),所述至少一个另外的输入装置54用于例如向第一CE装置44输入可听命令以控制装置44。示例性第一CE装置44还可包括一个或多个网络接口56,所述一个或多个网络接口56用于在一个或多个CE装置处理器58的控制下通过网络22进行通信。还可包括图形处理器58A。因此,接口56可以是但不限于Wi-Fi收发器,所述Wi-Fi收发器是无线计算机网络接口的示例,包括网状网络接口。应理解,处理器58控制第一CE装置44以实现本发明原理,包括本文描述的第一CE装置44的其他元件,诸如例如控制显示器50以在其上呈现图像和从其接收输入。此外,应注意,网络接口56可以是例如有线或无线调制解调器或路由器、或其他适当接口(诸如无线电话收发器、或如上文所提及的Wi-Fi收发器等)。

[0050] 除前述内容之外,第一CE装置44还可包括一个或多个输入端口60,诸如例如用于物理地连接(例如,使用有线连接)到另一个CE装置的HDMI端口或USB端口和/或用于将头戴式耳机连接到第一CE装置44以通过头戴式耳机将来自第一CE装置44的音频呈现给用户的头戴式耳机端口。第一CE装置44还可包括一个或多个有形计算机可读存储介质62,诸如基于磁盘的存储装置或固态存储装置。此外,在一些实施方案中,第一CE装置44可包括位置或定位接收器,诸如但不限于手机和/或GPS接收器和/或高度计64,所述位置或定位接收器被配置来例如使用三角测量从至少一个卫星和/或手机塔接收地理位置信息并且向CE装置处理器58提供所述信息和/或结合CE装置处理器58来确定第一CE装置44被设置处于的高度。然而,应理解,根据本发明原理,可使用除手机和/或GPS接收器和/或高度计之外的另一种合适的位置接收器来例如确定第一CE装置44在例如所有三个维度上的定位。

[0051] 继续对第一CE装置44的描述,在一些实施方案中,根据本发明原理,第一CE装置44可包括一个或多个相机66,所述一个或多个相机66可以是例如热成像相机、数字相机(诸如网络摄像头)和/或集成到第一CE装置44中并且可由CE装置处理器58控制以收集图片/图像和/或视频的相机。第一CE装置44上还可包括蓝牙收发器68和另一近场通信(NFC)元件70,所述蓝牙收发器68和另一近场通信(NFC)元件70用于分别使用蓝牙和/或NFC技术与其他装

置进行通信。示例性NFC元件可以是射频识别(RFID)元件。

[0052] 此外,第一CE装置44可包括向CE装置处理器58提供输入的一个或多个辅助传感器72(例如,运动传感器(诸如加速度计、陀螺仪、回转计)、或磁传感器、红外线(IR)传感器、光学传感器、速度和/或节奏传感器、手势传感器(例如,用于感测手势命令)等)。第一CE装置44可包括向CE装置处理器58提供输入的再一些传感器,诸如例如一个或多个气候传感器74(例如,气压计、湿度传感器、风传感器、光传感器、温度传感器等)和/或一个或多个生物识别传感器76。除前述内容之外,应注意,在一些实施方案中,第一CE装置44还可包括红外(IR)发射器和/或IR接收器和/或IR收发器78,诸如IR数据关联(IRDA)装置。可提供电池(未示出)以为第一CE装置44供电。CE装置44可通过上述通信模式和相关部件中的任一者与AVD 12进行通信。

[0053] 第二CE装置46可包括针对CE装置44示出的部件中的一些或全部。任一个或两个CE装置可由一个或多个电池供电。

[0054] 现在参考前面提及的至少一个服务器80,所述至少一个服务器80包括至少一个服务器处理器82、至少一个有形计算机可读存储介质84(诸如基于磁盘的存储装置或固态存储装置)和至少一个网络接口86,根据本原理,所述至少一个网络接口86在服务器处理器82的控制下实现通过网络22与图1的其他装置进行通信,并且实际上可促进服务器与客户端装置之间的通信。应注意,网络接口86可以是例如有线或无线调制解调器或路由器、Wi-Fi收发器或其他适当接口(诸如例如无线电话收发器)。

[0055] 因此,在一些实施方案中,服务器80可以是互联网服务器或整个服务器“农场”,并且在例如网络游戏应用的示例性实施方案中可包括并且执行“云”功能,使得系统10的装置可经由服务器80访问“云”环境。或者,服务器80可由与图1所示的其他装置位于同一房间中或附近的一个或多个游戏控制台或其他计算机来实现。

[0056] 本文中的方法可实现为由处理器、经适当配置的专用集成电路(ASIC)或现场可编程门阵列(FPGA)模块或如本领域技术人员将了解的任何其他便利方式执行的软件指令。在被采用的情况下,软件指令可以非暂态装置(诸如CD ROM或闪存驱动器)体现。软件代码指令可替代地以诸如无线电或光学信号的暂态布置或者经由通过互联网的下载来体现。

[0057] 现在参考图2至图6,示出用于确定图像是否是“伪造”(即,已从原始图像发生数字更改)的第一种技术。在图2中,可呈现在显示器202上的原始图像200显示具有面部204的人在说出可听短语206。图像200可以是诸如来自视频流的I帧的图像,并且视频流的一些或所有帧可如本文所公开的那样进行处理。

[0058] 利用输入装置210(诸如但不限于键盘)操作计算机208的人可更改图像和/或音频以产生此人的经更改图像212,此人可能被描绘为说出经更改的可听短语214。本发明原理涉及检测经更改图像212实际上已从原始图像200发生更改。

[0059] 图3示出可在第一种技术中执行的逻辑,而图6提供可体现图3的逻辑的示例性架构。在框300处开始,接收到图像。在框302处,可借由通过第一神经网络(NN)(诸如卷积NN(CNN))处理图像来直接分析图像。在框304处,第一NN输出表示图像的特征向量。

[0060] 另外,在框306处,可将图像输入到面部辨识模块以针对图像的面部和/或背景中的伪影(在本文也称为不规则性)以及图像中的照明不规则性进行分析。可采用一个或多个NN的面部辨识模块将特征向量输出到框304。

[0061] 通过举例方式,图像(空间域)中的面部的不规则性可包括具有棋盘状外观的小区域,指示由于数字更改导致的模糊分辨率。

[0062] 此外,在框308处,可使用例如输出频谱的离散傅里叶变换(DFT)来将图像转换到频域,在框310处,利用另一个NN(诸如CNN)分析所述频谱以检测频域中图像的不规则性。将表示频谱的特征向量提供到框304。

[0063] 通过举例方式,频域中的不规则性可包括沿着频域中图像的图形表示的外围的一个或多个亮斑。

[0064] 移动到决策菱形312,可包括一个或多个NN的检测模块分析来自框304的特征向量,以确定在空间域和/或频域中是否存在一个或多个不规则性。如果不存在不规则性,则过程可在状态314处结束,但在一些实现方式中,如果在任何域中存在任何不规则性,则可在框316处返回图像是伪造的指示。在其他实现方式中,只有在空间域中存在不规则性以及在频域中存在不规则性的两种情况下,可在框316处返回图像是伪造的指示。

[0065] 简要地参考图4,示出用于训练本文讨论的NN的过程。在框400处开始,将地面实况原始、未经更改的图像输入到NN。此外,在框402处,将地面实况经更改或伪造的图像输入到NN。设计者可使用“深度伪造”技术从地面实况原始图像生成伪造图像。NN可被编程为使用例如上文针对频域和空间域两者讨论的任何或示例性不规则性开始分析。在框404处,在地面实况输入上训练NN。在框404处,随后可应用强化学习来完善NN的训练。

[0066] 图5示出示例性空间域不规则性和频域不规则性。在原始空间域502和原始频域504中示出原始图像500。原始图像500的经更改图像506具有经更改空间域图像508和在510处描绘的经更改频域。

