(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118318435 A
(43)申请公布日 2024.07.09

(21)申请号 202280078658.7
(22)申请日 2022.10.14
(30)优先权数据
2151267-8 2021.10.15 SE
(85)PCT国际申请进入国家阶段日
2024.05.28
(86)PCT国际申请的申请数据
PCT/SE2022/050933 2022.10.14
(87)PCT国际申请的公布数据
WO2023/063870 EN 2023.04.20
(71)申请人 莱沃瑞纳技术公司
地址 瑞典斯德哥尔摩
(72)发明人 马格努斯·丹尼尔森
保罗·范登哈克
安德烈亚斯·比约克曼
(74)专利代理机构 北京汇思诚业知识产权代理有限公司 11444
专利代理师 刘晔 王刚
(51)Int.Cl.
H04N 7/15(2006.01)
H04L 12/18(2006.01)
H04L 65/403(2006.01)
H04N 21/2187(2006.01)
H04N 21/45(2006.01)

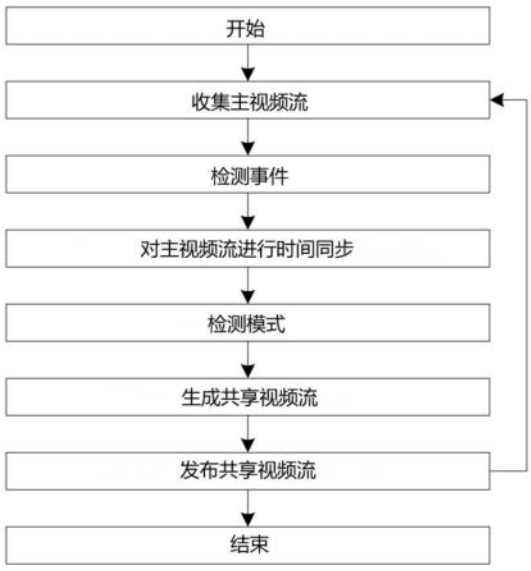
权利要求书4页 说明书19页 附图8页

(54)发明名称

用于生成共享视频流的系统和方法

(57)摘要

一种用于提供共享数字视频流的方法,包括以下步骤:收集步骤,从至少两个数字视频源(120)收集相应主数字视频流(210)。同步步骤,相对于公共时间基准对所述主数字视频流进行时间同步(260)。模式检测步骤,分析经时间同步的主数字视频流以检测至少一个模式(212)。生成步骤,基于所述经时间同步的主数字视频流的连续考虑的帧(213)和所述检测到的模式生成共享数字视频流作为输出数字视频流(230)。发布步骤,将所述输出数字视频流连续提供给共享数字视频流的消费者。同步步骤包括有意地引入延迟,使得输出数字视频流至少以所述延迟提供。模式检测步骤包括考虑所述主数字视频流的信息,该信息存在于还将用于生成输出数字视频流的经时间同步的主数字视频流的帧之后的帧(213)中。本发明还涉及一种系统和一种计算机软件产品。



CN 118318435 A

1.一种用于提供共享数字视频流的方法,所述方法包括以下步骤:

在收集步骤中,从至少两个数字视频源(120)收集相应的主数字视频流(210);

在同步步骤中,相对于公共时间基准(260)对所述主数字视频流(210)进行时间同步;

在模式检测步骤中,分析经时间同步的主数字视频流(210)以检测从第一模式集合中选择的至少一个模式(212);

在生成步骤中,基于所述经时间同步的主数字视频流(210)的连续考虑的帧(213)和所检测到的模式(212)生成所述共享数字视频流作为输出数字视频流(230);以及

在发布步骤中,将所述输出数字视频流(230)连续提供给所述共享数字视频流的消费者,

其中,所述同步步骤包括有意引入延迟,使得所述输出数字视频流(230)至少以所述延迟提供,并且其中,

所述模式检测步骤包括考虑所述主数字视频流(210)的信息,所述信息存在于还将用于生成所述输出数字视频流(230)的经时间同步的主数字视频流(210)的帧之后的帧(213)中。

2.根据权利要求1所述的方法,其中,

所述主数字视频流(210)中的至少两个被提供为涉及提供所讨论的所述主数字视频流(210)的相应远程连接的参与者客户端(121)的共享数字视频通信服务(110)的一部分。

3.根据权利要求2所述的方法,其中,

所述收集步骤包括从所述共享数字视频通信服务(110)收集所述主数字视频流(210)中的至少一个。

4.根据权利要求2或3所述的方法,其中,

所述收集步骤包括收集所述主数字视频流(210)中的至少一个作为从所述共享数字视频通信服务(110)外部的信息源(300)收集的外部数字视频流(301)。

5.根据前述权利要求中任一项所述的方法,其中,

所述主数字视频流(210)中的至少一个具有偏离的视频编码、帧速率、纵横比和/或分辨率。

6.根据前述权利要求中任一项所述的方法,其中,

所述收集步骤包括将所述主数字视频流(210)中的至少两个转换成公共协议(240),所述公共协议(240)规定在不执行任何数字视频解码或数字视频编码的情况下以原始二进制形式存储数字视频数据,所述公共协议(240)还规定存储与所存储的数字视频数据相关的指定时间点相关联的元数据(242)。

7.根据权利要求6所述的方法,其中,

所述元数据(242)包括关于所讨论的所述主数字视频流(210)的格式的信息,诸如关于所使用的数字视频编码、分辨率、帧速率或纵横比的信息。

8.根据权利要求6或7所述的方法,其中,

所述收集步骤包括对使用不同编码格式编码的主数字视频流(210)使用不同的格式特定的收集功能。

9.根据权利要求6至8中任一项所述的方法,其中,

所述转换包括将所述主数字视频流(210)中的每一个的原始二进制数据拆分为较小的

数据集(241),并将所述较小的数据集(241)中的每一个与所述公共时间基准(260)的相应时间相关联。

10.根据权利要求6至9中任一项所述的方法,其中,

所述转换包括根据需要将所述主数字视频流(210)的所述原始二进制数字视频数据下采样或上采样到公共帧速率和公共分辨率。

11.根据权利要求10所述的方法,其中,

所述主数字视频流(210)中的每一个作为各自与对应时间戳相关联的单独的帧或帧序列(213)存储在单独的缓冲器(250)中,所述对应时间戳又与所述公共时间基准(260)相关联。

12.根据权利要求10或11所述的方法,其中,

诸如覆盖物或效果的至少一条附加数字视频信息(220)也作为各自与对应时间戳相关联的单独的帧或帧序列而存储在相应单独的缓冲器(250)中,所述对应时间戳又与所述公共时间基准(260)相关联。

13.根据前述权利要求中任一项所述的方法,还包括:

在事件检测步骤中,单独分析所述主数字视频流(210)以检测从第一事件集合中选择的至少一个事件(211)。

14.根据权利要求6至12和13中任一项所述的方法,其中,

所述事件检测步骤包括使用所述公共协议(240)存储描述检测到的事件(211)的与其中检测到所讨论的所述事件(211)的所述主数字视频流(210)相关联的元数据(242)。

15.根据权利要求13或14所述的方法,其中,

所述事件检测步骤包括第一经训练的神经网络或其他机器学习部件(132a)分析所述主数字视频流(210)中的每一个,以便自动检测所述事件(211)。

16.根据权利要求15所述的方法,其中,

所述事件(211)是演示幻灯片的改变,并且其中,

所述事件检测步骤包括以下中的至少一个:

首先,基于幻灯片的第一图像和幻灯片的随后的第二图像之间的差异的图像分析来检测所述事件(211);以及

其次,基于所述第二图像的信息复杂度的图像分析来检测所述事件(211)。

17.根据权利要求15所述的方法,其中,

所述事件(211)是参与者客户端(121)到数字视频通信服务(110)的通信连接的丢失,并且其中,

所述检测步骤包括基于对应于所讨论的所述参与者客户端(121)的主数字视频流(210)的一系列后续视频帧(213)的图像分析来检测所述参与者客户端(121)已经失去通信连接。

18.根据前述权利要求中任一项所述的方法,其中,

所述公共时间基准(260)包括公共音频信号(111),所述公共音频信号(111)对于涉及各自提供所述主数字视频流(210)中的相应一个的至少两个远程连接的参与者客户端(121)的共享数字视频通信服务(110)是公共的。

19.根据权利要求1至17中任一项所述的方法,其中,

所述主数字视频流(210)中的至少两个由相应参与者客户端(121)提供给共享数字视频通信服务(110),每个这样的参与者客户端(121)具有被布置为检测作为提供给所讨论的所述参与者客户端(121)的所述输出数字视频流(230)的一部分而提供的时间同步元素(231)的到达时间的相应本地同步软件(125),并且其中,

所述公共时间基准(260)至少部分基于所述检测到的到达时间来确定。

20.根据前述权利要求中任一项所述的方法,其中,

所述公共时间基准(260)至少部分地基于所述主数字视频流(210)中的第一主数字视频流的音频部分(214)和所述第一主数字视频流(210)的图像部分(215)之间的所检测到的差异来确定,所述差异基于在所述第一主数字视频流(210)中观看的讲话参与者(122)的数字唇音同步视频分析。

21.根据前述权利要求中任一项所述的方法,其中,

所述延迟至多30秒,诸如至多5秒、所述至多1秒、所述至多0.5秒。

22.根据前述权利要求中任一项所述的方法,其中,

所述模式检测步骤包括第二经训练的神经网络或其他机器学习部件(134a)一起分析所述主数字视频流(120)以自动检测所述模式(212)。

23.根据前述权利要求中任一项所述的方法,其中,

所述检测到的模式(212)包括涉及共享视频通信服务(110)的各自与相应参与者客户端(121)相关联的至少两个不同说话参与者(122)的说话模式,所述说话参与者(122)中的每一个在所述主数字视频流(210)中的相应一个中观看。

24.根据前述权利要求中任一项所述的方法,其中,

所述生成步骤还包括基于关于所述输出数字视频流(230)中的所述主数字视频流(210)中的各个视频流的可见性的预定和/或动态可变参数的集合、视觉和/或听觉视频内容布置、所使用的视觉或听觉效果、和/或所述输出数字视频流(230)的输出模式,来生成所述输出数字视频流(230)。

25.根据前述权利要求中任一项所述的方法,其中,

所述主数字视频流(210)中的至少一个被提供给数字视频通信服务(110),并且其中,

所述发布步骤包括将所述输出数字视频流(230)提供给所述通信服务(110),诸如提供给所述通信服务(110)的参与者客户端(121)、或者提供给外部消费者(150)。

26.根据前述权利要求中任一项所述的方法,其中,

所述生成步骤由中央服务器(130)执行,从而经由应用编程接口API(137)将所述输出数字视频流(230)作为现场视频流提供给一个或几个并发消费者。

27.一种用于提供共享数字视频流的计算机软件产品,计算机软件功能被布置为在运行时执行:

收集步骤,其中从至少两个数字视频源(120)收集相应主数字视频流(210);

同步步骤,其中相对于公共时间基准(260)对所述主数字视频流(210)进行时间同步;

模式检测步骤,其中分析经时间同步的主数字视频流(210)以检测从第一模式集合中选择的至少一个模式(212);

生成步骤,其中基于所述经时间同步的主数字视频流(210)的连续考虑的帧(213)和所检测到的模式(212)生成所述共享数字视频流作为输出数字视频流(230);以及

发布步骤,其中将所述输出数字视频流(230)连续提供给所述共享数字视频流的消费者,

其中所述同步功能包括有意引入延迟,使得所述输出数字视频流(230)至少以所述延迟提供,并且其中,

所述模式检测步骤包括考虑所述主数字视频流(210)的信息,所述信息存在于还将用于生成所述输出数字视频流(230)的经时间同步的主数字视频流(210)的帧之后的帧(213)中。

28.一种用于提供共享数字视频流的系统(100),所述系统(100)包括中央服务器(130),所述中央服务器又包括:

收集功能(131),所述收集功能被布置为从至少两个数字视频源(120)收集相应主数字视频流(210);

同步功能(133),所述同步功能被布置为相对于公共时间基准(260)对所述主数字视频流(210)进行时间同步;

模式检测功能(134),所述模式检测功能被布置为分析经时间同步的主数字视频流(210)以检测从第一模式集合中选择的至少一个模式(212);

生成功能(135),所述生成功能被布置为基于所述经时间同步的主数字视频流(210)的连续考虑的帧(213)和所检测到的模式(212)生成所述共享数字视频流作为输出数字视频流(230);以及

