(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2019 年 11 月 28 日 (28.11.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/223040 A1

(51) 国际专利分类号:
H04N 21/433 (2011.01) *H04N 21/8547* (2011.01)
H04N 21/4402 (2011.01) *H04N 21/439* (2011.01)

(21) 国际申请号: PCT/CN2018/091206

(22) 国际申请日: 2018 年 6 月 14 日 (14.06.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810517553.2 2018年5月25日 (25.05.2018) CN

(71) 申请人: 网宿科技股份有限公司 (**WANGSU SCIENCE & TECHNOLOGY CO., LTD.**) [CN/CN]; 中国上海市徐汇区斜土路2899号甲光启文化广场A幢5楼, Shanghai 200030 (CN)。

(72) 发明人: 黄学辉(**HUANG, Xuehui**); 中国上海市徐汇区斜土路2899号甲光启文化广场A幢5楼, Shanghai 200030 (CN)。 吕士表(**LV, Shibiao**); 中国上海市徐汇区斜土路2899号甲光启文化广场A幢5楼, Shanghai 200030 (CN)。

(74) 代理人: 北京华智则铭知识产权代理有限公司 (**BEIJING HUA-WISDOM INTELLECTUAL PROPERTY INC.**); 中国北京市大兴区宏业路9号院7号楼1307室, Beijing 100162 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** METHOD AND DEVICE FOR SYNTHESIZING AUDIO AND VIDEO DATA STREAM

(54) 发明名称: 一种合成音视频数据流的方法和装置

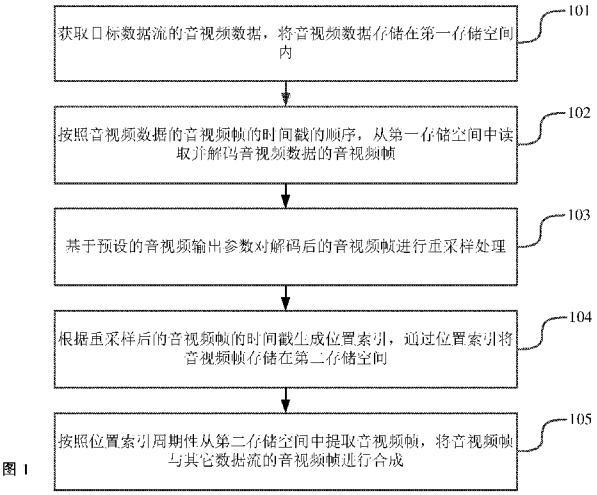


图 1

101 Acquire audio and video data of a target data stream, and store the audio and video data in a first storage space
102 Read and decode audio and video frames of the audio and video data from the first storage space according to a time stamp sequence of the audio and video frames of the audio and video data
103 Re-sample the decoded audio and video frames according to a pre-set audio and video output parameter
104 Generate a position index according to a time stamp of the re-sampled audio and video frames, and store the audio and video frames in a second storage space through the position index
105 Extract the audio and video frames from the second storage space periodically according to the position index, and synthesize the audio and video frames with audio and video frames of other data streams

(57) **Abstract:** Disclosed are a method and device for synthesizing an audio and video data stream, belonging to the technical field of multimedia. The method comprises: acquiring audio and video data of a target data stream, and storing the audio and video data in a first storage space; reading and decoding audio and video frames of the audio and video data from the first storage space according to a time stamp sequence of the audio and video frames of the audio and video data; re-sampling the decoded audio and video frames according to a pre-set audio and video output parameter; generating a position index according to a time stamp of the re-sampled audio and video frames, and storing the audio and video frames in a second storage space through the position index; and extracting the audio and video frames from the second storage space periodically according to the position index, and synthesizing the audio and video frames with audio and video frames of other data streams. By applying the present invention, data streams can be synthesized more conveniently and effectively.