[0067] 如图所示,经更改空间域图像508中的区域512具有放大的并且在514处描绘的棋盘模式。在原始图像与经更改图像之间也可存在照明不规则性。

[0068] 也可在频域510中的图像表示中检测到一个或多个频域不规则性516。频域不规则性516可包括沿着频域图形描绘的边缘或外围的亮斑,如图所示。在所示的示例中,每侧存在两个亮斑,指示频域中的图像更改产生的不规则性。

[0069] 图6示出可用于体现图3的逻辑的示例性架构。将要针对更改进行测试的图像600输入到面部检测模块602,所述面部检测模块602分析空间域中的图像以在模块602的神经网络(NN)604处检测图像中的照明不规则性以及在606处进行面部分辨率/不规则性检查。面部检测模块602可采用图像辨识原理并且可由一个或多个NN体现。

[0070] 此外,图像600可直接输入到NN 608以使用另外的规则来进行直接分析,所述NN 608可以是CNN。应注意,NN 608提取图像的特征向量。此外,NN 604执行图像处理并且在缺乏足够训练数据的情况下特别有利。然而,NN 604、608可由单个NN实现。

[0071] 此外,图像600通过离散傅里叶变换(DFT)610进行处理,所述DFT 610输出表示频域中的图像600的频谱612。将频谱612发送到CNN 614以分析频谱。

[0072] 面部辨识模块602(包括照明不规则性检查604和面部分辨率/伪影检查606)以及CNN 608和614产生表示空间域和频域两者中的图像600的一组特征向量616。可由一个或多个NN(诸如递归NN(RNN),诸如长短期模块(LSTM))实现的检测模块618根据本文提出的原理分析特征向量以查明图像600是否包含从原始图像的数字更改。如果包含,则在620处,生成图像600可能是伪造品的指示。

[0073] 图7示出用于检测经更改视频的第二种技术,并且图8提供用于体现图7的逻辑的示例性架构。在框700处,将视频序列(诸如视频短片或其他视频帧序列)输入到NN。在框702处,使用NN分析序列,在704处,所述NN输出表示视频序列的特征向量。

[0074] 在分析视频序列时,可训练NN来学习在例如说话期间的自然人类面部运动模式。如本文所理解,当更改视频序列时,更改者无法精确地模仿例如嘴唇的自然运动模式,并且因此NN可在伪造视频序列中检测到轻微不自然的运动模式。

[0075] 此外,在框706处,将与视频序列相关联的音频输入到频率变换。在框708处,将由频率变换706输出的频谱提供给NN以分析频谱,以便将表示音频的特征向量输出到框704。

[0076] 在分析伴随音频时,可训练NN来学习自然人类语音特性,诸如节奏、音调、音高模式和重读。如本文所理解,当更改音频(诸如声音音频)时,更改者无法精确地模仿自然人类语音模式。因此,NN可在伪造音频序列中检测到轻微不自然的语音模式,诸如不自然的节奏或音调或音高。可沿着图4所示的路线完成训练,其中地面实况音频和自原始地面实况音频衍生的伪造地面实况音频用作训练集。

[0077] 可将特征向量704提供给NN(诸如RNN 710)以分析特征向量,以便在决策菱形712处检测输入视频序列和/或伴随音频是否已从原始发生更改。如果未发现异常/不规则性,则过程可在状态714处结束,但如果检测到不规则性,则在框716处输出视频序列可能已更改的指示。

[0078] 在一些实施方案中,如果在音频或视频中检测到任何不规则性,则在框716处输出伪造品的指示。在其他实施方案中,只有在音频或视频两者中检测到不规则性的情况下,才在框716处输出伪造品的指示。

[0079] 图8示出可用于体现图7的逻辑的架构。将视频序列800输入到NN 802(诸如CNN)以从中提取特征向量804。另外,将音频(诸如语音806)输入到频率变换808(诸如短时傅里叶变换(STFT))以生成频域中的音频表示,由NN 810(诸如CNN)分析所述音频表示以从中提取特征向量。NN 212(诸如RNN(诸如LSTM))根据本文描述的原理分析特征向量以在框814处检测视频序列800和音频806中的任何不规则性。状态816指示输入可能是伪造品的指示的输出。

[0080] 现在转向图9,示出用于使用区块链技术和/或数字指纹技术来解决伪造视频生成的第三种技术。通常,视频的哈希/签名可使用例如网络浏览器集成到成像装置(诸如智能电话或其他记录装置)中,或编码到硬件中。数字指纹可从整个视频或子轨道中的数据位生成,使得如果视频内容改变,则指纹也将改变。数字指纹可与元数据(诸如最初制作视频所在的位置和时间戳)一起生成。每当试图再分销视频时,经销商必须向区块链上的原始视频请求许可并为新(复制)视频链接新区块,从而使得容易追溯到原始视频和黑链上的任何节点。在再次上传视频之前,可将视频的指纹与原始指纹进行匹配以确定试图添加的视频是否已被篡改。

[0081] 例如,视频网站可结合有视频指纹检测器,因此每当上传/下载视频时,所述视频将被记载和加时间戳。如果基于指纹不匹配原始视频指纹将视频分类为伪造,则所述视频可在整个链上被如此记载。这仿效杀毒软件,但在这种情况下,所有用户同时受到保护。

[0082] 在框900处开始,将原始(“真实”)视频连同视频的哈希一起添加到视频区块链,所述哈希可用作数字指纹并且通常是基于视频中的像素值或编码信息或其他图像相关值的。

可在框902处接收复制视频的请求,并且可在框904处准许所述请求。

[0083] 移动到框906,可接收重新将作为原始视频的复本的新视频添加到区块链中的请求。有效的请求可在新视频上伴随有哈希(指纹)。前进到决策菱形908,将试图添加到区块链的视频的哈希与所述视频从其复制的原始视频的哈希进行比较,并且如果哈希匹配,则可在框910处将新视频添加到区块链。

[0084] 另一方面,如果在决策菱形908处确定哈希不匹配,则逻辑可移动到框912以拒绝将新视频添加到区块链,或者将新视频与新视频已从原始视频发生更改并且因此可能是伪造品的指示一起添加到区块链。如果需要,可在框914处使经更改视频无法从区块链访问或以其他方式不可播放。

[0085] 此外,在检测到经更改视频时,逻辑可移动到框916,以向新的经更改视频的互联网服务提供商(ISP)或其他经销商报告所述视频实际上已从原始视频发生数字更改并且因此应进行检查以决定是否将新的(经更改)视频从公众视野移除。图10示出。

[0086] 如图所示,用户界面(UI)1000可呈现在执行图9的逻辑和/或从执行图9的逻辑的装置接收信息的装置的显示器1002上。UI 1000可包括已检测到潜在地伪造视频的提示1004。UI 1000还可包括选择器1006以使得用户能够将伪造品的存在连同识别信息一起报告给经销商或其他机构。

[0087] 图11和图12示出另外的指纹逻辑。在图11的框1100处开始,根据新原始视频的创建和/或在上传或下载视频时,对于所有视频帧的至少一些实施方案以及在所有视频帧的一些实施方案中,在帧中执行哈希。前进到框1002,然后将哈希嵌入所述哈希自其衍生的帧中。

[0088] 在示例中,视频帧的哈希可以肉眼不能检测的方式隐写地嵌入视频帧中并且可跨视频帧均匀地分布。例如,每片隐写哈希的像素可在已知位置中,这是因为所述已知位置始终是固定位置,或者是因为所述位置包含在帧的视频元数据中(从而允许每个帧不同)。已知此位置允许从被哈希的视频数据排除表示哈希的像素。换句话讲,原始哈希仅由非隐写更改的像素创建。视频压缩算法也可使用此位置以确保表示哈希的像素不以将影响哈希的方式压缩或更改。