发布功能(136),所述发布功能被布置为将所述输出数字视频流(230)连续提供给所述共享数字视频流的消费者,

其中,所述同步功能(133)被布置为有意引入延迟,使得所述输出数字视频流(230)至少以所述延迟提供,并且其中,

模式检测功能(134)被布置为考虑所述主数字视频流(210)的信息,所述信息存在于还将用于生成所述输出数字视频流(230)的经时间同步的主数字视频流(210)的帧之后的帧(213)中。

用于生成共享视频流的系统和方法

技术领域

[0001] 本发明涉及一种用于生成数字视频流,以及特别是用于基于两个或更多个不同的数字输入视频流生成数字视频流的系统、计算机软件产品和方法。在优选实施例中,数字视频流在数字视频会议或数字视频会议或会面系统的场景下生成,特别地涉及多个不同的并发用户。所生成的数字视频流可以在外部或在数字视频会议或数字视频会议系统内发布。

背景技术

[0002] 在其他实施例中,本发明应用于非数字视频会议的场景中,但是其中几个数字视频输入流被并发处理并组合为所生成的数字视频流。例如,这些场景可以是教育性的或指导性的。

[0003] 许多数字视频会议系统已经问世,诸如Microsoft®Teams®、Zoom®和Google® Meet®。通过这些数字视频会议系统,两个或更多个人可用本地录制且播送给全部参与者的数字视频和音频来进行虚拟会面,从而模拟实际会面。

[0004] 这些数字视频会议方案普遍需要改进,尤其是在观看内容的生成方面,诸如在什么时间向谁展示什么以及经由什么分布渠道。

[0005] 例如,一些系统自动检测当前正在讲话的参与者,并向其他参与者显示正在讲话的参与者的对应视频馈送。在许多系统中,可以共享图形,诸如当前显示的屏幕、视窗或数字演示。然而,随着虚拟会面变得越来越复杂,服务知道在每个时间点要向每个参与者示出全部当前可用信息中的哪些内容很快变得越来越困难。

[0006] 在其他示例中,演示参与者在台上四处走动,同时谈论数字演示中的幻灯片。然后系统需要决定是显示演示文稿、演示者还是两者,或者在两者之间进行切换。

[0007] 可能期望通过自动生成过程基于多个输入数字视频流生成一个或几个输出数字视频流,并将这样生成的一个或多个数字视频流提供给一个或几个消费者。

[0008] 然而,在许多情况下,由于这种数字视频会议系统面临许多技术困难,动态会议屏幕布局管理器或其他自动化生成功能难以选择要示出什么信息。

[0009] 首先,由于数字视频会面具有实时性方面,因此低延迟很重要。当不同的传入的数字视频流(诸如来自使用不同硬件加入的不同参与者)与不同的延迟、帧速率、纵横比或分辨率相关联时,这就生成了问题。很多时候,这些传入的数字视频流需要处理以便获得良好的用户体验。

[0010] 其次,存在时间同步的问题。由于各种输入数字视频流(诸如外部数字视频流或由参与者提供的数字视频流)通常被馈送到中央服务器等,因此不存在同步每个这种数字视频馈送的绝对时间。像过高的延迟一样,不同步的数字视频馈送将导致较差的用户体验。

[0011] 第三,多方数字视频会面可能涉及具有不同编码或格式的不同数字视频流,这些不同数字视频流需要解码和重新编码,进而生成延迟和同步方面的问题。这种编码涉及繁重的计算工作,因而造成很高的硬件成本。

[0012] 第四,不同的数字视频源可能与不同的帧速率、纵横比和分辨率相关联的事实也

可能导致存储器分配需求可能不可预测地变化,从而需要持续平衡。这可能导致附加延迟和同步问题。结果是较大的缓冲需求。

[0013] 第五,参与者可能在可变连接性、离开/重新连接等方面遇到各种挑战,从而在自动生成良好的用户体验方面提出了进一步的挑战。

[0014] 这些问题在更复杂的会面情况下被放大,例如涉及许多与会者、使用不同硬件和/或软件进行连接的参与者、外部提供的数字视频流、屏幕共享、或者多个主机。

[0015] 在将基于几个输入数字视频流生成输出数字视频流的其他场景中,诸如在用于教育和教学的数字视频生成系统中,会出现对应问题。

发明内容

[0016] 本发明解决了上述问题中的一个或几个。

[0017] 因此,本发明涉及一种用于提供共享数字视频流的方法,该方法包括以下步骤:在收集步骤中,从至少两个数字视频源收集相应主数字视频流;在同步步骤中,相对于公共时间基准对所述主数字视频流进行时间同步;在模式检测步骤中,分析经时间同步的主数字视频流以检测从第一模式集合中选择的至少一个模式;在生成步骤中,基于所述经时间同步的主数字视频流的连续考虑的帧和所述检测到的模式生成共享数字视频流作为输出数字视频流;以及在发布步骤中,将所述输出数字视频流连续提供给共享数字视频流的消费者,其中同步步骤包括有意地引入延迟,使得输出数字视频流至少以所述延迟提供,并且其中模式检测步骤包括考虑所述主数字视频流的信息,该信息存在于还将用于生成输出数字视频流的经时间同步的主数字视频流的帧之后的帧中。

[0018] 本发明还涉及一种用于提供共享数字视频流的计算机软件产品,该计算机软件功能被布置为在运行时执行:

[0019] 收集步骤,其中从至少两个数字视频源收集相应主数字视频流;同步步骤,其中相对于公共时间基准对所述主数字视频流进行时间同步;模式检测步骤,其中分析经时间同步的主数字视频流以检测从第一模式集合中选择的至少一个模式;生成步骤,其中基于所述经时间同步的主数字视频流的连续考虑的帧和所述检测到的模式,生成共享数字视频流作为输出数字视频流;以及发布步骤,其中将所述输出数字视频流连续地提供给共享数字视频流的消费者,其中同步步骤包括有意地引入延迟,使得输出数字视频流至少以所述延迟提供,并且其中模式检测步骤包括考虑所述主数字视频流的信息,该信息存在于还将用于生成输出数字视频流的经时间同步的主数字视频流的帧之后的帧中。

[0020] 而且,本发明涉及一种用于提供共享数字视频流的系统,该系统包括中央服务器,该中央服务器又包括:收集功能,该收集功能被布置为从至少两个数字视频源收集相应主数字视频流;同步功能,该同步功能被布置为相对于公共时间基准对所述主数字视频流进行时间同步;模式检测功能,该模式检测功能被布置为分析经时间同步的主数字视频流,以检测从第一模式集合中选择的至少一个模式;生成功能,该生成功能被布置为基于所述经时间同步的主数字视频流的连续考虑的帧和所述检测到的模式生成共享数字视频流作为输出数字视频流;以及发布功能,该发布功能被布置为向共享数字视频流的消费者连续提供所述输出数字视频流,其中同步功能被布置为有意地引入延迟,使得输出数字视频流至少以所述延迟提供,并且其中模式检测步骤被布置为考虑所述主数字视频流的信息,该信

息存在于还将用于生成输出数字视频流的经时间同步的主数字视频流的帧之后的帧中。

附图说明

[0021] 在下文中,将参照本发明的示例性实施例和附图详细描述本发明,其中:

[0022] 图1示出了根据本发明的第一系统;

[0023] 图2示出了根据本发明的第二系统;

[0024] 图3示出了根据本发明的第三系统;

[0025] 图4示出了根据本发明的中央服务器;

[0026] 图5示出了在用于在根据本发明的系统中使用的中央服务器;

[0027] 图6a至图6f示出了与图5中示出的方法中不同方法步骤相关的后续状态;以及

[0028] 图7概念性地示出了用于本发明中的公共协议。

[0029] 全部附图共享相同或相应部分的附图标记。

具体实施方式

[0030] 图1示出了根据本发明的被布置来执行根据本发明的用于提供共享数字视频流的方法的系统100。

[0031] 系统100可以包括视频通信服务110,但是在一些实施例中,视频通信服务110也可以在系统100外部。

[0032] 系统100可以包括一个或几个参与者客户端121,但是在一些实施例中,参与者客户端121中的一个、一些或全部也可以在系统100外部。

[0033] 系统100包括中央服务器130。

[0034] 如本文所用,术语“中央服务器”是被布置为以逻辑集中的方式被访问(诸如经由明确定义的API(Application Programming Interface,应用编程接口))的计算机实施的功能性。这种中央服务器的功能性可以完全以计算机软件来实施、或者以软件与虚拟和/或物理硬件的组合来实施。它可以在独立的物理或虚拟服务器计算机上实施,或者分布在几个互连的物理和/或虚拟服务器计算机上。

[0035] 中央服务器130在其上运行的物理或虚拟硬件(换句话说,限定中央服务器130的功能性的计算机软件)可以包括本身常规的CPU、本身常规的GPU、本身常规的RAM/ROM存储器、本身常规的计算机总线以及本身常规的外部通信功能,诸如互联网连接。

[0036] 视频通信服务110在其被使用的程度上也是所述意义上的中央服务器,其可以是与中央服务器130不同的中央服务器或者是中央服务器130的一部分。

[0037] 相应地,所述参与者客户端121中的每一个可以是在所述意义上的具有对应解释的中央服务器,并且每个参与者客户端121在其上运行的物理或虚拟硬件(换句话说,限定参与者客户端121的功能的计算机软件)还可以包括本身常规的CPU/GPU、本身常规的RAM/ROM存储器、本身常规的计算机总线以及本身常规的外部通信功能,诸如互联网连接。

[0038] 每个参与者客户端121通常还包括以下或与以下通信:计算机屏幕,该计算机屏幕被布置为显示作为正在进行的视频通信的一部分提供给参与者客户端121的视频内容;扬声器,该扬声器被布置为发出作为所述视频通信的一部分提供给参与者客户端121的声音内容;摄像机;以及麦克风,该麦克风被布置为将声音本地记录到所述视频通信的人类参与

者122,参与者122使用所讨论的参与者客户端121参与所述视频通信。

[0039] 换句话说,每个参与者客户端121的相应人机接口允许相应参与者122在与其他参与者和/或由各种源提供的音频/视频流的视频通信中与所讨论的客户端121进行交互。

[0040] 通常,参与者客户端121中的每一个包括相应输入装置123,该相应输入装置可以包括所述摄像机;所述麦克风;键盘;计算机鼠标或触控板;和/或用于接收数字视频流、数字音频流和/或其他数字数据的API。输入装置123特别地被布置为从中央服务器(诸如视频通信服务110和/或中央服务器130)接收视频流和/或音频流,这种视频流和/或音频流作为视频通信的一部分提供,并且优选地基于从这种数字数据输入流的至少两个源(例如参与者客户端121和/或外部源(参见下文))提供给所述中央服务器的对应数字数据输入流来生成。

[0041] 更一般地,参与者客户端121中的每一个包括相应输出装置124,该相应输出装置可以包括所述计算机屏幕;所述扬声器;以及用于发出数字视频和/或音频流的API,这种流代表使用所讨论的参与者客户端121在本地捕获到参与者122的视频和/或音频。

[0042] 实际上,每个参与者客户端121可以是布置有屏幕、扬声器、麦克风和互联网连接的移动设备,诸如移动电话,该移动设备本地运行计算机软件或访问远程运行的计算机软件以执行所讨论的参与者客户端121的功能性。相应地,参与者客户端121也可以是使用经由网络浏览器的远程访问功能性等(视情况而定)运行本地安装的应用的厚或薄的膝上型或固定计算机。

[0043] 可以存在在当前类型的同一视频通信中使用的多于一个(诸如至少三个)参与者客户端121。

[0044] 视频通信可以至少部分地由视频通信服务110提供,并且至少部分地由中央服务器130提供,如将在本文中描述和例示的那样。

[0045] 如本文中所用,术语“视频通信”是指交互式数字通信会话,该交互式数字通信会话涉及至少两个且优选地至少三个视频流并且优选地还与用于生成一个或几个混合或联合数字视频/音频流匹配,其中这些视频/音频流又由一个或几个消费者消费,且该一个或几个消费者可能或可能不经由视频和/或音频为视频通信做出贡献。这种视频通信是实时的、具有或不具有一定的延迟或延时。这种视频通信的至少一个且优选至少两个参与者122以交互方式参与视频通信,提供并消费视频/音频信息。