WO 2019/223040 A1

[见续页]

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：本发明公开了一种合成音视频数据流的方法和装置，属于多媒体技术领域。所述方法包括：获取目标数据流的音视频数据，将所述音视频数据存储在第一存储空间内；按照所述音视频数据的音视频帧的时间戳的顺序，从所述第一存储空间中读取并解码所述音视频数据的音视频帧；基于预设的音视频输出参数对解码后的音视频帧进行重采样处理；根据重采样后的音视频帧的时间戳生成位置索引，通过所述位置索引将所述音视频帧存储在第二存储空间；按照所述位置索引周期性从所述第二存储空间中提取所述音视频帧，将所述音视频帧与其它数据流的音视频帧进行合成。采用本发明，可以更加方便有效地实现数据流合成。

一种合成音视频数据流的方法和装置

技术领域

本发明涉及多媒体技术领域，特别涉及一种合成音视频数据流的方法和装置。

背景技术

伴随着互联网技术的发展以及宽带的不断提速，互联网与人们的生活联系的日益密切，观看直播已成为当下一种主流的娱乐形式。目前多人视频会话、直播连麦等合成有多路数据流的直播形式不断出现并受到广泛的欢迎。

在这些存在多路点播流或直播流同时输入的直播形式中，每路数据流可能具有不同的分辨率、码率、音频采样率、音视频编码格式，从而导致了传统单路数据流输入的直播系统中所没有遇到的多路数据流合成的问题。同时，每路数据流受主播端推流状态、网络传输质量等因素的影响，可能出现数据流拉流过程的不稳定，这使得输入的多路数据流之间的合成问题变得更加复杂，故而目前亟需一种适合多路数据流输入的，能够应对数据流网络波动影响的，方便有效地实现数据流合成的方法。

发明内容

为了解决现有技术的问题，本发明实施例提供了一种合成音视频数据流的方法和装置。所述技术方案如下：

第一方面，提供了一种合成音视频数据流的方法，所述方法包括：

获取目标数据流的音视频数据，将所述音视频数据存储在第一存储空间内；

按照所述音视频数据的音视频帧的时间戳的顺序，从所述第一存储空间中读取并解码所述音视频数据的音视频帧；

基于预设的音视频输出参数对解码后的音视频帧进行重采样处理；

根据重采样后的音视频帧的时间戳生成位置索引，通过所述位置索引将所述音视频帧存储在第二存储空间；

按照所述位置索引周期性从所述第二存储空间中提取所述音视频帧，将所述音视频帧与其它数据流的音视频帧进行合成。

可选的，所述将所述音视频数据存储在第一存储空间内，包括：

将所述音视频数据包含的音频数据和视频数据分别存储在第一存储空间的音频存储空间和视频存储空间内。

可选的，所述方法还包括：

如果所述目标数据流为直播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，删除所述第一存储空间中最早存储的音视频数据，继续存储所述音视频数据；

如果所述目标数据流为点播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，等待所述第一存储空间中存储的音视频数据被消耗后，继续存储所述音视频数据。

