



(12) 发明专利

(10) 授权公告号 CN 112153321 B

(45) 授权公告日 2022.04.05

(21) 申请号 201910577597.9

(22) 申请日 2019.06.28

(65) 同一申请的已公布的文献号  
申请公布号 CN 112153321 A

(43) 申请公布日 2020.12.29

(73) 专利权人 华为技术有限公司  
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72) 发明人 庄松海

(74) 专利代理机构 深圳市深佳知识产权代理事务所(普通合伙) 44285  
代理人 骆苏华

(51) Int. Cl.  
H04N 7/15 (2006.01)  
H04N 5/76 (2006.01)

(56) 对比文件

- CN 102055949 A, 2011.05.11
- CN 105100892 A, 2015.11.25
- CN 103475835 A, 2013.12.25
- CN 101771853 A, 2010.07.07
- CN 101662643 A, 2010.03.03
- CN 105141879 A, 2015.12.09
- CN 105512348 A, 2016.04.20
- CN 104767963 A, 2015.07.08
- CN 109597898 A, 2019.04.09
- CN 108305632 A, 2018.07.20

审查员 黄文波

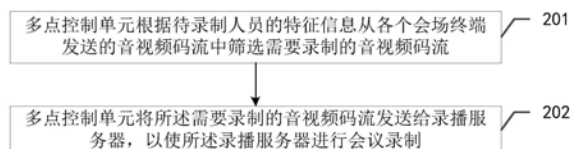
权利要求书4页 说明书16页 附图5页

(54) 发明名称

一种会议录制方法、装置及会议录制系统

(57) 摘要

本申请公开了一种会议录制方法、装置及会议录制系统。本申请方法包括：多点控制单元根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流；特征信息包括：图像信息或声音信息；多点控制单元将需要录制的音视频码流发送给录播服务器，以使录播服务器进行会议录制。由于待录制人员的特征信息(包括图像信息或声音信息)区别于其他参会人员的特征信息，该方法利用待录制人员的特征信息，实现对需要录制的音视频码流的自动筛选，从而无需人工筛选，节省会议录制的人工成本，同时大大提升会议录制效率。该方法提升会议录制的便捷性，促进视频会议功能的广泛应用。



1. 一种会议录制方法,其特征在于,包括以下步骤:

多点控制单元根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流;所述特征信息包括:图像信息或声音信息;

所述多点控制单元将所述需要录制的音视频码流发送给录播服务器,以使所述录播服务器进行会议录制;

当所述需要录制的音视频码流为所述待录制人员所在会场的整个音视频码流时,所述多点控制单元根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流,具体包括:

所述多点控制单元根据所述待录制人员的特征信息和所述各个会场终端发送的音视频码流筛选所述待录制人员对应的会场终端;

将筛选出的会场终端发送的音视频码流全部作为所述需要录制的音视频码流;

当所述需要录制的音视频码流为所述待录制人员个人的音视频码流时,所述多点控制单元根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流,具体包括:

所述多点控制单元根据所述待录制人员的特征信息和所述各个会场终端发送的音视频码流筛选所述待录制人员对应的会场终端;

从筛选出的会场终端发送的音视频码流中根据所述待录制人员的特征信息筛选所述待录制人员个人的音视频码流作为所述需要录制的音视频码流。

2. 根据权利要求1所述的会议录制方法,其特征在于,所述多点控制单元根据所述待录制人员的特征信息和所述各个会场终端发送的音视频码流筛选所述待录制人员对应的会场终端,具体包括:

所述多点控制单元将所述各个会场终端发送的音视频码流进行解码获得解码后的视频码流和音频码流;

所述多点控制单元根据所述待录制人员的图像信息与所述解码后的视频码流进行特征匹配,确定所述待录制人员对应的会场终端,或,将所述待录制人员的声音信息与所述解码后的音频码流进行特征匹配,确定所述待录制人员对应的会场终端。

3. 根据权利要求1或2所述的会议录制方法,其特征在于,所述会议终端为高级视频编码AVC会场终端,当所述待录制人员对应至少两个不同的AVC会场终端时,所述多点控制单元根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流,具体包括:

所述多点控制单元根据待录制人员的特征信息分别从所述至少两个不同的AVC会场终端发送的音视频码流中筛选需要录制的音视频码流。

4. 根据权利要求3所述的会议录制方法,其特征在于,所述需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流;所述多点控制单元将所述需要录制的音视频码流发送给录播服务器,具体包括:

所述多点控制单元将所述需要录制的视频码流进行画面合成获得合成画面,将所述合成画面发送给所述录播服务器,将所述需要录制的音频码流进行混音后发送给所述录播服务器。

5. 根据权利要求1或2所述的会议录制方法,其特征在于,所述会议终端为可伸缩视频

编码SVC会场终端,当所述待录制人员对应至少两个不同的SVC会场终端时,所述多点控制单元根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流,具体包括:

所述多点控制单元将适用于所述录播服务器的码流格式通知所有所述SVC会场终端;

所述多点控制单元接收所述至少两个不同的SVC会场终端发送的适用于所述录播服务器的码流格式的音视频码流;

所述多点控制单元根据待录制人员的特征信息从所述适用于所述录播服务器的码流格式的音视频码流中筛选需要录制的音视频码流。

6. 根据权利要求5所述的会议录制方法,其特征在于,所述需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流;所述多点控制单元将所述待录制人员需要录制的音视频码流发送给所述录播服务器,具体包括:

所述多点控制单元将所述至少两个不同的SVC会场终端对应的所述需要录制的视频码流发送给所述录播服务器,以使所述录播服务器对所述至少两个不同的SVC会场终端对应的所述需要录制的视频码流进行画面合成获得合成画面;

将所述至少两个不同的SVC会场终端对应的所述需要录制的音频码流进行混音后发送给所述录播服务器。

7. 根据权利要求1或2所述的会议录制方法,其特征在于,所述多点控制单元根据待录制人员的特征信息从所述音视频码流中筛选需要录制的音视频码流,具体包括:

所述多点控制单元根据所述待录制人员的特征信息利用预先训练的神经网络模型从所述音视频码流中筛选需要录制的音视频码流。

8. 一种会议录制装置,其特征在于,所述装置包括:

码流筛选模块,用于根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流;

码流发送模块,用于将所述需要录制的音视频码流发送给录播服务器,以使所述录播服务器进行会议录制;所述特征信息包括:图像信息或声音信息;

当所述需要录制的音视频码流为所述待录制人员所在会场的整个音视频码流时,所述码流筛选模块,具体包括:

会场终端筛选单元,用于根据所述待录制人员的特征信息和所述各个会场终端发送的音视频码流筛选所述待录制人员对应的会场终端;

码流第一筛选单元,用于将筛选出的会场终端发送的音视频码流全部作为所述需要录制的音视频码流;

当所述需要录制的音视频码流为所述待录制人员个人的音视频码流时,所述码流筛选模块,具体包括:

会场终端筛选单元,用于根据所述待录制人员的特征信息和所述各个会场终端发送的音视频码流筛选所述待录制人员对应的会场终端;

码流第二筛选单元,用于从筛选出的会场终端发送的音视频码流中根据所述待录制人员的特征信息筛选所述待录制人员个人的音视频码流作为所述需要录制的音视频码流。

9. 根据权利要求8所述的会议录制装置,其特征在于,所述会场终端筛选单元,具体包括:

解码子单元,用于将所述各个会场终端发送的音视频码流进行解码获得解码后的视频码流和音频码流;

会场终端确定子单元,用于根据所述待录制人员的图像信息与所述解码后的视频码流进行特征匹配,确定所述待录制人员对应的会场终端,或,将所述待录制人员的声音信息与所述解码后的音频码流进行特征匹配,确定所述待录制人员对应的会场终端。

10. 根据权利要求8或9所述的会议录制装置,其特征在于,所述会议终端为高级视频编码AVC会场终端,当所述待录制人员对应至少两个不同的AVC会场终端时,所述码流筛选模块,具体包括:

码流第三筛选单元,用于根据待录制人员的特征信息分别从所述至少两个不同的AVC会场终端发送的音视频码流中筛选需要录制的音视频码流。

11. 根据权利要求10所述的会议录制装置,其特征在于,所述需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流;所述码流发送模块,具体包括:

画面合成单元,用于将所述需要录制的视频码流进行画面合成获得合成画面;

画面发送单元,用于将所述合成画面发送给所述录播服务器;

第一混音单元,用于将所述需要录制的音频码流进行混音;

音频第一发送单元,用于将混音后的音频发送给所述录播服务器。

12. 根据权利要求8或9所述的会议录制装置,其特征在于,所述会议终端为可伸缩视频编码SVC会场终端,当所述待录制人员对应至少两个不同的SVC会场终端时,所述码流筛选模块,具体包括:

码流格式通知单元,用于将适用于所述录播服务器的码流格式通知所有所述SVC会场终端;

码流接收单元,用于接收所述至少两个不同的SVC会场终端发送的适用于所述录播服务器的码流格式的音视频码流;

码流第四筛选单元,用于根据待录制人员的特征信息从所述适用于所述录播服务器的码流格式的音视频码流中筛选需要录制的音视频码流。

13. 根据权利要求12所述的会议录制装置,其特征在于,所述需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流;所述码流发送模块,具体包括:

视频码流发送单元,用于将所述至少两个不同的SVC会场终端对应的所述需要录制的视频码流发送给所述录播服务器,以使所述录播服务器对所述至少两个不同的SVC会场终端对应的所述需要录制的视频码流进行画面合成获得合成画面;

第二混音单元,用于将所述至少两个不同的SVC会场终端对应的所述需要录制的音频码流进行混音;

音频第二发送单元,用于将混音后的音频发送给所述录播服务器。

14. 根据权利要求8或9所述的会议录制装置,其特征在于,所述码流筛选模块,具体包括:

码流第五筛选单元,用于根据所述待录制人员的特征信息利用预先训练的神经网络模型从所述音视频码流中筛选需要录制的音视频码流。

15. 一种会议录制系统,其特征在于,包括多点控制单元,录播服务器,以及至少两个会场终端;

所述会场终端,用于向所述多点控制单元发送音视频码流;

所述多点控制单元,用于根据待录制人员的特征信息从各个所述会场终端发送的音视频码流中筛选需要录制的音视频码流;所述特征信息包括:图像信息或声音信息;将所述需要录制的音视频码流发送给所述录播服务器;

所述录播服务器,用于根据所述需要录制的音视频码流进行会议录制;

