(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118743225 A
(43)申请公布日 2024.10.01

(21)申请号 202380020400.6
(22)申请日 2023.02.03
(30)优先权数据
2250113-4 2022.02.04 SE
(85)PCT国际申请进入国家阶段日
2024.08.05
(86)PCT国际申请的申请数据
PCT/SE2023/050090 2023.02.03
(87)PCT国际申请的公布数据
WO2023/149835 EN 2023.08.10
(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 安德烈亚斯·比约克曼
安德斯·尼尔森 拉尔斯·埃尔曼
佩尔·布兰肯斯
麦克斯·索尔奎斯特
马库斯·拉格纳森
(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 葛强
(51)Int.Cl.
H04N 21/234(2006.01)
H04L 12/18(2006.01)
H04N 7/15(2006.01)
H04N 21/2187(2006.01)

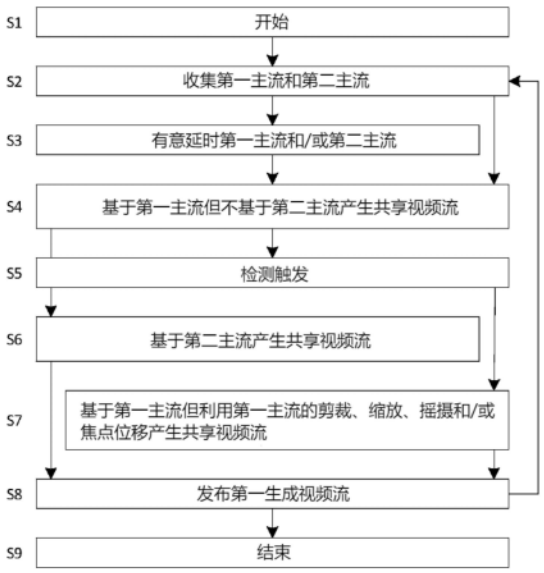
权利要求书3页 说明书32页 附图12页

(54)发明名称

用于产生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:在收集步骤(S2)中,收集第一数字视频流和第二数字视频流;在第一产生步骤(S4)中,基于所述第一流产生共享流,使得所述第一源在共享流中是可见的,但是使得所述第二源在共享流中是不可见的;在触发检测步骤(S5)中,基于预定图像和/或音频模式的检测,分析所述第一流和/或第二流以检测触发,触发是检测到的凝视或手势;在第二产生步骤(S6,S7)中,基于所述第二流和/或基于所述第一源但是以不同裁剪、不同缩放、不同摇摄或不同聚焦平面选择中的至少一个产生共享流;以及在发布步骤(S8)中,向消费者提供所述输出流。本发明还涉及一种计算机软件产品和一种系统。



CN 118743225 A

1.一种用于提供共享数字视频流的方法,所述方法包括:

在收集步骤(S2)中,从第一数字视频源收集第一数字视频流,并且从第二数字视频源收集第二数字流;

在第一产生步骤(S4)中,基于所述第一数字视频流的连续考虑的帧来产生所述共享数字视频流作为输出数字视频流,使得来自所述第一数字视频源的图像信息在所述共享数字视频流中是可见的,但是使得来自所述第二数字视频源的图像信息在所述共享数字视频流中是不可见的;

在触发检测步骤(S5)中,通过执行以下中的至少一个来自动检测触发:

a)数字分析所述第一数字视频流,以自动检测呈以下形式的触发:所述第一数字视频流中描绘的参与者(430)凝视朝向所述第二数字视频流中描绘的对象(440)或相对于所述对象做出预定手势,

所述检测基于关于第一相机(410)、所述参与者(430)和所述对象(440)的相对取向的信息,所述检测还基于对所讨论的所述参与者(430)的身体方向、头部方向、凝视方向和手势方向中的至少一个的基于数字图像的确定;以及

b)数字分析所述第一数字视频流和/或第二数字视频流,以自动检测呈现为所述第二数字视频流中描绘的所述参与者(430)的注视方向从不注视第二相机(420)到注视所述第二相机(420)的转变的形式的所述触发,

根据预定产生规则,所述触发是指示性的以改变所述共享数字视频流的产生模式;

在响应于检测到所述触发而启动的第二产生步骤(S6,S7)中,基于所述第二数字视频流的连续考虑的帧来产生所述共享数字视频流作为输出数字视频流,使得来自所述第二数字视频源的图像信息在所述共享数字视频流中是可见的,和/或基于所述第一数字视频源的连续考虑的帧但是与所述第一产生步骤(S4)相比以所述第一数字视频流的不同裁剪、不同缩放、不同摇摄或不同聚焦平面选择中的至少一个产生所述共享数字视频流作为输出数字视频流;以及

在发布步骤(S8)中,将所述输出数字视频流连续提供给所述共享数字视频流的消费者。

2.根据权利要求1所述的方法,其中,

所述第一数字视频流由第一数字相机(410)连续捕获,并且其中,

所述第二数字视频流由第二数字相机(420)捕获。

3.根据权利要求1或2所述的方法,其中,

所述第二对象(440)是人、人的群组、机器、白板或演示屏幕。

4.根据前述权利要求中任一项所述的方法,其中,

所述手势是指向所述对象(440)。

5.根据前述权利要求中任一项所述的方法,其中,

所述触发的检测包括检测预定图像模式,所述预定图像模式又包括所述第一数字视频流和/或第二数字视频流中的相对移动变化。

6.根据前述权利要求中任一项所述的方法,还包括

使用基于所述第一数字视频流和/或所述第二数字视频流的数字图像处理来确定所述参与者(430)和所述对象(440)相对于所述第一相机(410)和/或所述第二相机(420)的位

置。

7.根据前述权利要求中任一项所述的方法,其中,

所述第二产生步骤(S6,S7)包括确定当前示出所述对象(440)的几个可能的第二相机当中的特定一个第二相机(420),并且选择所述第二相机(420)作为提供所讨论的所述第二视频流的相机。

8.根据前述权利要求中任一项所述的方法,其中,

所述第二产生步骤(S6,S7)中的所述产生被配置为突出显示或移位所述输出数字视频流的焦点,以便向所述对象(440)提供增加的视觉焦点。

9.根据前述权利要求中任一项所述的方法,其中,所述方法还包括:

延时步骤(A3),其中至少对于所述第一数字视频流和所述第二数字视频流有意引入延迟,所述延迟存在于所述共享数字视频流中,其中,

在引入所述延迟之前,基于所述第一数字视频流和/或第二数字视频流来执行所述触发检测步骤。

10.根据权利要求9所述的方法,其中,

所述延迟至多30秒,诸如至多5秒、诸如至多1秒、诸如至多0.5秒。

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

收集步骤(S2),其中从第一数字视频源收集第一数字视频流,并且从第二数字视频源收集第二数字流;

第一产生步骤(S4),其中基于所述第一数字视频流的连续考虑的帧来产生所述共享数字视频流作为输出数字视频流,使得来自所述第一数字视频源的图像信息在所述共享数字视频流中是可见的,但是使得来自所述第二数字视频源的图像信息在所述共享数字视频流中是不可见的;

触发检测步骤(S5),其中通过以下中的至少一个来自动检测触发:

a)数字分析所述第一数字视频流,以自动检测呈以下形式的触发:所述第一数字视频流中描绘的参与者(430)凝视朝向所述第二数字视频流中描绘的对象(440)或相对于所述对象做出预定手势,

所述检测基于关于第一相机(410)、所述参与者(430)和所述对象(440)的相对取向的信息,所述检测还基于对所讨论的所述参与者(430)的身体方向、头部方向、凝视方向和手势方向中的至少一个的基于数字图像的确定;以及

b)数字分析所述第一数字视频流和/或第二数字视频流,以自动检测呈现为所述第二数字视频流中描绘的所述参与者(430)的注视方向从不注视第二相机(420)到注视所述第二相机(420)的转变的形式的所述触发,

根据预定产生规则,所述触发是指示性的以改变所述共享数字视频流的产生模式;

响应于检测到所述触发而启动的第二产生步骤(S6,S7),其中基于所述第二数字视频流的连续考虑的帧产生所述共享数字视频流作为输出数字视频流,使得来自所述第二数字视频源的图像信息在所述共享数字视频流中是可见的,和/或基于所述第一数字视频源的连续考虑的帧但是与所述第一产生步骤(S4)相比以所述第一数字视频流的不同裁剪、不同缩放、不同摇摄或不同聚焦平面选择中的至少一个产生所述共享数字视频流作为输出数字

视频流;以及

发布步骤(S8),其中将所述输出数字视频流连续提供给所述共享数字视频流的消费者。

12.一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,所述收集功能被布置为从第一数字视频源收集第一数字视频流,以及从第二数字视频源收集第二数字流;

第一产生功能(135,135',135",135"'),所述第一产生功能被布置为基于所述第一数字视频流的连续考虑的帧来产生所述共享数字视频流作为输出数字视频流,使得来自所述第一数字视频源的图像信息在所述共享数字视频流中是可见的,但是使得来自所述第二数字视频源的图像信息在所述共享数字视频流中是不可见的;

触发检测功能,所述触发检测功能被布置为通过执行以下中的至少一个来自动检测触发:

a)数字分析所述第一数字视频流,以自动检测呈以下形式的触发:所述第一数字视频流中描绘的参与者(430)凝视朝向所述第二数字视频流中描绘的对象(440)或相对于所述对象做出预定手势,

所述检测基于关于第一相机(410)、所述参与者(430)和所述对象(440)的相对取向的信息,所述检测还基于对所讨论的所述参与者(430)的身体方向、头部方向、凝视方向和手势方向中的至少一个的基于数字图像的确定;以及

b)数字分析所述第一数字视频流和/或第二数字视频流,以自动检测呈现为所述第二数字视频流中描绘的所述参与者(430)的注视方向从不注视第二相机(420)到注视所述第二相机(420)的转变的形式的所述触发,

根据预定产生规则,所述触发是指示性的以改变所述共享数字视频流的所述产生模式;

第二产生功能(135,135',135",135"'),所述第二产生功能被布置为响应于检测到所述触发而启动,并且基于所述第二数字视频流的连续考虑的帧产生所述共享数字视频流作为输出数字视频流,使得来自所述第二数字视频源的图像信息在所述共享数字视频流中是可见的,和/或基于所述第一数字视频源的连续考虑的帧但是与所述第一产生功能(135,135',135",135"')相比以所述第一数字视频流的不同裁剪、不同缩放、不同摇摄或不同聚焦平面选择中的至少一个产生所述共享数字视频流作为输出数字视频流;以及

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

用于产生视频流的系统和方法

技术领域

[0001] 本发明涉及一种用于产生数字视频流,并且特别是用于基于两个或更多个不同的数字输入视频流产生数字视频流的系统、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产生,特别地涉及多个不同的并发用户。产生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,本发明应用于不是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产生数字视频流。例如,这样的场景可以是教育性的或教学性的。

[0003] 存在许多已知的数字视频会议系统,诸如Microsoft®Teams®、Zoom®和Google® Meet®,向两个或更多参与者提供使用本地录制并向全部参与者广播的数字视频和音频进行虚拟会面以模拟实际会面。

[0004] 普遍需要改进这种数字视频会议解决方案,特别是在所观看的内容的产生方面,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] 可能期望通过自动产生过程基于多个输入数字视频流产生一个或几个输出数字视频流,并将这样一个或多个产生数字视频流提供给一个或几个消费实体。

[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产生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,因此重要的是延迟较低。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,这就产生了问题。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要求解码和重新编码,进而产生延迟和同步方面的问题。这种编码在计算上也是繁重的,并且因此在硬件要求方面成本很高。

[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,从而要求持续平衡。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(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)下被放大。

[0015] 对应问题出现在其中将基于几个输入数字视频流产生输出数字视频流的所述其他场景中,诸如在用于教育和教学的数字视频产生系统中。

[0016] 瑞典申请SE 2151267-8(其在本申请生效之日尚未公布)公开了以上讨论的问题的各种解决方案。

[0017] 瑞典申请2151461-7(其也没有在本申请的生效日期公开)公开了专门用于处理多参与者数字视频环境中的延迟的各种解决方案,诸如当不同的参与者群组与不同的总体延迟相关联时。

[0018] 仍然存在与实现以上讨论的类型的会面的自动化产生相关的问题。特别地,在几个相机在这种自动化产生中可用的情况下,已经证明难以以对会面的参与者来说感觉自然且直观的方式使用这些相机的相应图像输出。

发明内容

[0019] 本发明解决了上述问题中的一个或几个。

[0020] 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:在收集步骤中,从第一数字视频源收集第一数字视频流,以及从第二数字视频源收集第二数字流;在第一产生步骤中,基于所述第一数字视频流的连续考虑的帧产生共享数字视频流作为输出数字视频流,使得来自所述第一数字视频源的图像信息在共享数字视频流中是可见的,但是使得来自所述第二数字视频源的图像信息在共享数字视频流中是不可见的;在触发检测步骤中,通过执行以下中的至少一个来自动检测触发:a)数字分析所述第一数字视频流,以自动检测呈以下形式的触发:第一数字视频流中描绘的参与者凝视朝向所述第二数字视频流中描绘的对象或相对于该对象做出预定手势,检测基于关于第一相机、所述参与者和所述对象的相对取向的信息,检测还基于对所讨论的参与者的身体方向、头部方向、凝视方向和手势方向中的至少一个的基于数字图像的确定;以及b)数字分析所述第一数字视频流和/或第二数字视频流,以自动检测呈所述第二数字视频流中描绘的参与者的注视方向从不注视第二相机到注视第二相机的转变的形式的触发,根据预定产生规则,该触发是指示性的以改变共享数字视频流的产生模式;在响应于检测到所述触发而启动的第二产生步骤中,基于所述第二数字视频流的连续考虑的帧产生共享数字视频流作为输出数字视频流,使得来自所述第二数字视频源的图像信息在共享数字视频流中是可见的,和/或基于所述第一数字视频源的连续考虑的帧但是与第一产生步骤(S4)相比以所述第一数字视频流的不同裁剪、不同缩放、不同摇摄或不同聚焦平面选择中的至少一个产生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。

[0021] 此外,本发明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:收集步骤,其中从第一数字视频源收集第一数字视频流,并且从第二数字视频源收集第二数字流;第一产生步骤,其中基于所述第一数字视频流的连续考虑的帧产生共享数字视频流作为输出数字视频流,使得来自所述第一数字视频源的图像信息在共享数字视频流中是可见的,但是使得来自所述第二数字视频源的图像信息在共享数字视频流中是不可见的;触发检测步骤,其中通过执行以下中的至少一个来自动检测触发:a)数字分析所述第一数字视频流,以自动检测呈以下形式的触发:第一数字视频流中描绘的参与者凝视朝向所述第二数字视频流中描绘的对象或相对于该对象做出预定手势,检测基于关于第一相机、所述参与者和所述对象的相对取向的信息,检测还基于对所讨论的参与者的身体方向、头部方向、凝视方向和手势方向中的至少一个的基于数字图像的确定;以及b)数字分析所述第一数字视频流和/或第二数字视频流,以自动检测呈所述第二数字视频流中描绘的参与者的注视方向从不注视第二相机到注视第二相机的转变的形式的触发,根据预定产生规则,该触发是指示性的以改变共享数字视频流的产生模式;响应于检测到所述触发而启动的第二产生步骤,其中基于所述第二数字视频流的连续考虑的帧产生共享数字视频流作为输出数字视频流,使得来自所述第二数字视频源的图像信息在共享数字视频流中是可见的,和/或基于所述第一数字视频源的连续考虑的帧但是与第一产生步骤相比以所述第一数字视频流的不同裁剪、不同缩放、不同摇摄或不同聚焦平面选择中的至少一个产生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提供给共享数字视频流的消费者。