[0089] 图12示出视频回放软件随后将反转此过程。在框1200处开始,从视频帧提取所隐写嵌入的哈希。移动到框1202,对视频帧的其余像素进行哈希。前进到决策菱形1204,将新哈希与从帧提取的哈希进行比较。如果它们匹配,则帧尚未从原始源视频发生更改并因此逻辑移动到框1206以指示这一点,并且如果需要,将视频添加到区块链(假设所有或至少阈值数量的帧哈希匹配)。如果哈希不匹配,则逻辑移动到框1208以指示正在观看的视频已从原始视频发生更改,并且(例如)已发生更改的帧周围具有红色边框或突出显示。甚至可勾勒出框架的经更改部分。

[0090] 此相同验证过程可在后端服务器上执行,所述后端服务器检测伪造品并且主动阻止伪造品被发布或向视频附加警告。

[0091] 如果任何恶意行为者以任何有意义的方式更改源视频,则帧将以不同方式哈希并且/或者所嵌入的隐写哈希将被破坏。只要此活动的两端都有善意行为者,就可检测到视频的更改。

[0092] 图13示出可采用结合上述原理的混合技术。框1300指示图像处理/视频序列与频

域分析相结合可用于识别视频中的伪影/不规则性。框1302进一步指示语音处理可与上述技术中的任一种结合使用以识别视频中的伪影/不规则性。框1304指示对视频中伪影/不规则性的识别可与区块链技术结合来跟踪原始(真实)视频及其已经更改的复本(伪造品)。

[0093] 图14至图16提供经更改图像(在图中标记为“伪造”图像)中可出现的伪影或不规则性的另外的示例。图14中的第一真实图像1400已被更改以产生对应的经更改图像1402,其中在区域1404中,照明看起来比第一真实图像1400中的对应区域更亮。同样地,第二真实图像1406已被更改以产生经更改图像1408,其中面部上的区域1410中的照明看起来比真实图像1406中的更亮。经更改图像1402、14108的分辨率也小于对应的真实图像1400、1406的分辨率,意指NN可基于照明不规则性和分辨率降低中的一者或两者学习区分经更改图像。

[0094] 图15示出经更改图像1500,其中在小区域1502中,由于执行生成对抗网络(GAN)上采样以产生经更改图像1500而存在图像不规则性或伪影。如区域1502的分解图1504所示,GAN不规则性可包括图像区域的均匀纯色,在所述区域中,非均匀纯色主题(在所示的示例中,具有各种阴影程度的草)显现在原始图像中。

[0095] 图16示出真实图像1600和通过将另一个人的面部叠加到真实图像1600中的对象的面部上而自真实图像1600衍生的经更改图像1602。如1604处所示,这种叠加导致面部与头部或身体其余部位不对准,在这种情况下,鼻子与描绘头部所处的角度不对准。

[0096] 将了解,虽然已经参考一些示例性实施方案描述了本发明原理,但是这些实施方案并不意图是限制性的,并且可使用各种替代布置来实现本文所要求保护的主题。

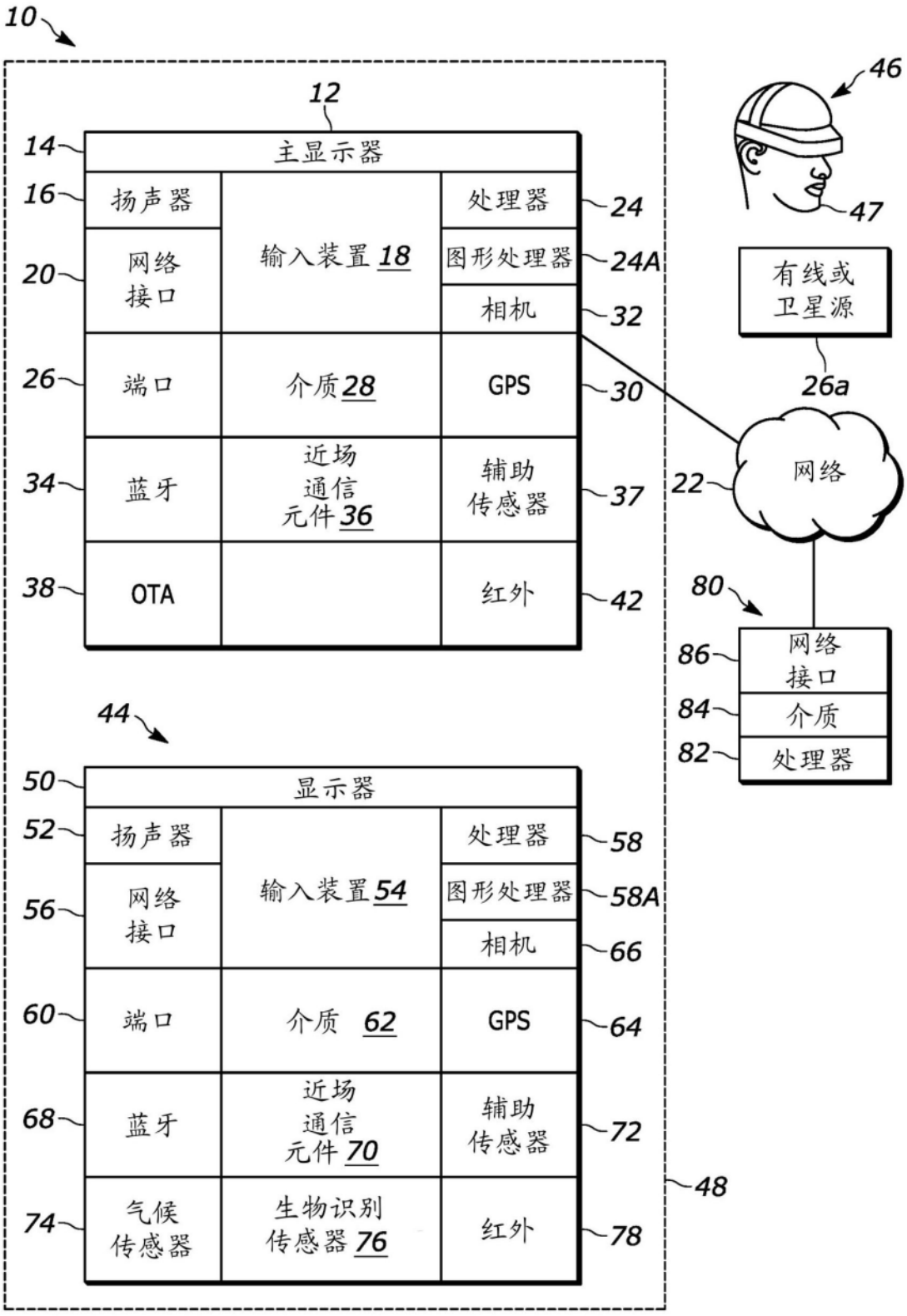

10
12
主显示器
14
扬声器
16
网络接口
20
输入装置 18
处理器
24
图形处理器
24A
相机
32
端口
26
介质 28
GPS
30
蓝牙
34
近场通信元件 36
辅助传感器
37
OTA
38
红外
42
46
47
有线或卫星源
26a
网络
22
80
网络接口
86
介质
84
处理器
82
44
显示器
50
扬声器
52
网络接口
56
输入装置 54
处理器
58
图形处理器
58A
相机
66
端口
60
介质 62
GPS
64
蓝牙
68
近场通信元件 70
辅助传感器
72
气候传感器
74
生物识别传感器 76
红外
78
48