[0046] 参与者客户端121中的至少一个或参与者客户端121中的全部包括本地同步软件功能125,这将在下文中更详细地描述。

[0047] 视频通信服务110可以包括或访问公共时间基准,如也将在下文中更详细地描述那样。

[0048] 中央服务器130可以包括API 137,用于与中央服务器130外部的实体进行数字通信。这种通信可能涉及输入和输出两者。

[0049] 系统100(诸如中央服务器130)可以被布置为与外部信息源300(诸如外部提供的视频流)进行数字通信,并且特别地从该外部信息源接收数字信息(诸如音频和/或视频流数据)。外部信息源300中的“外部”意指该信息源不是由中央服务器130提供的或者不是中央服务器的一部分。优选地,由外部信息源300提供的数字数据独立于中央服务器130,并且中央服务器130不能影响其信息内容。例如,外部信息源130可以是现场捕获的视频和/或音

频,诸如公共体育赛事或正在进行的新闻事件或报道的视频和/或音频。外部信息源130也可以由网络相机等捕获,但不是由参与者客户端121中的任何一个捕获。因此,这种捕获的视频可以描绘与参与者客户端121中的任何一个相同的地点,但不是作为参与者客户端121本身的活动的一部分而被捕获的。外部提供的信息源300和内部提供的信息源120之间的一个可能的区别是,内部提供的信息源可以作为并且在其能力方面作为以上限定的类型的视频通信的参与者来提供,而外部提供的信息源300不是,而是作为所述视频会议外部的场景的一部分来提供。

[0050] 还可以存在并行地向中央服务器130提供所述类型的数字信息(诸如音频和/或视频流)的几个外部信息源300。

[0051] 如图1所示,参与者客户端121中的每一个构成了如所描述的由所讨论的参与者客户端121提供给视频通信服务110的信息(视频和/或音频)流120的源。

[0052] 系统100(诸如中央服务器130)还可以被布置为与外部消费者150进行数字通信,以及特别地向该外部消费者发出数字信息。例如,由中央服务器130生成的数字视频和/或音频流可以通过所述API 137实时或接近实时地连续提供给一个或几个外部消费者150。同样,外部消费者150中的“外部”意指消费者150不作为中央服务器130的一部分提供,和/或它不是所述视频通信的一方。

[0053] 除非另外说明,否则本文中的全部功能和通信以数字地和电子方式提供的,由在适当的计算机硬件上运行的计算机软件实现,并且通过数字通信网络或信道(诸如因特网)进行通信。

[0054] 因此,在图1中示出的系统100的配置中,多个参与者客户端121参与由视频通信服务110提供的数字视频通信。每个参与者客户端121因此可以具有与视频通信服务110正在进行的登录、会话等,并且可以参与由视频通信服务110提供的同一个正在进行的视频通信。换句话说,视频通信在参与者客户端121之间“共享”,并且因此也被对应人类参与者122“共享”。

[0055] 在图1中,中央服务器130包括自动参与者客户端140,该自动参与者客户端是对应于参与者客户端121但不与人类参与者122相关联的自动客户端。相反,自动参与者客户端140作为参与者客户端被添加到视频通信服务110,以参加与参与者客户端121相同的共享视频通信。作为这样的参与者客户端,自动参与者客户端140被授权访问由视频通信服务110作为正在进行的视频通信的一部分提供并且可以经由自动参与者客户端140被中央服务器130消费的(多个)连续生成的数字视频和/或音频流。优选地,自动参与者客户端140从视频通信服务110接收:被分布到或者可以被分布到每个参与者客户端121的公共视频和/或音频流;从参与者客户端121中的一个或几个中的每一个提供给视频通信服务110并由视频通信服务110以原始或修改的形式中继给全部或请求的参与者客户端121的相应视频和/或音频流;和/或公共时间基准。

[0056] 中央服务器130包括收集功能131,该收集功能被布置为从自动参与者客户端140并且还可能地从(多个)所述外部信息源300接收所述类型的视频和/或音频流,用于如下所述的处理,并且然后经由API 137提供共享视频流。例如,这个共享视频流可以由外部消费者150和/或由视频通信服务110消费以便由视频通信服务110依次分布到参与者客户端121中的全部或任何请求的一个。

[0057] 图2类似于图1,但是未使用自动客户端参与者140,而是中央服务器130经由视频通信服务110的API 112从正在进行的视频通信接收视频和/或音频流数据。

[0058] 图3也类似于图1,但是没有示出视频通信服务110。在这种情况下,参与者客户端121直接与中央服务器130的API 137通信,从而例如向中央服务器130提供视频和/或音频流数据和/或从中央服务器130接收视频和/或音频流数据。然后,所生成的共享流可以被提供给外部消费者150和/或提供给客户端参与者121中的一个或几个。

[0059] 图4更详细地示出了中央服务器130。如所示出那样,所述收集功能131可以包括一个或者优选地几个格式特定的收集功能131a。所述格式特定的收集功能131a中的每一个格式特定的收集功能被布置为接收具有预定格式的视频和/或音频流,诸如预定的二进制编码格式和/或预定的流数据容器,并且被特别布置为将所述格式的二进制视频和/或音频数据解析为单独的视频帧、视频帧序列和/或时隙。

[0060] 中央服务器130还包括事件检测功能132,该事件检测功能被布置为从收集功能131接收视频和/或音频流数据(诸如二进制流数据),并对所接收的数据流中的每个单独的数据流执行相应事件检测。事件检测功能132可以包括用于执行所述事件检测的AI(Artificial Intelligence,人工智能)部件132a。该事件检测可以在没有首先对各个收集的流进行时间同步的情况下进行。

[0061] 中央服务器130还包括同步功能133,该同步功能被布置为对由收集功能131提供的并由事件检测功能132处理的数据流进行时间同步。同步功能133可以包括用于执行所述时间同步的AI部件133a。

[0062] 中央服务器130还包括模式检测功能134,该模式检测功能被布置为基于所接收的数据流中的至少一个、但在许多情况下至少两个、诸如至少三个、例如全部的组合来执行模式检测。模式检测还可以基于由事件检测功能132针对所述数据流中的每一个单独数据流检测到的一个或在某些情况下至少两个或更多个事件。由所述模式检测功能134考虑的这种检测到的事件可以相对于每个单独的所收集的流跨时间分布。模式检测功能134可以包括用于执行所述模式检测的AI部件134a。

[0063] 中央服务器130还包括生成功能135,该生成功能被布置为基于从收集功能131提供的数据流并且还基于任何检测到的事件和/或模式来生成共享数字视频流。共享视频流至少包括被生成以包括由收集功能131提供的一个或几个视频流(原始的、重新格式化的或经变换)的视频流,并且还可以包括对应音频流数据。

[0064] 中央服务器130还包括发布功能136,该发布功能被布置为例如经由如上所述的API 137发布所生成的共享数字视频流。

[0065] 注意,图1、图2和图3示出了可以如何使用中央服务器130来实施本文中描述的原理并且特别地提供根据本发明的方法的三个不同示例,但是在使用或不使用一个或多个视频通信服务110的情况下的其他配置也是可能的。

[0066] 因此,图5示出了根据本发明的用于提供所述共享数字视频流的方法。图6a至图6f示出了由图5中示出的方法步骤生成的不同数字视频/音频数据流状态。

[0067] 在第一步骤中,该方法开始。

[0068] 在随后的收集步骤中,诸如通过所述收集功能131从至少两个所述数字视频源120、300收集相应主数字视频流210、301。每个这样的主数据流210、301可以包括音频部分

214和/或视频部分215。应当理解的是,在这个场景中,“视频”指的是这种数据流的移动和/或静止图像内容。每个主数据流210、301可以根据任何视频/音频编码规范进行编码(使用提供所讨论的主数据流210、301的实体所使用的相应编解码器),并且在同一视频通信中并发使用的所述主数据流210、301中的不同流之间,编码格式可以不同。优选的是,主数据流210、301中的至少一个(诸如全部)被提供为二进制数据流,可能地以本身常规的数据容器数据结构提供。优选的是,主数据流210、301中的至少一个(诸如至少两个)或者甚至全部被提供为相应现场视频记录。

[0069] 注意,当通过收集功能131接收主流210、301时,这些主流在时间上可能是不同步的。这可能意味着它们与相对于彼此的不同延迟或延时相关联。例如,在两个主视频流210、301是现场记录的情况下,这可能意味着它们在通过收集功能131接收时与相对于记录时间的不同延迟相关联。

[0070] 还要注意的是,主流210、301本身可以是来自网络相机的相应现场相机馈送、当前共享的屏幕或演示文稿、观看的电影剪辑或类似物、或者以各种方式布置在相同屏幕中的这些的任意组合。

[0071] 收集步骤在图6a和图6b中示出。在图6b中,还示出了收集功能131可以如何将每个主视频流210、301存储为捆绑的音频/视频信息或存储为与相关联的视频流数据单独的音频流数据。图6b示出了主视频流210、301数据如何被存储为单独的帧213或帧的集合/簇,“帧”在此指的是图像数据和/或任何相关联的音频数据的时间受限部分,诸如每个帧是单独的静止图像或一起形成移动图像视频内容的连续图像系列(诸如构成最多1秒的移动图像的系列)。

[0072] 在由事件检测功能132执行的后续事件检测步骤中,诸如通过所述事件检测功能132并且特别是所述AI部件132a分析所述主数字视频流210、301,以检测从第一事件集合中选择的至少一个事件211。这在图6c中示出。

[0073] 优选的是,对至少一个、诸如至少两个、诸如全部主视频流210、301执行这个事件检测步骤,并且对每个这样的主视频流210、301单独执行这个事件检测步骤。换句话说,事件检测步骤优选地在仅考虑作为所讨论的这个特定主视频流210、301的一部分而包含的信息,并且特别是不考虑作为其他主视频流的一部分而包含的信息的情况下针对所述单独的主视频流210、301进行。另外,事件检测优选地在不考虑与几个主视频流210、301相关联的任何公共时间基准260的情况下进行。

[0074] 另一方面,事件检测优选地考虑在某个时间间隔(诸如长于0秒,诸如至少0.1秒,诸如至少1秒的主视频流的历史时间间隔)内作为所讨论的单独分析的主视频流的一部分而包含的信息。

[0075] 事件检测可以考虑包含在作为所述主视频流210、301的一部分而包含的音频和/或视频数据中的信息。

[0076] 所述第一事件集合可以包含任意数量的事件类型,诸如构成或作为所讨论的主视频流210、301的一部分的幻灯片演示文稿中的幻灯片的改变;导致图像质量改变、图像数据的丢失或图像数据恢复的、提供所讨论的主视频流210、301的源120、300的连接性质量方面的改变;以及所讨论的主视频流210、301中的检测到的移动物理事件,诸如视频中的人或物体的移动、视频中的照明的改变、音频中的突然尖锐噪声、或音频质量的改变。应当理解,提

出这些示例并不旨在穷尽地列举,而是便于理解目前描述的原理的适用性。

[0077] 在由同步功能133执行的后续同步步骤中,相对于公共时间基准260对主数字视频流210进行时间同步。如图6d所示,这种时间同步涉及使用所述公共时间基准260将主视频流210、301相互对准,使得它们可以被组合以形成时间同步的场景。公共时间基准260可以是数据流、心跳信号或其他脉冲数据,或者适用于各个主视频流210、301中的每一个的时间锚。重要的是,公共时间基准可以以某种方式应用于各个主视频流210、301中的每一个,使得所讨论的主视频流210、301的信息内容可以相对于公共时间轴明确地与公共时间基准相关。换句话说,公共时间基准允许主视频流210、301通过时移来对准,以便在当前意义上时间同步。

[0078] 如图6d所示,时间同步可以包括针对每个主视频流210、301确定与公共时间基准260相关的一个或几个时间戳261。

[0079] 在由模式检测功能134执行的随后模式检测步骤中,分析由此经时间同步的主数字视频流210、301以检测从第一模式集合中选择的至少一个模式212。这在图6e中示出。

[0080] 与事件检测步骤相反,模式检测步骤可以优选地基于作为共同考虑的经时间同步的主视频流210、301中的至少两个的一部分而包含的视频和/或音频信息来执行。

[0081] 所述第一模式集合可以包含任意数量的类型模式,诸如可互换地或同时交谈的几个参与者、或者作为不同事件(诸如不同的参与者讲话)并发发生的演示幻灯片改变。这个列表并不详尽,而是说明性的。

[0082] 在替代性实施例中,检测到的模式212可能不涉及包含在所述主视频流210、301中的若干主视频流中的信息,而涉及仅包含在所述主视频流210、301中的一个中的信息。在这种情况下,优选的是,基于跨越至少两个检测到的事件211(例如两个或更多个连续检测到的演示幻灯片改变或连接质量改变)的单个主视频流210、301中包含的视频和/或音频信息来检测这种模式212。作为示例,与针对每个检测到的幻灯片改变事件的一个单独的幻灯片改变模式相反,随着时间彼此快速跟随的几个连续的幻灯片改变可以被检测为一个单独幻灯片改变模式。