[0022] 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第一数字视频源收集第一数字视频流,以及从第二数字视频源收集第二数字流;第一产生功能,该第一产生功能被布置为基于所述第一数字视频流的连续考虑的帧产生共享数字视频流作为输出数字视频流,使得来自所述第一数字视频源的图像信息在共享数字视频流中是可见的,但是使得来自所述第二数字视频源的图像信息在共享数字视频流中是不可见的;触发检测功能,该触发检测功能被布置为通过执行以下中的至少一个来自动检测触发:a)数字分析所述第一数字视频流,以自动检测呈以下形式的触发:第一数字视频流中描绘的参与者凝视朝向所述第二数字视频流中描绘的对象或相对于该对象做出预定手势,检测基于关于第一相机、所述参与者和所述对象的相对取向的信息,检测还基于对所讨论的参与者的身体方向、头部方向、凝视方向和手势方向中的至少一个的基于数字图像的确定;以及b)数字分析所述第一数字视频流和/或第二数字视频流,以自动检测呈所述第二数字视频流中描绘的参与者的注视方向从不注视第二相机到注视第二相机的转变的形式的触发,根据预定产生规则,该触发是指示性的以改变共享数字视频流的产生模式;第二产生功能,该第二产生功能被布置为响应于检测到所述触发而启动,并且基于所述第二数字视频流的连续考虑的帧产生共享数字视频流作为输出数字视频流,使得来自所述第二数字视频源的图像信息在共享数字视频流中是可见的,和/或基于所述第一数字视频源的连续考虑的帧但是与第一产生功能相比以所述第一数字视频流的不同裁剪、不同缩放、不同摇摄或不同聚焦平面选择中的至少一个产生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。

附图说明

[0023] 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:

[0024] 图1示出了第一例示性系统;

[0025] 图2示出了第二例示性系统;

[0026] 图3示出了第三例示性系统;

[0027] 图4示出了中央服务器;

[0028] 图5示出了第一方法;

[0029] 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随后状态;

[0030] 图7从概念上示出了通用协议;

[0031] 图8示出了第四例示性系统;

[0032] 图9示出了第二方法;以及

[0033] 图10a至图11b从上方并以各种配置示出了相机和参与者。

[0034] 所有附图具有相同或相应部分的附图标记。

具体实施方式

[0035] 图1示出了根据本发明被布置为执行根据本发明的用于提供数字视频流(诸如共享数字视频流)的方法的系统100。

[0036] 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。如将要讨论的那样,可能存在多于一个视频通信服务110。

[0037] 系统100可以包括一个或几个参与客户端121,但是在一些实施例中,参与客户端121中的一个、一些或全部也可以在系统100外部。

[0038] 系统100可以包括中央服务器130。

[0039] 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。

[0040] 中央服务器130在其上运行的物理或虚拟硬件(换言之,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。

[0041] 每个视频通信服务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。

[0042] 相应地,所述参与客户端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,并且每个参与客户端121在其上运行的物理或虚拟硬件(换言之,限定参与客户端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。

[0043] 每个参与客户端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使用所讨论的参与客户端121参与所述视频通信。

[0044] 换言之,每个参与客户端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。

[0045] 通常,参与客户端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产生。

[0046] 更普遍地,参与客户端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。

[0047] 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还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。

[0048] 可以存在在本发明类型的同一视频通信中使用的多于一个(诸如至少三个或者甚至至少四个)参与客户端121。

[0049] 可能存在至少两个不同的参与客户端群组。参与客户端中的每一个可以被分配到这样的相应群组。这些群组可以反映参与客户端的不同角色、参与客户端的不同虚拟或物理位置、和/或参与客户端的不同交互权限。

[0050] 各种可用的这种角色可以是例如“领导者”或“与会者”、“说话者”、“小组参与者”、“互动观众”或“远程听众”。

[0051] 各种可用的这种物理位置可以是例如“在舞台上”、“在小组中”、“在物理上在场的观众中”或“在物理上远程的观众中”。

[0052] 虚拟位置可以根据物理位置来限定,但是也可以涉及可能与所述物理位置部分重叠的虚拟分组。例如,物理上在场的观众可以被划分为第一虚拟群组和第二虚拟群组,并且一些物理上在场的观众参与者可以与一些物理上较远的观众参与者被分组在同一个虚拟群组中。

[0053] 各种可用的这种交互权限可以是例如,“完全交互”(没有限制)、“可以讲话但仅在请求麦克风之后”(诸如在视频会议服务中举起虚拟手)、“不能讲话但在普通聊天中写入”或“仅观看/收听”。

[0054] 在一些情况下,可以根据某些预定的交互权限来限定所限定的每个角色和/或物理/虚拟位置。在其他情况下,具有相同交互权限的全部参与者形成群组。因此,任何所限定的角色、位置和/或交互权限可以反映各种群组分配,并且不同的群组可以是不相交的或重叠的,视情况而定。

[0055] 这将在下面例示。

[0056] 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。

[0057] 如本文中使用术语那样,“视频通信”是涉及至少两个、优选地至少三个或者甚至至少四个视频流,并且优选地还涉及用于产生一个或几个混合或联合数字视频/音频流的匹配的音频流的交互式数字通信会话,这些视频/音频流又被一个或几个消费者(诸如所讨论类型的参与客户端)消费,该一个或几个消费者可能或可能不经由视频和/或音频对视频通信做出贡献。这种视频通信是实时的、具有或没有一定的延迟或延时。这种视频通信的至少一个、优选至少两个或者甚至至少四个参与者122以交互方式参与视频通信,提供且消费视频/音频信息。

[0058] 参与客户端121中的至少一个或参与客户端121中的全部可以包括本地同步软件功能125,这将在下文中更详细地描述。

[0059] 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。

[0060] 至少一个中央服务器130中的每一个可以包括相应API 137,用于与所讨论的中央服务器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。

[0061] 系统100(诸如所述中央服务器130)另外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)。信息源300是“外部的”意味着它不是从中央服务器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客户端121中的任何一个相同的地点,但不是作为参与客户端121本身的活动的一部分而被捕获的。外部提供的信息源300与内部提供的信息源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。

[0062] 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。

[0063] 如图1所示,参与客户端121中的每一个可以构成如所描述的由所讨论的参与客户端121提供给视频通信服务110的相应信息(视频和/或音频)流120的源。

[0064] 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产生的数字视频和/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,消费者150是“外部的”意味着消费者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。

[0065] 除非另外说明,否则本文中的全部功能和通信是数字地和电子地提供的,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。

[0066] 因此,在图1中示出的系统100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。换言之,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”。

[0067] 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)连续产生的数字视频和/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客户端121的相应视频和/或音频流;和/或公共时间基准。

[0068] 中央服务器130可以包括收集功能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提供产生(诸如共享)视频流。例如,这个产生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中的所有请求或任何请求之一。

[0069] 图2类似于图1,但是不是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。

[0070] 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,所产生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。

[0071] 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可以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,并且可以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。

[0072] 中央服务器130还可以包括事件检测功能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。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。

[0073] 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提供的并可能已经由事件检测功能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。

[0074] 中央服务器130还可以包括模式检测功能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、诸如至少三个活着甚至至少四个、诸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。模式检测还可以基于上面讨论的分组,并且特别地被设置为检测仅相对于一个群组、仅相对于某些群体而非全部群组、或者相对于所有群组出现的特定模式。

[0075] 中央服务器130还包括产生功能135,其被设置为基于从收集功能131提供的数据流并且可能地还基于任何检测到的事件和/或模式,来产生产生数字视频流,诸如共享数字视频流。这种产生视频流可以至少包括被产生以包括由收集功能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。如将在下面例示的那样,可以存在几个产生视频流,其中一个这样产生视频流可以以上面讨论的方式但是进一步基于另一已经产生的视频流产生。

[0076] 所有产生视频流优选地被连续产生,并且优选地接近实时(在扣除下文中讨论的类型的任何延迟和延时之后)地产生。

[0077] 中央服务器130还可以包括发布功能136,该发布功能被布置为例如经由如上所述的API 137发布产生数字视频流。

[0078] 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。

[0079] 因此,图5示出了用于提供所述产生数字视频流的方法。图6a至图6f示出了由图5中示出的方法步骤产生的不同数字视频/音频数据流状态。

[0080] 在第一步骤中,方法开始。

[0081] 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。

[0082] 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。

[0083] 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同一屏幕中的

这些的任意组合。

[0084] 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分离的音频流数据。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)。

[0085] 在由事件检测功能132执行的随后的事件检测步骤中,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。

[0086] 优选的是,可以对至少一个、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,并且可以对每个这样的主视频流210、301单独执行这个事件检测步骤。换言之,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。

[0087] 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。

[0088] 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。

[0089] 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;导致图像品质改变、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的连接性品质方面的改变;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、或音频品质的改变。应该认识到,这并不旨在是详尽列表,而是提供这些示例以便理解目前描述的原理的适用性。

[0090] 在由同步功能133执行的随后的同步步骤中,对主数字视频流210进行时间同步。这种时间同步可以相对于公共时间基准260。如图6d所示,时间同步可能涉及例如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。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。换言之,公共时间基准可以允许主视频流210、301通过时移来对准,以便在当前意义上时间同步。在其他实施例中,时间同步可以基于关于所讨论的主视频流210、301之间的时间差的已知信息,诸如基于测量。

[0091] 如图6d所示,时间同步可以包括针对每个主视频流210、301确定一个或几个时间戳261,诸如相对于公共时间基准260;或者相对于另一个视频流210、301或者相对于其他视频流210、301针对每个视频流210、301确定一个或几个时间戳。

[0092] 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。

[0093] 与事件检测步骤相比,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。

[0094] 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。

[0095] 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之一中的信息。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品质改变)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对比,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。

[0096] 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。

[0097] 在由产生功能135执行的随后的产生步骤中,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,将共享数字视频流产生为输出数字视频流230。

[0098] 如将在下面解释并详述的那样,本发明允许视频流的完全自动产生,诸如输出数字视频流230的完全自动产生。

[0099] 例如,这种产生可以涉及选择来自什么主视频流210、301的什么视频和/或音频信息在这种输出视频流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。

[0100] 这在图6f中示出,该图还示出了可以(诸如与所述公共时间基准260)时间同步并且在产生输出视频流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(其可能与公共时间基准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。

[0101] 在由发布功能136执行的随后的发布步骤中,所产生的输出数字视频流230被连续提供给产生数字视频流的消费者110、150,如上所述。产生数字视频流可以被提供给一个或几个参与客户端121,诸如经由视频通信服务110。

[0102] 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,以产生作为连续提供的流的输出视频流230。优选地,输出视频流230被产生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或延时)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的产生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。

[0103] 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(诸如至少三个,诸如至少四个,或者甚至至少五个)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)。

[0104] 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。

[0105] 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。

[0106] 优选的是,收集功能131被预先配置为读取并解释出现在全部所收集的主视频流210、301中的所有编码格式、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。

[0107] 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。换言之,为了低延迟,整体视频和音频品质将根据需要降低。

[0108] 另一方面,外部视频馈送301通常将具有更稳定的帧率、更高的品质,但因此可能具有更高的延迟。

[0109] 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,在这种情况下,本文中描述的分析和视频产生过程需要将这些不同格式的流210、301组合成新的流以获得组合式体验。

[0110] 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。

[0111] 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。

[0112] 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。换言之,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。

[0113] 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。

[0114] 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。

[0115] 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。

[0116] 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,诸如关于用于产生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。元数据242还可以包括关于所存储的数据(诸如与由此所讨论的主视频流210、301的时间基准相关或者与如上所讨论的不同视频流相关)的时间戳的信息。

[0117] 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。

[0118] 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。

[0119] 每个由此收集的主视频流210、301可以存储在中央服务器130中的其自己分离的存储缓冲器(诸如RAM存储缓冲器)中。

[0120] 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。

[0121] 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的相应时间线规则分布的子集)与沿着共享时间线(诸如与所述公共时间基

准260相关)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的随后的时间同步。取决于所使用的公共时间基准的类型,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由同步功能133来执行。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。

[0122] 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品质和/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。

[0123] 所述主数字视频流210、301中的每一个可以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。

[0124] 在被提供来说明这些原理的具体示例中,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。

[0125] 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。

[0126] 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)拾取原始数据。每个Teems®会面参与者以及相关联的音频数据与分离的端口相关联。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。

[0127] 如所提及的那样,公共协议240可以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要求昂贵的硬件。而且,这个问题与参与者的数量成比例。

[0128] 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。

[0129] 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
			}

[0130]	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		保留用于将来使用
	8 字节	1000000	以字节为单位的二进制数据的长度（长 int）
[0131]	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

[0132] 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)可以替代地被放在分离的存储缓冲器中。

[0133] 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。

[0134] 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。

[0135] 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。

[0136] 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。

[0137] 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,并且可能在比通常情况下长一些的时间段期间进行演示。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。

[0138] 实际上,事件检测步骤可以包括以下中的至少一个:

[0139] 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随后的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。

[0140] 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。

[0141] 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,从而以更大的具体性(specificity)确定事件的类型。

[0142] 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。

[0143] 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。

[0144] 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客户端121的主数字视频流210的一系列随后视频帧213的图像分析来检测所述参与客户端121已经失去通信连接。

[0145] 因为参与客户端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或与中央服务器130的连接。在这种情况下,期望避免在所产生的输出视频流230中显示黑屏或空白屏幕。

[0146] 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。

[0147] 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并且期望的各种行动。

[0148] 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相关或者在时域中彼此相关,这使得同步功能133可以对它们彼此时间同步。

[0149] 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客户端121的共享数字视频通信服务110是公共的。

[0150] 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产生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分离的(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。

[0151] 换言之,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中的所有主视频流210(但可能不是外部主视频流301)的心跳。换言之,所有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。

[0152] 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客户端121中的一个或几个单独参与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客户端121和所讨论的本地软件功能

125可用。

[0153] 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)。

[0154] 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。

[0155] 这种通信可以经由所讨论的参与客户端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,参与客户端121可以在由所讨论的参与客户端121产生的主视频流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。

[0156] 在再附加的示例中,每个参与客户端121可以在可用于由视频通信服务110的全部参与客户端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,从而导致不同参与客户端121之间的相对时间偏移的估计集合。