可选的，所述方法还包括：

根据预设的最大播放延迟和最大网络延迟确定所述第一存储空间的大小；

在获取目标数据流的音视频数据过程中，根据检测到的播放延迟需求和/或实时网络延迟调整所述第一存储空间的大小。

可选的，所述方法还包括：

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的音视频帧的时间戳时长；

如果所述时间戳时长大于预设播放速率与读取并解码所述音视频数据所用的物理时长的乘积，则在当前周期暂停从所述第一存储空间读取所述音视频数据的音视频帧。

可选的，所述周期性检测从所述第一存储空间中读取并解码的所述音视频数据的时间戳时长，包括：

记录从所述第一存储空间中读取并解码的所述音视频数据的初始音频帧时间戳和初始视频帧时间戳；

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

将所述最新音频帧时间戳和所述最新视频帧时间戳的较小值与所述初始音频帧时间戳和所述初始视频帧时间戳的较小值之差确定为所述音视频数据的音

视频帧的时间戳时长。

可选的，所述按照所述音视频数据的音视频帧的时间戳的顺序，从所述第一存储空间中读取并解码所述音视频数据的音视频帧，包括：

检测已解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

如果所述最新音频帧时间戳大于等于所述最新视频帧时间戳，则按照所述视频帧的时间戳的顺序，从所述视频存储空间中读取并解码所述音视频的视频帧；

如果所述最新音频帧时间戳小于所述最新视频帧时间戳，则按照所述音频帧的时间戳的顺序，从所述音频存储空间中读取并解码所述音视频数据的音频帧。

可选的，所述基于预设的音视频输出参数对解码后的音视频帧进行重采样处理，包括：

基于预设的标准视频帧率和解码后的视频帧的时间戳确定所述视频帧对应的位置索引；

根据所述位置索引对所述视频帧的时间戳进行更新。

可选的，所述方法还包括：

如果存在多个视频帧对应同一个位置索引，则保留所述多个视频帧中最后进行重采样处理的视频帧，并删除所述多个视频帧中的其它视频帧；

如果所述视频帧对应的位置索引不连续，则确定所有空缺位置索引，并将每个空缺位置索引的相邻位置索引对应的视频帧复制为所述空缺位置索引对应的视频帧。

可选的，所述基于预设的音视频输出参数对解码后的音视频帧进行重采样处理，包括：

基于预设的音频采样率和声道数对解码后的音频帧进行转换；

按照预设的采样点数对所述音频帧进行拆分重组，并将重组后的音频帧的第一个采样点的时间戳确定为所述音频帧的时间戳。

可选的，所述方法还包括：

根据预设的音频采样率、采样点数和音频帧的时间戳确定所述音频帧对应的位置索引；

如果存在多个音频帧对应同一个位置索引，则保留所述多个音频帧中最后

进行重采样处理的音频帧，并删除所述多个音频帧中的其它音频帧；

如果所述音频帧对应的位置索引不连续，则确定所有空缺位置索引，并为每个空缺位置索引对应设置静音帧。

可选的，所述根据重采样后的音视频帧的时间戳生成位置索引，包括：

根据服务启动时间、所述第二存储空间的首个音视频帧的时间戳和对应的存储时间，确定所述目标数据流的时间戳偏移量；

根据所述时间戳偏移量调整重采样后的音视频帧的时间戳；

基于调整后的音视频帧的时间戳重新生成所述音视频帧的位置索引。

可选的，所述通过所述位置索引将所述音视频帧存储在第二存储空间，包括：

将所述第二存储空间分为包括预设数目个缓存位置的音频存储空间和视频存储空间；

根据所述音视频帧的位置索引将所述音视频数据的音频帧和视频帧分别存储至所述音频存储空间和所述视频存储空间。

可选的，所述按照所述位置索引周期性从所述第二存储空间中提取所述音视频帧，包括：

基于预设的音视频输出参数确定音视频帧的标准帧时长；

以所述标准帧时长为间隔，周期性根据当前时间和所述标准帧时长确定待提取索引；

从所述第二存储空间中提取所述位置索引为所述待提取索引的音视频帧。

可选的，所述方法还包括：

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的音频帧，则创建静音帧；

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的视频帧，则复制最近提取的视频帧。

第二方面，提供了一种合成音视频数据流的装置，所述装置包括：

第一存储模块，用于获取目标数据流的音视频数据，将所述音视频数据存储在第一存储空间内；

解码模块，用于按照所述音视频数据的音视频帧的时间戳的顺序，从所述

第一存储空间中读取并解码所述音视频数据的音视频帧；

重采样模块，用于基于预设的音视频输出参数对解码后的音视频帧进行重采样处理；

第二存储模块，用于根据重采样后的音视频帧的时间戳生成位置索引，通过所述位置索引将所述音视频帧存储在第二存储空间；

合成模块，用于按照所述位置索引周期性从所述第二存储空间中提取所述音视频帧，将所述音视频帧与其它数据流的音视频帧进行合成。

可选的，所述第一存储模块，用于：

将所述音视频数据包含的音频数据和视频数据分别存储在第一存储空间的音频存储空间和视频存储空间内。

可选的，所述第一存储模块还用于：

如果所述目标数据流为直播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，删除所述第一存储空间中最早存储的音视频数据，继续存储所述音视频数据；

如果所述目标数据流为点播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，等待所述第一存储空间中存储的音视频数据被消耗后，继续存储所述音视频数据。