[0083] 认识到的是第一事件集合和所述第一模式集合可以包括使用相应的参数组和参数间隔来定义的预定类型的事件/模式。如下将解释那样,所述集合中的事件/模式也可以或附加地使用各种AI工具来定义和检测。

[0084] 在由生成功能135执行的后续生成步骤中,基于经时间同步的主数字视频流210、301的连续考虑的帧213和所述检测到的模式212,将共享数字视频流生成为输出数字视频流230。

[0085] 如将在下文中详细所述的那样,本发明允许输出数字视频流230的完全自动生成。

[0086] 例如,这种生成可以涉及选择来自什么主视频流210、301的什么视频和/或音频信息在所述输出视频流230中使用到什么程度、输出视频流230的视频屏幕布局、随着时间的不同的这种用途或布局之间的切换模式等。

[0087] 这在图6f中示出,该图还示出了可以与所述公共时间基准260时间同步并且在生成输出视频流230中与经时间同步的主视频流210、301一起使用的一条或几条附加时间相关(与公共时间基准260相关)的数字视频信息220,诸如附加的数字视频信息流。例如,附加流220可以包括关于要使用的任何视频和/或音频效果的信息,诸如动态地基于检测到的模

式、用于视频通信的所计划的时间表等。

[0088] 在由发布功能136执行的后续发布步骤中,所生成的输出数字视频流230被连续提供给共享数字视频流的消费者110、150,如上所述。

[0089] 在随后的步骤中,方法结束。然而,首先,方法可以迭代任意次,如图5所示,以生成作为连续提供的流的输出视频流230。优选地,输出视频流230被生成为实时或接近实时地(考虑到由沿途的全部步骤增加的总延迟)并且连续地(当更多信息可用时发布立即进行,然而不计数以下描述的有意增加的延迟)消费。这样,输出视频流230可以以交互的方式被消费,使得输出视频流230可以被反馈到视频通信服务110中或被反馈到形成用于再次被馈送到收集功能131的主视频流210的生成的基础的任何其他上下文中以便形成闭环反馈;或者使得输出视频流230可以被消费到不同的(系统100外部或至少中央服务器130外部)场景中、但是在那里形成实时、交互式视频通信的基础。

[0090] 如上所提及那样,在一些实施例中,所述主数字视频流210、301中的至少两个被提供为共享数字视频通信(诸如由所述视频通信服务110提供)的一部分,视频通信涉及提供所讨论的主数字视频流210的相应远程连接的参与者客户端121。在这种情况下,收集步骤可以包括从共享数字视频通信服务110本身收集所述主数字视频流210中的至少一个(诸如经由自动参与者客户端140,该自动参与者客户端进而被授权访问来自所讨论的视频通信服务110内的视频和/或音频流数据、和/或经由视频通信服务110的API 112)。

[0091] 而且,在这种情况和其他情况下,收集步骤可以包括收集所述主数字视频流210、301中的至少一个作为从共享数字视频通信服务110外部的信息源300收集的相应外部数字视频流301。注意,一个或几个所使用的这种外部视频源300也可以在中央服务器130的外部。

[0092] 在一些实施例中,主视频流210、301不以相同的方式格式化。这种不同的格式化可以呈它们在不同类型的数据容器(诸如AVI或MPEG)中被递送到收集功能131的形式,但是在优选实施例中,就所述偏离的主数字视频流210、301具有偏离的视频编码、偏离的固定或可变帧速率、偏离的纵横比、偏离的视频分辨率、和/或偏离的音频采样率而言,主视频流210、301中的至少一个根据偏离的格式(与所述主视频流210、301中的至少一个其他主视频流相比)被格式化。

[0093] 优选的是,收集功能131被预先配置为读取和解释出现在全部所收集的主视频流210、301中的全部编码格式、容器标准等。这使得可以执行本文中所述的处理,而不需要任何解码,直到过程中相对较晚时(诸如不需要任何解码直到所讨论的主流被放入相应缓冲器之后、不需要任何解码直到事件检测步骤之后、或者甚至不需要任何解码直到事件检测步骤之后)。然而,在其中主视频馈送210、301中的一个或几个使用收集功能131在不解码的情况下无法解释的编解码器进行编码的罕见情况下,收集功能131可以被布置为执行这种主视频流210、301的解码和分析,随后转换成可以通过例如事件检测功能处理的格式。注意,即使在这种情况下,优选的是在这个阶段不执行任何重新编码。

[0094] 例如,从多方视频事件(诸如由视频通信服务110提供的多方视频事件)中获取的主视频流220通常具有关于低延迟的要求,并且因此通常与可变帧率和可变像素分辨率相关联,以使参与者122能够进行有效通信。换句话说,为了低延迟,整体视频和音频质量将根据需要降低。

[0095] 另一方面,外部视频馈送301通常将具有更稳定的帧率、更高的质量,但因此可能具有更高的延迟。

[0096] 因此,视频通信服务110可以在每个时刻使用与外部视频源300不同的编码和/或容器。因此,本文中描述的分析和视频生成过程需要将这些不同格式的流210、301组合成新的流以以获得组合式体验。

[0097] 如上所提及那样,收集功能131可以包括格式特定的收集功能131a的集合,每个格式特定的收集功能被布置为处理特定类型格式的主视频流210、301。例如,这些格式特定的收集功能131a中的每一个格式特定的收集功能可以被布置为处理已经使用不同的视频相应编码方法/编解码器(诸如Windows® Media®或DivX)编码的主视频流210、301。

[0098] 然而,在优选实施例中,收集步骤包括将主数字视频流210、301中的至少两个(例如全部)转换为公共协议240。

[0099] 如这个场景中使用的那样,术语“协议”指的是指定如何存储包含在数字视频/音频流中的信息的信息结构化标准或数据结构。然而,公共协议优选地不指定如何在二进制级别上如此存储数字视频和/或音频信息(即,指示声音和图像本身的经编码/压缩的数据),而是替代地形成用于存储这种数据的预定格式的结构。换句话说,公共协议规定在不执行与这种存储相关的任何数字视频解码或数字视频编码的情况下以原始的二进制形式存储数字视频数据,可能地通过除了可能地级联和/或拆分二进制形式的字节序列之外,根本不修正现有的二进制形式。相反,所讨论的主视频流210、301的原始(经编码/压缩的)二进制数据内容被保留,同时以由协议限定的数据结构重新打包这个原始二进制数据。在一些实施例中,公共协议限定了视频文件容器格式。

[0100] 作为示例,图7示出了通过相应格式特定的收集功能131a重构并使用所述公共协议240的图6a中示出的主视频流210、301。

[0101] 因此,公共协议240规定将数字视频和/或音频数据存储在数据集241中,该数据集优选地沿着与所讨论的主视频流210、301相关的时间线被划分离散的连续数据集。每个这样的数据集可以包括一个或几个视频帧、以及相关联的音频数据。

[0102] 公共协议240还可以规定存储与所存储的数字视频和/或音频数据集241相关的指定时间点相关联的元数据242。

[0103] 元数据242可以包括关于所讨论的主数字视频流210的原始二进制格式的信息,诸如关于用于生成所述原始二进制数据的数字视频编码方法或编解码器的信息、视频数据的分辨率、视频帧速率、帧速率可变性标志、视频分辨率、视频纵横比、音频压缩算法、或者音频采样率。由此元数据242还可以包括与所讨论的主视频流210、301的时间基准相关的关于所存储数据的时间戳的信息。

[0104] 结合所述公共协议240使用所述格式特定的收集功能131a使得可以通过解码/重新编码所接收的视频/音频数据来快速收集主视频流210、301的信息内容而不增加延迟。

[0105] 因此,收集步骤可以包括使用所述格式特定的收集功能131a中的不同的格式特定的收集功能来收集正在使用不同二进制视频和/或音频编码格式编码的主数字视频流210、301,以便解析所讨论的主视频流210、301,并使用公共协议将经解析的原始的二进制数据以及任何相关元数据存储在数据结构中。不言而喻,关于对什么主视频流210、301使用什么格式特定的收集功能131a的确定可以通过收集功能131基于所讨论的每个主视频流210、

301的预定和/或动态检测到的属性来执行。

[0106] 每个由此收集的主视频流210、301可以存储在中央服务器130中的其自己单独的存储缓冲器(诸如RAM存储缓冲器)中。

[0107] 通过每个格式特定的收集功能131a执行的主视频流210、301的转换因此可以包括将每个这样经转换的主数字视频流210、301的原始二进制数据拆分为所述较小数据集241的有序集合。

[0108] 而且,转换还可以包括将所述较小集合241中的每一个(或子集,诸如沿着所讨论的主流210、301的相应时间线的规则分布的子集)与所述公共时间基准260的相应时间相关联。这种关联可以通过分析原始二进制视频和/或音频数据以下文中描述的原理方式中的任何一个或者以其他方式来执行,并且可以被执行为以便能够执行主视频流210、301的后续时间同步。取决于所使用的公共时间基准260的类型,数据集241中的每一个的这种关联的至少一部分也可以或替代地通过同步功能133执行。在后一情况下,收集步骤可以替代地包括将较小集合241中的每一个或子集与所讨论的主流210、301特定的时间线的相应时间相关联。

[0109] 在一些实施例中,收集步骤还包括将从主视频流210、301收集的原始二进制视频和/或音频数据转换为统一质量和/或更新频率。这可以涉及根据需要将主数字视频流210、301的所述原始二进制数字视频和/或音频数据下采样或上采样到公共视频帧速率、公共视频分辨率、或者公共音频采样率。注意,这种重新采样可以在不执行完全解码/重新编码的情况下执行,或者甚至在根本不执行任何解码的情况下执行,因为所讨论的格式特定的收集功能131a可以根据正确的二进制编码目标格式直接处理原始二进制数据。

[0110] 优选地,所述主数字视频流210、301中的每一个作为如上所述并且各自与对应时间戳相关联的单独的帧213或帧序列213存储在单独的数据存储缓冲器250中,该对应时间戳又与所述公共时间基准260相关联。

[0111] 在出于说明目的而提供的具体示例中,视频通信服务110是运行涉及并发参与者122的视频会议的Microsoft® Teams®。自动参与者客户端140被注册为Teems®会面中的会面参与者。

[0112] 然后,主视频输入信号210可用于经由自动参与者客户端140的收集功能130并且经由自动参与者客户端通过收集功能获得。这些是呈H264格式的原始信号,并且包含针对每个视频帧的时间戳信息。

[0113] 相关的格式特定的收集功能131a在可配置的预定义TCP端口上通过IP(云中的LAN网络)采集原始数据。每个Teems®会面参与者以及相关联的音频数据与单独的端口相关联。收集功能131然后使用来自音频信号(其呈50Hz)的时间戳,并且在将视频流220存储在其相应的单独缓冲器250中之前将视频数据下采样为25Hz的固定输出信号。

[0114] 如所提及的那样,公共协议240以原始二进制形式存储数据。它可以被设计为示非常低级的并用于处理视频/音频数据的原始位和字节。在优选实施例中,数据作为简单的字节数组或对应数据结构(诸如切片)存储在公共协议240中。这意味着根本不需要将数据放入常规视频容器中(所述公共协议240在这个场景中不构成常规容器)。而且,编码和解码视频在计算量方面很大,这意味着它导致延迟并需要昂贵的硬件。而且,这个问题与参与者的数量成比例。

[0115] 使用公共协议240,变得可以在收集功能131中为与每个Teams®会面参与者122相

关联的主视频流210以及为任何外部视频源300预留存储器，并且然后在过程期间动态地改变所分配的存储器的量。这样，变得可以改变输入流的数量，并且结果保持每个缓冲器有效。例如，由于像分辨率、帧速率等信息可能是可变的，但在公共协议240中被存储为元数据，因此这个信息可以用于如可能需要那样快速调整每个缓冲区的大小。

[0116] 以下是当前类型的公共协议240的规范的示例：

字节	示例	描述
1 字节	1	0=视频；1=音频
4 字节	1234567	缓冲器长度（int）
8 字节	424234234	来自传入的音频/视频缓冲器的时间戳 以 tick 进行测量，1tick = 100 ns。（长 int）

1 字节	0	VideoColorFormat
		{
		NV12 = 0,
		Rgb24 = 1,
		Yuy2 = 2,
		H264 = 3
		}

4 字节	720	视频帧像素高度（int）
4 字节	640	视频帧像素宽度（int）
4 字节	25.0	视频帧率 每秒帧数量（浮点）
1 字节	0	音频是否静音？1 = 真； 0 = 假