[0157] 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。

[0158] 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。

[0159] 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)(在本上下文中,术语“延时”和“延迟”旨在是指相同事物),使得输出数字视频流230至少以所述延时提供。无论如何,有意引入的延时是至少几个视频帧,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”意味着引入延时,而不管对基于同步问题或类似问题引入这种延时的任何需要。换言之,除了作为主视频流210、301的同步的一部分而引入的任何延时之外,还引入了有意引入的延时,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,有意引入的延时可以是预定的、固定的或可变的。延时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中的更多延时的流与相对较小的有意增加的延迟相关联。

[0160] 在一些实施例中,引入了相对较小的延时,诸如0.5秒或更小的延时。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时。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,可能引入更大的延时。

[0161] 这种有意引入的延迟可能足以实现用于同步功能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。它还可能足以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,有意引入所述延时可能足以允许改进的模式检测功能134,如将在下文中描述的那样。

[0162] 认识到的是所述延时的引入可能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。换言之,主视频流210、301中的至少一个、几个或甚至全部的视频和/或音频数据然后可以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)。

[0163] 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、诸如至少四个、或甚至全部)的特定信息,该信息存在于还将用于产生输出数字视频流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,以产生输出视频流230的当前帧。一旦输出视频流230时间线到达所讨论的帧213,它将被用于产生输出视频流230的对应帧并且此后可以被丢弃。

[0164] 换言之,模式检测功能134能够具有仍未用于产生输出视频流230的视频/音频帧213的集合,并可以所使用这个数据来检测所述模式。

[0165] 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、诸如至少四个、或甚至全部)以自动检测所述模式212。

[0166] 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(诸如至少三个、诸如至少四个)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。

[0167] 产生步骤优选地包括确定、跟踪并更新输出视频流230的当前产生状态。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,产生功能135可以被视为关于所产生的输出视频流230的状态机。

[0168] 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。

[0169] 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,因为演示参与客户端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。

[0170] 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。换言之,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检测到的模式的产生功能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。

[0171] 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并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。

[0172] 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产生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,所述模式对于产生平滑的输出视频流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。

[0173] 然后,产生功能135可以在采取与所述产生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。

[0174] 在一些实施例中,可以存在并行检测到的多个模式,并形成对产生功能135状态机的输入。这样的多个模式可以由产生功能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,从而产生要在所述产生中使用的所确定的期望状态变化。例如,可以产生不同的这种预定分类器,特别地被布置为根据各种不同的产生风格和期望自动产生输出视频流230。训练可以基于作为期望输出的已知产生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产生这样的分类器。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,我从不直接从说话者A切换到说话者B,而是总是在我聚焦于另一说话者之前首先示出概述,除非另一说话者非常占优势并且说话声音很大”的输入。然后,这个产生逻辑被表示为呈总体形式“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。

[0175] 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产生格式,以便在视频流的自动化产生中使用。

[0176] 总之,使用基于各个主视频流210、301的事件检测的组合、有意引入的延时、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产生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的自动产生。这个结果在由事件检测功能132、模式检测功能134和产生功能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。另外,它在以下描述的实施例中是有效的,这些实施例的特征是第一产生视频流被用于第二产生视频流的自动产生;以及对不同参与客户端群组使用不同的有意添加的延时。特别地,它在下面描述的实施例中也是有效的,其中所检测到的触发导致在所产生的输出视频流中使用什么视频流的切换,或者导致输出视频流中的所使用的视频流的自动裁剪或缩放。

[0177] 如上所例示那样,产生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的输出模式来产生输出数字视频流230。这些参数可以由所述产生功能135状态机自动确定和/或由控制产生的操作者设置(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。

[0178] 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。各种可用产生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,一个产生形式可能是“小组讨论”,而另一产生形式是“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产生格式,系统100的操作者可以快速选择预定义的这种产生格式集合中的一个产生格式,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产生格式完全自动地产生输出视频流230。

[0179] 另外,在产生期间,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的产生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、对当前使用的产生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。

[0180] 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。

[0181] 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。

[0182] 总体而言,产生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。

[0183] 如上所提及那样,参与客户端121可以被组织为两个或更多个参与客户端121群组。图8是在存在这样的群组的情况下处于用于执行输出视频流的自动产生的配置的系统100的简化视图。

[0184] 在这个图8中,中央服务器130包括收集功能131,该收集功能可以如上所述。

[0185] 中央服务器130还包括第一产生功能135’、第二产生功能135”和第三产生功能135”’。每个这样的产生功能135’、135”、135”’对应于产生功能135,并且上面关于产生功能135的所述的内容同样适用于产生功能135’、135”’和135”’。根据中央服务器130的详细配置,产生功能135′、135″、135”’可以是不同的,或者可以与几个功能共同布置在一个单个逻辑功能中,并且也可以是三个以上的产生功能。产生功能135′、135″、135”’在某些情况下可以是同一产生功能135的不同功能方面,视情况而定。产生功能135’、135”、135”’和其他实体之间的各种通信可以经由合适的API进行。

[0186] 还认识到的是,根据详细的配置,对于产生功能135’、135”、135”’中的每一个或这种产生功能的群组,可以存在分离的收集功能131,并且可以存在几个逻辑上分离的中央服务器130,其中每个中央服务器具有相应收集功能131。

[0187] 而且,中央服务器130包括第一发布功能136’、第二发布功能136”和第三发布功能136”’。每个这样的发布功能136’、136”、136”’对应于发布功能136,并且上面关于发布功能136的所述的内容同样适用于发布功能136’、136”’和136”’。根据中央服务器130的详细配置,发布功能136′、136″、136”’可以是不同的,或者可以与几个功能共同布置在一个单个逻辑功能中,并且也可以是三个以上的发布功能。发布功能136′、136″、136”’在某些情况下可以是同一发布功能136的不同功能方面,视情况而定。

[0188] 在图8中,为了说明性目的,示出了三个参与客户端集合或群组,每个集合或群组对应于上述参与客户端121。因此,存在这样的参与客户端121’的第一群组、这样的参与客户端121”的第二群组、以及这样的参与客户端121”’的第三群组。这些群组中的每一个可以包括一个或者优选地至少两个参与客户端。根据详细配置,可以只有两个这样的群组,或者多于三个这样的群组。群组121′、121″、121”’之间的分配在每个参与客户端121被分配给最多一个这样的群组121′、121″、121”’的意义上可以是排他性的。在替代性配置中,至少一个

参与客户端121可以同时被分配给多于一个这样的群组121’、121”、121”’。

[0189] 图8还示出了外部消费者150,并且认识到的是可能存在多于一个如上所述的外部消费者150。

[0190] 为了简单起见,图8没有示出视频通信服务110,但是应该认识到的是,以上讨论的通用类型的视频通信服务可以与中央服务器130一起使用,诸如以以上讨论的方式使用中央服务器130向每个参与客户端121提供共享视频通信服务。

[0191] 可以通过收集功能131从相应参与客户端121(例如,所述群组121’、121”、121”’的参与客户端)收集相应的主视频流。基于所提供的主视频流,产生功能135’、135”、135”’可以产生相应的数字视频输出流。

[0192] 如图8所示,一个或几个这样的所产生的输出流可以作为相应的输入数字视频流从一个或几个相应的产生功能135’、135”’提供给另一产生功能135”,该另一个产生功能又可以产生用于通过发布功能136”发布的辅数字输出视频流,辅数字输出视频流因此基于一个或几个输入主数字视频流以及基于一个或几个预先产生的数字输入数字视频流产生。

[0193] 两个或更多个不同的产生步骤135’、135”、135”’可以包括引入相应的时间延时。在一些实施例中,由于所述时间延时引入,来自这些产生步骤135’、135”、135”’的相应的所产生的输出数字视频流中的一个或几个可以与在所述发布步骤中可以被提供给其他参与客户端的视频流中的任何其他视频流时间不同步。这个时间延时可以以本文中描述的方式中的任一个被有意添加和/或作为所讨论的第二产生数字视频流的产生的直接结果。结果,消费所述时间不同步的所产生的输出数字视频流的任何参与客户端将在相对于所述其他参与客户端的视频流消费“时区”稍微偏移(时间方面)的“时区”中这样做。

[0194] 例如,参与客户端121的群组121’、121”、121”’中的一个可以在第一这样的“时区”中消费相应的产生视频流,而所述群组121’、121”、121”’中的另一群组的参与客户端121可以在第二这样的“时区”中消费相应的产生视频流。因为两个这些相应的产生视频流可以至少部分地基于相同的主视频流产生,所以全部这样的参与客户端121将在相同的视频通信中但是在彼此不同的“时区”中是活跃的。换言之,产生视频流消费的相应时间线可以在不同群组121’、121”、121”’之间在时间方面偏移。

[0195] 例如,一些产生步骤(诸如135’、135”’)可以是直接的(不使用任何有意引入的时间延时)和/或在提供用于发布之前仅涉及计算上相对轻量的处理;而其他产生步骤(诸如135”)可能涉及有意引入的时间延时和/或相对较重的处理,从而导致所讨论的产生数字视频流相对于用于先前产生步骤135’、135”’的相应数字视频流的发布的最早延时以一定延时被产生用于最早发布。

[0196] 因此,在所述群组121’、121”、121”’中的一个或几个中的每个参与客户端121可能能够以相同的感知时间延时彼此交互。同时,与较大的相应时间延时相关联的群组可以使用来自具有较小时间延时的群组的产生视频流作为输入视频流以便在产生较大时间延时的群组在其随后的“时区”观看的输出视频流时使用。

[0197] 因此,这个第一较大时间延时产生的结果(在步骤135”中)是上述类型的产生数字视频流,该产生的数字视频流可以例如在视觉上包括呈处理或未处理的形式的所讨论的主视频流中的一个或几个作为子部分。这个所讨论的产生视频流可以包括现场捕获的视频流、幻灯片、外部提供的视频或图像等,如上文关于由中央服务器130产生的视频输出流一

般描述的那样。所讨论的产生视频流也可以以上描述的一般方式基于由参与者客户端121提供的有意延迟或实时输入的主视频流的检测到的事件和/或模式来产生。

[0198] 在说明性示例中,第一群组121'参与客户端是使用视频通信服务110以相对低的延迟进行通信的辩论小组的一部分,这些参与客户端中的每一个被连续地馈送来自发布步骤136'的产生视频流(或彼此的相应主视频流)。辩论小组的观众由第二群组121"参与客户端构成,这些参与客户端被连续馈送来自产生步骤135"的产生视频流,该产生视频流又与稍高的延迟相关联。来自产生步骤135"的产生视频流能够以以上讨论的一般方式自动产生,以在各个辩论小组说话者(分配给第一群组121'的参与客户端,这样的视图直接从收集功能131提供)的视图和示出全部辩论小组说话者的产生视图(这个视图是第一产生视频流)之间自动切换。因此,观众可以接收良好体验,同时小组说话者可以以最小的延迟彼此交互。

[0199] 有意添加到在产生步骤136"中使用的相应主视频流的延时可以是至少0.1s,诸如至少0.2s,诸如至少0.5s;并且可以是至多5s,诸如至多2s,诸如至多1s。它还可以取决于与每个相应的主视频流相关联的固有延迟,以便实现所使用的主视频流中的每一个与还有从产生步骤135'进入到产生步骤135"的产生视频流之间的完全时间同步。

[0200] 应当理解的是,所讨论的主视频流以及来自产生步骤135'的产生视频流可以全部被附加地有意延迟,以便以上述一般方式改进用于在第二产生功能135"中使用的模式检测。

[0201] 图8另外示出了发布由中央服务器130产生的各种产生视频流的多种替代性或并发方式。

[0202] 总体而言,在由被布置为从第一产生功能135'接收第一产生视频流的第一发布功能136'执行的发布步骤中,所述第一产生视频流可以被连续地提供给第一参与客户端121和第二参与客户端121中的至少一个。例如,这个第一参与客户端可以是来自向第一产生功能135'提供相应的主要数字视频流的群组121'的参与客户端。

[0203] 在一些实施例中,群组121'的参与客户端中的一个或几个也可以通过第二发布功能136"接收第二产生视频流,该第二发布功能又被布置为从第二产生功能135"接收第二产生视频流。

[0204] 因此,被分配给第一群组121'的主视频流提供参与客户端中的每一个如果没有被直接提供彼此的主数字视频流的话可以被提供第一产生视频流,这涉及由于所述主视频流之间的同步而导致的特定延时或延迟以及还有可能有意地添加的延时或被添加以便允许足够的时间用于事件和/或模式检测的延迟,如上所述。

[0205] 相应地,被分配到第二群组121"的参与客户端中的每一个可以被提供第二产生视频流,也包括与第二产生步骤相关的所述有意添加的延迟,添加的目的是使第一产生视频流与所述第一主视频流和第二主视频流时间同步。这种额外的延时可能或可能不导致第二群组121"的参与客户端之间的通信困难,例如因为它们以与第一群组121'参与者不同的方式与视频通信服务110进行交互。

[0206] 因此,第一群组121'参与客户端形成当前参与所讨论的视频通信服务110的、在比任何参与客户端稍微提前(诸如提前1至3秒)的"时区"中存在并使用所讨论的服务,而不是被连续提供产生视频流(诸如第一产生视频流或第二产生视频流)的全部参与客户端121的

子群组。然而,其他参与客户端(未被分配给第一群组121’而是被分配给第二群组121”)将被连续地提供第二产生视频流,该第二产生视频流基于第二产生视频流基于其产生的主视频流中的至少一个(并且在每个时间点可以包含该至少一个主视频中的一个或两者)但是在稍晚的“时区”中产生。由于第一产生视频流直接基于这些至少一个主视频流产生,而没有添加任何延时或延迟来将它们与基于主视频流本身的已经产生的视频流进行时间同步,因此这些参与客户端121然后可以获得更直接、低延迟的视频通信服务110体验。

[0207] 同样,这也可能意味着分配给第一群组121’的参与客户端121未被提供对第二产生视频流的访问。

[0208] 如图8所示,还可以附加地基于第三产生步骤135”’的所产生的输出视频(第三产生输出数字视频流)将第二产生视频流产生为数字视频流。因此,产生步骤135’、135”、135”’中的每一个可以基于至少部分重叠的主输入数字视频流,但是使用相对于所述主视频流的并且被提供给参与客户端121的不同的相应群组121’、121”、121”’的不同的相应有意添加的延迟(“时区”)产生。

[0209] 分配给第三群组121”的参与客户端121可能比分配给第一群组121’的参与客户端121具有较不严格的延迟要求。例如,第一群组121’参与客户端121可以是以上讨论的辩论小组的成员(彼此实时交互,因此需要低延迟),而第三群组121”参与客户端121可以构成但以更结构化的方式(诸如使用清晰的问题/答案)与小组交互的专家小组等,因此能够容忍比第一群组121’高的延迟。

[0210] 第一产生视频流和第三产生视频流两者可以被馈送到第二产生功能135”以作为第二产生视频流的产生的基础,视情况而定。

[0211] 因此,第一产生步骤135’可以涉及引入以上讨论的类型的故意延时或延迟,该延时或延迟除了作为第一主视频流和第二主视频流的同步的一部分引入的任何延迟之外被引入,并且被引入为以便获得足够的时间来例如执行有效的事件和/或模式检测。有意延时或延迟的这种引入可以作为由所述同步功能133执行的同步的一部分而发生(为了简单起见,在图8中未示出)。对于第三产生步骤135”’对应内容可能如此,但是引入了与用于第一产生步骤135’的有意引入的延时或延迟不同的有意引入的延时或延迟。