可选的，所述第一存储模块还用于：

根据预设的最大播放延迟和最大网络延迟确定所述第一存储空间的大小；

在获取目标数据流的音视频数据过程中，根据检测到的播放延迟需求和/或实时网络延迟调整所述第一存储空间的大小。

可选的，所述解码模块，还用于：

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的时间戳时长；

如果所述时间戳时长大于预设播放速率与读取并解码所述音视频数据所用的物理时长的乘积，则在当前周期暂停从所述第一存储空间读取所述音视频数据的音视频帧。

可选的，所述解码模块，用于：

记录从所述第一存储空间中读取并解码的所述音视频数据的初始音频帧时间戳和初始视频帧时间戳；

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

将所述最新音频帧时间戳和所述最新视频帧时间戳的较小值与所述初始音频帧时间戳和所述初始视频帧时间戳的较小值之差确定为所述音视频数据的音视频帧的时间戳时长。

可选的，所述解码模块，用于：

检测已解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

如果所述最新音频帧时间戳大于等于所述最新视频帧时间戳，则按照所述视频帧的时间戳的顺序，从所述视频存储空间中读取并解码所述音视频的视频帧；

如果所述最新音频帧时间戳小于所述最新视频帧时间戳，则按照所述音频帧的时间戳的顺序，从所述音频存储空间中读取并解码所述音视频数据的音频帧。

可选的，所述重采样模块，用于：

基于预设的标准视频帧率和解码后的视频帧的时间戳确定所述视频帧对应的位置索引；

根据所述位置索引对所述视频帧的时间戳进行更新。

可选的，所述重采样模块，还用于：

如果存在多个视频帧对应同一个位置索引，则保留所述多个视频帧中最后进行重采样处理的视频帧，并删除所述多个视频帧中的其它视频帧；

如果所述视频帧对应的位置索引不连续，则确定所有空缺位置索引，并将每个空缺位置索引的相邻位置索引对应的视频帧复制为所述空缺位置索引对应的视频帧。

可选的，所述重采样模块，用于：

基于预设的音频采样率和声道数对解码后的音频帧进行转换；

按照预设的采样点数对所述音频帧进行拆分重组，并将重组后的音频帧的第一个采样点的时间戳确定为所述音频帧的时间戳。

可选的，所述重采样模块，还用于：

根据预设的音频采样率、采样点数和音频帧的时间戳确定所述音频帧对应的位置索引；

如果存在多个音频帧对应同一个位置索引，则保留所述多个音频帧中最后进行重采样处理的音频帧，并删除所述多个音频帧中的其它音频帧；

如果所述音频帧对应的位置索引不连续，则确定所有空缺位置索引，并为每个空缺位置索引对应设置静音帧。

可选的，所述第二存储模块，用于：

根据服务启动时间、所述第二存储空间的首个音视频帧的时间戳和对应的存储时间，确定所述目标数据流的时间戳偏移量；

根据所述时间戳偏移量调整重采样后的音视频帧的时间戳；

基于调整后的音视频帧的时间戳重新生成所述音视频帧的位置索引。

可选的，所述第二存储模块，用于：

将所述第二存储空间分为包括预设数目个缓存位置的音频存储空间和视频存储空间；

根据所述音视频帧的位置索引将所述音视频数据的音频帧和视频帧分别存储至所述音频存储空间和所述视频存储空间。

可选的，所述合成模块，用于：

基于预设的音视频输出参数确定音视频帧的标准帧时长；

以所述标准帧时长为间隔，周期性根据当前时间和所述标准帧时长确定待提取索引；

从所述第二存储空间中提取所述位置索引为所述待提取索引的音视频帧。

可选的，所述合成模块，还用于：

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的音频帧，则创建静音帧；

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的视频帧，则复制最近提取的视频帧。

本发明实施例提供的技术方案带来的有益效果是：

本发明实施例中，获取目标数据流的音视频数据，将音视频数据存储在第一存储空间内；按照音视频数据的音视频帧的时间戳的顺序，从第一存储空间中读取并解码音视频数据的音视频帧；基于预设的音视频输出参数对解码后的音视频帧进行重采样处理；根据重采样后的音视频帧的时间戳生成位置索引，