1 字节	0	AudioFormat
		{
		0 = Pcm16K
		1 = Pcm44KStereo
		}

1 字节	0	检测到的事件（如果有的话） 0 = 无事件 1、2、3 等 = 检测到的指定类型的事件
30 字节		保留以供将来使用

[0117]

[0118]

8 字节	1000000	以字节为单位的二进制数据的长度（长 int）
可变	0x87A879…	这个（多个）帧的原始二进制视频/音频数据
4 字节	1234567	主扬声器端口
4 字节	1234567	有源扬声器

[0119] 在上文中,“检测到的事件(如果有的话)”数据被包括作为公共协议260规范的一部分。然而,在一些实施例中,这个信息(关于检测到的事件)可以替代地被放在单独的存储缓冲器中。

[0120] 在一些实施例中,可以是覆盖物或效果的所述至少一条附加数字视频信息220也作为各自与对应时间戳相关联的单独的帧或帧序列存储在相应单独的缓冲器250中,该对应时间戳又与所述公共时间基准260相关联。

[0121] 如上所例示那样,事件检测步骤可以包括使用所述公共协议240存储描述检测到的事件211的与其中检测到所讨论的事件211的主数字视频流210、301相关联的元数据242。

[0122] 事件检测可以以不同的方式执行。在一些实施例中,由AI部件132a执行的事件检测步骤包括第一经训练的神经网络或其他机器学习部件单独分析所述主数字视频流210、301中的至少一个(诸如几个甚至全部),以便自动检测所述事件211中的任何一个。这可以涉及AI部件132a在受管理的分类中将主视频流210、301数据分类到预定义事件的集合中,和/或在非受管理的分类中分类为动态确定的事件集合中。

[0123] 在一些实施例中,检测到的事件211是作为所讨论的主视频流210、301或者包括在所讨论的主视频流中的演示文稿中的演示幻灯片的改变。

[0124] 例如,如果演示文稿的演示者决定改变他/她当时给予观众的演示文稿中的幻灯片,这意味着给定观众感兴趣的内容可以改变。可能的是,新示出的幻灯片可能只是在所谓的“蝴蝶”模式下能够最好地简洁地看的高级图片(例如,在输出视频流230中与演示者的视频并排显示幻灯片)。替代性地,幻灯片可能包含许多细节、具有较小字体大小的文本等。在这个后一情况下,幻灯片应该替代地以全屏演示,并且可能演示比通常情况下长一些的时间。蝴蝶模式可能不是合适的,因为在这种情况下,对于演示文稿的观看者幻灯片可能比演示者的脸更感兴趣。

[0125] 实际上,事件检测步骤可以包括以下中的至少一个:

[0126] 首先,可以基于对检测到的幻灯片的第一图像和检测到的幻灯片的后续第二图像之间的差异的图像分析来检测事件211。可以使用本身常规的数字图像处理(诸如结合OCR(Optical Character Recognition,光学字符识别)使用运动检测)来自动确定作为示出幻灯片的主视频流220、301的性质。

[0127] 这可能涉及使用自动计算机图像处理技术来检查所检测到的幻灯片是否已经足够大地改变,以实际将其归类为幻灯片改变。这可以通过检查当前幻灯片和前一张幻灯片之间的关于RGB颜色值的差值来完成。例如,人们可以评估在所讨论的幻灯片覆盖的屏幕区域中RGB值全局改变程度,以及是否有可能找到属于一起并且一致地改变的像素群组。这样,可以检测到相关的幻灯片改变,同时例如过滤掉不相关改变,诸如屏幕上示出的计算机鼠标移动。这种方法还允许完全可配置性——例如,有时期望能够捕获计算机鼠标移动,例如当演示者希望使用计算机鼠标指向不同的事物来详细演示某个事物时。

[0128] 其次,可以基于所述第二图像本身的信息复杂性的图像分析来检测事件211,从而以更大的特异性来确定事件的类型。

[0129] 例如,这可能涉及评估所讨论的幻灯片上的文本信息总量以及相关联的字体大小。这可以通过使用常规的OCR方法(诸如基于深度学习的字符识别技术)来完成。

[0130] 注意,由于所评估的视频流210、301的原始二进制格式是已知的,所以这可以直接在二进制域中执行,而无需首先解码或重新编码视频数据。例如,事件检测功能132可以调用用于图像解释服务的相关格式特定的收集功能,或者事件检测功能132本身可以包括用于针对多个不同的所支持的原始二进制视频数据格式评估图像信息(诸如在单个像素级别下)的功能性。

[0131] 在另一示例中,所检测到的事件211是参与者客户端121到数字视频通信服务110的通信连接的丢失。然后,检测步骤可以包括基于对应于所讨论的参与者客户端121的主数字视频流210的一系列后续视频帧213的图像分析来检测所述参与者客户端121已经失去通信连接。

[0132] 因为参与者客户端121与不同的物理位置和不同的互联网连接相关联,所以可能发生的是有人将失去与视频通信服务110或中央服务器130的连接。在这种情况下,期望避免在所生成的输出视频流230中显示黑屏或空白屏幕。

[0133] 相反,可以由事件检测功能132将这种连接丢失检测为事件,诸如通过应用其中所使用的2个类是连接/未连接(无数据)的2类分类算法。在这种情况下,可以理解的是“无数据”不同于演示者故意发出黑屏。因为短暂的黑屏(诸如仅1或2帧的黑屏)在最终产品流230中可能不明显,所以人们可以随时间应用所述2类分类算法来创建时间序列。然后,指定用于连接中断的最小长度的阈值可以用于决定是否存在丢失的连接。

[0134] 如将在下文中解释的那样,可以通过模式检测功能134使用这些例示性类型的检测到的事件来采取适当的和期望的各种行动。

[0135] 如上所提及那样,各个主视频流210、301各自与公共时间基准260相关,这使得同步功能133可以对它们彼此时间同步。

[0136] 在一些实施例中,公共时间基准260基于或包括公共音频信号111(参见图1至图3),公共音频信号111对于涉及如上所述的各自提供所述主数字视频流210中的相应一个的至少两个远程连接的参与者客户端121的共享数字视频通信服务110是公共的。

[0137] 在上面讨论的Microsoft® Teams®的示例中,公共音频信号经由自动参与者客户端140和/或经由API 112生成并可以由中央服务器130捕获。在这个示例和其他示例中,这种公共音频信号可以用作心跳信号,以通过基于这个心跳信号将各个主视频流220绑定到特定时间点来对各个主视频流进行时间同步。这种公共音频信号可以作为单独的(相对于其他主视频流210中的每一个)信号来提供,由此其他主视频流210可以各自基于所讨论的其他主视频流210中包含的音频或者甚至基于其中包含的图像信息(诸如使用基于自动图像处理的唇音同步技术)而单独地与公共音频信号时间相关。

[0138] 换句话说,为了处理与各个主视频流210相关联的任何可变和/或不同的延迟,并且为了实现针对组合式视频输出流230的时间同步,这种公共音频信号被用作用于中央服务器130中的全部主视频流210(但可能不是外部主视频流301)的心跳。换句话说,全部其他信号映射到这个公共音频时间心跳,以确保一切处于时间同步。

[0139] 在不同的示例中,使用被引入到输出数字视频流230中并通过被提供为参与者客户端121中的一个或几个单独参与者客户端的一部分的相应本地时间同步软件功能125检测的时间同步元素231来实现时间同步,本地软件功能125被布置为检测输出视频流230中的时间同步元素231的到达时间。如所理解的那样,在这样的实施例中,输出视频流230被反馈到视频通信服务110中,或者以其他方式使其对每个参与者客户端121和所讨论的本地软件功能125可用。

[0140] 例如,时间同步元件231可以是视觉标记,诸如以预定顺序或方式改变颜色的、以规则的时间间隔放置或更新在输出视频230中的像素、在输出视频230中更新和显示的可视时钟、被添加到形成输出视频流230的一部分的音频中的声音信号(其可以通过例如具有足够低的幅值和/或足够高的频率而被设计为对于参与者122是听不见的)。本地软件功能125被布置为使用合适的图像和/或音频处理来自动检测(多个)时间同步元件231中的每一个的相应到达时间(或检测(多个)时间同步元件中的每一个)。

[0141] 然后,可以至少部分地基于所述检测到的到达时间来确定公共时间基准260。例如,本地软件功能125中的每一个可以向中央服务器130通信传送表示所述检测到的到达时间的相应信息。

[0142] 这种通信可以经由所讨论的参与者客户端121和中央服务器130之间的直接通信链路来进行。然而,通信也可以经由与所讨论的参与者客户端121相关联的主视频流210进行。例如,参与者客户端121可以在由所讨论的参与者客户端121生成的主视频流210中引入诸如以上讨论的类型的可视或可听代码,用于由中央服务器130进行自动检测并用于确定公共时间基准260。

[0143] 在再附加的示例中,每个参与者客户端121可以在可用于由视频通信服务110的全部参与者客户端121观看的公共视频流中执行图像检测,并且以对应于以上讨论的方式的方式将这种图像检测的结果中继到中央服务器130,以在那里用于随时间确定每个参与者客户端121相对于彼此的相应偏移。这样,公共时间基准260可以被确定为各个相对偏移的集合。例如,通常可用的视频流的所选择的参考像素可以由几个或全部参与者客户端121监控,诸如由所述本地软件功能125监控,并且这个像素的当前颜色可以被通信传送到中央服务器130。中央服务器130可以基于从多个(或全部)参与者客户端121中的每一个连续接收的这种颜色值来计算相应时间序列,并执行交叉相关,从而导致不同参与者客户端121之间的相对时间偏移的估计集合。

[0144] 实际上,被馈送到视频通信服务110中的输出视频流230可以被包括作为所讨论的视频通信的每个参与者客户端的共享屏幕的一部分,并且因此可以被用于评估与参与者客户端121相关联的这种时间偏移。特别地,馈送到视频通信服务110的输出视频流230可以经由自动参与者客户端140和/或API 112再次可用于中央服务器。

[0145] 在一些实施例中,公共时间基准260可以至少部分地基于所述主数字视频流210、301中的第一主数字视频流的音频部分214和所述第一主数字视频流210、301的图像部分215之间的检测到的差异来确定。例如,这种差异可以基于在所讨论的所述第一主数字视频流210、301中观看的讲话参与者122的数字唇音同步视频图像分析。这种唇音同步分析同样是常规的,并且可以例如使用经过训练的神经网络。可以通过同步功能133针对与可用公共音频信息相关的每个主视频流210、301执行分析,并且可以基于这个信息确定各个主视频

流210、301之间的相对偏移。

[0146] 在一些实施例中,同步步骤包括有意地引入至多30秒的延迟(诸如至多5秒,诸如至多1秒,诸如至多0.5秒,但长于0秒的延迟),使得输出数字视频流230至少以所述延迟提供。无论如何,有意引入的延迟是至少几个视频帧,诸如至少三个,或者甚至至少五个或者甚至10个视频帧,诸如在收集步骤中的任何重新采样之后存储的这个数量的帧(或者单个图像)。如本文所使用那样,术语"有意地"无论是否需要基于同步问题等引入该延迟,均引入该延迟。换句话说,除了作为主视频流210、301的同步的一部分而引入的任何延迟之外,还引入了有意引入的延迟,以便使主视频流相对于彼此时间同步。相对于公共时间基准260,有意引入的延迟可以是预定的、固定的或可变的。延迟时间可以相对于主视频流210、301中最少延迟的一个来测量,使得作为所述时间同步的结果,这些流210、301中的更多延迟的流与相对较小的有意增加的延迟相关联。

[0147] 在一些实施例中,引入了相对较小的延迟,诸如0.5秒或更小的延迟。使用输出视频流230的视频通信服务110的参与者将几乎察觉不到这种延迟。在其他实施例中,诸如当上输出视频流230将不在交互式场景中使用而是替代地以单向通信发布给外部消费者150时,可能引入更大的延迟。

[0148] 这种有意引入的延迟实现了用于同步功能133将所收集的各个主流210、301视频帧映射到正确的公共时间基准260时间戳261上的足够的时间。它还允许足够的时间来执行上述事件检测,以便检测丢失的主流210、301信号、幻灯片改变、分辨率改变等。另外,有意引入所述延迟允许改进的模式检测功能134,如将在下文中描述的那样。