当所述需要录制的音视频码流为所述待录制人员所在会场的整个音视频码流时,所述多点控制单元具体用于:根据所述待录制人员的特征信息和所述各个会场终端发送的音视频码流筛选所述待录制人员对应的会场终端;将筛选出的会场终端发送的音视频码流全部作为所述需要录制的音视频码流;

当所述需要录制的音视频码流为所述待录制人员个人的音视频码流时,所述多点控制单元具体用于:根据所述待录制人员的特征信息和所述各个会场终端发送的音视频码流筛选所述待录制人员对应的会场终端;从筛选出的会场终端发送的音视频码流中根据所述待录制人员的特征信息筛选所述待录制人员个人的音视频码流作为所述需要录制的音视频码流。

## 一种会议录制方法、装置及会议录制系统

### 技术领域

[0001] 本申请涉及多媒体技术领域,尤其涉及一种会议录制方法、装置及会议录制系统。

### 背景技术

[0002] 会议录制是多媒体技术领域的一个重要功能,通过录播服务器将多点控制单元(multipoint control unit,MCU)召开的会议内容录制下来,以用于会议转播、重播、后期制作等。随着视频会议越来越广泛的应用,在视频会议时进行会议录制也成为一种频繁且重要的需求。

[0003] 目前,会议录制一般是录制整个会议的音视频,但是后期需要查找某一个人的发言或者画面时,需要人工进行浏览和剪辑。这样的方式造成人力成本较高,同时,录制效率低下。

### 发明内容

[0004] 为了解决以上技术问题,本申请提供一种会议录制方法、装置及会议录制系统,能够自动筛选需要录制的音视频码流,提升会议录制的效率,节省人力成本。

[0005] 第一方面,本申请提供一种会议录制方法,包括以下步骤:

[0006] 首先,多点控制单元依据待录制人员的特征信息,从各个会场终端发送的音视频码流中筛选需要录制的音视频码流。其中,特征信息可以包括:图像信息或声音信息。其后,多点控制单元将上一步筛选出的需要录制的音视频码流发送给录播服务器,由录播服务器进行会议录制。

[0007] 因为每个人的特征信息互不相同,因此,利用待录制人员的特征信息所筛选出的需要录制的音视频码流与该待录制人员准确对应。因此,应用该方法利用待录制人员的特征信息,可以将需要录制的音视频码流自动地筛选出来以进行录制。省去了人工筛选的环节,因此会议录制人工成本降低。另外,自动筛选也有效提升会议录制效率。会议录制的便捷性更高,为会议录制功能的应用拓展提供高效的实现方案。

[0008] 在实际应用中,录制需求是多样化的。例如,录制需求可能是1)录制待录制人员所在会场的整个音视频码流,录制需求也可能是2)录制待录制人员个人的音视频码流。下面针对以上两种不同的录制需求,描述具体筛选方式。

[0009] 对于录制需求1),多点控制单元首先筛选待录制人员的会场终端,具体实现时,根据待录制人员的特征信息和各个会场终端发送的音视频码流筛选待录制人员对应的会场终端;其后,多点控制单元将筛选出的该会场终端发送的音视频码流,即整个音视频码流,全部作为需要录制的音视频码流。从而满足录制需求1)。

[0010] 对于录制需求2),多点控制单元也需要首先筛选待录制人员的会场终端,此步骤与需求1)中实现方式类似。其后,多点控制单元从筛选出的会场终端发送的音视频码流中根据待录制人员的特征信息筛选该人员个人的音视频码流。例如,在会议过程中,某一待录制人员共发言三次,每次存在一定的时间间隔,如果其特征信息为声音信息,则本申请中可

以具体筛选出其发言的三段音视频码流,以此作为需要录制的音视频码流。从而满足录制需求2)。

[0011] 可见,本申请提供的会议录制方法能够满足多种录制需求,因此适用性较强,满足会议录制场景下的多种录制应用。尤其对于录制需求2) 无需人工进行剪辑,提升所录制会议后续应用的便利性。

[0012] 在筛选会场终端时,具体可以采用以下方式进行筛选:

[0013] 首先,多点控制单元将各个会场终端发送的音视频码流解码,得到解码后的视频码流和音频码流;如果特征信息包括图像信息,则多点控制单元利用图像信息与解码后的视频码流进行特征匹配,确定待录制人员对应的会场终端;如果特征信息包括声音信息,则多点控制单元利用声音信息与解码后的音频码流进行特征匹配,确定待录制人员对应的会场终端。

[0014] 可见,通过特征信息能够唯一且准确地确定待录制人员对应的会场终端,即其所在的会场配置的会场终端。因此可知待录制人员一定不在其他会场终端,这些其他会场终端发送的音视频码流即可高效过滤,减轻多点控制单元的分析与处理负担。

[0015] 实际应用中,会议场景下的多个会议终端可以统一为高级视频编码AVC会场终端,也可以统一为可伸缩视频编码SVC会场终端。下面针对不同的会议终端,对本申请方法的具体实现进行描述。

[0016] 可选地,会议终端均为AVC会场终端,当待录制人员对应至少两个不同的AVC会场终端时,多点控制单元筛选音视频码流,首先根据待录制人员的特征信息分别从至少两个不同的AVC会场终端发送的音视频码流中筛选需要录制的音视频码流。若需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流,则多点控制单元将需要录制的视频码流进行画面合成获得合成画面,将合成画面发送给录播服务器,将需要录制的音频码流进行混音后发送给录播服务器。

[0017] 当需要录制的音视频码流为多个待录制人员个人的音视频码流时,利用本实施例提供的会议录制方法由多点控制单元进行画面合成和混音并发送给录播服务器,大大减少网络带宽,并且节省了录播服务器的存储空间。此外,按照指定的待录制人员进行实时录制,避免人工后期制作的音频码流和视频码流的裁剪过程,节约人力成本,提高会议录制的效率。

[0018] 可选地,会议终端为SVC会场终端,当待录制人员对应至少两个不同的SVC会场终端时,多点控制单元筛选音视频码流,首先将适用于录播服务器的码流格式通知所有SVC会场终端;其后接收至少两个不同的SVC会场终端发送的适用于录播服务器的码流格式的音视频码流;最后根据待录制人员的特征信息从适用于录播服务器的码流格式的音视频码流中筛选需要录制的音视频码流。也就是说,筛选出的音视频码流的码流格式对于录播服务器来说,是可接收,可处理的。若需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流,则多点控制单元将至少两个不同的SVC会场终端对应的需要录制的视频码流发送给录播服务器,从而,由录播服务器对至少两个不同的SVC会场终端对应的需要录制的视频码流进行画面合成获得合成画面。混音则由多点控制单元进行,其将至少两个不同的SVC会场终端对应的需要录制的音频码流进行混音后发送给录播服务器。

[0019] 当需要录制的音视频码流为多个待录制人员个人的音视频码流时,利用本实施例

提供的会议录制方法由多点控制单元进行混音并发送给录播服务器,由录播服务器对MCU筛选的对于多个待录制人员的需要录制的视频码流进行画面合成,最终进行会议录制,大大减少网络带宽,并且节省了录播服务器的存储空间。此外,本实施例按照指定的待录制人员进行实时录制,避免人工后期制作的音频码流和视频码流的裁剪过程,节约人力成本,提高会议录制的效率。

[0020] 可选地,多点控制单元根据待录制人员的特征信息从音视频码流中筛选需要录制的音视频码流,具体包括:

[0021] 多点控制单元根据待录制人员的特征信息利用预先训练的神经网络模型从音视频码流中筛选需要录制的音视频码流。

[0022] 利用神经网络模型进行音视频码流的筛选,提升了音视频码流的筛选效率,进而提升会议录制过程的整体速度。提升用户的会议录制体验。

[0023] 第二方面,本申请提供一种会议录制装置,装置包括:码流筛选模块和码流发送模块。其中码流筛选模块依据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流;码流发送模块将需要录制的音视频码流发送给录播服务器,以使录播服务器进行会议录制。

[0024] 该装置利用特征信息准确筛选出需要录制的音视频码流,该音视频码流与待录制人员相互匹配,实现音视频码流的自动筛选。相比于人工筛选和录制,效率大大提升,节省人力成本。

[0025] 可选地,当需要录制的音视频码流为待录制人员所在会场的整个音视频码流时,码流筛选模块,具体包括:

[0026] 会场终端筛选单元,用于根据待录制人员的特征信息和各个会场终端发送的音视频码流筛选待录制人员对应的会场终端;

[0027] 码流第一筛选单元,用于将筛选出的会场终端发送的音视频码流全部作为需要录制的音视频码流。

[0028] 可选地,当需要录制的音视频码流为待录制人员个人的音视频码流时,码流筛选模块,具体包括:

[0029] 会场终端筛选单元,用于根据待录制人员的特征信息和各个会场终端发送的音视频码流筛选待录制人员对应的会场终端;

[0030] 码流第二筛选单元,用于从筛选出的会场终端发送的音视频码流中根据待录制人员的特征信息筛选待录制人员个人的音视频码流作为需要录制的音视频码流。

[0031] 可选地,会场终端筛选单元,具体包括:

[0032] 解码子单元,用于将各个会场终端发送的音视频码流进行解码获得解码后的视频码流和音频码流;

[0033] 会场终端确定子单元,用于根据待录制人员的图像信息与解码后的视频码流进行特征匹配,确定待录制人员对应的会场终端,或,将待录制人员的声音信息与解码后的音频码流进行特征匹配,确定待录制人员对应的会场终端。

[0034] 可选地,会议终端为高级视频编码AVC会场终端,当待录制人员对应至少两个不同的AVC会场终端时,码流筛选模块,具体包括:

[0035] 码流第三筛选单元,用于根据待录制人员的特征信息分别从至少两个不同的AVC



会场终端发送的音视频码流中筛选需要录制的音视频码流。

[0036] 可选地,需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流;码流发送模块,具体包括:

[0037] 画面合成单元,用于将需要录制的视频码流进行画面合成获得合成画面;

[0038] 画面发送单元,用于将合成画面发送给录播服务器;

[0039] 第一混音单元,用于将需要录制的音频码流进行混音;

[0040] 音频第一发送单元,用于将混音后的音频发送给录播服务器。

[0041] 可选地,会议终端为可伸缩视频编码SVC会场终端,当待录制人员对应至少两个不同的SVC会场终端时,码流筛选模块,具体包括:

[0042] 码流格式通知单元,用于将适用于录播服务器的码流格式通知所有SVC会场终端;

[0043] 码流接收单元,用于接收至少两个不同的SVC会场终端发送的适用于录播服务器的码流格式的音视频码流;

[0044] 码流第四筛选单元,用于根据待录制人员的特征信息从适用于录播服务器的码流格式的音视频码流中筛选需要录制的音视频码流。

[0045] 可选地,需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流;码流发送模块,具体包括:

[0046] 视频码流发送单元,用于将至少两个不同的SVC会场终端对应的需要录制的视频码流发送给录播服务器,以使录播服务器对至少两个不同的SVC会场终端对应的需要录制的视频码流进行画面合成获得合成画面;

[0047] 第二混音单元,用于将至少两个不同的SVC会场终端对应的需要录制的音频码流进行混音;

[0048] 音频第二发送单元,用于将混音后的音频发送给录播服务器。

[0049] 可选地,码流筛选模块,具体包括:

[0050] 码流第五筛选单元,用于根据待录制人员的特征信息利用预先训练的神经网络模型从音视频码流中筛选需要录制的音视频码流。

[0051] 第三方面,本申请提供一种会议录制系统,包括多点控制单元,录播服务器,以及至少两个会场终端;

[0052] 会场终端,用于向多点控制单元发送音视频码流;

[0053] 多点控制单元,用于根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流;特征信息包括:图像信息或声音信息;将需要录制的音视频码流发送给录播服务器;

[0054] 录播服务器,用于根据需要录制的音视频码流进行会议录制。

[0055] 该系统利用特征信息准确地从会场终端提供的众多音视频码流中筛选出需要录制的音视频码流,该音视频码流与待录制人员相互匹配,实现音视频码流的自动筛选。相比于人工筛选和录制,效率大大提升,节省人力成本。

[0056] 从以上技术方案可以看出,本申请实施例具有以下优点:

[0057] 由于待录制人员的特征信息(包括图像信息或声音信息)区别于其他参会人员的特征信息,即待录制人员的特征信息是待录制人员唯一对应的,因此,多点控制单元MCU根据待录制人员的特征信息能够从各个会场发送的音视频码流中准确筛选出需要录制的音

视频码流。其后,MCU将需要录制的音视频码流发送给录播服务器,录播服务器即可根据从MCU接收的音视频码流进行会议录制。该方法利用待录制人员的特征信息,实现对需要录制的音视频码流的自动筛选,从而无需人工筛选,节省会议录制的人工成本,同时大大提升会议录制效率。该方法提升会议录制的便捷性,促进视频会议功能的广泛应用。

#### 附图说明

- [0058] 图1为本申请实施例提供的一种会议录制场景示意图;
- [0059] 图2为本申请实施例提供的一种会议录制方法的流程图;
- [0060] 图3为本实施例提供的一种多点控制单元获得需要录制的音视频码流的流程图;
- [0061] 图4为本申请实施例提供的另一种多点控制单元获得需要录制的音视频码流的流程图;
- [0062] 图5为本申请实施例提供的一种会议录制方法的信令图;
- [0063] 图6为本申请实施例提供的另一种会议录制方法的信令图;
- [0064] 图7为本申请实施例提供的一种会议录制装置的结构示意图;
- [0065] 图8为本申请实施例提供的一种会议录制系统的结构示意图;
- [0066] 图9为本申请实施例提供的另一种会议录制系统的结构示意图。

#### 具体实施方式

[0067] 当前视频会议进行会议录制时,往往需要人工查找某一人员所在的会议的音视频码流。假设该人员在会议末尾露面或发言,则为录制该人员参加的会议,需要人工地查找整段会议才能确定该人员参加了此会议。显然,当前的会议录制方法很大程度上依靠人工操作,不但消耗较高的人力成本,并且效率低下。

[0068] 基于此问题,发明人经过研究,提供一种会议录制方法、装置及会议录制系统。在本申请中,利用待录制人员的特征信息对多个会场的音视频码流进行筛选,由于每个人的特征信息(包括图像信息或声音信息)互不相同,因此,利用待录制人员的特征信息能够准确地识别出包含待录制人员的特征信息的音视频码流,从而实现音视频码流的自动筛选。MCU将需要录制的音视频码流筛出并发送给录播服务器,服务器即可根据从MCU接收的音视频码流进行会议录制。本申请提供的技术方案节省人力成本,同时提升会议录制效率。

[0069] 为便于理解本申请技术方案,下面结合附图对本申请提供的会议录制方法的应用场景进行描述和说明。参见图1,该图为本申请实施例提供的一种会议录制场景示意图。

[0070] 如图1所示,本申请提供的会议录制方法的应用场景中,包括:多点控制单元MCU,录播服务器,会议应用服务器(application server,AS),以及多个会场终端。在实际应用中,会场终端的数量可以是两个或两个以上,图1中仅以三个会场终端为示例,本实施例对于会议录制场景下会场终端的具体数量不进行限定。图1中三个会场终端分别为会场终端1、会场终端2和会场终端3,不同的会场终端分属于不同的会场。

[0071] MCU与录播服务器位于同一局域网内。在会议录制场景中,会议AS作为视讯业务管理平台,用户通过会议AS预约会议,提供或指定包含待录制人员的特征信息的材料。此处,特征信息可以包括图像信息或声音信息。也就是说,特征信息可以仅包括图像信息,可以仅包括声音信息,还可以既包括图像信息又包括声音信息。

[0072] 作为一示例,用户可以向会议AS上传包含待录制人员的特征信息的材料,例如待录制人员的图片或音频文件。会议AS对图片进行处理,得到待录制人员的图像信息,图像信息具体可以为图像特征,例如人脸特征;会议AS对音频文件进行处理,得到待录制人员的声音信息,声音信息具体可以为声纹特征。可以理解的是,待录制人员的特征信息区别于其他人员的特征信息,即待录制人员的特征信息是待录制人员唯一对应的,可由待录制人员的特征信息唯一地确定待录制人员。

[0073] 作为另一示例,会议AS中存储有多个人员各自对应的材料,例如各个人员的图片或音频文件,并且会议AS已经预先处理获得了每个人员对应的图像信息和声音信息。用户在会议AS中能够根据需求从多个待选的人员的材料中选择包含待录制人员特征信息的材料,例如,选中待录制人员的图片或音频文件。会议AS根据用户的选择可以确定用户指定的待录制人员。

[0074] 会议AS向多点控制单元MCU召集会议,并在召集时向MCU下发待录制人员的特征信息。由MCU呼叫会场终端1、会场终端2和会场终端3入会,各个终端入会后,向MCU发送音视频码流。MCU根据会议AS所下发的待录制人员的特征信息,从各个会场终端发送的音视频码流中筛选需要录制的音视频码流。

[0075] 例如,实际情况下待录制人员在会场终端2参加会议,因此会场终端2向MCU发送的音视频码流中,视频码流带有待录制人员的图像信息,而音频码流中带有待录制人员的声音信息。如果待录制人员的特征信息具体为图像信息,则MCU从各个会场终端发送的视频码流中能够通过图像信息的匹配,确定待录制人员在会场终端2参加会议,从而将会场终端2发送的音视频码流筛选出来,将其发送给录播服务器以使得录播服务器进行会议录制。如果待录制人员的特征信息具体为声音信息,则MCU从各个会场终端发送的音频码流中能够通过声音信息的匹配,确定待录制人员在会场终端2参加会议,从而将会场终端2发送的音视频码流筛选出来,将其发送给录播服务器以使得录播服务器进行会议录制。

[0076] 下面结合附图和实施例,对本申请提供的会议录制方法进行描述和说明。

[0077] 方法实施例一

[0078] 参见图2,该图为本实施例提供的会议录制方法的流程图。该应用于会议录制场景下的多点控制单元MCU。

[0079] 如图2所示,本实施例提供的会议录制方法包括:

[0080] 步骤201:多点控制单元根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流。

[0081] 在前文对该方法应用场景的描述中已经介绍,本实施例中,特征信息包括:图像信息或声音信息。也就是说,MCU能够单独根据待录制人员的图像信息筛选需要录制的音视频码流,能够单独根据待录制人员的声音信息筛选需要录制的音视频码流,同样也能够综合利用待录制人员的图像信息和声音信息筛选需要录制的音视频码流。可以理解的是,综合利用待录制人员的图像信息和声音信息筛选需要录制的音视频码流可以提升音视频码流筛选的准确性,降低筛选失误率。

[0082] 在实际应用中,待录制人员的数量可以为一个或多个。如果MCU从会议AS接收的待录制人员的特征信息属于同一待录制人员,则表示待录制人员的数量仅有一个;如果接收的待录制人员的特征信息属于多个不同的待录制人员,则表示待录制人员数量为多个。在

实际应用中,多个待录制人员可能位于同一个会场,即共同对应同一个会场终端,也可能位于不同的会场,即分别对应不同的会场终端。

[0083] 当待录制人员有多个时,实际录制需求可能为:录制多画面的会议,其中每个画面对应不同的待录制人员。

[0084] 当仅有一个待录制人员时,实际录制需求可能有多种,下面示例性地提供几种录制需求:

[0085] (1) 录制待录制人员参加的完整会议;

[0086] (2) 仅录制待录制人员发言时的会议;

[0087] (3) 录制出现待录制人员时的会议。

[0088] 可以理解的是,在实际应用中对于不同的录制需求,筛选出的需要录制的音视频码流也不同。作为一示例,对于录制需求(1),则筛选出的音视频码流具体为待录制人员所在会场在会议过程中的整个音视频码流。作为另一示例,对于录制需求(3),则筛选出的音视频码流具体为待录制人员发言时的音视频码流。

[0089] 作为一种可能的实现方式,本实施例中MCU可以利用预先训练的神经网络模型,根据待录制人员的特征信息从音视频码流中筛选需要录制的音视频码流。该神经网络模型为利用大量不同人员的特征信息以及包含不同人员特征信息的材料(例如待录制人员的图片或音频文件)的训练得到。对于本领域技术人员来说,训练能够准确识别带有某种图像信息的视频码流或能够准确识别带有某种声音信息的音频码流的神经网络模型,属于比较成熟的技术,因此本实施例对于该神经网络模型的具体训练过程不进行详述。