通过位置索引将音视频帧存储在第二存储空间；按照位置索引周期性从第二存储空间中提取音视频帧，将音视频帧与其它数据流的音视频帧进行合成。这样，视频合成服务器通过两级的数据缓存，平滑了网络波动的影响，同时引入重采样技术，并利用时间戳和位置索引对音视频数据进行存取，实现了多数据流的同步，可以更加方便有效地实现数据流合成。本发明提供的方法同时能够在输入的各数据流之间、对数据流的拉流处理和合成处理之间进行解耦合，降低了拉流处理对合成处理的影响，即使某路数据流出现拉流问题或者数据流本身出现问题，也不会影响其它路的数据流，能够保证合成的画面和声音的稳定。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 是本发明实施例提供的一种合成音视频数据流的方法流程图；

图 2 是本发明实施例提供的一种重采样后的视频帧示意图；

图 3 是本发明实施例提供的一种重采样后的视频帧示意图；

图 4 是本发明实施例提供的一种重采样后的音频帧示意图；

图 5 是本发明实施例提供的一种第二存储空间的存储示意图；

图 6 是本发明实施例提供的一种合成音视频数据流的装置结构示意图；

图 7 是本发明实施例提供的一种视频合成服务器的结构示意图。

具体实施方式

为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进一步地详细描述。

本发明实施例提供了一种合成音视频数据流的方法，该方法的执行主体可以视频合成服务器，视频合成服务器可以是直播平台的后台服务器，主要用于对多路音视频数据流进行合成输出。视频合成服务器可以同时接收多路音视频数据流的输入，然后针对每路音视频数据流执行初次缓存、解码、重采样、再次缓存的处理，再将处理后的多路音视频数据流进行合成。视频合成服务器可

以支持多个视频处理进程同时运行，每个视频处理进程负责一路音视频数据流的下述处理。视频合成服务器可以包括处理器、存储器、收发器，处理器可以用于进行下述流程中对于合成音视频数据流的处理，存储器可以用于存储处理过程中需要的数据以及产生的数据，如存储音视频数据流的音视频帧、合成配置文件等，收发器可以用于接收和发送处理过程中的相关数据，如用于接收原始音视频数据，输出合成后的音视频数据等。

下面将结合具体实施方式，对图 1 所示的处理流程进行详细的说明，内容可以如下：

步骤 101，获取目标数据流的音视频数据，将音视频数据存储在第一存储空间内。

在实施中，视频合成服务器可以通过网络拉取音视频数据流或者从本地存储设备中读取音视频数据流，然后按照需求将音视频数据流与其它音视频数据流进行合成。具体的，以目标数据流为例，当视频合成服务器中新增了目标数据流，或者目标数据流的拉流地址发生变化，视频合成服务器可以基于目标数据流的拉流地址获取目标数据流的音视频数据，并针对目标数据流创建视频处理进程。之后，视频合成服务器可以先通过视频处理进程将目标数据流的音视频数据存储在第一存储空间内。此处，第一存储空间可以理解为预设的第一级缓存空间，用于存储新拉取到，未经处理的原始数据。此外，如果目标数据流需要延后播放，则可以在第一存储空间内存储延后时间内的帧数据。

可选的，为了便于后续处理，可以将目标数据流的音视频数据进行分离，相应的，步骤 101 的部分处理可以如下：将音视频数据包含的音频数据和视频数据分别存储在第一存储空间的音频存储空间和视频存储空间内。

在实施中，技术人员可以将第一存储空间可以具体分为用于存储音频数据的音频存储空间和用于存储视频数据的视频存储空间，其中，音频存储空间和视频存储空间的结构、属性均一致，仅存储的数据类型不同。这样，视频合成服务器在获取到目标数据流的音视频数据后，可以对音视频数据按照数据类型分离为音频数据和视频数据，然后可以将分离得到的音频数据和视频数据分别存储在第一存储空间的音频存储空间和视频存储空间，从而可以便于后续自由地选择对音频数据或者视频数据进行处理。