图1

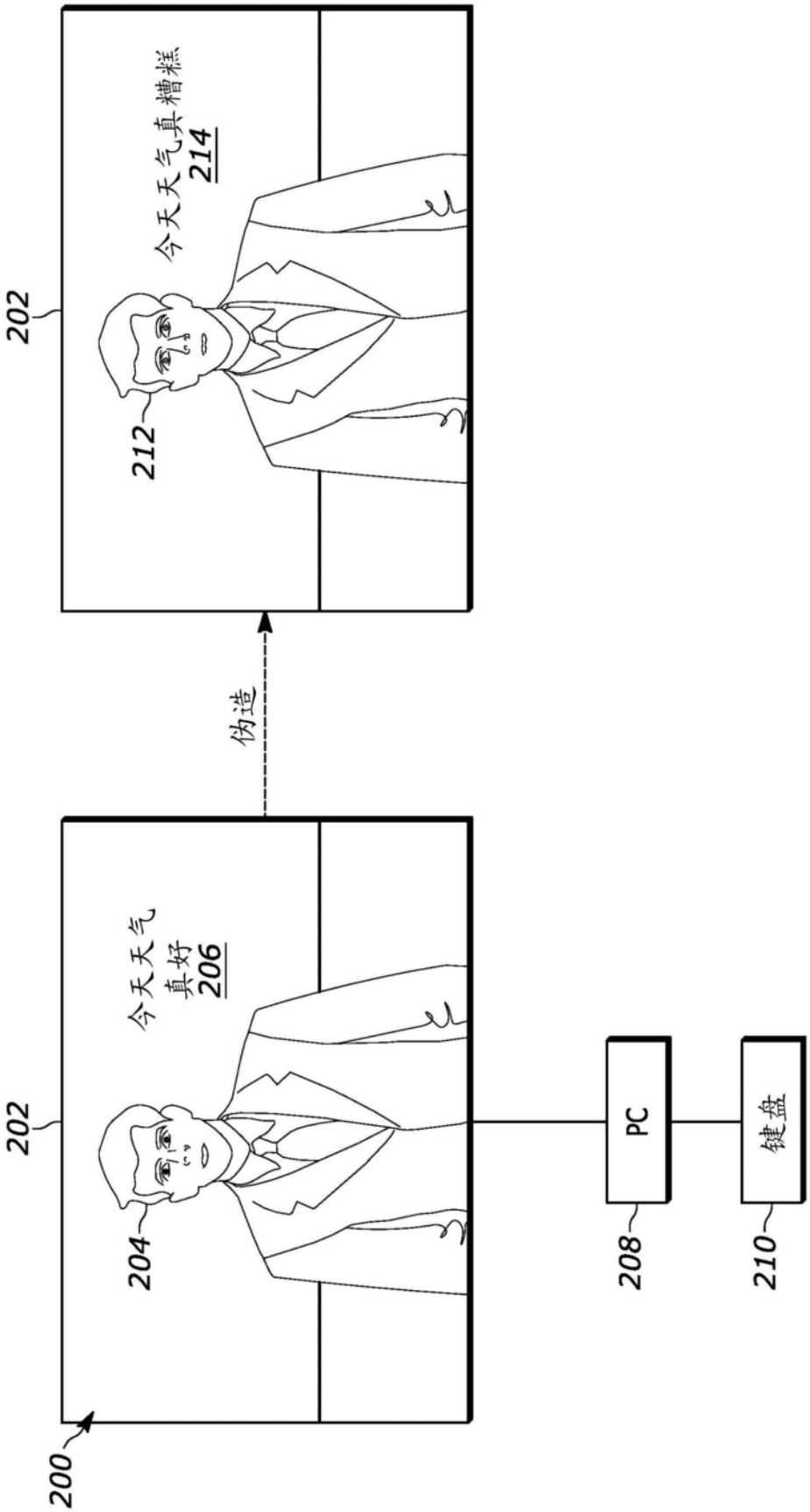


图2

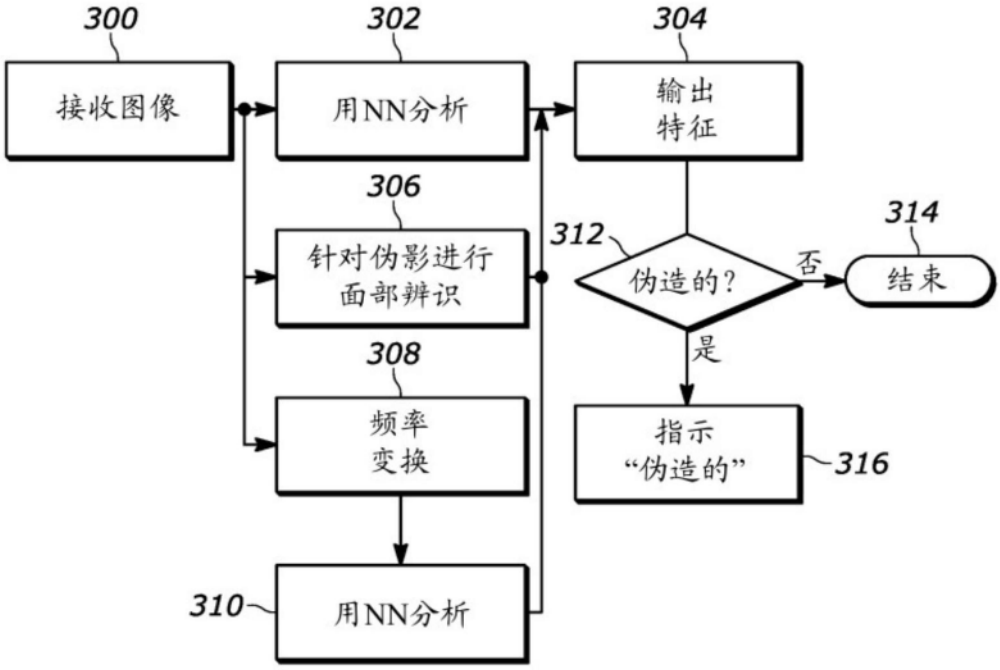

300
接收图像
302
用NN分析
304
输出
特征
306
针对伪影进行
面部辨识
312
伪造的?
否
314
结束
是
308
频率
变换
指示
“伪造的”
316
310
用NN分析