[0090] 步骤202:多点控制单元将需要录制的音视频码流发送给录播服务器,以使录播服务器进行会议录制。

[0091] 通过筛选,MCU已经得到需要录制的音视频码流。录播服务器是一种能够配合MCU和会场终端使用的服务器,能够将视频、音频和计算机屏幕信号等进行一体化同步录制,因此本步骤MCU将需要录制的音视频码流发送至录播服务器后,录播服务器即可根据需要录制的音视频码流进行会议录制。

[0092] 以上即为本申请实施例提供的会议录制方法。由于待录制人员的特征信息(包括图像信息或声音信息)区别于其他参会人员的特征信息,即待录制人员的特征信息是待录制人员唯一对应的,因此,多点控制单元MCU根据待录制人员的特征信息能够从各个会场发送的音视频码流中准确筛选出需要录制的音视频码流。其后,MCU将需要录制的音视频码流发送给录播服务器,录播服务器即可根据从MCU接收的音视频码流进行会议录制。该方法利用待录制人员的特征信息,实现对需要录制的音视频码流的自动筛选,从而无需人工筛选,节省会议录制的人工成本,同时大大提升会议录制效率。该方法提升会议录制的便捷性,促进视频会议功能的广泛应用。

[0093] 在实际应用中,上述示例录制需求中(1)表示需要录制的音视频码流为待录制人员所在会场的整个音视频码流;(2)和(3)表示需要录制的音视频码流为待录制人员个人的音视频码流。针对不同的需要录制的音视频码流,步骤201的实现方式存在差异。下面结合图3描述需要录制的音视频码流为待录制人员所在会场的整个音视频码流时,步骤201的详细流程;结合图4描述需要录制的音视频码流为待录制人员个人的音视频码流时,步骤201的详细流程。

[0094] 参见图3,该图为本实施例提供的一种多点控制单元获得需要录制的音视频码流的流程图。

[0095] 如图3所示,当需要录制的音视频码流为待录制人员所在会场的整个音视频码流时,多点控制单元获得需要录制的音视频码流具体包括:

[0096] 步骤301:多点控制单元将各个会场终端发送的音视频码流进行解码获得解码后的视频码流和音频码流。

[0097] 在本实施例中,会场终端向MCU发送的音视频码流具体可以为音视频实时传输协议(real-time transport protocol,RTP)码流。为便于后续将需要录制的音视频码流筛选出来,需要对音视频码流预先进行解码。对于本领域技术人员,对音视频码流解码属于比较成熟的技术,所以此处对于解码过程不进行详述。通过解码,MCU得到可单独处理的视频码流和音频码流。可以理解的是,由音视频码流解码得到的视频码流和音频码流存在时序联系。例如,会场终端1向MCU发送的T1时刻至T2时刻会场终端1所在会场的音视频码流,则MCU解码得到的是T1时刻至T2时刻会场终端1所在会场的视频码流和音频码流。

[0098] 步骤302:多点控制单元根据待录制人员的图像信息与解码后的视频码流进行特征匹配,确定待录制人员对应的会场终端,或,将待录制人员的声音信息与解码后的音频码流进行特征匹配,确定待录制人员对应的会场终端。

[0099] 可以理解的是,不同的会场所处地理位置不同,同一人员在会议召开时仅处于一个固定的会场,而不会出现在其他的会场。作为一示例,如果会场终端1属于待录制人员所在的会场,则发送带有该待录制人员的特征信息的音视频码流的会场终端必然是会场终端1,而不可能是其他的会场终端。因此,为筛选需要录制的音视频码流,本实施例中只需确定传输带有待录制人员的特征信息的会场终端。

[0100] 如果MCU从会议AS接收的待录制人员的特征信息仅包括图像信息,则本步骤中,由MCU根据待录制人员的图像信息与解码后的视频码流进行特征匹配,确定待录制人员对应的会场终端。如果MCU从会议AS接收的待录制人员的特征信息仅包括声音信息,则本步骤中,由MCU根据待录制人员的声音信息与解码后的音频码流进行特征匹配,确定待录制人员对应的会场终端。可以理解的是,如果MCU从会议AS接收的待录制人员的特征信息既包括图像信息又包括声音信息,则综合利用图像信息和声音信息匹配录制人员对应的会场终端,能够提升匹配结果的准确性和可信度,降低失误率。

[0101] 以上步骤301-302实现多点控制器对待录制人员对应的会场终端的筛选。也就是说,步骤302最终匹配确定的会场终端为多点控制单元根据待录制人员的特征信息和各个会场终端发送的音视频码流,从多个会场终端筛选得到的。

[0102] 步骤303:将筛选出的会场终端发送的音视频码流全部作为需要录制的音视频码流。

[0103] 由于录制需求是前述录制需求(1),即录制待录制人员参加的完整会议,因此在筛选需要录制的音视频码流时,直接将步骤302筛选出的会场终端发送的音视频码流全部作为需要录制的音视频码流。

[0104] 通过执行图3所示的流程,MCU筛选得到待录制人员所在会场的整个音视频码流,通过将该整个音视频码流发送给录播服务器以进行会议录制,满足上述录制需求(1)。

[0105] 参见图4,该图为本实施例提供的另一种多点控制单元获得需要录制的音视频码

流的流程图。

[0106] 如图4所示,当需要录制的音视频码流为待录制人员个人的音视频码流时,多点控制单元获得需要录制的音视频码流具体包括:

[0107] 步骤401:多点控制单元将各个会场终端发送的音视频码流进行解码获得解码后的视频码流和音频码流。

[0108] 步骤402:多点控制单元根据待录制人员的图像信息与解码后的视频码流进行特征匹配,确定待录制人员对应的会场终端,或,将待录制人员的声音信息与解码后的音频码流进行特征匹配,确定待录制人员对应的会场终端。

[0109] 本实施例中,步骤401-402与前述步骤301-302的实现方式相同,步骤401-402的相关描述可参照步骤301-302,此处不再赘述。

[0110] 步骤403:从筛选出的会场终端发送的音视频码流中根据待录制人员的特征信息筛选待录制人员个人的音视频码流作为需要录制的音视频码流。

[0111] 如果MCU从会议AS接收的待录制人员的特征信息仅包括声音信息,则实际的录制需求很可能是录制需求(2),即仅录制待录制人员发言时的会议。本步骤对于该录制需求,可以利用待录制人员的声音信息从步骤401解码得到的T1时刻至T2时刻的音频码流中确定出包含该待录制人员的声音信息的T3时刻至T4时刻的音频码流(会议起始时刻T1早于会议结束时刻T2,待录制人员发言起始时刻T3早于待录制人员发言结束时刻T4,T3晚于或等于T1且早于T2,T4早于或等于T2)。由于音视频码流解码得到的视频码流和音频码流存在时序联系,因此,根据T3时刻至T4时刻的音频码流也可以相应地得到T3时刻至T4时刻的视频码流。T3时刻至T4时刻的音频码流和视频码流统称为待录制人员个人的音视频码流,该音视频码流满足录制需求(2)。

[0112] 如果MCU从会议AS接收的待录制人员的特征信息仅包括图像信息,则实际的录制需求很可能是录制需求(3),即录制出现待录制人员时的会议。本步骤对于该录制需求,可以利用待录制人员的声音信息从步骤401解码得到的T1时刻至T2时刻的视频码流中确定出包含该待录制人员的图像信息的T5时刻至T6时刻的视频码流(会议起始时刻T1早于会议结束时刻T2,待录制人员在T5时刻至T6时刻出现,T5早于T6,T5晚于或等于T1,T6早于或等于T2)。由于音视频码流解码得到的视频码流和音频码流存在时序联系,因此,根据T5时刻至T6时刻的视频码流也可以相应地得到T5时刻至T6时刻的音频码流。T5时刻至T6时刻的音频码流和视频码流统称为待录制人员个人的音视频码流,该音视频码流满足录制需求(3)。

[0113] 下面示例性地提供一种本实施例方法的应用场景。在该示例场景中,待录制人员有多个,并且多个待录制人员位于同一会场,即多个待录制人员对应于同一会场终端。具体的录制需求是,仅录制待录制人员发言时的会议;多点控制单元获得的待录制人员的特征信息包括:每个待录制人员的声音信息。本实施例提供的会议录制方法在具体实现时,由多点控制单元MCU将各个会场终端发送的音视频码流进行解码获得解码后的视频码流和音频码流;将各个待录制人员的声音信息与解码后的音频码流进行特征匹配,确定各个待录制人员对应的同一会场终端;MCU从筛选出的会场终端发送的音视频码流中根据各个待录制人员的声音信息,筛选各个待录制人员其个人的音视频码流,这些音视频码流共同作为需要录制的音视频码流。MCU最后将这些需要录制的音视频码流发送给录播服务器以进行会议录制。

[0114] 按照会议能力的单流和多流,可将会场终端划分两类,一类是单流的高级视频编码(advanced video coding,AVC)会场终端,另一类是多流的可伸缩视频编码(scalable video coding,SVC)会场终端。在实际应用中,会议录制场景中的各个会场终端可能统一是AVC会场终端,也可能统一是SVC会场终端。鉴于AVC会场终端和SVC会场终端的会议能力不同,本申请实施例提供的会议录制方法中多点控制单元MCU执行的操作也相应存在差别。下面以两个实施例分别描述AVC会场终端场景下的会议录制方法以及SVC会场终端场景下的会议录制方法。

[0115] 方法实施例二 (AVC会场终端场景)

[0116] 参见图5,该图为本实施例提供的一种会议录制方法的信令图。在图5示意的会议录制场景中,包括多点控制单元MCU,会议AS,录播服务器,以及多个AVC会场终端,分别为AVC 1、AVC 2和AVC 3。

[0117] 在图5所示的会议录制方法中,包括以下步骤:

[0118] 步骤501:会议AS根据用户的预约向多点控制单元MCU召集会议,下发待录制人员的特征信息。

[0119] 在本实施例中,待录制人员对应至少两个不同的AVC会场终端,也就是说,会议AS下发的待录制人员的特征信息属于至少两个待录制人员。作为示例,会议AS下发了待录制人员Role1的特征信息和待录制人员Role2的特征信息。

[0120] 步骤502:多点控制单元呼叫所有AVC会场终端入会,包括:AVC 1、AVC 2和AVC 3。

[0121] 步骤503:AVC 1、AVC 2和AVC 3入会,分别向MCU发送音视频码流。

[0122] 步骤504:MCU呼叫录播服务器入会。

[0123] 步骤505:MCU将AVC 1、AVC 2和AVC 3发送的音视频码流分别解码。

[0124] 本步骤MCU将AVC 1、AVC 2和AVC 3发送的音视频码流分别解码,解码能够获得AVC 1、AVC 2和AVC 3各自发送的音频码流和视频码流。通过解码得到音频码流和视频码流,以便于后续利用待录制人员Role1和Role2的特征信息进行音视频码流的筛选。