可选的，第一存储空间可以支持阻塞模式和覆盖模式两种存储模式，对于

点播流和直播流，可以选用不同的存储模式，相应的处理可以如下：如果目标数据流为直播流，则在第一存储空间存储音视频数据的过程中，当第一存储空间已满时，删除第一存储空间中最早存储的音视频数据，继续存储音视频数据；如果目标数据流为点播流，则在第一存储空间存储音视频数据的过程中，当第一存储空间已满时，等待第一存储空间中存储的音视频数据被消耗后，继续存储音视频数据。

在实施中，视频合成服务器在第一存储空间内存储目标数据流的音视频数据的过程中，可以检测第一存储空间是否已满。如果已满，则可以视目标数据流的数据流类型，采用不同的处理。具体的，如果目标数据流为直播流，则可以选用覆盖模式，即删除第一存储空间中最早存储的音视频数据，继续存储音视频数据，避免输出的音视频数据存在较高的延时；而如果目标数据流为点播流，用户对数据实时性无较大需求，则可以选用阻塞模式，即等待第一存储空间中存储的音视频数据被消耗后，继续存储音视频数据，保证观看的音视频数据连续无间断。另外，当数据流的存储操作被阻塞时，数据流的获取操作也可以同步阻塞，避免继续获取过多的音视频数据。

可选的，可以基于能容忍的网络延迟和播放延迟来设置第一存储空间的大小，相应的处理可以如下：根据预设的最大播放延迟和最大网络延迟确定第一存储空间的大小；在获取目标数据流的音视频数据过程中，根据检测到的播放延迟需求和/或实时网络延迟调整第一存储空间的大小。

在实施中，视频合成服务器可以针对每路音视频数据流设置最大播放延迟（即数据从产生到播放的间隔时间）和最大网络延迟（即前一帧数据到达至后一帧数据到达的间隔时间），然后基于该最大播放延迟和最大网络延迟设置第一存储空间的大小。可以理解，视频合成服务器在第一存储空间存入某一帧数据后，需要处理完前面已存储的所有帧数据，才能开始处理该帧数据，故而如果第一存储空间过大，则播放延迟较高，而如果第一空间过小，则在网络延迟较高的情况下，第一存储空间内已存储的帧数据均已处理完，后续的帧数据却还未到达，将会大大降低音视频数据流输出的连贯性。进一步的，视频合成服务器在获取目标数据流的音视频数据过程中，还可以实时检测播放延迟需求和实时网络延迟，然后根据播放延迟需求和实时网络延迟对第一存储空间的大小进行调整。

步骤 102，按照音视频数据的音视频帧的时间戳的顺序，从第一存储空间中读取并解码音视频数据的音视频帧。

在实施中，由于网络波动等情况的影响，可能会造成数据流获取以及添加到第一存储空间中的速率是波动的，同时音频数据和视频数据之间相同时间戳的帧数据的到达时间也可能不一致，故而视频合成服务器将目标数据流的音视频数据存入第一存储空间后，可以先确定音视频数据的音视频帧的时间戳，然后按照时间戳递增的顺序，从第一存储空间中读取音视频数据的音视频帧，并对音视频帧进行解码。

可选的，可以按照实时标准对音视频数据流的解码速度进行控制，相应的处理可以如下：周期性检测从第一存储空间中读取并解码的音视频数据的音视频帧的时间戳时长；如果时间戳时长大于预设播放速率与读取并解码的所述音视频数据所用的物理时长的乘积，则在当前周期暂停从第一存储空间读取音视频数据的音视频帧。