[0212] 特别地,有意引入的延时或延迟导致第一产生视频流和第三产生视频流之间的时间不同步。这意味着在第一产生视频流和第三产生视频流两者在产生每个单独帧时被立即和连续地发布的情况下,它们不遵循公共时间线。

[0213] 因此,可以同时但是在不同的“时区”产生和消费/发布三个分离的产生视频流。即使它们至少部分地基于相同的主视频材料,产生视频流也以不同的延迟发布。要求最低延迟的第一群组121’可以使用第一产生视频流进行交互,从而提供非常低的延迟;愿意接受稍大延迟的第三群组121”’可以使用第二产生视频流进行交互,从而提供更大的延迟,但另一方面在有意添加的延迟方面提供更大的灵活性以便实现更好的自动产生,如本文别处所述;而对延迟不太敏感的第二群组121”可以享受使用第二产生视频流的交互,该第二产生视频流既可以结合来自第一群组121’和第三群组121”’的材料,也可以以非常灵活的方式自动产生。特别要注意的是,尽管使用了所述变化的延迟并因此在不同的“时区”中起作用,这些参与者用户群组121’、121”、121”’中的全部使用所述视频通信服务110彼此交互。然而,由于每个产生功能中的各个输入视频流的同步,参与者用户121将不从他们相应角度注

意到不同的延迟。

[0214] 如上所述,分配给所述群组121’、121”、121”’中的每一个的相应参与客户端121可以参与在其中连续发布第二产生视频流的同一个视频通信服务110。

[0215] 然后,所述群组121’、121”、121”’中的不同群组可以与所述视频通信服务110中的不同参与者交互权限相关联。在这些和其他实施例中,所述群组121’、121”、121”’中的不同群组还可以与用于产生向分配给所讨论的群组121’、121”、121”’的参与客户端121发布的相应产生视频流的不同的最大时间延时(延迟)相关联。

[0216] 例如,小组辩论参与者客户端的第一群组121’可以与完全交互权限相关联,并且可以在他们希望的任何时候说话。参与客户端的第三群组121”’可以与例如他们需要在能够通过视频通信服务110取消其麦克风静音来讲话之前请求发言权的稍微受限的交互权限相关联。观众参与者用户的第二群组121”可以与更多受限的交互权限(诸如只能在公共聊天室中以书写形式提出问题但不能讲话)相关联。

[0217] 因此,对于向其发布的相应产生视频流,不同的参与用户群组可以以这样的方式与不同的交互权限和不同的延迟相关联,即,延迟是减少交互权限的递增函数。视频通信服务110允许所讨论的参与用户121与其他用户交互越自由,可接受的延迟就越低。可接受的延迟越短,对应自动产生功能考虑诸如检测到的事件或模式的可能性就越小。

[0218] 具有最大延迟的群组可以是除了被动参与视频通信服务之外没有交互权限的仅观看者群组。

[0219] 特别地,用于所述群组121’、121”、121”’中的每一个的相应最大时间延时(延迟)可以被确定为全部主视频流和被连续发布给所讨论的群组中的参与客户端的任何产生视频流之间的最大延迟差。为此,可以添加有意添加的任何附加时间延时,目的是如上所述检测事件和/或模式。

[0220] 如本文所用,术语“产生”和“产生数字视频流”可以指不同类型的产生。在一个示例中,由中央实体(诸如中央服务器130)产生单个明确限定的数字视频流,以形成所讨论的产生数字视频流,用于提供给要消费所讨论的产生数字视频流的参与客户端的特定集合121中的每一个并在该处发布。在其他情况下,不同的各个这样的参与客户端121可以观看所讨论的产生数字视频流的稍微不同的版本。例如,产生数字视频流可以包括几个单独的或组合的数字视频流,参与客户端121的本地软件功能125可以允许所讨论的用户122在这些数字视频流之间切换;这些数字视频流布置在屏幕124上;或者这些数字视频流以任何其他方式配置或处理。很多时候,重要的在于提供产生数字视频流(包括任何时间同步的子成分)的“时区”(即,以什么延迟)。因此,上面结合图8描述的向第一群组121’的不同参与客户端120提供彼此的主视频流的情况,可以被视为向这些参与客户端提供第一产生数字视频流(在时间同步的原始或经处理的第一主数字视频流和第二主数字视频流的集合对第一参与客户端和第二参与客户端两者可用的意义上)。

[0221] 为了进一步阐明和例示上述参与客户端群组121’、121”、121”’的使用,以涉及三个不同的并发“时区”的视频通信服务会面的形式提供了以下示例:

[0222] 第一参与客户端121’群组实时地或至少接近实时地(取决于不可避免的硬件和软件延迟)体验与彼此的交互。向这些参与客户端提供来自彼此的视频(包括声音)以实现所讨论的用户122之间的这种交互和通信。第一群组121’可以在其他(非第一群组121’)参与

客户端可能对参与其交互感兴趣的会议的核心为用户122服务。

[0223] 这样的其他参与客户端121”的第二群组参与相同的会面,但是在比第一群组参与客户端121’距实时更远的不同的“时区”。例如,第二群组121”可以是具有互动特权(诸如可以向第一群组121’提问的可能性)的观众。第二群组121”的“时区”可以相对于第一群组121’的“时区”具有延时,使得提出的问题和回答与明显但短暂的延时相关联。另一方面,这个稍大的延迟允许这个第二群组121”参与客户端体验以更复杂的方式自动产生的产生数字视频流,从而提供更令人满意的用户体验。

[0224] 这样的其他参与客户端121”’的第三群组也参与相同的会面,但是仅作为观看者。这个第三群组121”’消耗产生数字视频流,这个产生数字视频流可以以更加精细和复杂的方式自动产生、在比第二“时区”具有更多延时的第三“时区”中被消耗。然而,由于第三群组121”’不能以影响第一群组121’和第二群组121”的方式向通信服务提供输入,所以第三群组121”’将如“实时”执行的那样体验会面、具有令人满意的结果。

[0225] 当然,使用本文中描述的原理,可以存在与越来越大的时间延时和越来越大的产生复杂性的相应会面“时区”相关联的多于三个这样的参与客户端群组。

[0226] 图9示出了根据本发明的方法,该方法用于提供共享数字视频流。

[0227] 在第一步骤S1中,方法开始。

[0228] 在随后的收集步骤S2中,以以上总体描述的方式,从第一数字视频源收集第一数字视频流,以及从第二数字视频源收集第二数字流。因此,第一数字视频流和/或第二数字视频流可以各自从相应的参与客户端121或从外部源300收集,并且可以由中央服务器130的收集功能131执行。

[0229] 在随后的第一产生步骤S4中,所述共享数字视频流被产生为输出数字视频流。这种产生可以如以上总体描述的那样执行,例如通过产生步骤135、135’、135”、135”’。

[0230] 在第一产生步骤S4中,基于所述第一数字视频流的连续考虑的帧产生共享数字视频流,使得来自所述第一数字视频源的图像信息在共享数字视频流中是可见的,但是使得来自所述第二数字视频源的图像信息在共享数字视频流中是不可见的。换言之,共享视频流将至少在某种程度上包含源自第一数字视频流的可视材料,但是将不包含源自第二数字视频流的任何可视材料。

[0231] 在随后的触发检测步骤S5中,数字分析所述第一数字视频流和/或第二数字视频流,以便检测至少一个触发。

[0232] 这种分析和检测可以由产生所讨论的共享视频流的相同产生步骤135、135’、135”、135”’来执行,并且基于预定类型的图像和/或音频模式的自动检测。

[0233] 触发可以是上面描述和例示的类型的事件或模式(分别由事件检测功能132和模式检测功能134执行),并且其检测通常将使用包含在所讨论的(多个)数字视频流中的音频和/或图像/视频数据的自动数字处理来进行。例如,并且如上所例示那样,可以采用自动图像处理算法(诸如使用经训练的神经网络或其他机器学习工具)来基于包含在所述第一视频馈送和/或第二视频馈送中的图像自动检测特定触发的存在。对应地,本身对应类型的常规自动音频处理算法可以用于基于包含在所述第一视频馈送和/或第二视频馈送中的音频来检测特定触发的存在。

[0234] 图像和/或音频模式是“预定类型”意味着所讨论的模式可以根据在所讨论的检测

之前定义的一个或几个绝对或相对参数值的集合进行表征。这将在下面例示。

[0235] 一般而言,所述音频或图像模式的存在构成了所讨论的对应触发。而且,触发被特别地预先确定以便根据预定产生规则对于产生共享视频流的自动产生步骤135、135’、135”、135”’是指示性,以在检测到所讨论的触发时改变共享数字视频流的产生模式(规则)。

[0236] 因此,产生步骤135、135’、135”、135”’可以包括根据表征对应图像和/或音频模式的对应参数值(瞬时或随时间变化)来定义一个或几个触发的数据库或访问该数据库。

[0237] 在响应于所述触发的检测而启动的随后的第二产生步骤S6或S7中,然后通过相同(或不同)产生步骤135、135’、135”、135”’但不以与第一产生步骤S4中相同的方式再次产生共享数字视频流。

[0238] 在第一替代性第二产生步骤S6中,基于所述第二数字视频流的连续考虑的帧产生共享数字视频流作为输出数字视频流,使得来自所述第二数字视频源的图像信息在共享数字视频流中是可见的。注意的是,在这种情况下,输出数字视频流还可以基于或者不基于第一数字视频流的连续考虑的帧来产生,使得来自第一数字视频源的图像信息在共享数字视频流中是可见的。换言之,当从第一产生步骤S4切换到第二产生步骤S6时,共享视频流可以从显示来自第一视频流而不是第二视频流的内容,到显示来自第二视频流而不是第一视频流的内容、或者显示来自第一视频流和第二视频流两者的内容。

[0239] 在第二替换性第二产生步骤S7中,基于所述第一数字视频源的连续考虑的帧但是与第一产生步骤S4相比以所述第一数字视频流的不同裁剪、不同缩放、不同摇摄和不同聚焦平面的选择中的至少一个产生共享数字视频流作为输出数字视频流。换言之,相对于在第一产生步骤S4中使用的视频流的当前裁剪/缩放/平移/聚焦平面状态,正在共享视频流中显示的视频流的内容被裁剪、去裁剪、放大、缩小、竖直和/或水平平移、和/或移位所讨论的视频流的聚焦平面。

[0240] 认识到的是,这种裁剪/摇摄/缩放/聚焦平面可以由基于现有视频流(该现有视频流本身可以包含在像素级上具有不同图像信息的多个可能的聚焦平面)由所讨论的产生步骤135、135’、135”、135”’来执行,和/或由向捕获所讨论的视频流的视频源(诸如数字相机)通信传送命令以相应地修改对应捕获参数的产生步骤135、135’、135”、135”’来执行。例如,这然后可能需要根据由产生步骤135、135’、135”、135”’提供的指令来捕获所讨论的视频流缩放、摇摄和/或移位其焦平面的对应相机。

[0241] 在随后的发布步骤S8中,以以上描述的一般方式,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,诸如提供给参与客户端121和/或外部消费者150。

[0242] 该方法然后可以迭代回到步骤S2,如图9所示。

[0243] 在随后的步骤S9中,方法结束。

[0244] 第一数字视频流可以由第一数字相机连续捕获,以及第二数字视频流可以由不同的第二数字相机连续捕获(因此使用本文中先前使用的术语构成主视频流)。替代性地,在使用与以上描述的不同相应延迟(“时区”)相关联的参与客户端121的几个不同群组121’、121”、121”’的情况下,第一数字视频流和/或第二数字视频流可以构成相应的先前产生的数字视频流,与当前产生的共享视频流相比,该先前产生的数字视频流又以不同的延迟(“时区”)产生(参见上文以获得关于这种“时区”的细节)。

[0245] 在第一视频流是已经产生的视频流的情况下,优选的是(尽管不是必须的),基于已经存在的第一视频流执行任何裁剪/缩放/摇摄,而不是指令任何上游相机修改其裁剪/缩放/摇摄设置。

[0246] 使用本方法,可以实现共享视频流的自动产生,这可以向产生共享视频的消费者提供更直观且自然的体验,因为相应视频流的实际音频/视频内容被用于检测触发,这些触发又导致自动产生从一个自动产生格式移动到不同的这种格式。

[0247] 可以使用合适的参数预定义触发,以满足各种需求。例如,在第一视频流和/或第二视频流中描绘的各个人的行为可以相对于这种触发被自动评估,并且被允许根据如何进行这种行为来改变产生格式。在其他示例中,特定的预定义触发可以被用作由第一视频流和/或第二视频流中描绘的人递送的手动提示,以在正在进行的产生期间动态地改变所讨论的产生格式。

[0248] 在下文中,将描述这种触发和对应产生格式变化的多个示例。

[0249] 在图10a和图10b中示出的第一示例中,预定模式包括在所述第一数字视频流(在图10a中被示为由第一数字相机410捕获)中描绘的第一(人或例如机械机器)参与者430。第一参与者430凝视朝向对象440(在图10a中示出的示例中是第二(人或机器)参与者),第二参与者440又在第二数字视频流中进行描绘。在这个示例中,第二数字视频流由第二数字相机420捕获。

[0250] 认识到的是,第二对象440可以是其他事物,诸如人类参与者的群组或者对于正在进行的通信有普遍兴趣的任何物理对象。例如,共享视频流可以是医疗程序的文档,由此对象440可以是患者的一部分。在其他示例中,对象440可以是教学会话或销售演示的对象。对象440还可以是例如白板或幻灯片演示屏幕。

[0251] 第一视频流和第二视频流可以是上文中描述的类型中的任何一个。

[0252] 在图10a和图10b中示出的情况下,基于关于所述第一相机410、所述参与者430和所述对象440的相对取向的信息,检测预定图像和/或音频模式。这个信息可以存在于系统100中(并且特别是在中央服务器130中),诸如先前在设置/配置期间提供的和/或在正在进行的产生期间自动检测的。

[0253] 例如,第一相机410和第二相机420的相应位置可以从所讨论的相应相机410、420被包括测量装置(诸如包括加速度计和陀螺仪的MEMS电路)、或者被布置为基于一些合适的几何参考连续或间歇地测量所讨论的相机410、420的当前位置其他本身常规定位装置(诸如步进电机)的相机馈送到中央服务器130。在其他示例中,所讨论的相机410、420的取向可以由第三相机(未示出)使用合适的自动数字图像处理算法来检测,第三相机捕获图像中的相机410、420中的至少一个并使用数字图像处理基于这个捕获的图像信息来确定所讨论的相对取向。

[0254] 注意,在该上下文中,“取向”可以涵括位置和取向分量两者。

[0255] 基于由第一相机410和/或第二相机420捕获的(多个)视频流,可以使用数字图像处理来确定第一参与者430和对象440相对于一些合适的参考系统(诸如相对于第一相机410和/或第二相机420)的位置。