图3

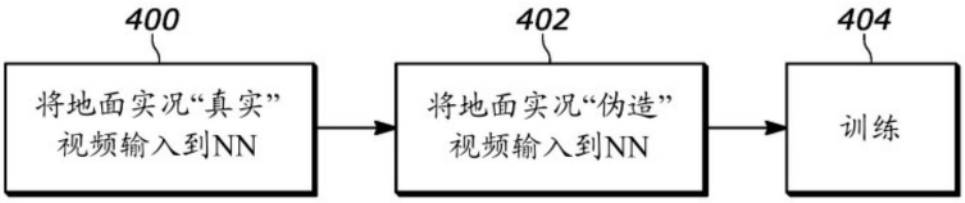

400
将地面实况“真实”
视频输入到NN
402
将地面实况“伪造”
视频输入到NN
404
训练

图4

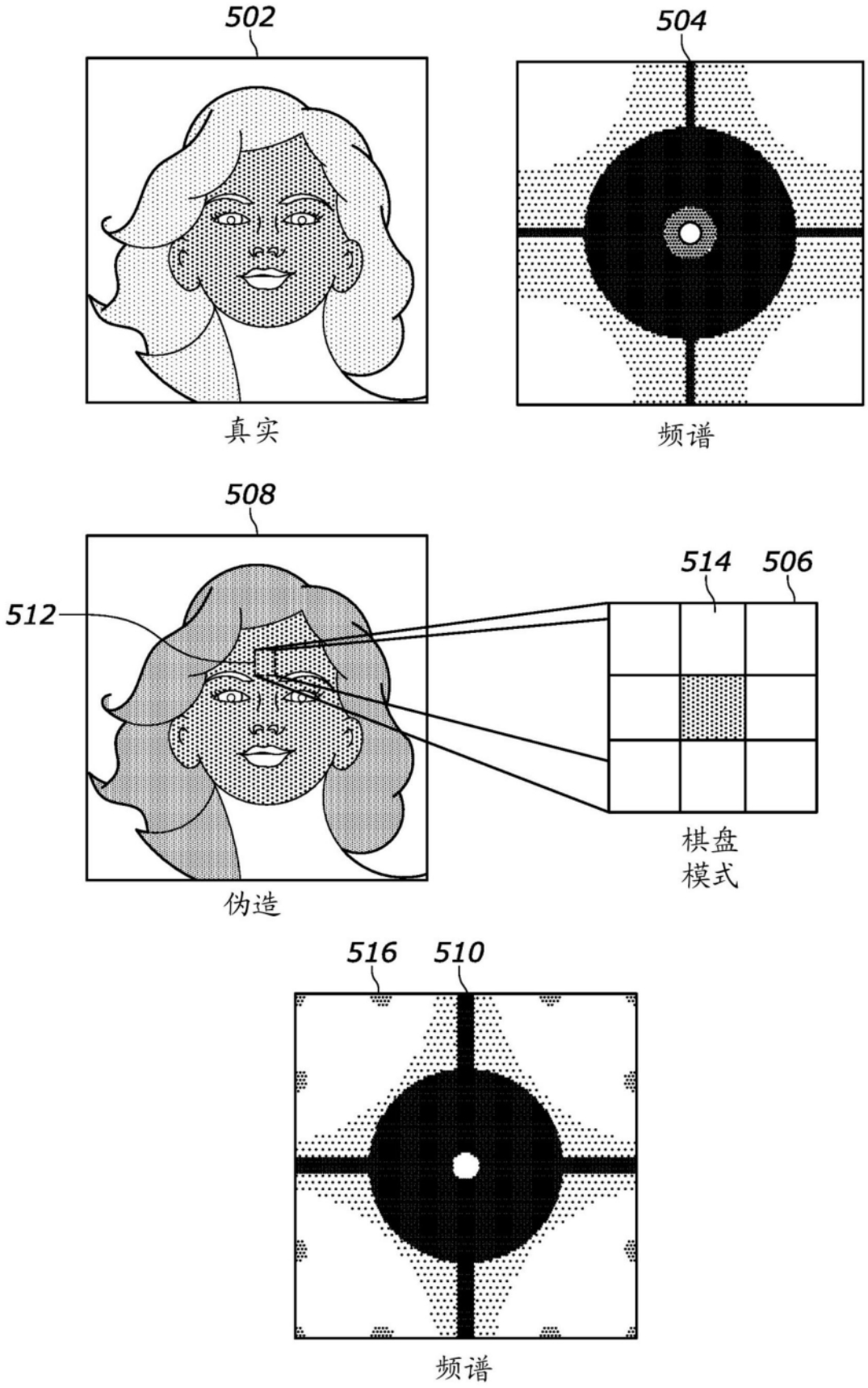


图5

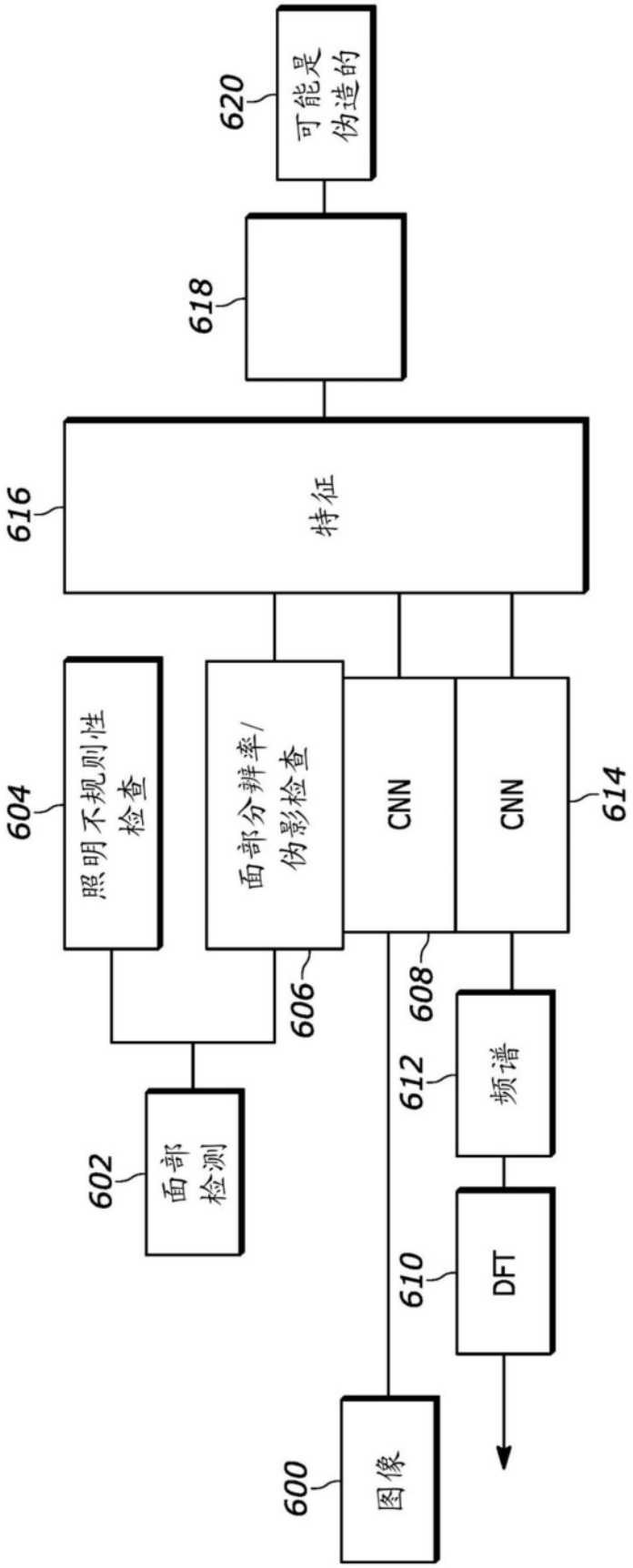

620
可能是伪造的
618
616
特征
604
照明不规则性检查
面部分辨率/伪影检查
CNN
CNN
614
606
608
602
面部检测
612
频谱
610
DFT
600
图像