在实施中，视频合成服务器可以按照一定的速率要求解码音视频数据，即在一定物理时间内解码的音视频数据所对应的时间戳时长应当是固定的，当时间戳时长大于 N 倍物理时间时，意味着解码速度过快，此时应当降低解码速度，视频合成服务器可以适当休眠，此处，N 为设置的音视频数据流的播放速率。故而，视频合成服务器可以周期性检测从第一存储空间中读取并解码的目标数据流的时间戳时长，然后判断该时间戳时长是否大于预设播放速率与读取并解码的所述音视频数据所用的物理时长的乘积，如果大于，则可以在当前周期暂停从第一存储空间中读取音视频帧。例如，开始解码的物理时间为 T1，首个音视频帧的时间戳为 TS1，在解码过程中的预设周期开始时，获取当前物理时间为 T2，已解码音视频帧的时间戳为 TS2，当 T2-T1<TS2-TS1 时，视频合成服务器可以在本周期暂停读取音视频帧。

可选的，上述确定时间戳时长的处理可以具体如下：记录从第一存储空间中读取并解码的音视频数据的初始音频帧时间戳和初始视频帧时间戳；周期性检测从第一存储空间中读取并解码的音视频数据的最新音频帧时间戳和最新视频帧时间戳；将最新音频帧时间戳和最新视频帧时间戳的较小值与初始音频帧时间戳和初始视频帧时间戳的较小值之差确定为音视频数据的音视频帧的时间戳时长。

在实施中，视频合成服务器可以在开始读取并解码目标数据流的音视频数据时，记录音视频数据的初始音频帧时间戳和初始视频帧时间戳。在解码过程中，视频合成服务器可以周期性检测从第一存储空间中读取并解码的音视频数据的最新音频帧时间戳和最新视频帧时间戳。进而，可以分别确定最新音频帧时间戳和最新视频帧时间戳的较小值，以及初始音频帧时间戳和初始视频帧时间戳的较小值，再将两个较小值之差确定为音视频数据的音视频帧的时间戳时长。

可选的，基于上述在第一存储空间中将音频数据和视频数据分开存储的处理，可以按照一定规则，自由地选择从音频存储空间或视频存储空间中读取解码数据，相应的，步骤 102 的处理可以如下：检测已解码的音视频数据的最新音频帧时间戳和最新视频帧时间戳；如果最新音频帧时间戳大于等于最新视频帧时间戳，则按照视频帧的时间戳的顺序，从视频存储空间中读取并解码音视频的视频帧；如果最新音频帧时间戳小于最新视频帧时间戳，则按照音频帧的时间戳的顺序，从音频存储空间中读取并解码音视频数据的音频帧。

在实施中，视频合成服务器从第一存储空间读取并解码音视频帧的过程中，可以检测已解码的音视频数据的最新音频帧时间戳和最新视频帧时间戳，然后比较最新音频帧时间戳和最新视频帧时间戳的大小，如果最新音频帧时间戳大于等于最新视频帧时间戳，则按照视频帧的时间戳的顺序，从视频存储空间中读取并解码音视频数据的视频帧；如果最新音频帧时间戳小于最新视频帧时间戳，则按照音频帧的时间戳的顺序，从音频存储空间中读取并解码音视频数据的音频帧。这样可以使得解码完成的相邻音频帧和视频帧之间的时间戳差距尽可能小，从而保证它们逻辑上所对应的时间是一致的。

步骤 103 ,基于预设的音视频输出参数对解码后的音视频帧进行重采样处理。

在实施中 ,由于不同的数据流具有不同的视频帧率、音频采样率和声道数，视频合成服务器在对目标数据流的音视频帧解码之后，可以根据预先配置的视频帧率和音频采样率等音视频输出参数，对已解码的音视频帧进行重采样处理，从而使得不同数据流经过重采样操作后的音视频帧具有相同的帧率、音频采样率和声道数等参数，从而方便进一步的合成控制操作。

可选的，对于视频帧的重采样处理可以如下：基于预设的标准视频帧率和解码后的视频帧的时间戳确定视频帧对应的位置索引；根据位置索引对视频帧

的时间戳进行更新。

在实施中，视频合成服务器在对目标数据流的视频帧的解码之后，可以先基于预设的标准视频帧率和解码后的视频帧的时间戳确定视频帧对应的位置索引，然后根据位置索引对视频帧的时间戳进行更新。具体的，首先根据预设的标准视频帧率 F_r，计算出一帧视频帧持续的时间 $T_{vf} = 1 / F_r$，然后，将解码后的每一个视频帧的时间戳除以 T_{vf} 并向下取整，得到视频帧对应的位置索引，最后更新视频帧的时间戳为对应的位置索引乘以 T_{vf}。