[0125] 步骤506:MCU根据待录制人员的特征信息分别从AVC 1、AVC 2和AVC 3发送的音视频码流中筛选需要录制的音视频码流。

[0126] 本步骤在具体实现时,可以根据待录制人员的特征信息和解码得到的各个会场终端的音频码流和视频码流,筛选待录制人员对应的会场终端。具体筛选过程可参照前述实施例,此处不再赘述。

[0127] 作为示例,最终根据Role1的特征信息筛选Role1对应的会场终端为AVC 1,根据Role2的特征信息筛选Role2对应的会场终端为AVC 2。可以理解的是,为筛选需要录制的音视频码流,MCU可以通过指定视频源的方式进行筛选。例如,MCU指定AVC 1的视频源名称,根据该视频源名称筛除不符的音频码流和视频码流,最终筛选得到的音频码流和视频码流即作为对Role1需要录制的音频码流和视频码流;MCU指定AVC 2的视频源名称,根据该视频源名称筛除不符的音频码流和视频码流,最终筛选得到的音频码流和视频码流即作为对Role2需要录制的音频码流和视频码流。

[0128] 步骤507:MCU将需要录制的视频码流进行画面合成获得合成画面,将合成画面发送给录播服务器,将需要录制的音频码流进行混音后发送给录播服务器。

[0129] 在实际应用中,MCU具备对AVC会场终端发送的视频码流进行画面合成的功能,以

及对AVC会场终端发送的音频码流进行混音的功能。由于本实施例中,待录制人员对应至少两个不同的AVC会场终端,因此MCU可根据前一步骤得到的对于多个待录制人员(例如Role1和Role2)的需要录制的视频码流进行合成,并根据前一步骤得到的对于多个待录制人员的需要录制的音频码流进行混音。

[0130] 步骤508:录播服务器收到码流后进行会议录制。

[0131] 在本实施例中,待录制人员对应至少两个AVC会场终端,即待录制人员共有至少两个。会议AS通过下发待录制人员的特征信息,实现对会议录制的待录制人员的指定。当需要录制的音视频码流为多个待录制人员个人的音视频码流时,利用本实施例提供的会议录制方法由MCU进行画面合成和混音并发送给录播服务器,大大减少网络带宽,并且节省了录播服务器的存储空间。此外,本实施例按照指定的待录制人员进行实时录制,避免人工后期制作的音频码流和视频码流的裁剪过程,节约人力成本,提高会议录制的效率。

[0132] 方法实施例三(SVC会场终端场景)

[0133] 参见图6,该图为本实施例提供的另一种会议录制方法的信令图。在图6示意的会议录制场景中,包括多点控制单元MCU,会议AS,录播服务器,以及多个SVC会场终端,分别为SVC 1、SVC 2和SVC 3。

[0134] 在图6所示的会议录制方法中,包括以下步骤:

[0135] 步骤601:会议AS根据用户的预约向多点控制单元MCU召集会议,下发待录制人员的特征信息。

[0136] 在本实施例中,待录制人员对应至少两个不同的SVC会场终端,也就是说,会议AS下发的待录制人员的特征信息属于至少两个待录制人员。作为示例,会议AS下发了待录制人员Role3的特征信息和待录制人员Role4的特征信息。

[0137] 步骤602:多点控制单元呼叫所有SVC会场终端入会,包括SVC 1、SVC 2和SVC 3,并且多点控制单元将适用于录播服务器的码流格式通知SVC 1、SVC 2和SVC 3。

[0138] 实际应用中,每个SVC会场终端均可向MCU提供不同码流格式的音视频码流。然而录播服务器通常只能基于其中一种码流格式的音视频码流进行会议录制,其他码流格式的音视频码流不适用于录播服务器。为提升音视频码流的传输效率,避免SVC会场终端向MCU发送不适用于录播服务器的码流格式的音视频码流,本实施例中需要预先向各个SVC会场终端通知适用于录播服务器的码流格式。

[0139] 步骤603:SVC 1、SVC 2和SVC 3入会,分别向MCU发送适用于录播服务器的码流格式的音视频码流。

[0140] 步骤604:MCU呼叫录播服务器入会。

[0141] 步骤605:MCU将SVC 1、SVC 2和SVC 3发送的音视频码流分别解码。

[0142] 本步骤MCU将SVC 1、SVC 2和SVC 3发送的音视频码流分别解码,解码能够获得SVC 1、SVC 2和SVC 3各自发送的音频码流和视频码流。通过解码得到音频码流和视频码流,以便于后续利用待录制人员Role3和Role4的特征信息进行音视频码流的筛选。

[0143] 步骤606:MCU根据待录制人员的特征信息分别从SVC 1、SVC 2和SVC 3发送的适用于录播服务器的码流格式的音视频码流中筛选需要录制的音视频码流。

[0144] 本步骤在具体实现时,可以根据待录制人员的特征信息和解码得到的各个会场终端的音频码流和视频码流,筛选待录制人员对应的会场终端。具体筛选过程可参照前述实



施例,此处不再赘述。

[0145] 作为示例,最终根据Role3的特征信息筛选Role3对应的会场终端为SVC 3,根据Role4的特征信息筛选Role4对应的会场终端为SVC 2。可以理解的是,为筛选需要录制的音视频码流,MCU可以通过指定视频源的方式进行筛选。例如,MCU指定SVC 3的视频源名称,根据该视频源名称筛除不符的音频码流和视频码流,最终筛选得到的音频码流和视频码流即作为对Role3需要录制的音频码流和视频码流;MCU指定SVC 2的视频源名称,根据该视频源名称筛除不符的音频码流和视频码流,最终筛选得到的音频码流和视频码流即作为对Role4需要录制的音频码流和视频码流。

[0146] 步骤607:MCU将至少两个不同的SVC会场终端对应的需要录制的视频码流发送给录播服务器,将至少两个不同的SVC会场终端对应的需要录制的音频码流进行混音后发送给录播服务器。

[0147] 沿用前述示例,本步骤即是将前一步骤SVC 3和SVC 2发送的需要录制的视频码流发送给录播服务器,这是由于录播服务器具备将多个SVC会场终端的视频码流进行画面合成的功能。MCU具备对SVC会场终端发送的音频码流进行混音的功能,因此SVC 3和SVC 2发送的需要录制的音频码流由MCU负责混音。

[0148] 步骤608:录播服务器对至少两个不同的SVC会场终端对应的需要录制的视频码流进行画面合成获得合成画面,根据合成画面和混音后的音频码流进行会议录制。

[0149] 在本实施例中,待录制人员对应至少两个SVC会场终端,即待录制人员共有至少两个。会议AS通过下发待录制人员的特征信息,实现对会议录制的待录制人员的指定。当需要录制的音视频码流为多个待录制人员个人的音视频码流时,利用本实施例提供的会议录制方法由MCU进行混音并发送给录播服务器,由录播服务器对MCU筛选的对于多个待录制人员的需要录制的视频码流进行画面合成,最终进行会议录制,大大减少网络带宽,并且节省了录播服务器的存储空间。此外,本实施例按照指定的待录制人员进行实时录制,避免人工后期制作的音频码流和视频码流的裁剪过程,节约人力成本,提高会议录制的效率。

[0150] 基于前述实施例提供的会议录制方法,相应地,本申请还提供一种会议录制装置。下面结合实施例和附图对该装置的具体实现进行描述。

#### [0151] 装置实施例

[0152] 参见图7,该图所示为本实施例提供的会议录制装置的结构示意图。如图7所示,该装置包括:

[0153] 码流筛选模块701,用于根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流;

[0154] 码流发送模块702,用于将需要录制的音视频码流发送给录播服务器,以使录播服务器进行会议录制;特征信息包括:图像信息或声音信息。

[0155] 在本实施例中,会议录制装置利用待录制人员的特征信息,实现对需要录制的音视频码流的自动筛选,从而无需人工筛选,节省会议录制的人工成本,同时大大提升会议录制效率。应用该装置能够提升会议录制的便捷性,促进视频会议功能的广泛应用。

[0156] 在实际应用中,根据录制需求,需要录制的音视频码流可以是待录制人员所在会场的整个音视频码流,也可以是待录制人员个人的音视频码流。下面基于这两种情况分别对码流筛选模块701的实现方式进行描述。

[0157] 当需要录制的音视频码流为待录制人员所在会场的整个音视频码流时,码流筛选模块701,具体包括:

[0158] 会场终端筛选单元,用于根据待录制人员的特征信息和各个会场终端发送的音视频码流筛选待录制人员对应的会场终端;

[0159] 码流第一筛选单元,用于将筛选出的会场终端发送的音视频码流全部作为需要录制的音视频码流。

[0160] 当需要录制的音视频码流为待录制人员个人的音视频码流时,码流筛选模块701,具体包括:

[0161] 会场终端筛选单元,用于根据待录制人员的特征信息和各个会场终端发送的音视频码流筛选待录制人员对应的会场终端;

[0162] 码流第二筛选单元,用于从筛选出的会场终端发送的音视频码流中根据待录制人员的特征信息筛选待录制人员个人的音视频码流作为需要录制的音视频码流。

[0163] 通过上述描述可知,作为一种可能的实现方式,无论需要录制的音视频码流是待录制人员所在会场的整个音视频码流,还是待录制人员个人的音视频码流,会议录制装置的会场终端筛选单元均可以首先通过匹配,确定出待录制人员对应的会场。具体地,会场终端筛选单元,具体包括:

[0164] 解码子单元,用于将各个会场终端发送的音视频码流进行解码获得解码后的视频码流和音频码流;

[0165] 会场终端确定子单元,用于根据待录制人员的图像信息与解码后的视频码流进行特征匹配,确定待录制人员对应的会场终端,或,将待录制人员的声音信息与解码后的音频码流进行特征匹配,确定待录制人员对应的会场终端。

[0166] 按照会议能力的单流和多流,可将会场终端划分两类,一类是单流的高级视频编码(advanced video coding,AVC)会场终端,另一类是多流的可伸缩视频编码(scalable video coding,SVC)会场终端。在实际应用中,会议录制场景中的各个会场终端可能统一是AVC会场终端,也可能统一是SVC会场终端。鉴于AVC会场终端和SVC会场终端的会议能力不同,本申请实施例提供的会议录制装置具体实现方式也相应存在差别。下面分别描述AVC会场终端场景下以及SVC会场终端场景下会议录制装置的实现方式。