[0149] 认识到的是所述延迟的引入涉及在使用所讨论的所缓冲的帧213发布输出视频流230之前缓冲250所收集的和经时间同步的主视频流210、301中的每一个。换句话说,主视频流210、301中的至少一个、几个或甚至全部的视频和/或音频数据将以缓冲的方式存在于中央服务器130中,很像以能够处理变化的带宽情况为目的但出于上述原因使用(特别地要由模式检测功能134使用)的缓存但不是(像常规缓存缓冲器)。

[0150] 因此,在一些实施例中,所述模式检测步骤包括考虑主数字视频流210、301中的至少一个(诸如几个或甚至全部)的特定信息,该信息存在于还将用于生成输出数字视频流230的经时间同步的主数字视频流210的帧之后的帧213中。因此,在形成输出视频流230的一部分(或用于成输出视频流的基础)之前的特定延迟期间,新添加的帧213将存在于所讨论的缓冲器250中。在这个时间段期间,所讨论的帧213中的信息将构成与当前使用的帧相关的"未来"的信息,以生成输出视频流230的当前帧。一旦输出视频流230时间线到达所讨论的帧213,它将被用于生成输出视频流230的对应帧并且此后可以被丢弃。

[0151] 换句话说,模式检测功能134能够具有仍未用于生成输出视频流230的视频/音频帧213的集合,并使用这个数据来检测所述模式。

[0152] 模式检测可以以不同的方式执行。在一些实施例中,由AI部件134a执行的模式检测步骤包括第二经训练的神经网络或其他机器学习部件一起分析所述主数字视频流120、301中的至少两个(诸如至少三个或甚至全部)以自动检测所述模式212。

[0153] 在一些实施例中,检测到的模式212包括涉及共享视频通信服务110的至少两个不同说话参与者122(各自与相应参与者客户端121相关联)的说话模式,所述说话参与者122中的每一个在所述主数字视频流210、301中的相应一个中被可视地观看。

[0154] 生成步骤优选地包括确定、跟踪和更新输出视频流230的当前生成状态。例如,这样的状态可以指示在输出视频流230中什么(如果有的话)参与者122是可见的、以及在屏幕上的什么位置;任何外部视频流300在输出视频流230中是否可见,以及在屏幕上的什么位置;任何幻灯片或共享屏幕以全屏模式示出还是以与任何实时视频流结合示出等等。因此,生成功能135可以被视为关于所生成的输出视频流230的状态机。

[0155] 为了生成输出视频流230作为将由例如终端消费者150观看的组合式视频体验,有利的是中央服务器130能够理解比仅仅检测到与各个主视频流210、301相关联的各个事件更深层次上发生的事情。

[0156] 在第一示例中,演示参与者客户端121正在改变当前观看的幻灯片。如上所述,这个幻灯片改变通过事件检测功能132检测,并且元数据242被添加到所讨论的帧,从而指示已经发生了幻灯片改变。这发生多次,因为演示参与客户机121结果是快速连续地向前跳过多个幻灯片,从而导致由事件检测功能132检测到并与对应元数据242一起存储在所讨论的主视频流210的单独缓冲器250中的一系列“幻灯片改变”事件。实际上,每个这样快速向前跳过的幻灯片可能仅在几分之一秒内可见。

[0157] 查看所讨论的缓冲器250中的信息、跨越这些检测到的幻灯片改变中的几个的模式检测功能134将检测对应于单个幻灯片改变(即,对应于向前跳过中的最后幻灯片,一旦快速跳过完成,幻灯片保持可见)的模式,而不是多个或快速执行的幻灯片改变。换句话说,模式检测功能134将注意到,例如,在非常短的时间段内存在十个幻灯片改变,为什么它们将被处理为表示一个单个幻灯片改变的所检测到的模式。结果,访问由模式检测功能134检测到的模式的生成功能135可以选择以全屏模式在输出视频流230中示出最后的幻灯片持续几秒钟,因为它确定这个幻灯片在所述状态机中潜在地是重要的。它也可以选择在输出流230中根本不示出中间观看的幻灯片。

[0158] 具有几个快速改变的幻灯片的模式的检测可以通过简单的基于规则的算法来检测,但是可以替代性地使用被设计和训练为通过分类来检测移动图像中的这种模式的经训练的神经网络来检测。

[0159] 在不同的示例中,这例如在视频通信是脱口秀、小组辩论等的情况下可能是有用的,可能期望在一方面当前说话者之间快速切换视觉注意力,而另一方面仍然通过生成和发布平静且平滑的输出视频流230来给予消费者150相关的观看体验。在这种情况下,事件检测功能132可以连续地分析每个主视频流210、301,以一直确定在这个特定主视频流210、301中被观看的人当前是否正在说话。例如,这可以如上所述使用本身长谷的图像处理工具来执行。然后,模式检测功能134可以可操作以检测涉及所述主视频流210、301中的几个的特定整体模式,所述模式对于生成平滑的输出视频流230是有用的。例如,模式检测功能134可以检测当前说话者之间的非常频繁切换的模式和/或涉及几个并发说话者的模式。

[0160] 然后,生成功能135可以在采取与所述生成状态相关的自动决策时考虑这种检测到的模式,例如通过在再次沉默之前不自动将视觉焦点切换到仅说话持续半秒钟的说话者,或者切换到其中在当两个说话者交替或并发地说话时的某个时间段期间并排显示几个说话者的状态。这个状态判定过程本身可以使用时间序列模式识别技术或者使用训练过的神经网络来执行,但是也可以至少部分地基于预定的规则集合。

[0161] 在一些实施例中,可以存在并行检测到的多个模式,并形成对生成功能135状态机

的输入。这样的多个模式可以由生成功能135通过不同的AI部件、计算机视觉检测算法等使用。作为示例,可以在并发地检测一些参与者客户端121的不稳定连接的同时检测永久幻灯片改变,而其他模式检测当前主要说话参与者122。使用全部这种可用的模式数据,可以训练分类器神经网络,和/或可以开发规则集合,用于分析这种模式数据的时间序列。这种分类可以至少部分地(诸如完全地)被监督,以导致要在所述生成中使用的所确定的期望状态变化。例如,可以生成不同的这种预定分类器,特别地被布置为根据各种不同的生产风格和期望自动生成输出视频流230。训练可以基于作为期望输出的已知生产状态改变序列和作为训练数据的已知模式时间序列数据。在一些实施例中,贝叶斯模型可以用于生成这样的分类器。在具体示例中,可以从有经验的制作人先验地收集信息,从而提供诸如“在脱口秀中,我从不直接从讲话者A切换到讲话者B,而是总是在我聚焦于另一讲话者之前首先示出概述,除非另一讲话者非常占优势并且说话声音很大”的输入。然后,这个生成逻辑被表示为呈一般形式“如果X为真|假定Y为真的事实|执行Z”的贝叶斯模型。实际检测(某人是否大声说话等的实际检测)可以使用分类器或基于阈值的规则来执行。

[0162] 通过较大数据集(模式时间序列数据的较大数据集),人们可以使用深度学习方法来开发正确且有吸引力的生成格式,以便在视频流的自动化生成中使用。

[0163] 总之,使用基于各个主视频流210、301的事件检测的组合、有意引入的延迟、基于几个经时间同步的主视频流210、301和检测到的事件的模式检测、以及基于检测到的模式的生成过程使得可以根据多种可能的品味和风格选择来实现输出数字视频流230的自动生成。这个结果在由事件检测功能132、模式检测功能134和生成功能135所使用的各种可能的神经网络和/或基于规则的分析技术中都是有效的。

[0164] 如上所例示那样,生成步骤可以包括基于关于所述输出数字视频流230中的所述主数字视频流210、301中的各个视频流的可见性的预定和/或动态可变参数的集合、视觉和/或听觉视频内容布置、所使用的视觉或听觉效果、和/或输出数字视频流230的输出模式来生成输出数字视频流230。这些参数可以由所述生成功能135状态机自动确定和/或由控制生成的操作者设置(使其半自动)和/或基于某些先验配置期望(诸如输出视频流230布局=改变或以上例示类型的状态改变之间的最短时间)预先确定。

[0165] 在实际示例中,状态机可以支持可以应用于输出视频流230的预定标准布局的集合,诸如全屏演示者视图(全屏示出当前说话的参与者122)、幻灯片视图(全屏示出当前共享的演示幻灯片)、“蝴蝶视图”(以并排视图示出当前说话的参与者122以及当前共享的演示幻灯片)、多说话者视图(并排或以矩阵布局示出全部参与者122或参与者的所选择的子集)等。各种可用生成格式可以由状态机状态改变规则的集合(如上所述)和可用状态集合(诸如所述标准布局集合)来限定。例如,一个生成格式可以是“小组讨论”,另一生成格式可以是“演示”等。通过经由到中央服务器130的GUI或其他接口选择特定生成格式,系统100的操作者可以快速选择预定义的这种生成格式集合中的一个生成格式,并且然后允许中央服务器130基于如上所述的可用信息根据所讨论的生成格式完全自动地生成输出视频流230。

[0166] 另外,在生成期间,如上所述,针对每个会面参与者客户端121或外部视频源300创建并保持相应存储器内缓冲器。这些缓冲器可以很容易地现场移除、添加和改变。然后,中央服务器130可以被布置为在输出视频流230的生成期间接收关于所添加/减少的参与者客户端121和被安排用于递送语音的参与者122、演示的所计划的或意外的暂停/恢复、对当前

使用的生产格式的期望改变等的信息。如上所述,这种信息例如可以经由操作者GUI或界面被馈送到中央服务器130。

[0167] 如上所例示那样,在一些实施例中,主数字视频流210、301中的至少一个被提供给数字视频通信服务110,并且然后发布步骤可以包括将所述输出数字视频流230提供给相同的通信服务110。例如,输出视频流230可以被提供给视频通信服务110的参与者客户端121,或者经由API 112作为外部视频流被提供给视频通信服务110。这样,使得输出视频流230可以可用于当前由视频通信服务110实现的视频通信事件的参与者中的几个或全部。

[0168] 同样如上所讨论那样,附加地或替代性地,输出视频流230可以被提供给一个或几个外部消费者150。

[0169] 一般而言,生成步骤可以由中央服务器130执行,从而经由API 137将所述输出数字视频流230作为现场视频流提供给一个或几个并发消费者。

[0170] 本发明还涉及一种用于根据以上已经描述的内容提供共享数字视频流的计算机软件功能。这样的计算机软件功能然后被布置为在运行时执行上述收集步骤、事件检测步骤、同步步骤、模式检测步骤、生成步骤和发布步骤。计算机软件功能被布置为在中央服务器130的物理或虚拟硬件上运行,如上所述。

[0171] 本发明还涉及系统100,其是用于提供共享数字视频流的系统并且又包括中央服务器130。中央服务器103又被布置为执行所述收集步骤、事件检测步骤、同步步骤、模式检测步骤、生成步骤和发布步骤。例如,这些步骤由运行所述计算机软件功能的中央服务器130执行,以执行如上所述的所述步骤。

[0172] 在上文中,已经描述了优选实施例。然而,对于本领域技术人员来说显而易见的是,在不脱离本发明的基本思想的情况下,可以对所公开的实施例进行许多修改。

[0173] 例如,许多附加功能可以作为本文中描述的系统100的一部分来提供,并且这些附加功能在本文中没有描述。总的来说,目前描述的解决方案提供了一种框架,在该框架之上可以构建详细的功能性和特征,以满足其中视频数据流用于通信的各种不同的具体应用。

[0174] 一个示例是展示情形,其中主视频流包括演示者的视图、共享的基于数字幻灯片的演示以及正在展示的产品的现场视频。

[0175] 另一示例是教学情形,其中主视频流包括教师的视图、作为教学的主题的物理实体的现场视频、以及可能提出问题并与教师进行对话的几个学生的各自相应视频。

[0176] 在这两个示例的任一个中,视频通信服务(其可以是或不是系统的一部分)可以提供主视频流中的一个或几个,和/或几个主视频流可以作为本文中讨论的类型的外部视频源来提供。