图6

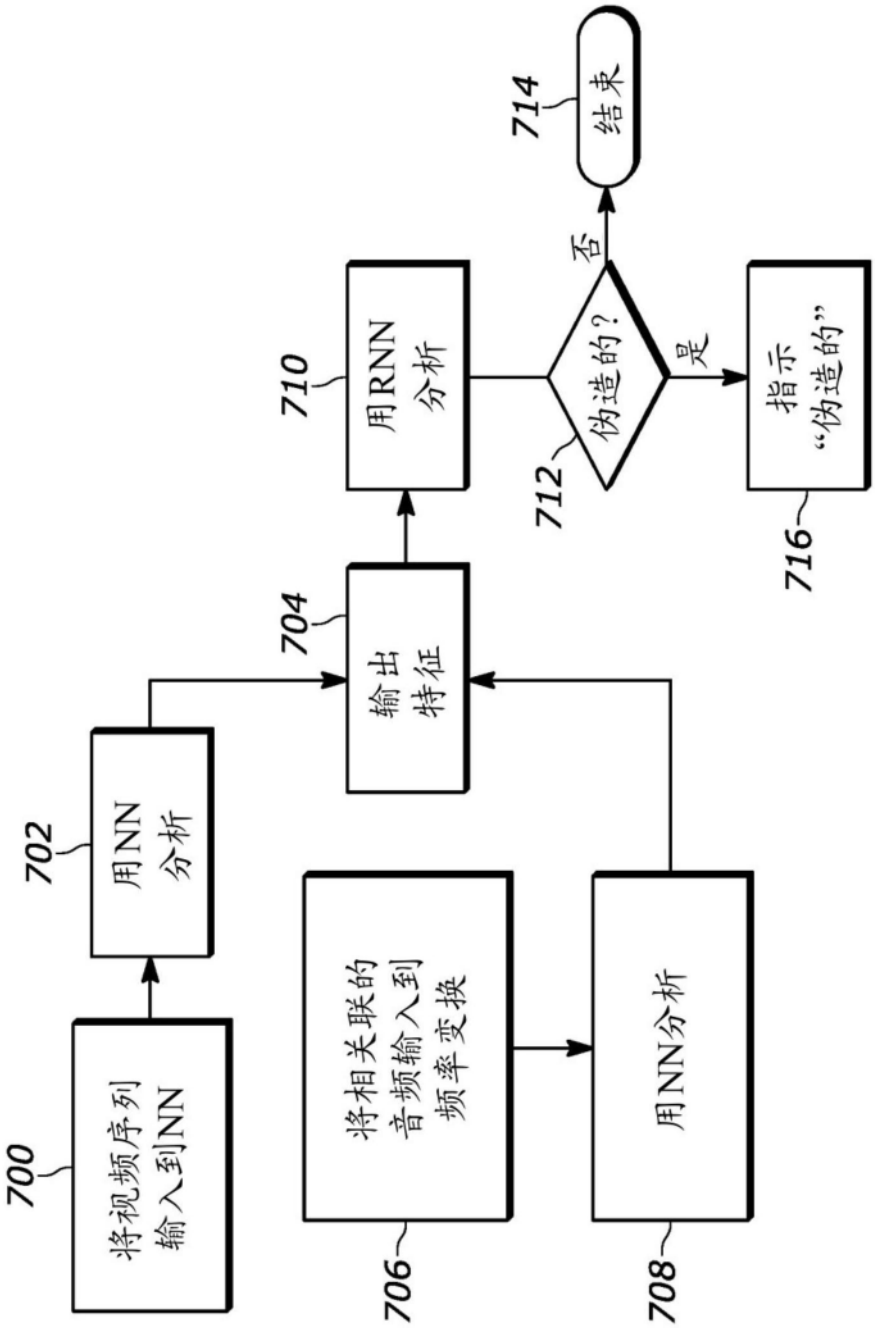

700
将视频序列输入到NN
702
用NN分析
704
输出特征
706
将相关联的音频输入到频率变换
708
用NN分析
710
用RNN分析
712
伪造的?
否
是
714
结束
716
指示“伪造的”