[0256] 如图10a所示,在图中第一参与者430向下凝视,而不是朝向第二参与者430凝视。

[0257] 然而,在这个示例中,基于第一数字视频流、基于对第一参与者430的身体方向、头

部方向和注视方向中的至少一个的基于数字图像的确定,进一步检测预定图像和/或音频模式。

[0258] 如图10b所示,第一参与者430已经转向朝向第二参与者440,从而面向(并凝视)朝向第二参与者440。

[0259] 可以基于由第一相机410捕获的第一视频流的数字图像处理来确定第一参与者430的身体和头部方向。这样的算法由此本身是常规的,并且可以例如使用关于当在各个方向上转动时由第一相机410捕获的视频流中的第一参与者430的预期形状的先验知识。这可以使用经训练的神经网络或其他机器学习部件来完成。连同第一相机410的相对取向、第一参与者430和对象440的相对位置以及第一参与者430的所确定的身体或头部方向,中央服务器130可以确定第一参与者430是否转向(头部或身体)朝向所讨论的对象440。

[0260] 第一参与者430的凝视方向可以以类似的方式实现,诸如基于由第一相机410捕获的图像。这种凝视跟踪技术本身是众所周知的,并且可以例如基于辨别瞳孔的位置和第一参与者430的眼中可见的光反射。

[0261] 所讨论的触发可以被定义为检测到转变模式,诸如由第一参与者430从没有被引导朝向或凝视朝向对象440到第一参与者430实际上面向和/或凝视朝向对象440的转变。因此,基于适当设置的对应绝对或相对参数值,中央服务器130可以连续监测这种转变,并且一旦这种转变发生,就可以检测到触发。

[0262] 认识到的是,在这个和其他实施例中,可以同时监视几个不同的预定图像和/或音频模式,并且任何这种检测到的预定图像和/或音频模式然后将构成所检测到的对应触发,这又导致根据预定参数化产生逻辑而自动产生切换到不同的相应模式。

[0263] 图10c示出了类似于图10b中示出的示例的第二示例,但是其中对应于所讨论的触发的预定图像和/或音频模式不是第一参与者430转向或凝视对象440。相反,预定图像和/或音频模式包括第一数字视频流中描绘的(诸如由第一相机410描绘的)、做出预定手势的参与者(诸如,第一参与者430)。

[0264] 手势可以是预定的参数化类型的任何手势,诸如与所述第二数字视频流中描绘的所述对象440成几何关系的手势。具体地,手势可以是第一参与者430指向朝向所述对象440(如示出了第一参与者430的手臂431指向朝向对象440的图10c所描绘那样)。然而,手势也可以仅基于例如第一参与者430的手或手指。

[0265] 可以以对应于结合图10b描述的情况的方式、因此基于关于所述第一相机410、所述第一参与者430和所述对象440的相对取向的信息、并且还基于根据第一数字视频流的第一参与者430的手势方向的基于数字图像的确定,来检测这个手势类型的预定模式(以及特别是其在空间中的方向,如果有的话)。

[0266] 对于图10b中示出的情况和图10c中示出的情况两者,也可以检测第二相机420的取向,并将该取向用于确定是否检测到触发。例如,它可以用于确定对象440在第二相机420中是可见的,这可以构成在触发与在所讨论的检测到的触发时切换到第二相机420相关联的情况下检测到触发的条件。在一些实施例中,第二产生步骤S6可以包括确定当前示出对象440的一个第二相机420(几个可能的第二相机当中),并且选择这个第二相机420作为在第二产生步骤S6中提供所讨论的第二视频流的相机。

[0267] 在图10d中示出的示例中,只有一个相机,即第一相机410(尽管可以理解的是,在

各种实施例中当然可以存在更多的相机和其他视频流生成部件)。第一相机410捕获第一参与者430和对象440两者。在第一参与者430将她的身体、头部或凝视转向朝向对象440时,这个检测到的图像模式构成了所检测到的触发。在这种情况下,第二产生步骤可以包括由第一相机410捕获的视频流在所产生的输出视频流中被摇摄和/或缩放和/或裁剪,以便将所产生的输出视频流的观看者的注意力集中到对象440。

[0268] 在实际示例中,所检测到的第一参与者430的注意力(体现在第一参与者430的身体、头部或凝视方向方面,如上所例示那样)使得自动产生(在所述第二产生步骤中)基于现有视频流或通过指令产生所讨论的主视频流的单元以某种方式突出显示或移位焦点,以便为所讨论的对象440提供增加的视觉焦点。

[0269] 在图11a和图11b中示出的另一示例中,预定图像和/或音频模式包括在所述第二数字视频流中描绘的第一参与者430凝视朝向连续捕获第二数字视频流的第二相机420。具体而言,在图11a中,第一参与者430没有凝视第二相机420,而图11a中的第一参与者430凝视第二相机420。因此,这种检测到的凝视方向的切换可以构成检测到所讨论的对应触发。

[0270] 在其他示例中,预定图像和/或音频模式包括所述第一数字视频流和/或所述第二数字视频流中的相对移动变化。例如,第一视频流中示出的总体移动的量可以被参数化,并且第一视频流和/或第二视频流的缩放然后可以作为减少总体移动的函数而增加,反之亦然;或者在检测到增加的总体移动的情况下,第二产生步骤可以切换到第二图像流,第二视频流由示出了由第一相机410描绘的场景的更广角或更远的视图的第二相机420捕获。相应地,参与者430的语音的记录的声音可以用于确定这种放大/缩小/相机切换性能。例如,如果记录到参与者430说话更大声,则可能存在输出的共享视频流中的第一视频流的缩小,反之亦然。

[0271] 在另外的示例中,预定图像和/或音频模式包括预定声音模式,该预定声音模式又通过例如由特定声音捕获麦克风的相对麦克风音量来确定的其频率和/或幅值和/或幅值时间导数和/或绝对幅值变化和/或声音位置来表征。这种麦克风可以是例如第一相机410的一部分,或者可以是独立的麦克风。然后,这样的麦克风被布置来记录在第一视频流中正在示出的场景处产生的或者与该场景直接相关的声音。

[0272] 例如,预定声音模式可以包括包含至少一个口头说出的单词的预定短语。声音可以被记录并作为馈送到第一产生步骤的视频流的一部分被提供,并且预定模式可以被中央服务器130检测到,并且然后,当检测到对应触发时,产生可以切换到第二产生步骤。音频分析可以使用合适的数字音频处理算法(诸如使用各种声音信息(诸如音调、幅值和模式匹配)来确定是否检测到预定声音模式的基于规则的决策引擎)、经训练的神经网络等。

[0273] 如上面详细描述的那样,本方法还可以包括延时步骤(参见图9),其中至少对于第一数字视频流和第二数字视频流有意引入延迟,延迟存在于共享数字视频流中。然后,在引入所述延迟之前,可以基于第一数字视频流和/或第二数字视频流来执行触发检测步骤。

[0274] 延迟可以至多30秒,诸如至多5秒、诸如至多1秒、诸如至多0.5秒。

[0275] 使用这种有意添加的延时,自动产生可以例如基于检测到的触发(诸如参与者430看向第二相机420)来计划从第一相机410到第二相机420的自动切换,该计划发生在事件参与所产生的共享视频流之前的某个时间(诸如,0.5s)。这样,这种切换可以在触发实际发生的精确时间执行,或者在与其他参数化的产生参数最佳拟合的时间(诸如利用所讨论的参

与者430的语音节奏定时)执行。在存在以上讨论的类型的几个群组121'、121"、121"'的情况下,这种计划可以相对于关于这些群组121'、121"、121"'中的不同群组的相应参与客户端121产生(被产生以便由其消费)的所产生的输出视频流以不同的时间范围发生。

[0276] 预定图像和/或音频模式可以构成(并被确定为)上面结合事件检测功能132和模式检测功能132描述的类型的"事件"和/或"模式"。

[0277] 以下是如何在实践中实现本发明的多个示例:

[0278] 在多相机演示、脱口秀或小组讨论中,不同的相机可以呈现演示者的不同相机角度。然后,自动产生功能可以被配置为根据演示者正在看相机中的哪一个在不同的相机角度之间自动选择、和/或由手势或音频提示触发。

[0279] 在视频播客或说话人头像视频中,自动产生功能可以被配置为根据当前说话者面向什么相机在不同相机之间自动切换。

[0280] 在市政厅会面中,自动产生功能可以被配置为切换到指向观众的相机,用于添加来自与会者的输入或问题。这可以通过监测与这个相机相关联的音频馈送来触发,并且当其达到特定水平时,这个相机开始直播;或者通过语音命令(例如"来自观众的问题")来触发。

[0281] 在产品呈现或回顾中,自动产生功能可以被配置为当其感测到这个源中的运动时、或基于另一触发,自动切换到指向产品的相机。

[0282] 在机器人手术中,在检测到预定手势、音频提示时、或者在识别到外科医生没有使用外科医生控制台或者没有从控制台向上看时,由捕获外科手术的机器人相机捕获的视频流可以被正常的信息呈现所代替。

[0283] 在教育场景中,相机可以被配置为指向常规白板或黑板,并且自动产生功能可以被配置为当教师朝向白板做出手势或通过语音命令时切换到这个相机。

[0284] 在音乐会或其他文化事件中,自动产生功能可以被配置为在指向歌手、乐队成员或管弦乐队的相机之间切换。这可以通过手势来完成,或者通过人正在看哪个相机来触发。

[0285] 在戏剧表演中,自动产生功能可以被配置为根据谁在说话、通过面部跟踪、音频提示、手势、或通过遵循预定时间表(纲要)在不同的相机角度之间切换。

[0286] 因此,除了上文描述的类型的触发检测之外,自动产生还可以基于预定时间表(纲要)从一个格式(产生规则)切换到另一格式,并且当然也可以手动超控,视情况而定。

[0287] 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可以被布置在运行时执行以上描述的收集、延迟、第一产生、触发检测、第二产生和发布步骤中的至少一些。计算机软件功能可以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。

[0288] 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130又可以被布置为执行以上描述的所述收集、延迟、第一产生、触发检测、第二产生和发布步骤中的至少一些。例如,这些步骤可以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。收集可以由收集功能131执行。以上描述的预定图像和/或音频模式和触发的检测可以由中央服务器130的事件和模式检测功能132、134或者产生功能135来执行。任何有意延时能够以以上总体描述的方式执行,诸如通过收集功能131。发布可以由发布功能136执行。

[0289] 应当理解的是,以上描述的基于输入视频流的可用集合的自动产生的原理(诸如涉及这种输入视频流的时间同步、事件和/或模式检测以及触发检测等)可以并发地在不同的级别下应用。因此,一个这样的自动产生的视频流可以形成到下游自动产生功能的可用输入视频流,该下游自动产生功能进而产生视频流。

[0290] 中央服务器130可以被布置来控制到各个参与客户端121的群组121'、121"、121"'分配。例如,在现场视频通信服务会话的过程期间动态地改变对特定的这种参与客户端的群组分配可以是由中央服务器130自动产生所述视频通信服务的一部分。这种重新分配可以基于预定的时间表或动态地触发,例如作为可以随时间动态变化的参数数据的函数,例如在各个参与者客户端用户122的请求(经由所讨论的客户端121提供)时。

[0291] 相应地,中央服务器130可以被布置为在视频通信服务的过程期间动态地改变群组结构,诸如仅在预定的时间时隙内使用某个群组(诸如在计划的小组辩论期间)。

[0292] 以预定方式改变群组分配可以是以对应于上面结合图9至图11b描述的内容的方式进行的特定触发的检测的自动结果。

[0293] 在所有上述方面中,本发明还可以包括交互步骤,在该交互步骤中,第一群组的至少一个参与客户端以双向(双向性)方式与相同的第一群组的或第二群组的至少一个参与客户端进行交互,第一群组与第一延迟相关联,第二群组与第二延迟相关联,第二延迟不同于第一延迟。应当理解的是,这些参与客户端可以全部是上述类型的同一个通信服务的参与者。

[0294] 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。

[0295] 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未进行描述。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。