[0177] 一般而言,与本方法相关所述的全部内容适用于本系统和计算机软件产品,反之亦然。

[0178] 因此,本发明不限于所描述的实施例,而是可以在所附权利要求的范围内变化。

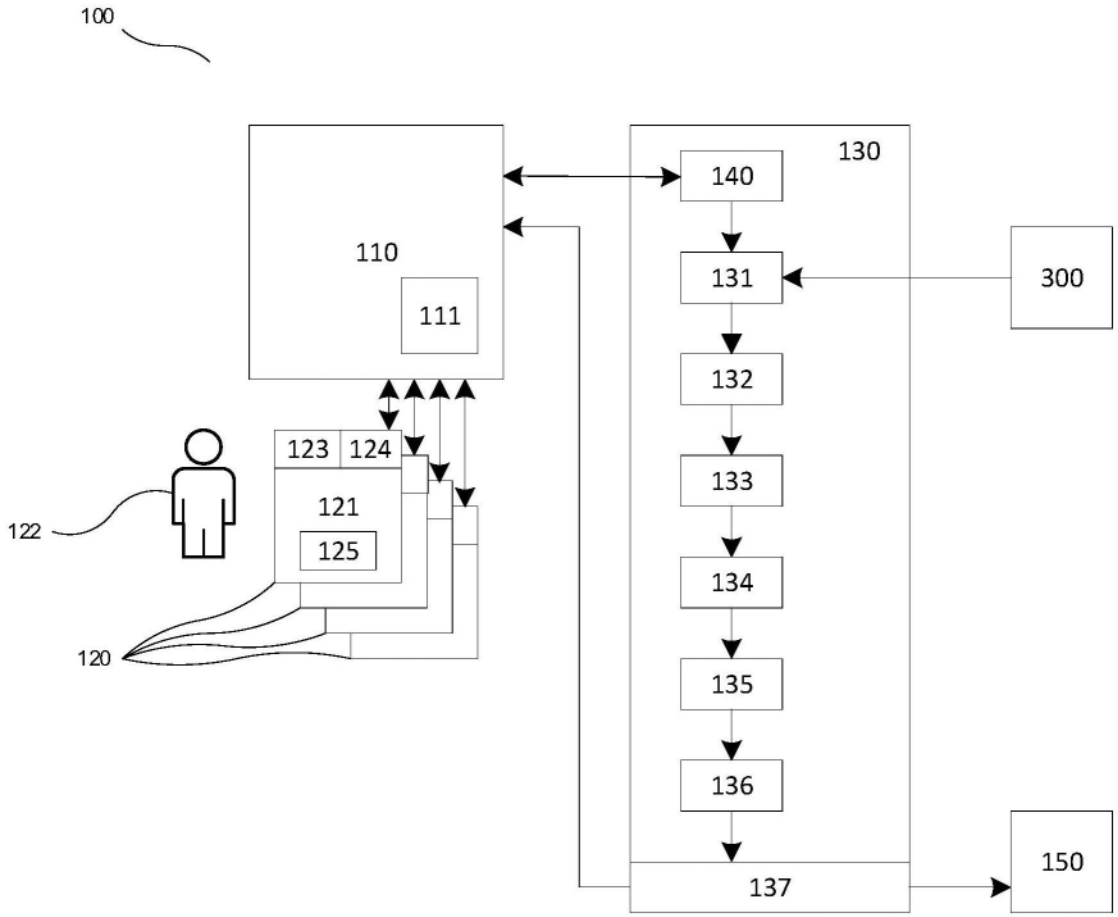

100
110
111
130
140
131
300
132
133
134
135
136
137
150
123
124
121
125
122
120

图1

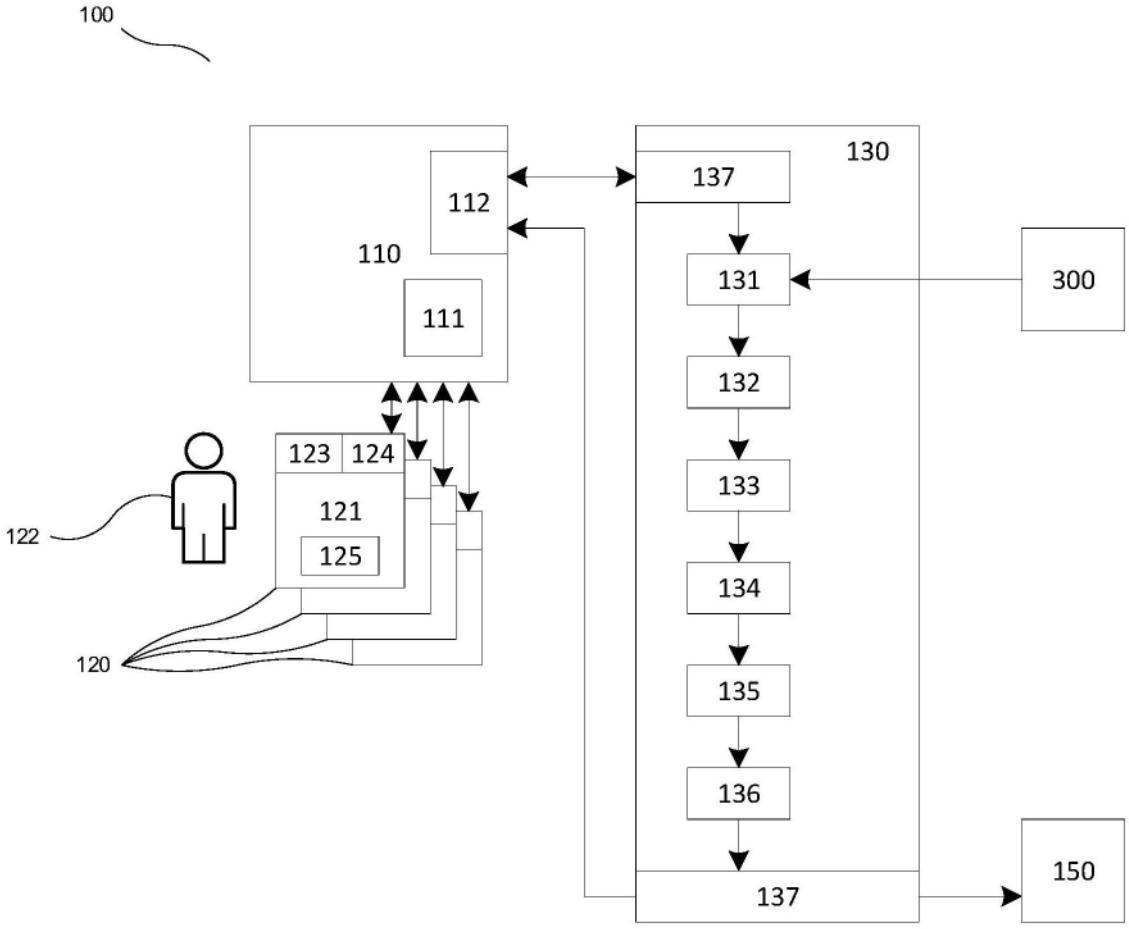

100
112
110
111
130
137
131
300
132
133
134
135
136
137
150
123
124
121
125
122
120