图7

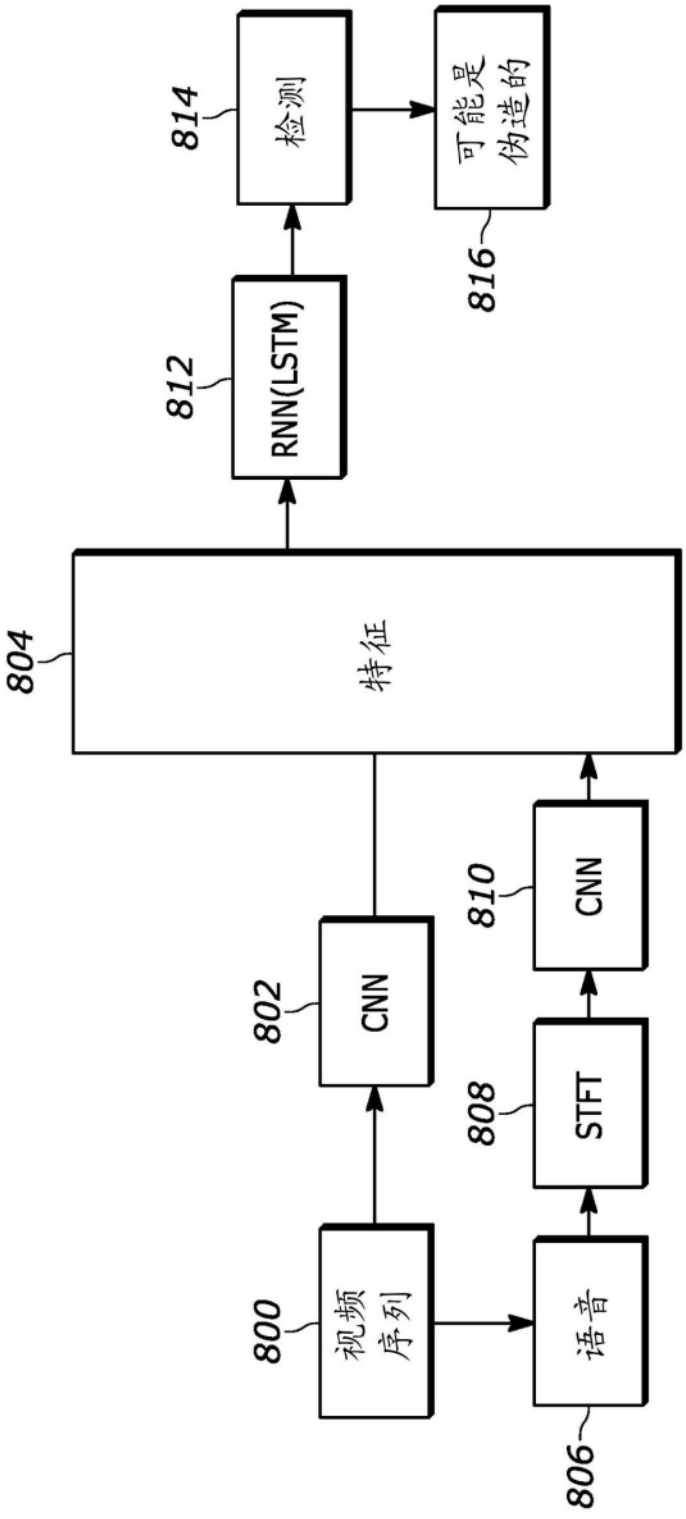

800
视频序列
802
CNN
804
特征
806
语音
808
STFT
810
CNN
812
RNN(LSTM)
814
检测
816
可能是伪造的

图8

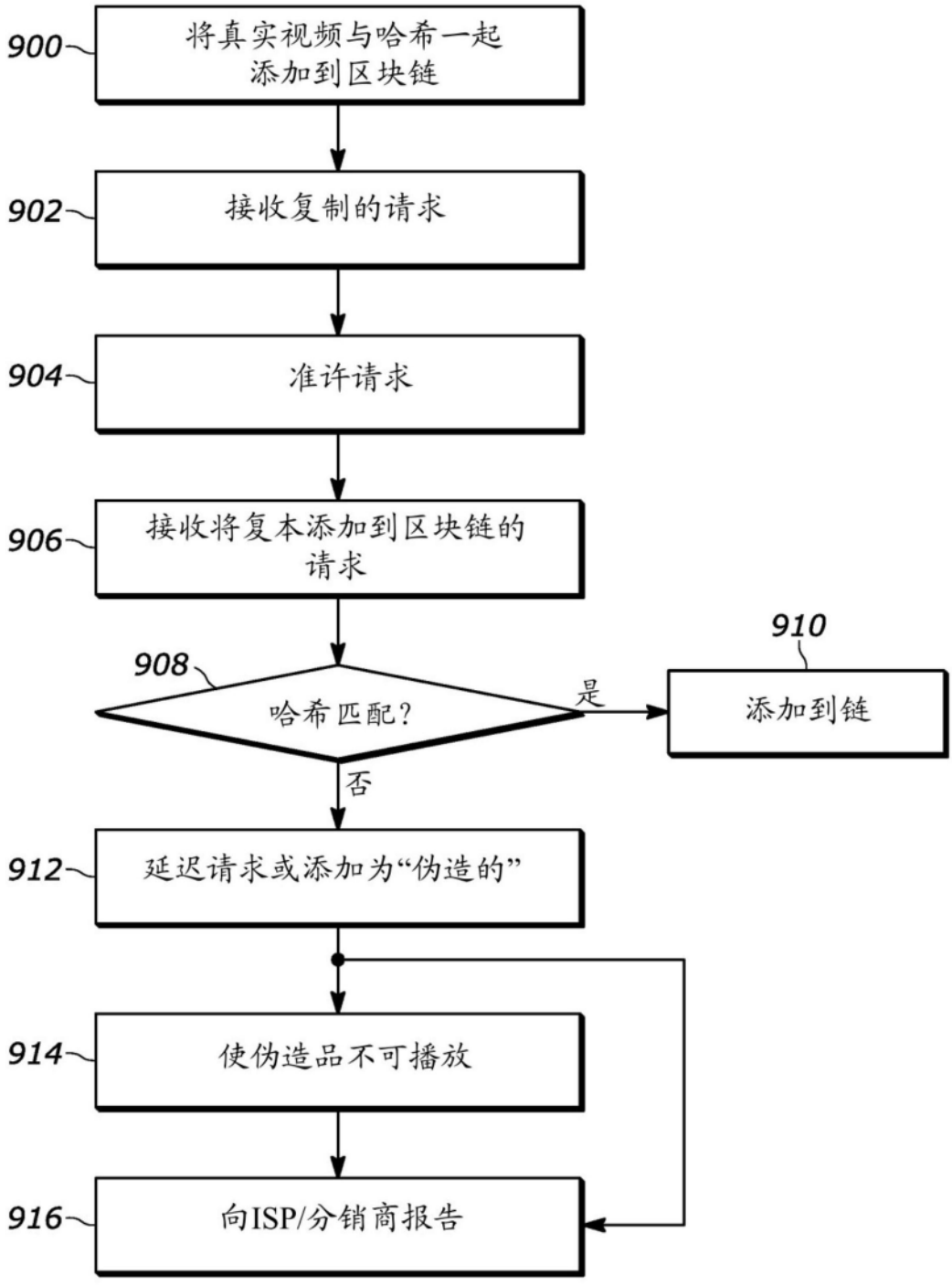

900
将真实视频与哈希一起添加到区块链
902
接收复制的请求
904
准许请求
906
接收将复本添加到区块链的请求
908
哈希匹配?
是
910
添加到链
否
912
延迟请求或添加为“伪造的”
914
使伪造品不可播放
916
向ISP/分销商报告

图9

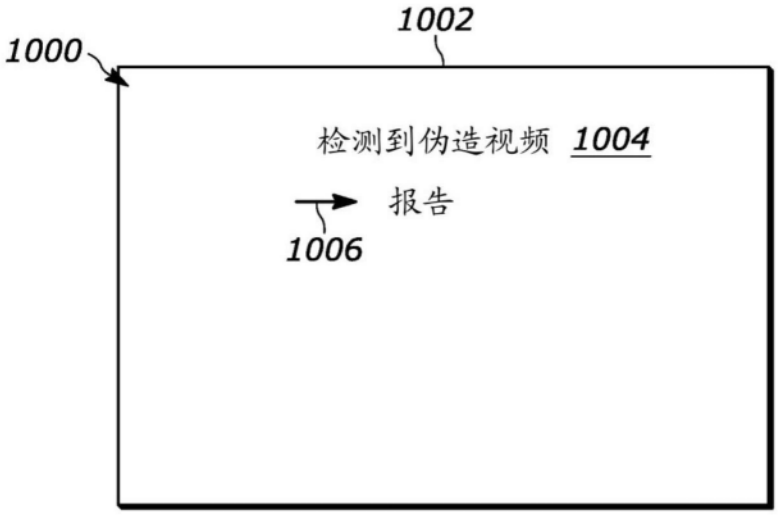

1000
1002
检测到伪造视频 1004
报告
1006

图10

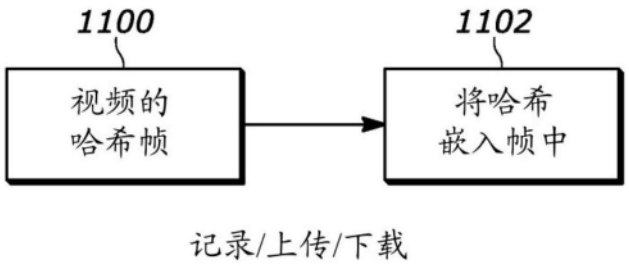

1100
视频的
哈希帧
1102
将哈希
嵌入帧中
记录/上传/下载

图11

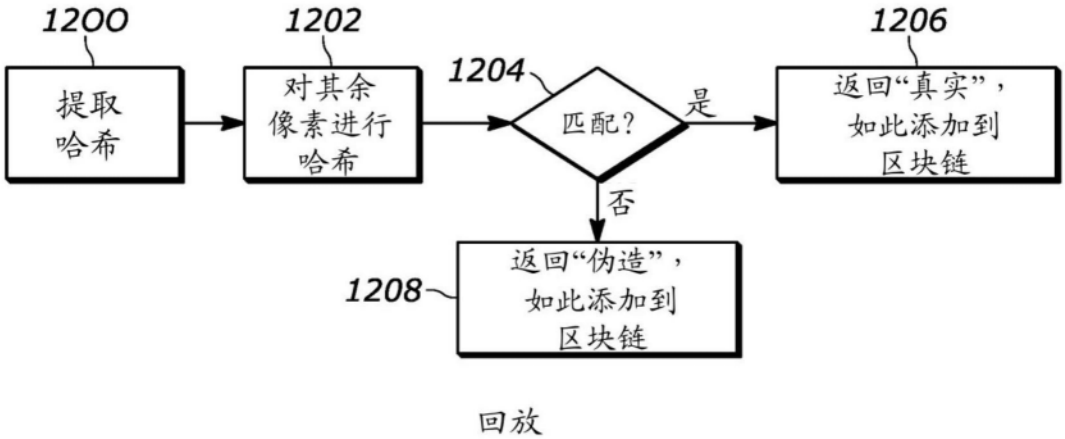

1200
提取
哈希
1202
对其余
像素进行
哈希
1204
匹配?
是
否
1206
返回“真实”，
如此添加到
区块链
1208
返回“伪造”，
如此添加到
区块链
回放