[0296] 总体而言,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。

[0297] 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进行变化。

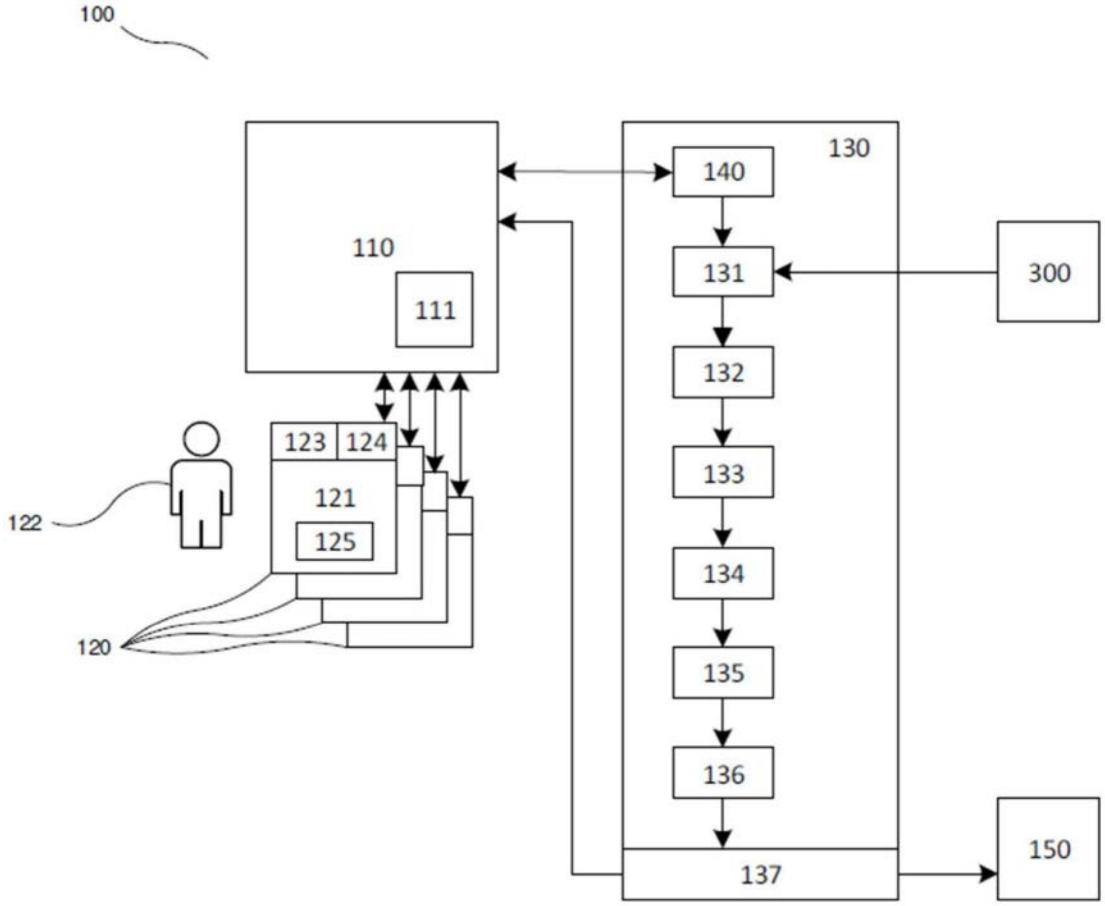

100
110
111
130
140
131
300
132
133
134
135
136
137
150
123
124
121
125
122
120

图1

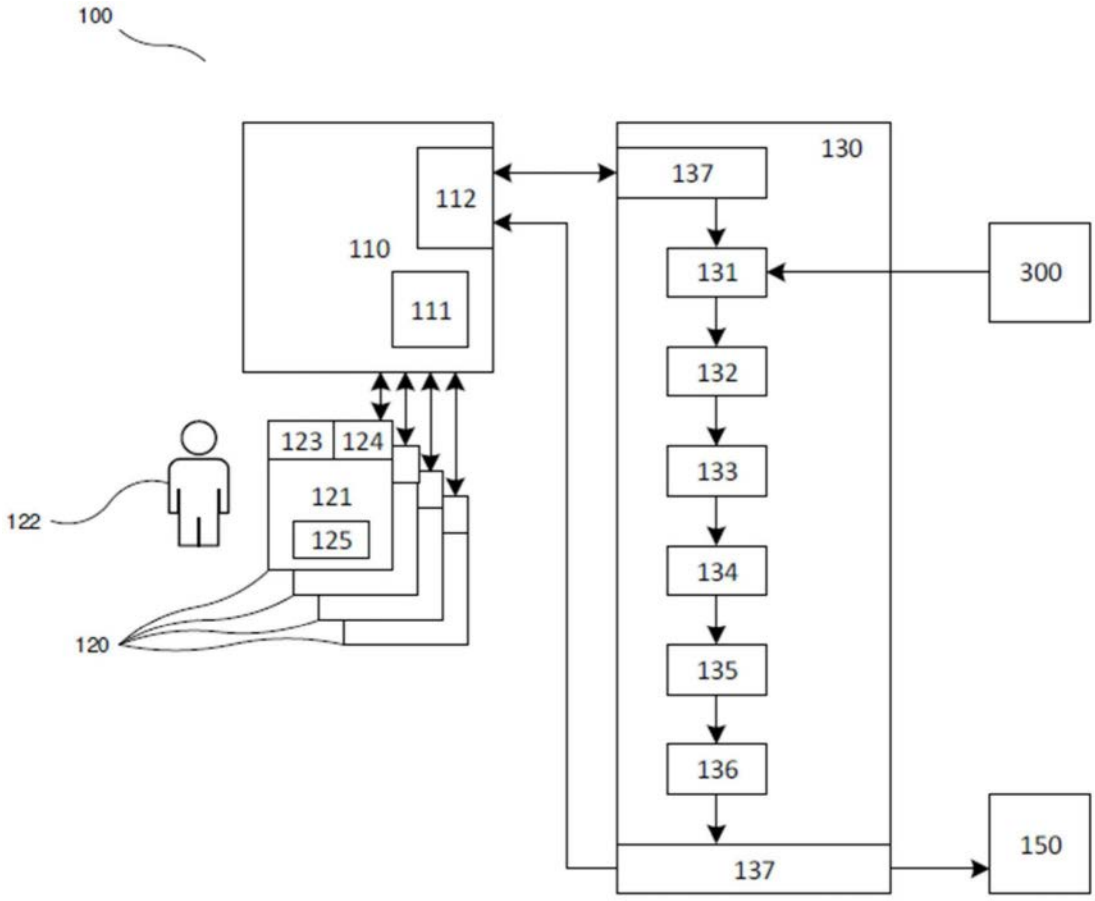

100
112
110
111
123
124
121
125
122
120
130
137
131
132
133
134
135
136
137
300
150