[0167] 当待录制人员对应至少两个不同的高级视频编码AVC会场终端时,码流筛选模块701,具体包括:

[0168] 码流第三筛选单元,用于根据待录制人员的特征信息分别从至少两个不同的AVC会场终端发送的音视频码流中筛选需要录制的音视频码流。

[0169] 需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流;码流发送模块702,具体包括:

[0170] 画面合成单元,用于将需要录制的视频码流进行画面合成获得合成画面;

[0171] 画面发送单元,用于将合成画面发送给录播服务器;

[0172] 第一混音单元,用于将需要录制的音频码流进行混音;

[0173] 音频第一发送单元,用于将混音后的音频发送给录播服务器。

[0174] 在本实施例中,待录制人员对应至少两个AVC会场终端,即待录制人员共有至少两个。会议AS通过下发待录制人员的特征信息,实现对会议录制的待录制人员的指定。当需要

录制的音视频码流为多个待录制人员个人的音视频码流时,利用本实施例提供的会议录制装置由MCU进行画面合成和混音并发送给录播服务器,大大减少网络带宽,并且节省了录播服务器的存储空间。此外,应用本实施例提供的装置可按照指定的待录制人员进行实时录制,避免人工后期制作的音频码流和视频码流的裁剪过程,节约人力成本,提高会议录制的效率。

[0175] 当待录制人员对应至少两个不同的可伸缩视频编码SVC会场终端时,码流筛选模块701,具体包括:

[0176] 码流格式通知单元,用于将适用于录播服务器的码流格式通知所有SVC会场终端;

[0177] 码流接收单元,用于接收至少两个不同的SVC会场终端发送的适用于录播服务器的码流格式的音视频码流;

[0178] 码流第四筛选单元,用于根据待录制人员的特征信息从适用于录播服务器的码流格式的音视频码流中筛选需要录制的音视频码流。

[0179] 需要录制的音视频码流包括需要录制的视频码流和需要录制的音频码流;码流发送模块702,具体包括:

[0180] 视频码流发送单元,用于将至少两个不同的SVC会场终端对应的需要录制的视频码流发送给录播服务器,以使录播服务器对至少两个不同的SVC会场终端对应的需要录制的视频码流进行画面合成获得合成画面;

[0181] 第二混音单元,用于将至少两个不同的SVC会场终端对应的需要录制的音频码流进行混音;

[0182] 音频第二发送单元,用于将混音后的音频发送给录播服务器。

[0183] 在本实施例中,待录制人员对应至少两个SVC会场终端,即待录制人员共有至少两个。会议AS通过下发待录制人员的特征信息,实现对会议录制的待录制人员的指定。当需要录制的音视频码流为多个待录制人员个人的音视频码流时,利用本实施例提供的会议录制装置由MCU进行混音并发送给录播服务器,由录播服务器对MCU筛选的对于多个待录制人员的需要录制的视频码流进行画面合成,最终进行会议录制,大大减少网络带宽,并且节省了录播服务器的存储空间。此外,应用本实施例提供的装置可按照指定的待录制人员进行实时录制,避免人工后期制作的音频码流和视频码流的裁剪过程,节约人力成本,提高会议录制的效率。

[0184] 作为一种可能的实现方式,本实施例中码流筛选模块701,具体包括:

[0185] 码流第五筛选单元,用于根据待录制人员的特征信息利用预先训练的神经网络模型从音视频码流中筛选需要录制的音视频码流。

[0186] 该神经网络模型为利用大量不同人员的特征信息以及包含不同人员特征信息的材料(例如待录制人员的图片或音频文件)的训练得到。

[0187] 对于本领域技术人员来说,训练能够准确识别带有某种图像信息的视频码流或能够准确识别带有某种声音信息的音频码流的神经网络模型,属于比较成熟的技术,因此本实施例对于该神经网络模型的具体训练过程不进行详述。

[0188] 基于前述实施例提供的会议录制方法及会议录制装置,相应地,本申请还提供一种会议录制系统。下面结合附图和实施例对该系统的具体实现方式进行描述。

[0189] 系统实施例

[0190] 参见图8,该图为本申请实施例提供的一种会议录制系统的结构示意图。

[0191] 如图8所示,本实施例提供的会议录制系统,包括:多点控制单元MCU,录播服务器801和至少两个会场终端。

[0192] 在本实施例中,多点控制单元MCU具体可以执行前述实施例提供的会议录制方法。

[0193] 会场终端,用于向多点控制单元MCU发送音视频码流;

[0194] 多点控制单元MCU,用于根据待录制人员的特征信息从各个会场终端发送的音视频码流中筛选需要录制的音视频码流;特征信息包括:图像信息或声音信息;将需要录制的音视频码流发送给录播服务器801;

[0195] 录播服务器801,用于根据需要录制的音视频码流进行会议录制。

[0196] 本实施例中会场终端的数量为至少两个。所有会场终端可以统一是AVC会场终端,还可以统一是SVC会场终端。各个会场终端可以是采用会话初始协议(Session Initiation Protocol, SIP)的终端,也可以是采用H.323协议的终端。此处对于会场终端采用的通信协议不进行限定。如图8中所示,会场终端811采用SIP协议与MCU进行通信,会场终端812和813分别采用H.323协议与MCU进行通信。

[0197] 本实施例提供的会议录制系统利用待录制人员的特征信息,实现对需要录制的音视频码流的自动筛选,从而无需人工筛选,节省会议录制的人工成本,同时大大提升会议录制效率。该系统提升会议录制的便捷性,促进视频会议功能的广泛应用。

[0198] 可选地,本实施例提供的会议录制系统中,MCU还具备转发音视频码流的功能。下面结合图8描述该功能的实现场景。

[0199] MCU转发音视频码流的示例性场景:一个会场终端向MCU请求转播另一会场终端的会议。会场终端813用于向MCU发送转播会场终端812的会议的请求。在实际应用中,会场终端813能够播放的音视频码流的格式可能为会场终端812能够提供的多种码流格式之一,因此,会场终端813发送的请求中携带会场终端813能够播放的音视频码流的码流格式。MCU用于根据请求中的码流格式,向会场终端812发送通知,以使会场终端812根据该通知向MCU发送该格式的音视频码流。MCU还用于将会场终端812发送的会场终端813能够播放的码流格式的音视频码流转发给会场终端813,以便于会场终端813能够根据音视频码流进行播放。从而,会场终端813所在会场的所有参会人员能够观看到会场终端812召开的会议。

[0200] 可选地,本实施例提供的会议录制系统中,录播服务器801还具备转发音视频码流的功能。下面描述该功能的实现场景。

[0201] 录播服务器801转发音视频码流的示例性场景:其他服务器向录播服务器请求点播或直播会议。录播服务器801还用于接收其他服务器的点播请求或直播请求。在实际应用中,其他服务器能够播放的音视频码流的格式可能为各个会场终端能够提供的多种码流格式之一,因此,其他服务器发送的点播请求或直播请求中可以携带其他服务器能够播放的音视频码流的码流格式。录播服务器801还用于将该码流格式通知MCU,以便MCU将该码流格式通知各个会场终端。当MCU接收到各个会场终端发送的符合该码流格式的音视频码流后,将其发送给录播服务器。录播服务器801还用于基于点播请求或直播请求,以及MCU发送的音视频码流,向提出该请求的其他服务器进行转发。作为一示例,其他服务器向录播服务器801请求点播会场终端811的会议,则录播服务器801向其他服务器转发来自会场终端811的符合该其他服务器的码流格式的音视频码流。作为另一示例,其他服务器向录播服务器801

请求直播所有入会的会场终端的会议,则录播服务器801向其他服务器转发来自各个会场终端的符合其他服务器的码流格式的音视频码流;如果适用于录播服务器801的码流格式与发出请求的其他服务器能够播放的码流格式,则录播服务器还可以将一边录制一边向其他服务器转发,转发内容即为多画面(即合成画面)的音视频码流。

[0202] 通过以上场景示例可知,在本实施例提供的会议录制系统中,MCU以及录播服务器801均能够提供音视频码流的转发服务,从而丰富了会议录制系统的整体功能,提升用户的使用体验。

[0203] 可选地,本实施例提供的会议录制系统还可进一步包括:会议应用服务器(会议AS)802。如图9所示,该图为本实施例提供的另一种会议录制系统的结构示意图。

[0204] 在图9所示意的系统中,会场终端与会议AS802之间的连接虚线表示会场终端注册到会议AS802;录播服务器801与会议AS802之间的连接虚线表示录播服务器801注册到会议AS802;会议AS802与MCU之间的连接虚线表示会议AS802向MCU下发待录制人员的特征信息。会议AS802用于向用户提供上传带有待录制人员特征信息的材料(例如待录制人员的图片或音频文件)的功能;对材料进行处理得到待录制人员的特征信息;将待录制人员的特征信息下发至MCU。会议AS802还用于存储多个人员各自对应的材料并处理获得每个人员的特征信息,当用户登录会议AS802后,将各个人员对应的材料提供给用户,并接收用户对待录制人员的材料的选定消息,根据选定消息向MCU下发被选定的材料对应的待录制人员的特征信息。在实际应用中,会议AS802可以在向MCU召集会议时,向MCU下发待录制人员的特征信息。

[0205] 通过上文描述可知,本实施例提供的会议录制系统中,会议AS802能够向用户提供指定待录制人员的服务,从而增加后续会议录制的便捷性,只对待录制人员进行录制,节省录播服务器801的存储空间,减小通信带宽,提升用户的会议录制体验。

[0206] 应当理解,在本申请中,“至少一个(项)”是指一个或者多个,“多个”是指两个或两个以上。“以下至少一项(个)”或其类似表达,是指这些项中的任意组合,包括单项(个)或复数项(个)的任意组合。例如,a,b或c中的至少一项(个),可以表示:a,b,c,“a和b”,“a和c”,“b和c”,或“a和b和c”,其中a,b,c可以是单个,也可以是多个。