图12

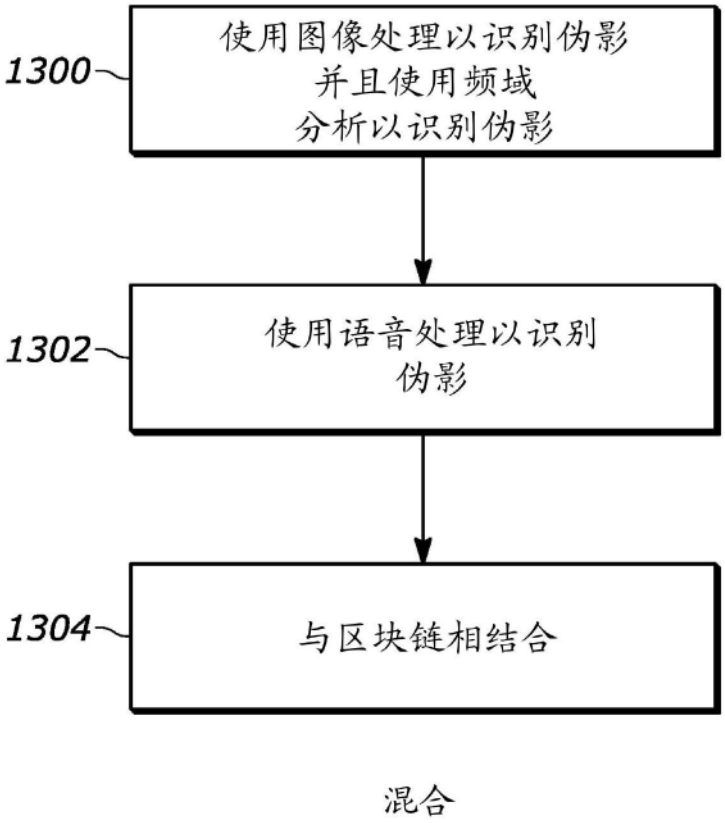

1300
使用图像处理以识别伪影
并且使用频域
分析以识别伪影
1302
使用语音处理以识别
伪影
1304
与区块链相结合
混合

图13

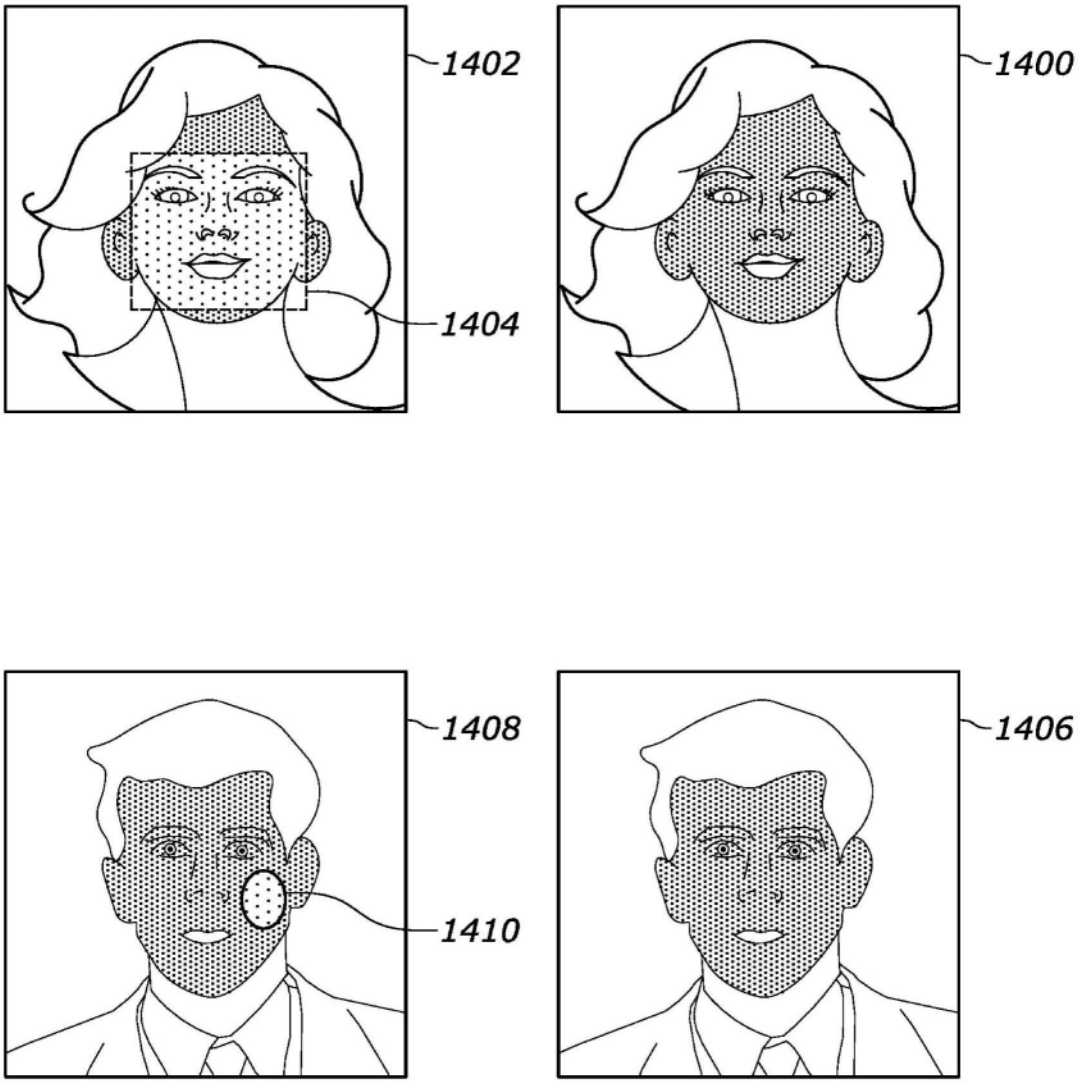


图14

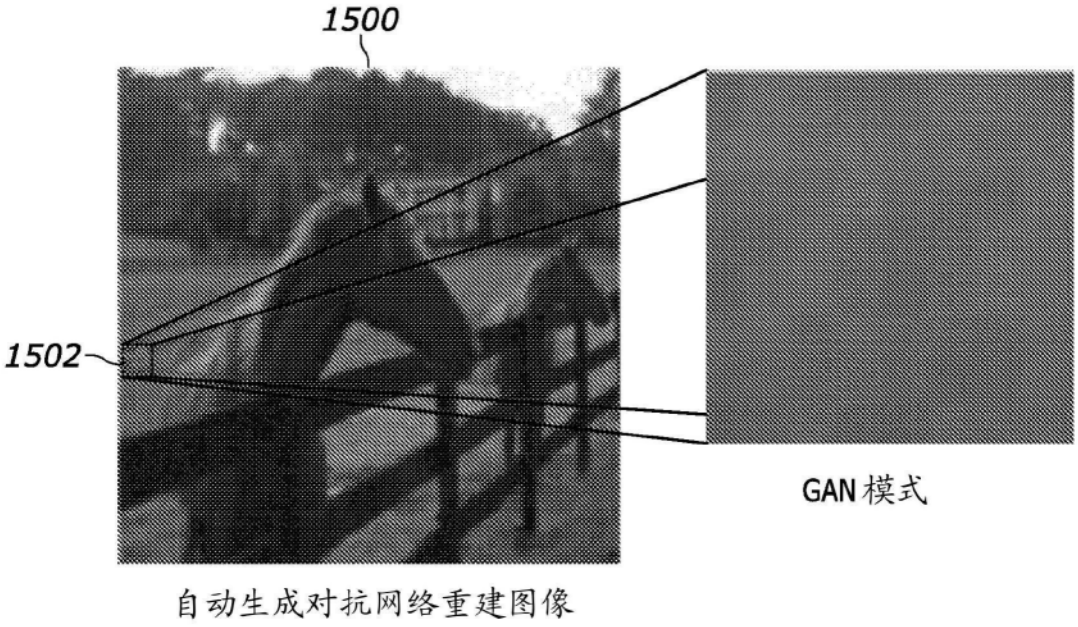


图15

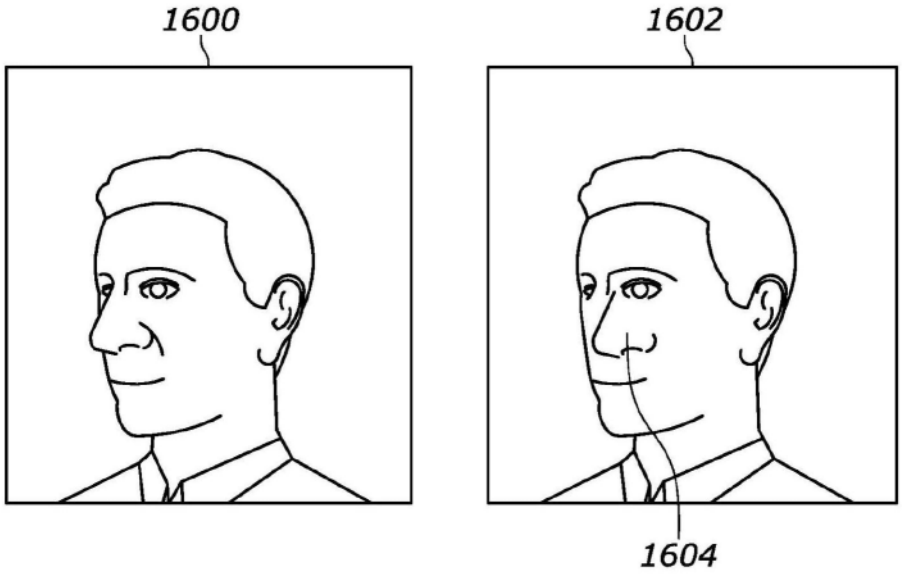


图16