图2

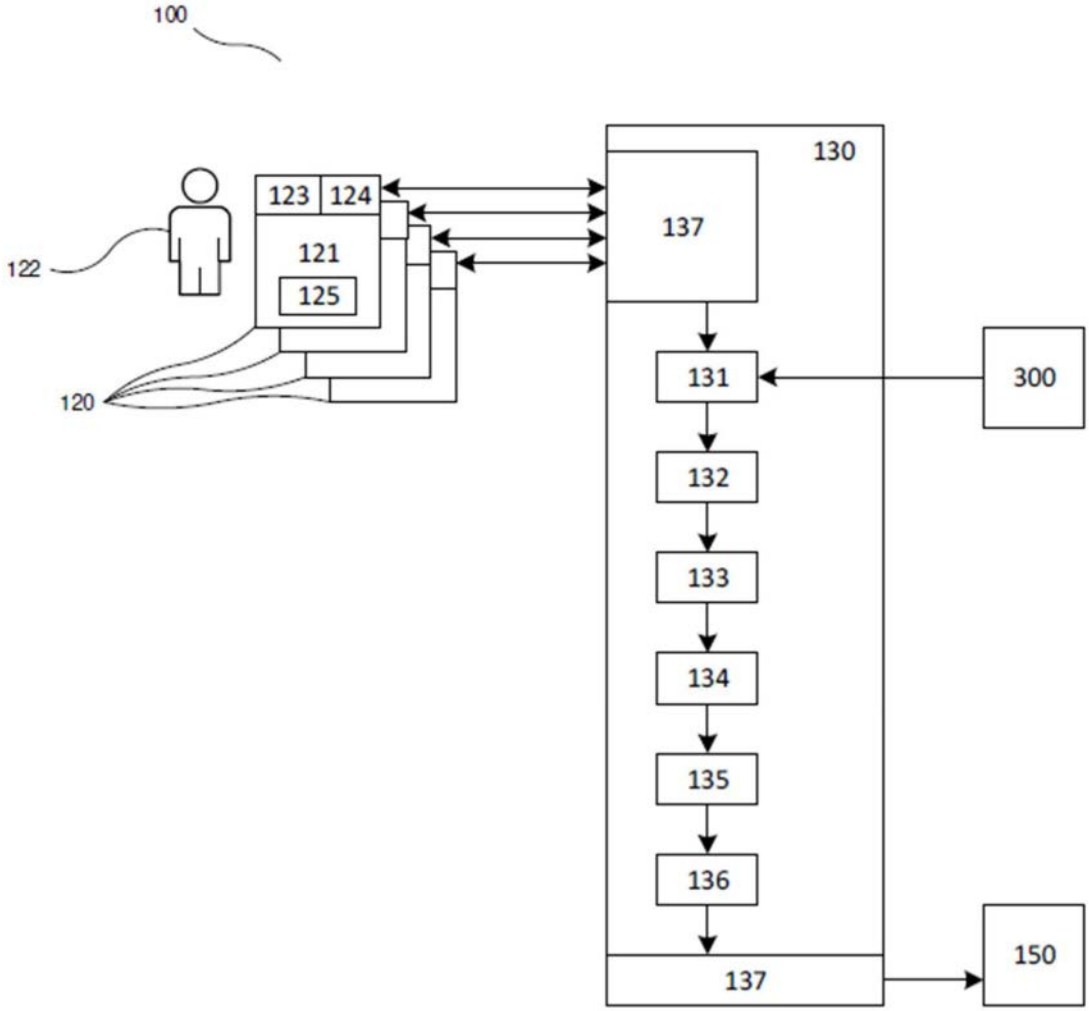

100
130
137
123
124
121
125
122
120
131
300
132
133
134
135
136
137
150

图3

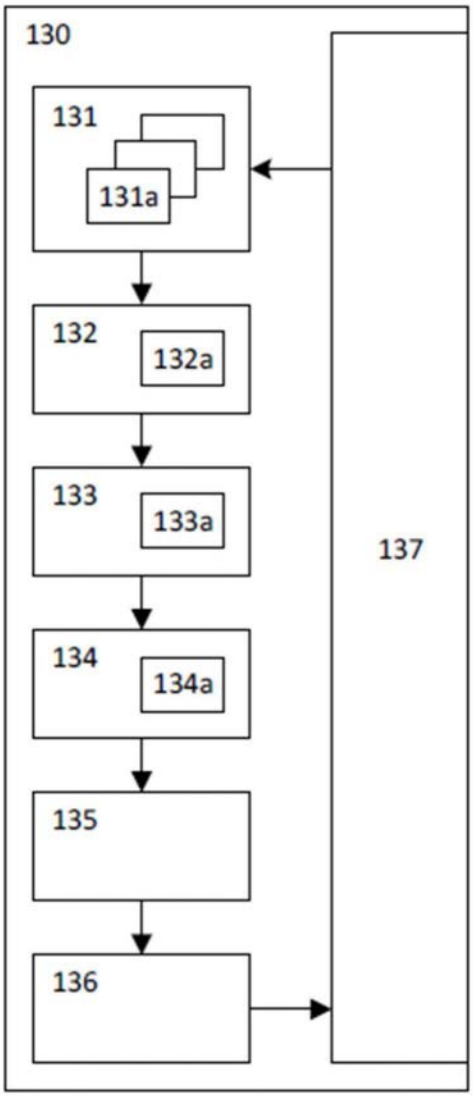


图4

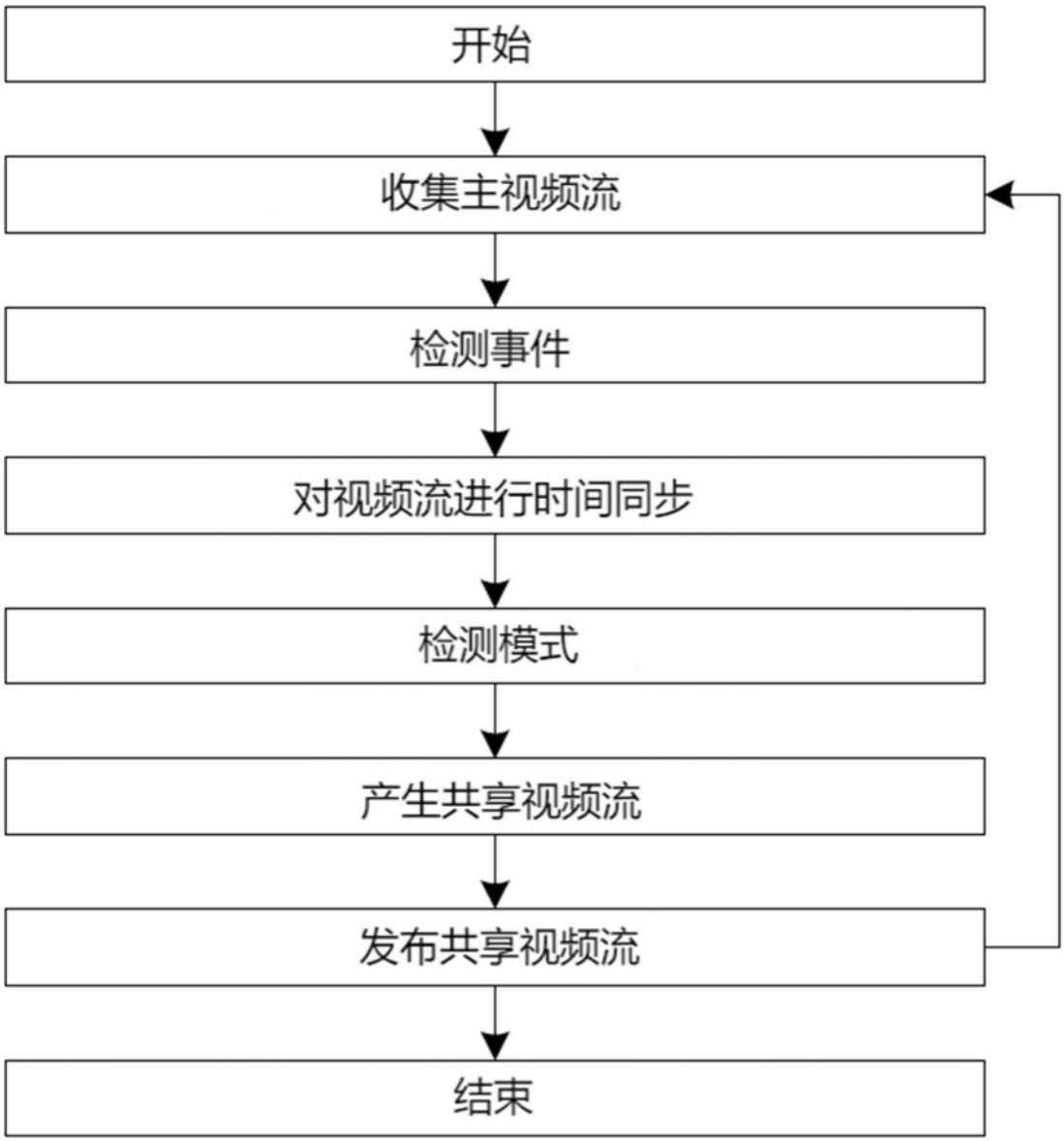


图5

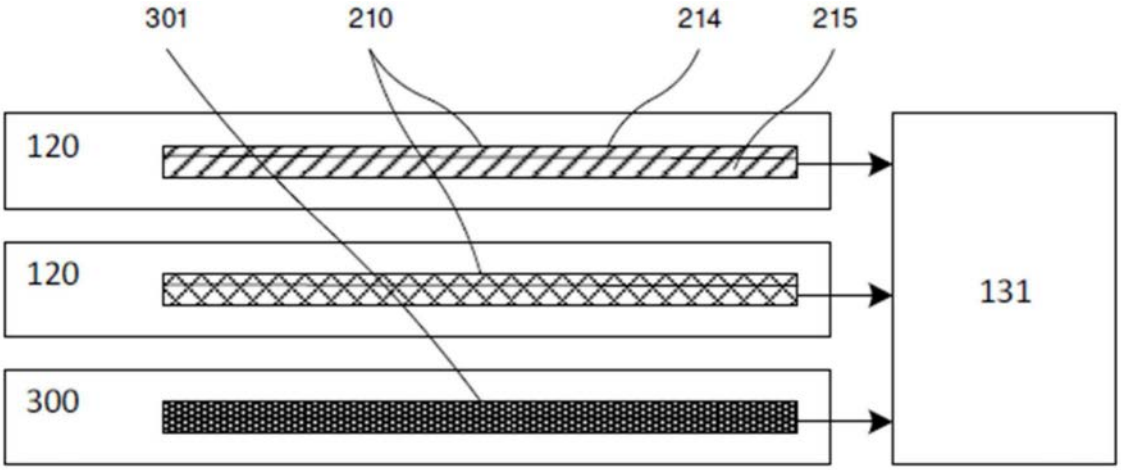


图6a

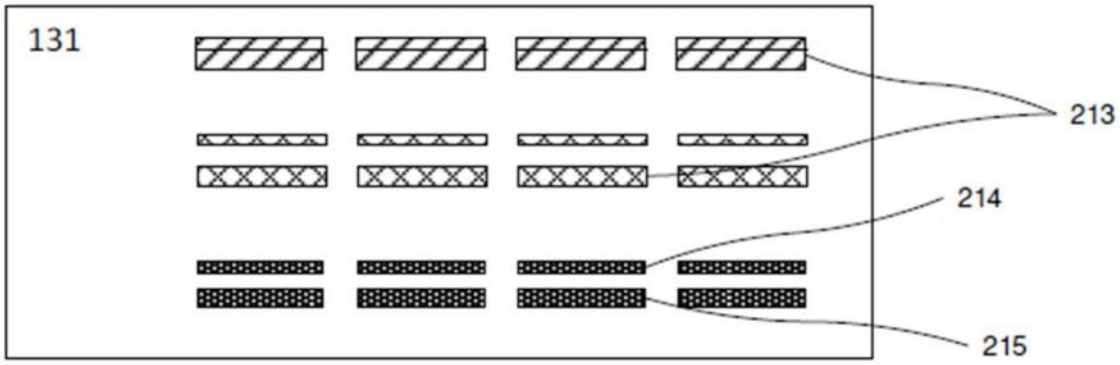


图6b

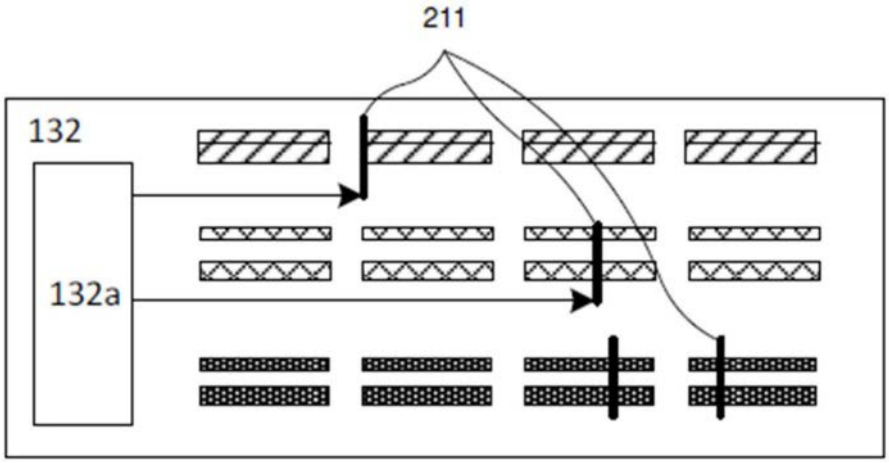


图6c

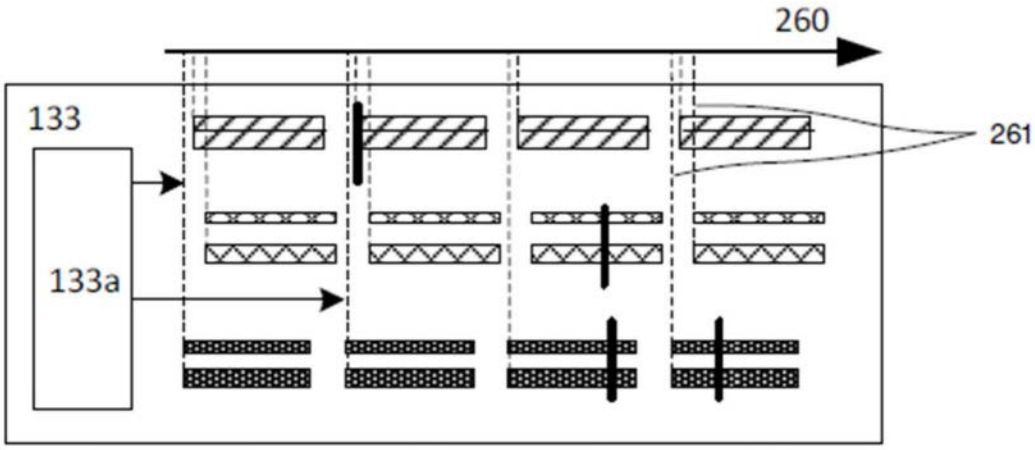


图6d

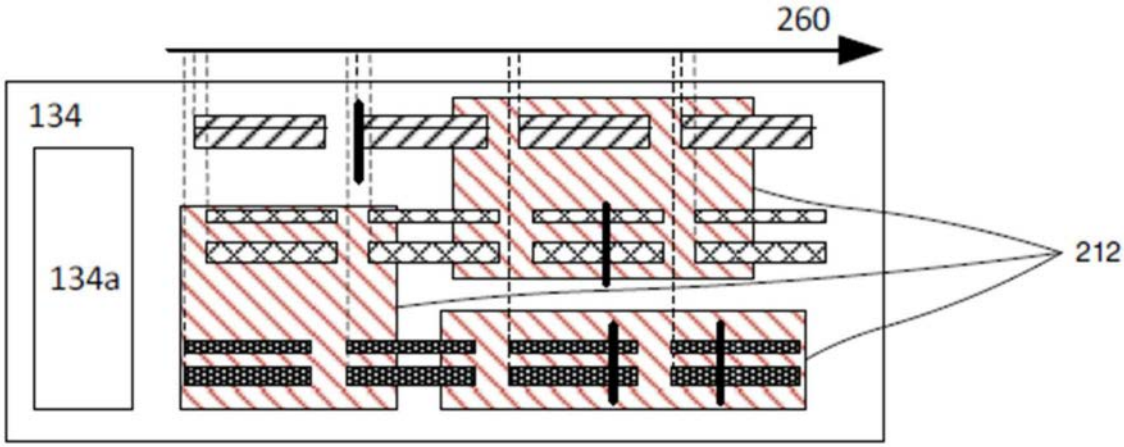


图6e

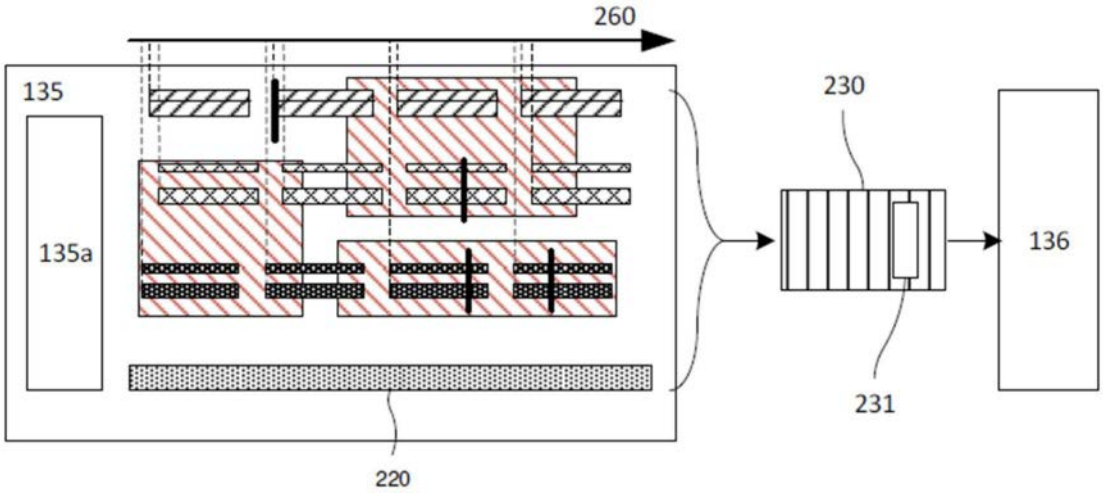


图6f

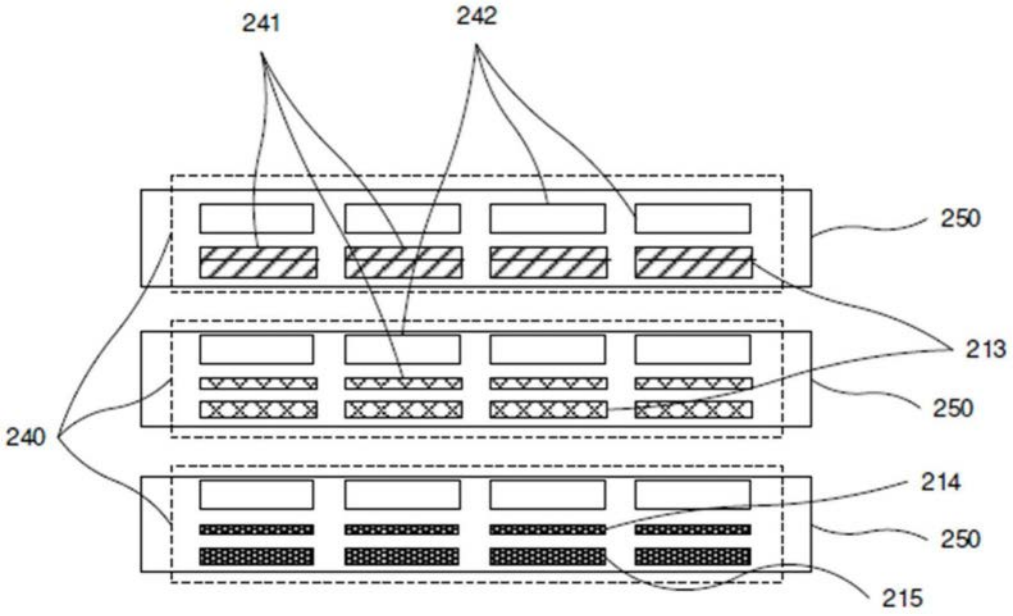

241
242
250
213
250
240
214
250
215

图7

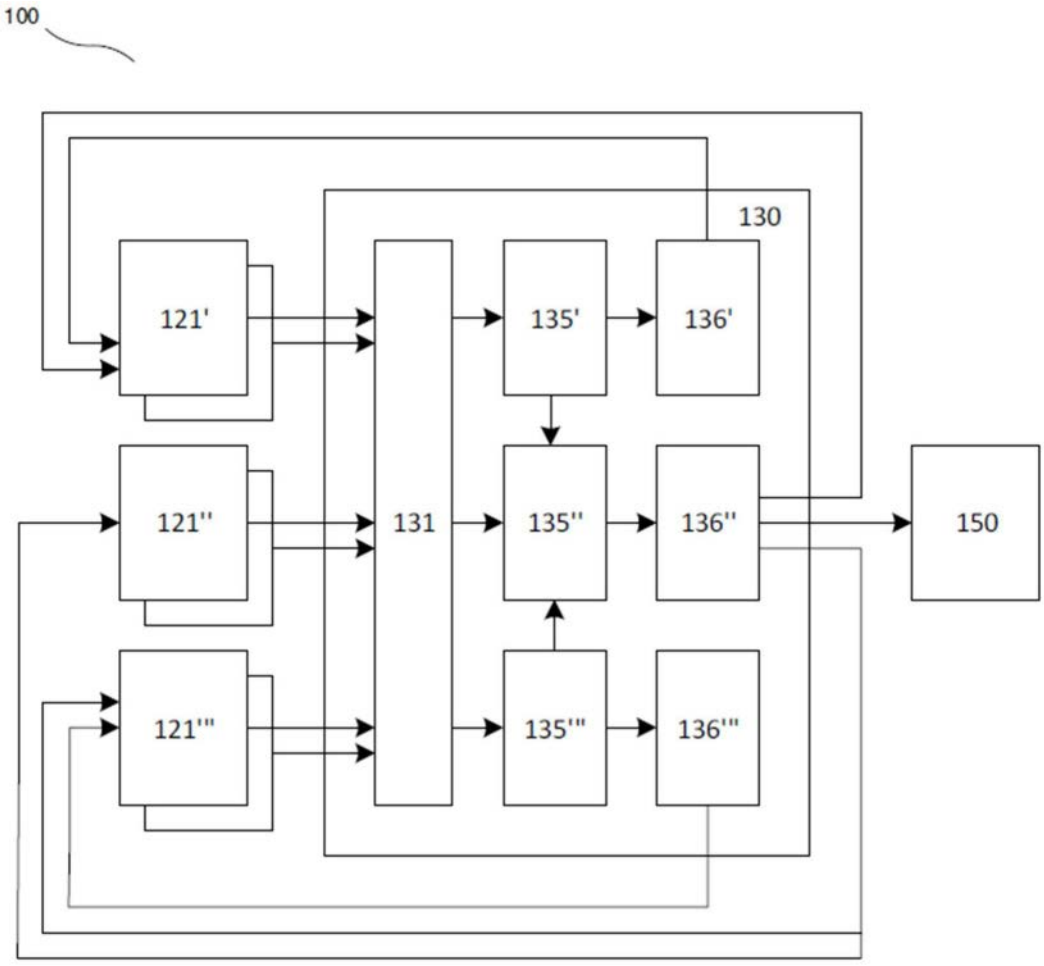

100
130
121'
135'
136'
121''
131
135''
136''
150
121'''
135'''
136'''

图8

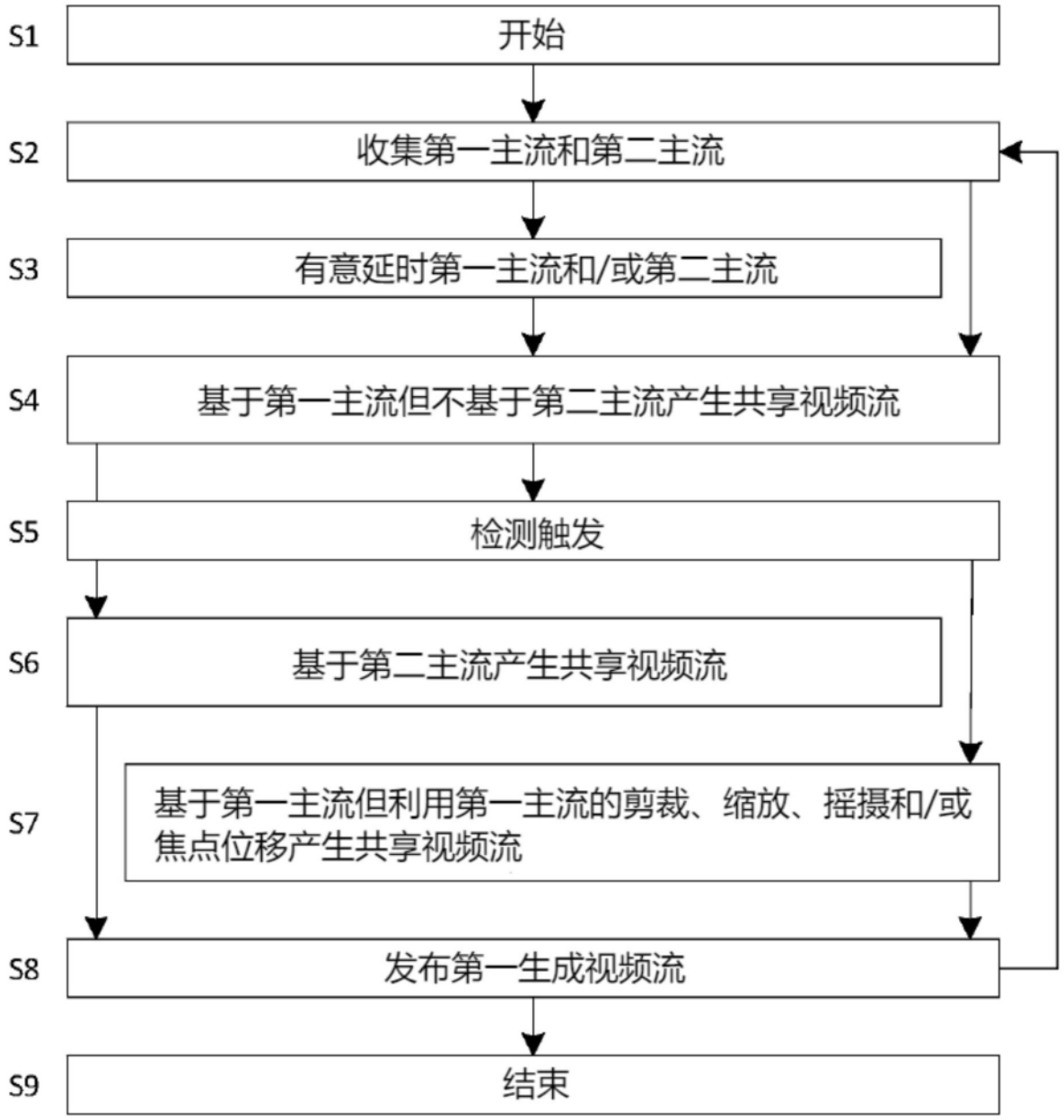

S1
开始
S2
收集第一主流和第二主流
S3
有意延时第一主流和/或第二主流
S4
基于第一主流但不基于第二主流产生共享视频流
S5
检测触发
S6
基于第二主流产生共享视频流
S7
基于第一主流但利用第一主流的剪裁、缩放、摇摄和/或焦点位移产生共享视频流
S8
发布第一生成视频流
S9
结束