[0207] 以上所述,以上实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。

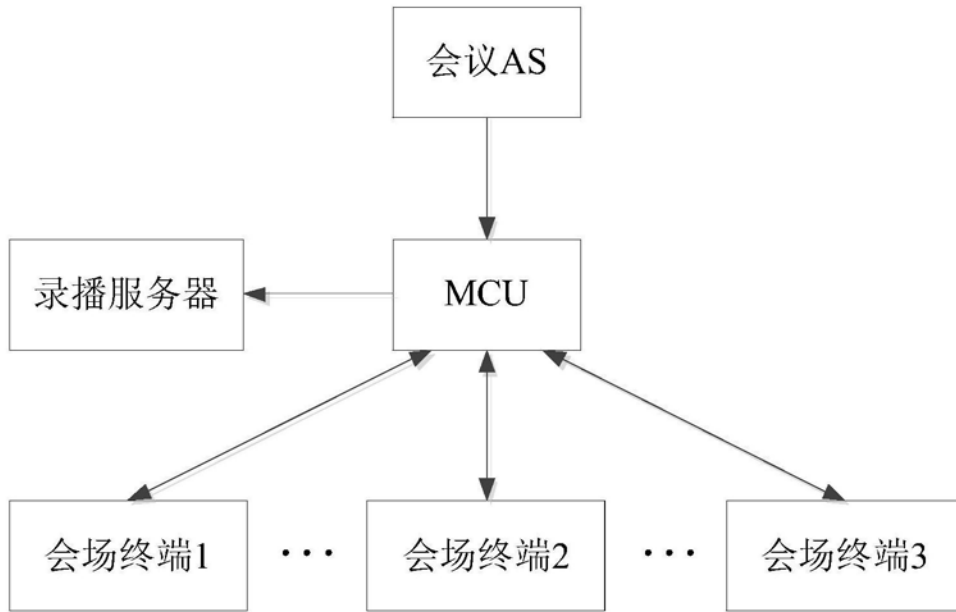


图1

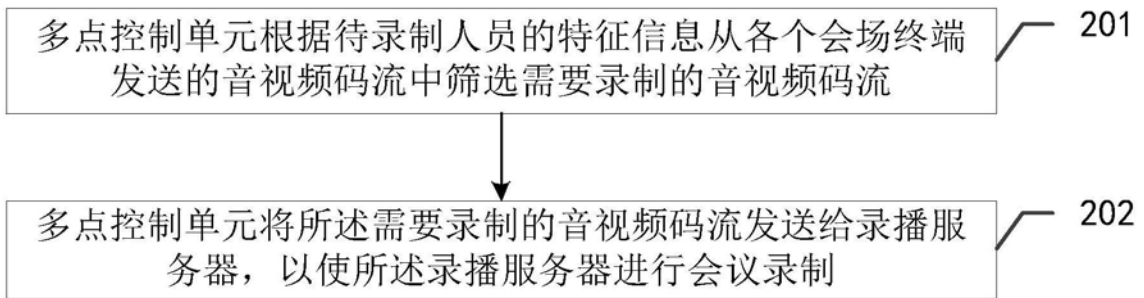


图2

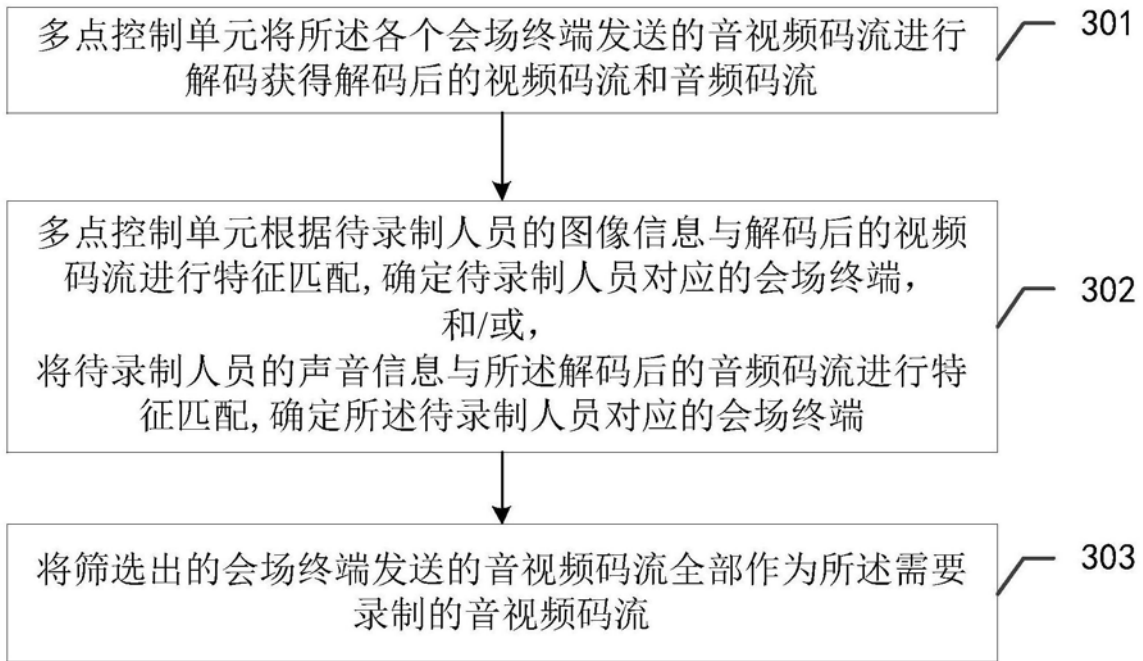


图3

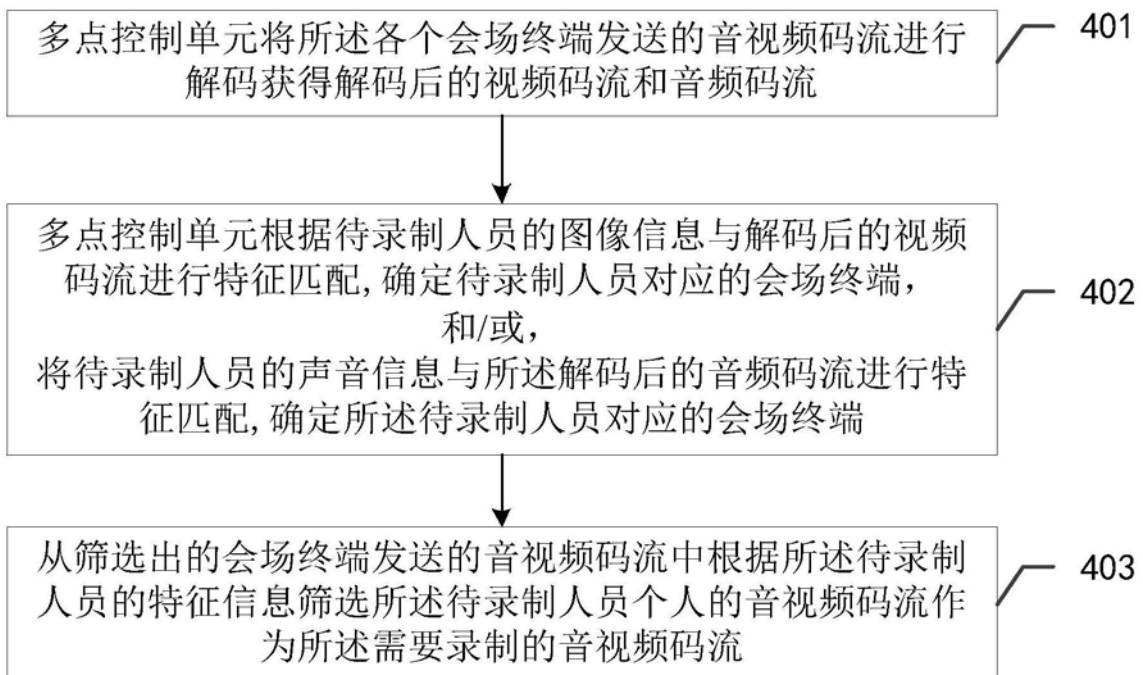