图2

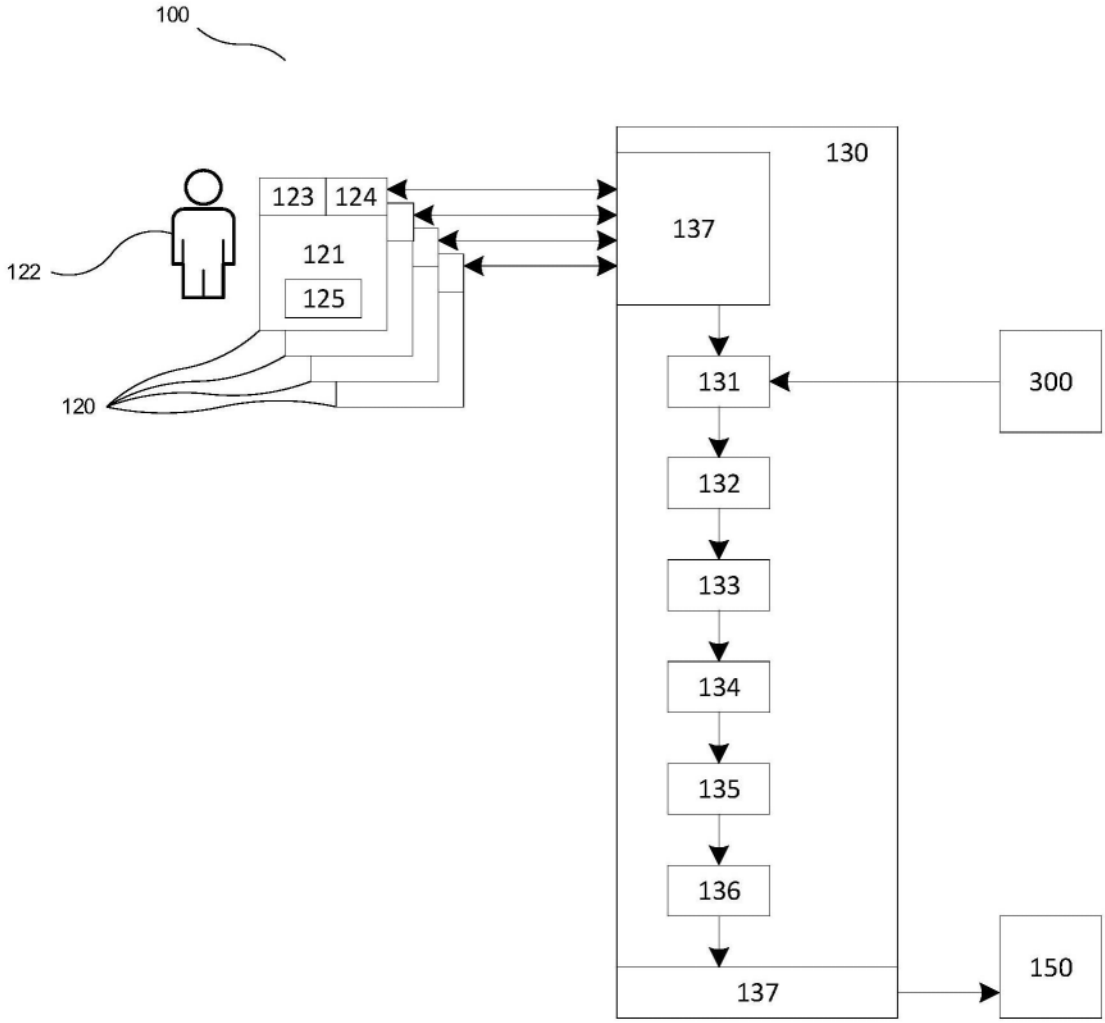

100
130
137
123
124
121
125
122
120
131
300
132
133
134
135
136
137
150

图3

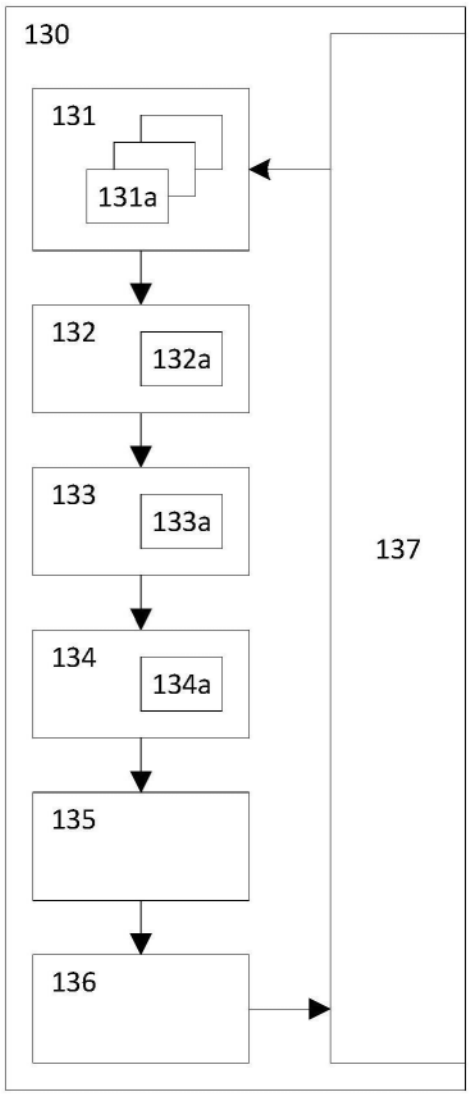


图4

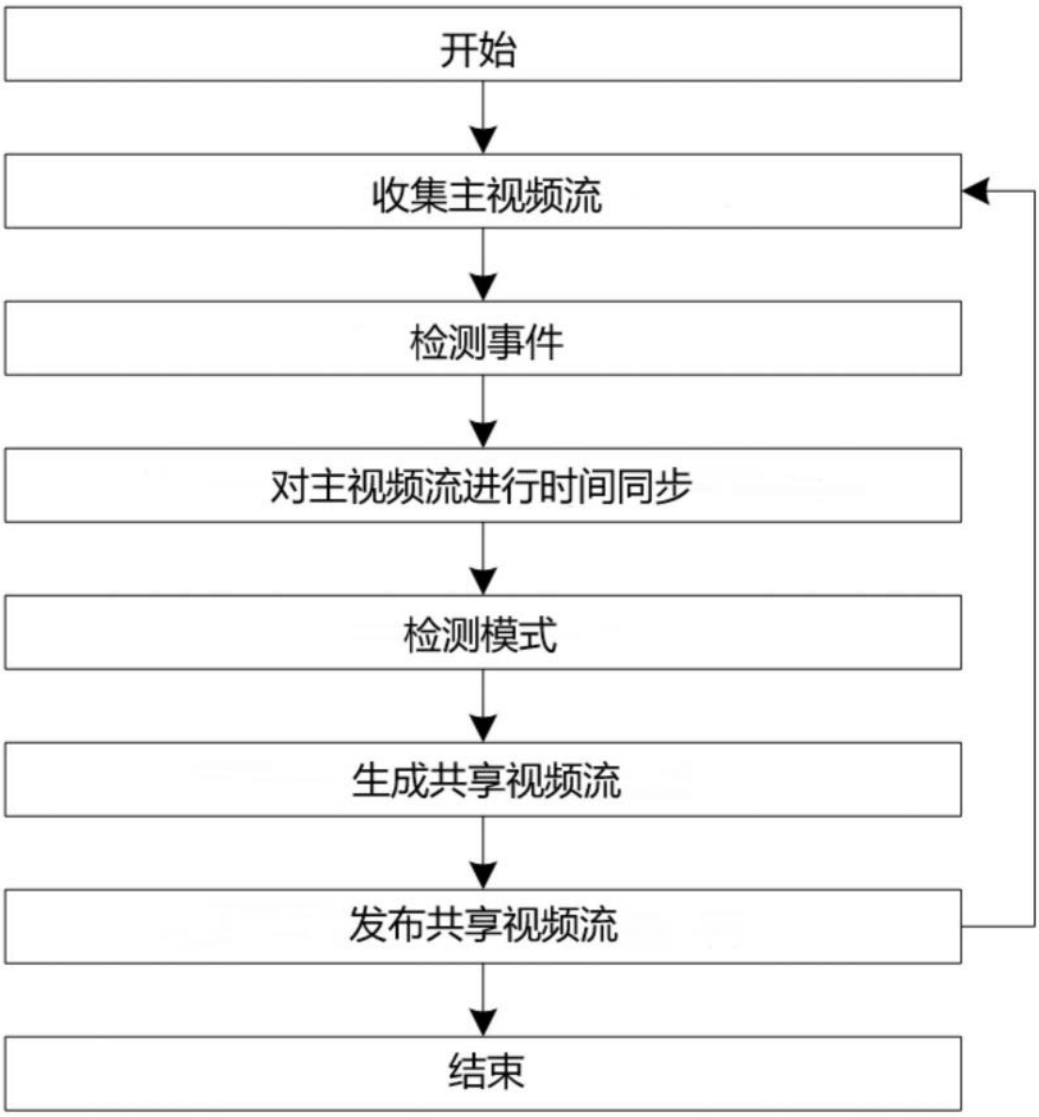

开始
收集主视频流
检测事件
对主视频流进行时间同步
检测模式
生成共享视频流
发布共享视频流
结束

图5

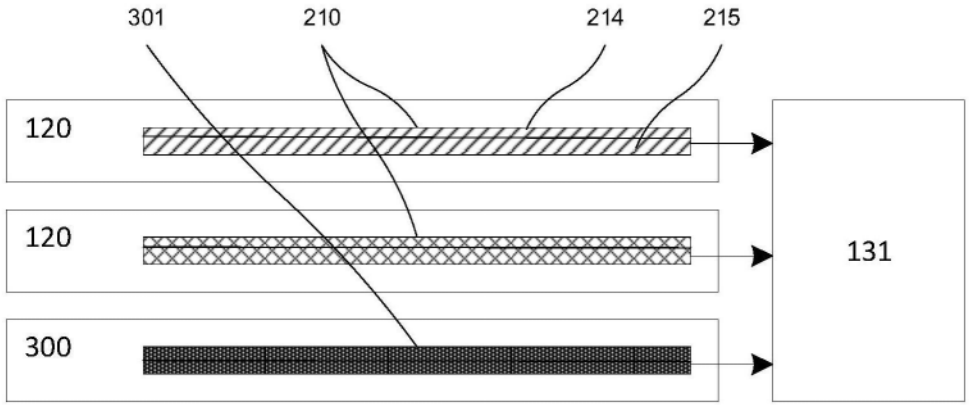

301
210
214
215
120
120
300
131

图6a

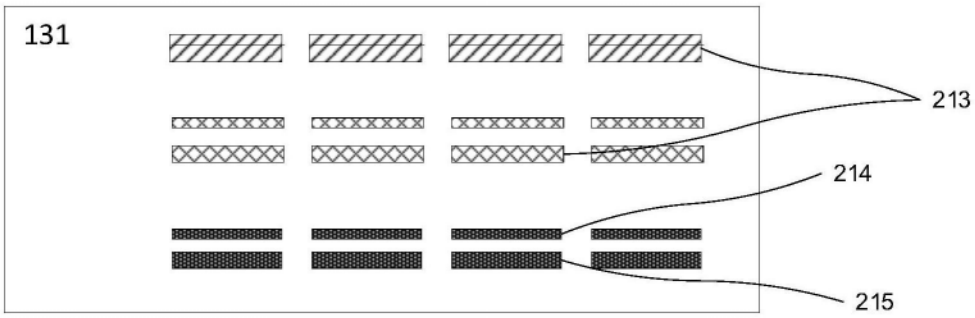


图6b

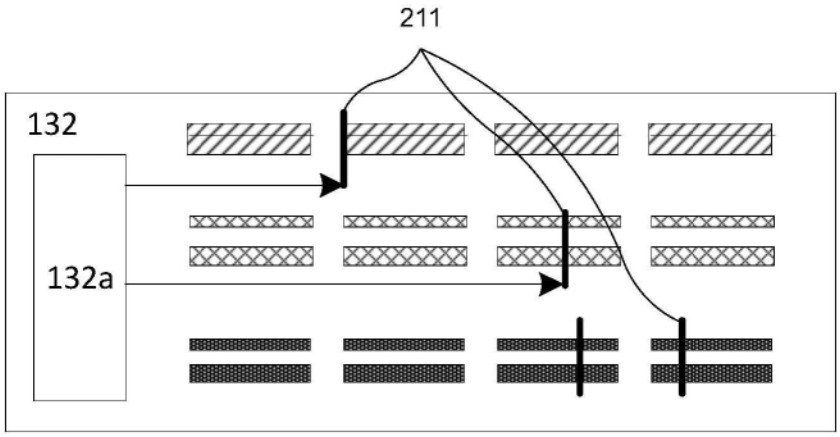


图6c

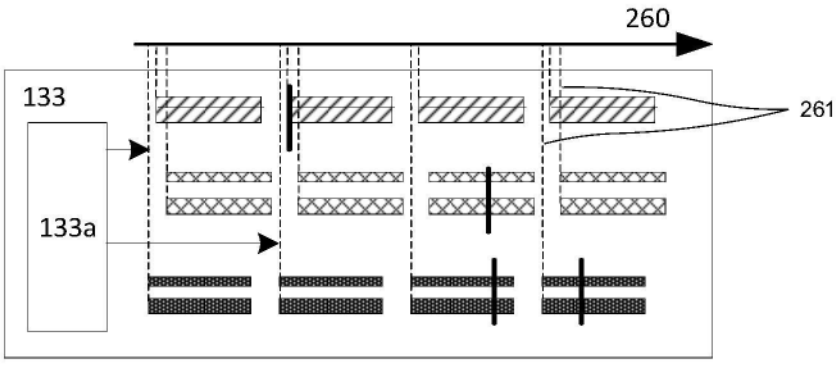


图6d

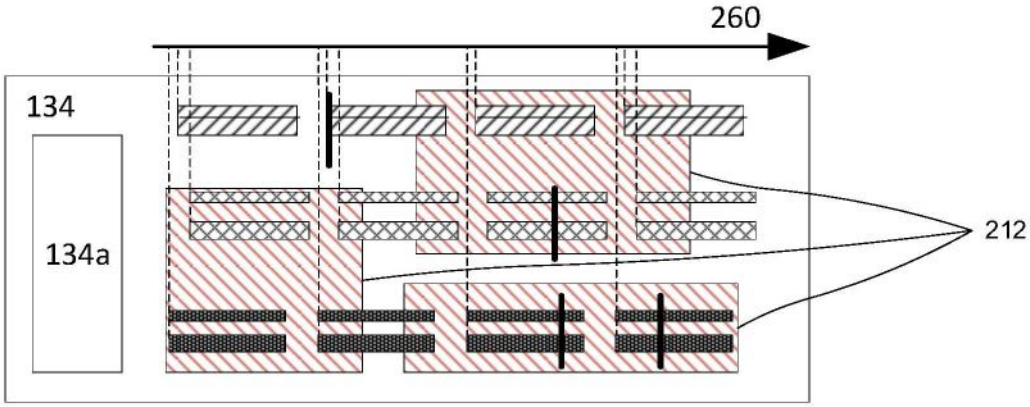


图6e

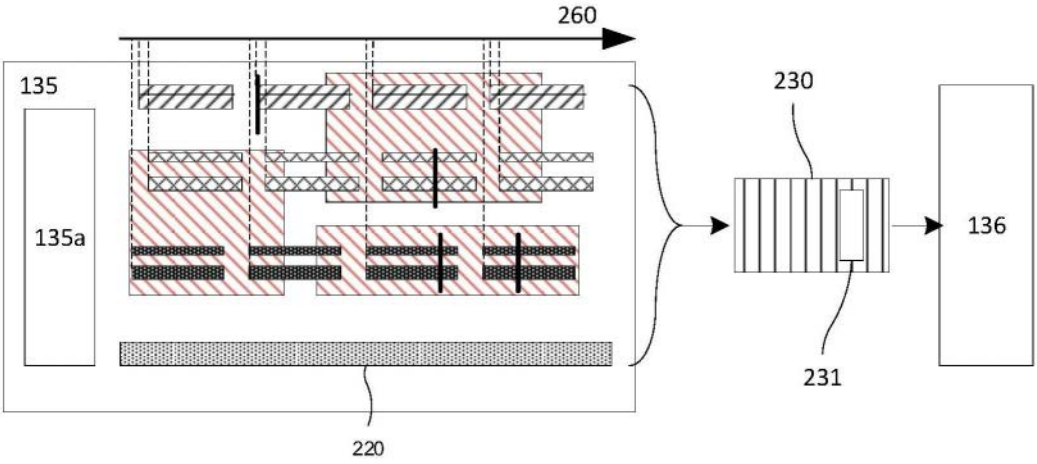


图6f

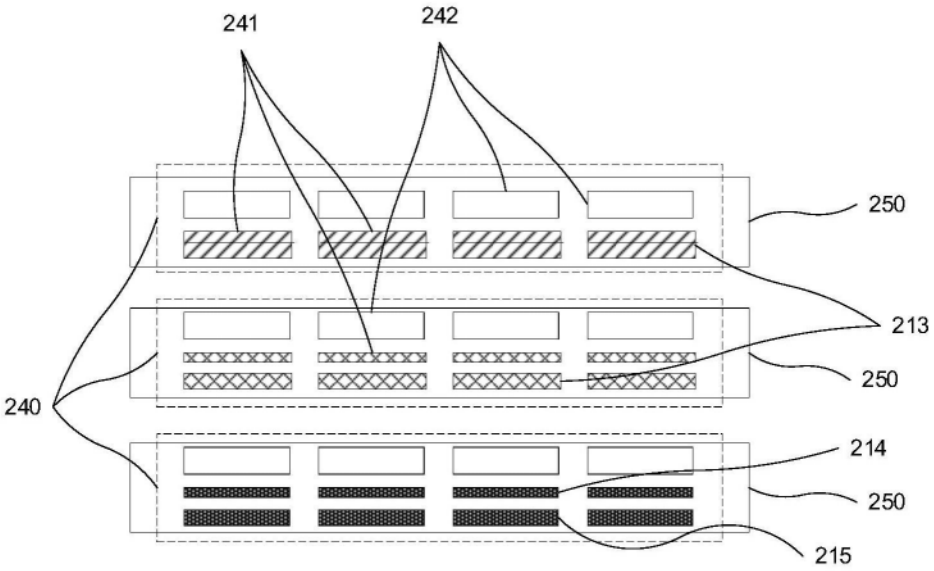


图7