图9

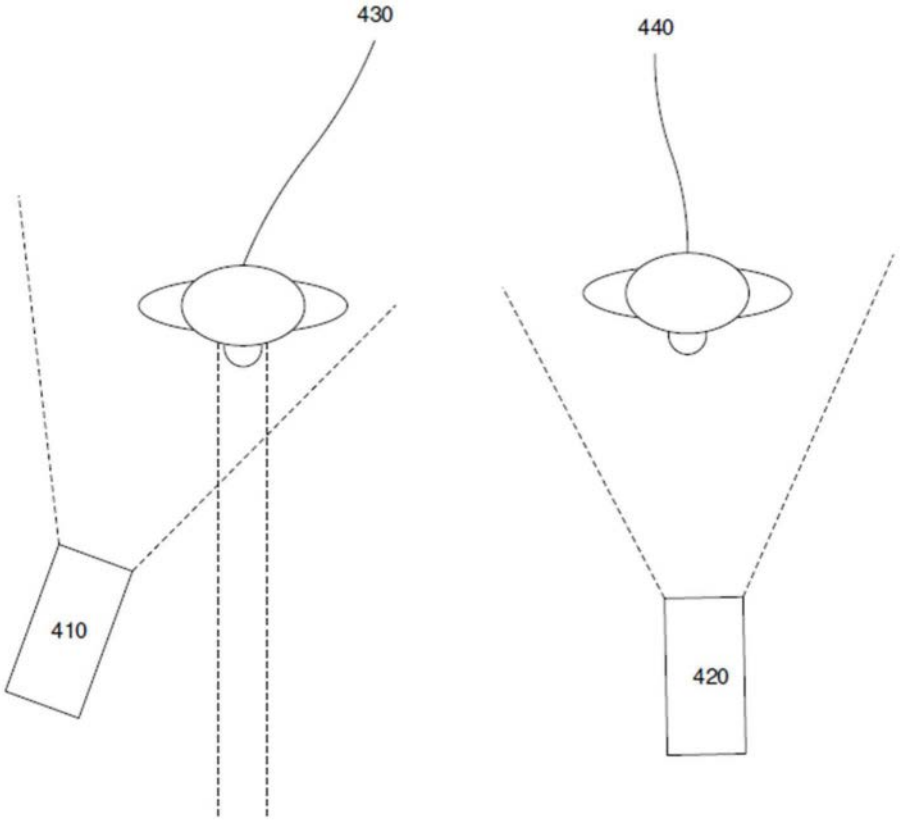


图10a

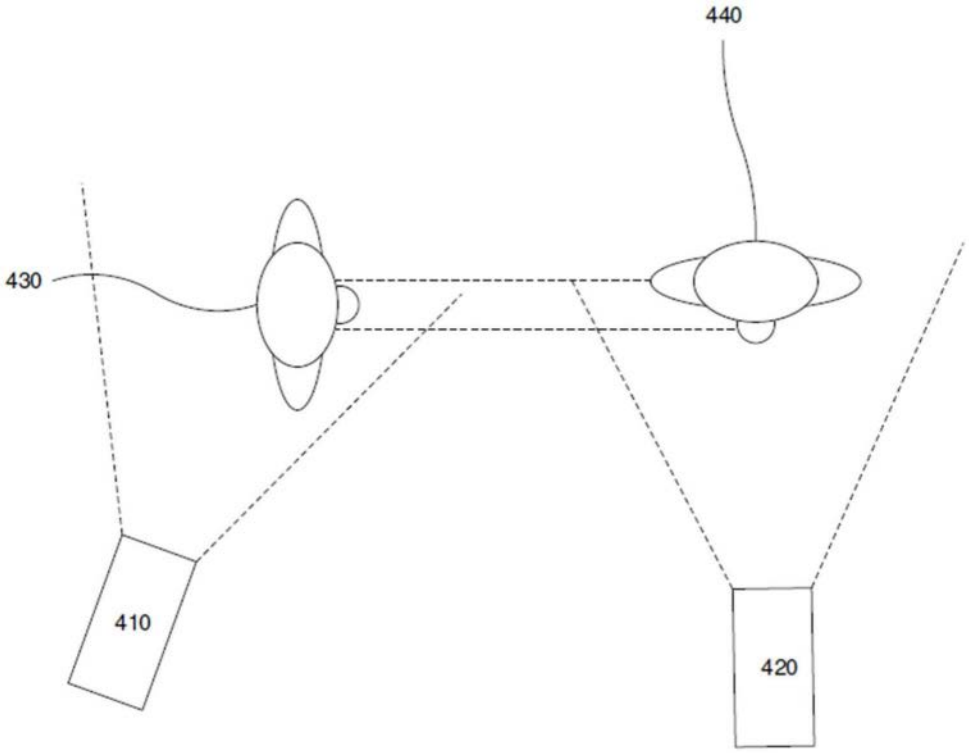


图10b

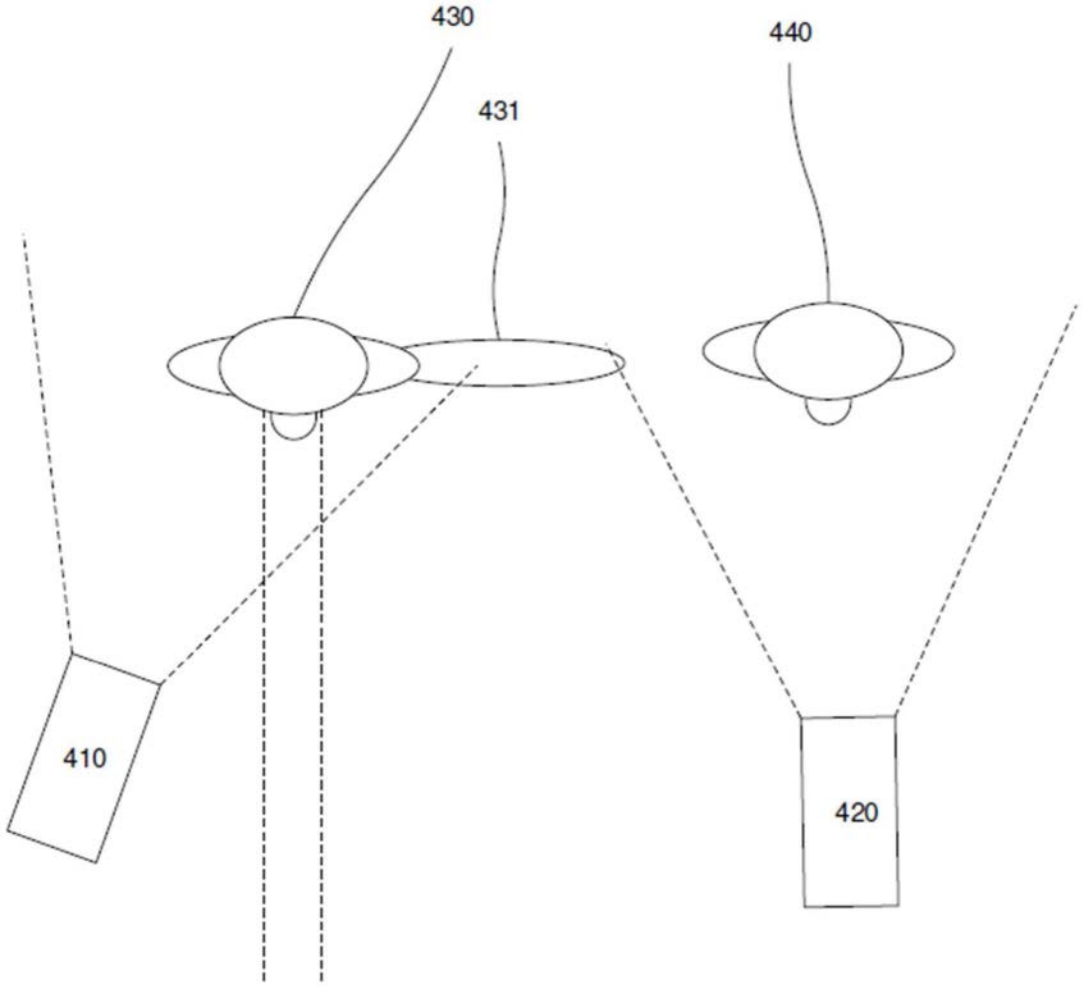


图10c

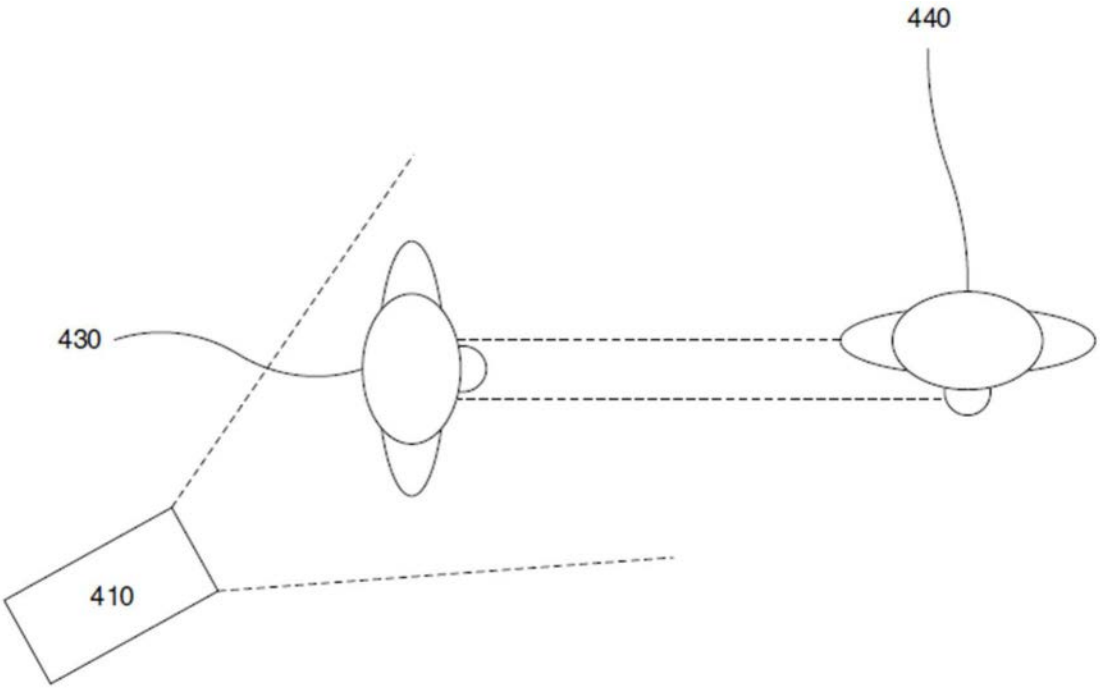


图11a

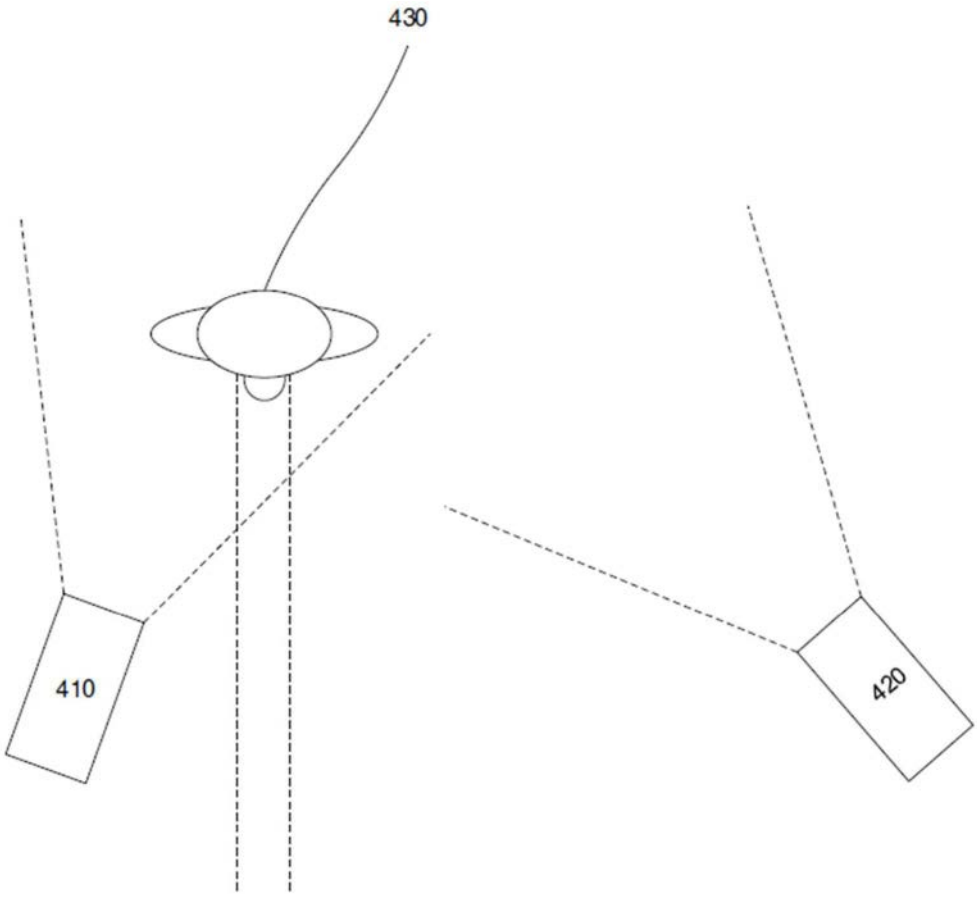


图11b