图4

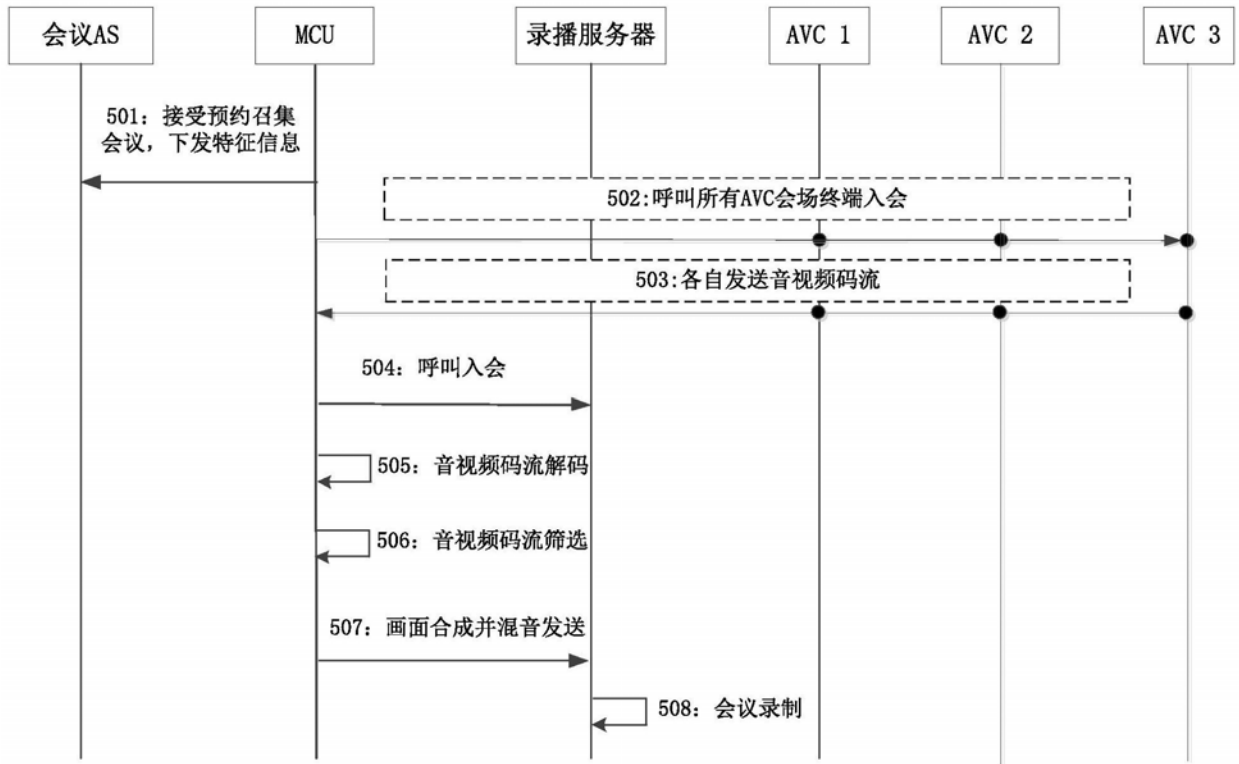


图5

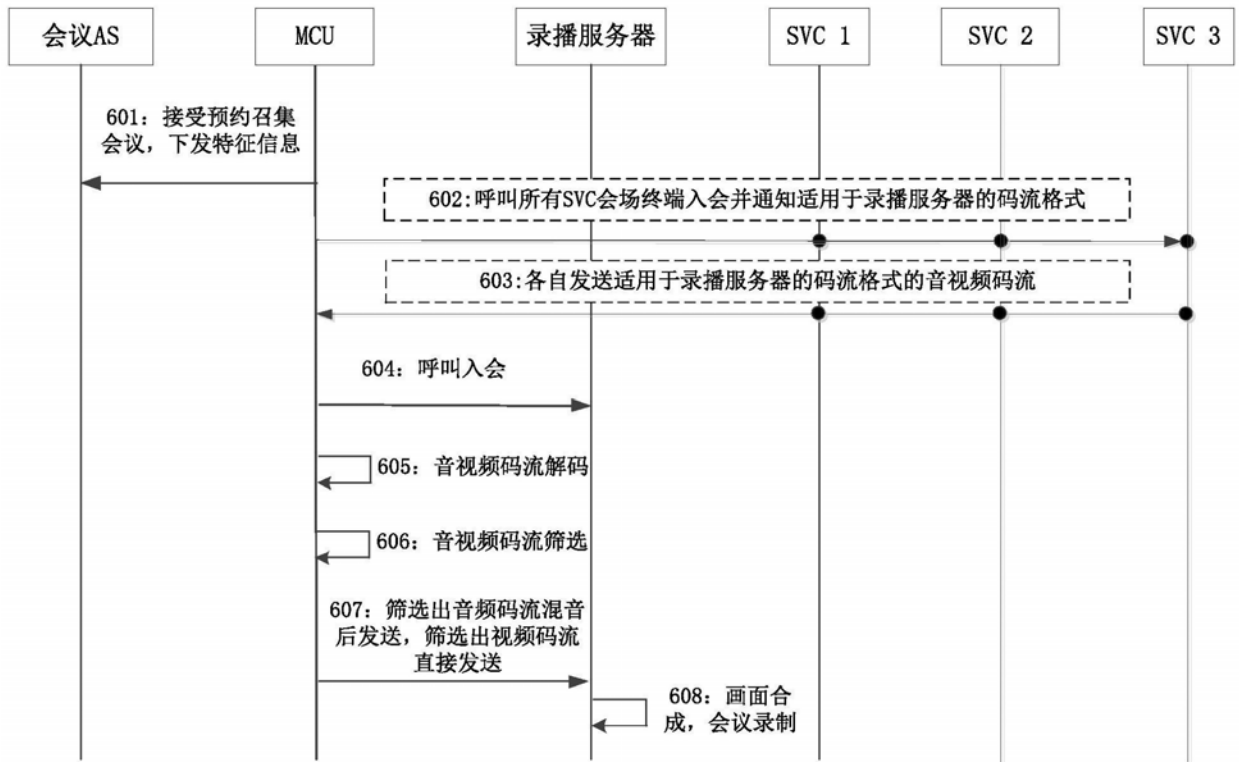


图6



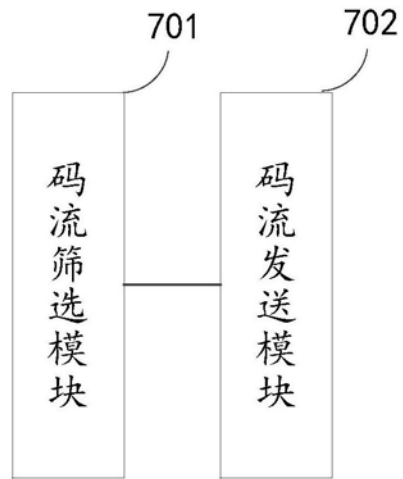


图7

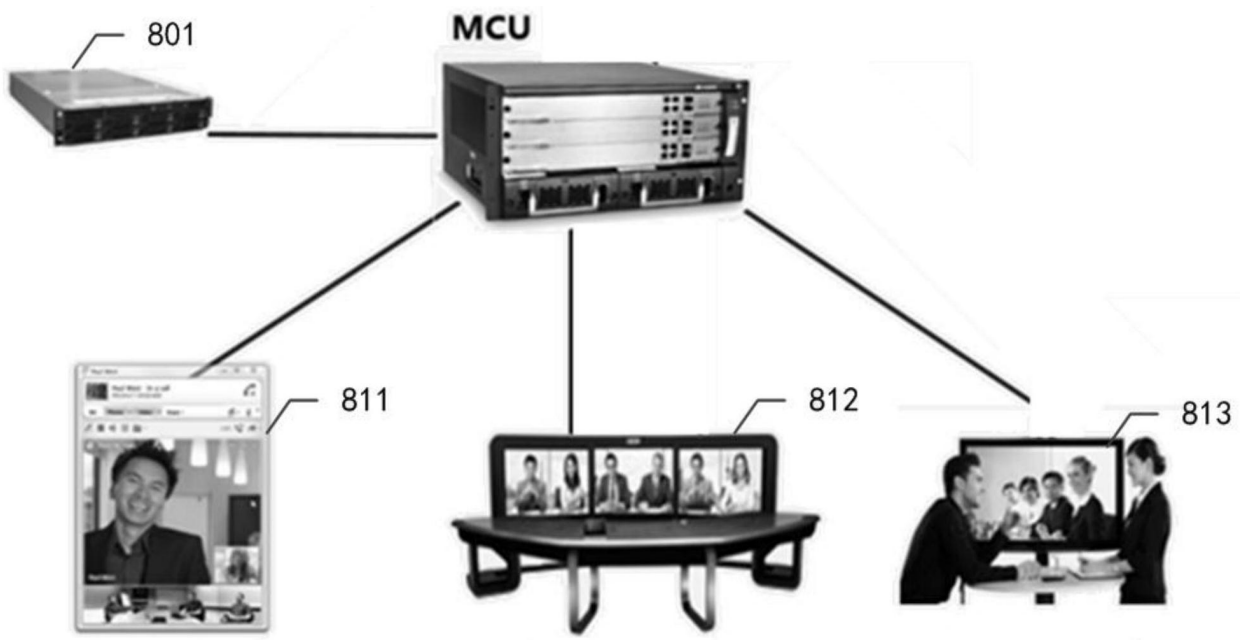


图8

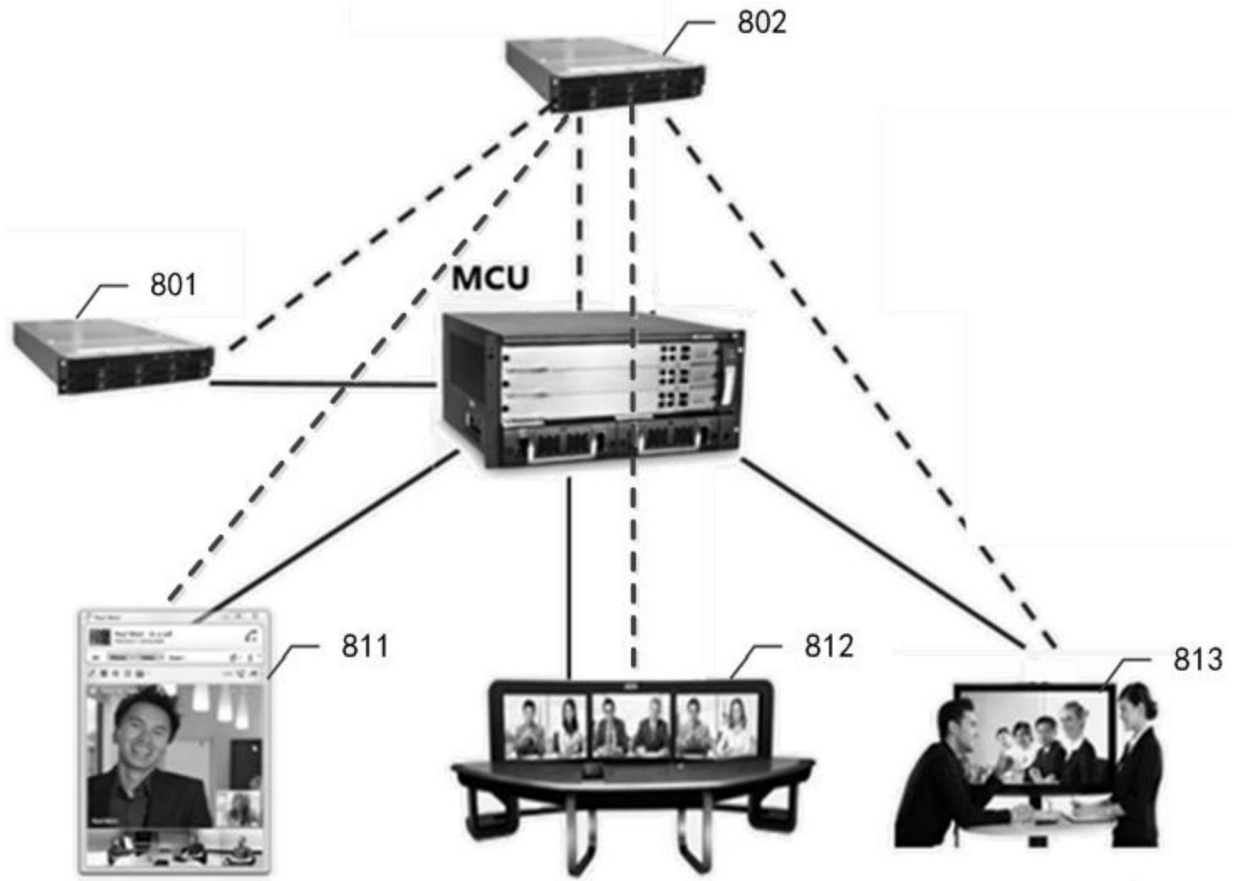


图9