可选的，如果存在多个视频帧对应同一个位置索引，则保留多个视频帧中最后进行重采样处理的视频帧，并删除多个视频帧中的其它视频帧；如果视频帧对应的位置索引不连续，则确定所有空缺位置索引，并将每个空缺位置索引的相邻位置索引对应的视频帧复制为空缺位置索引对应的视频帧。

在实施中，视频合成服务器在确定视频帧对应的位置索引后，可以对位置索引进行进一步的处理，如果存在多个视频帧对应同一个位置索引，那么则保留这些具有相同位置索引的最后一个视频帧，并删除其它视频帧。如果因为相邻视频帧的时间戳差距比较大而导致位置索引不连续，可以先确定所有空缺位置索引，然后将空缺位置索引的相邻位置索引的视频帧复制为空缺位置索引对应视频帧，使得位置索引连续。

上述处理可以参考图 2 和图 3，图中下方的箭头代表时间方向，平行四边形代表视频帧，四边形内部的数值代表视频帧的时间戳，标准视频帧率是 $F_r=1/T$，则视频帧的持续时间 T_{vf} 为 T。

图 2 给出了帧率为 2/T 的数据流在经过重采样后的视频帧和对应的时间戳。可以看到，时间戳为 T 和 1.5T 的视频帧对应相同的位置索引 1，因此只保留了 1.5T 这个视频帧，同时把 1.5T 的视频帧的时间戳改为了 T。经过重采样处理后，把帧率为 2/T 的视频数据转换为帧率为 1/T 的视频数据。

图 3 给出了帧率接近 1/T 的数据流，在经过重采样后的视频帧和对应的时间戳。可以看到，时间戳为 T 和 1.5T 的视频帧的处理同图 2 一样。时间戳为 6.3T 的视频帧，计算出其位置索引为 6，并将该视频帧的时间戳改为了 6T。同时，输出的视频帧中缺少了位置索引为 5 的视频帧，故而可以将位置索引为 4 的视频帧复制为位置索引 5 对应的视频帧。

可选的，对于音频帧的重采样处理可以如下：基于预设的音频采样率和声

道数对解码后的音频帧进行转换；按照预设的采样点数对音频帧进行拆分重组，并将重组后的音频帧的第一个采样点的时间戳确定为音频帧的时间戳。

在实施中，由于许多音频编码器对于输入的音频帧有固定的采样点数要求，同时为了便于对不同数据流进行音频混音和音效处理，视频合成服务器在对目标数据流的音频帧的解码之后，可以先基于预设的音频采样率和声道数对该解码后的音频帧进行采样率和声道数的转换，再按照预设的采样点数，对音频帧进行拆分重组，并将重组后的音频帧的第一个采样点的时间戳确定为音频帧的时间戳。图 4 给出了预设的采样点数为 1024，对音频帧进行重采样的示意图，其中，箭头处代表音频帧的拆分点，上面一行的平行四边形为重组前的音频帧，下面一行的平行四边形为重组后的音频帧，四边形内部的数字为该音频帧所包含的采样点数。由图 4 可以看出，经过音频采样率、声道数转换后的音频帧可能具有不同的采样点数，拆分重组后每个音频帧包含 1024 个采样点。

可选的，可以同样基于位置索引对重采样后的音频帧进行调整，具体的处理可以如下：根据预设的音频采样率、采样点数和音频帧的时间戳确定音频帧对应的位置索引；如果存在多个音频帧对应同一个位置索引，则保留多个音频帧中最后进行重采样处理的音频帧，并删除多个音频帧中的其它音频帧；如果音频帧对应的位置索引不连续，则确定所有空缺位置索引，并为每个空缺位置索引对应设置静音帧。

在实施中，与视频重采样类似，视频合成服务器可以根据预设的音频采样率、采样点数和音频帧的时间戳确定重采样后的音频帧对应的位置索引。如果存在多个音频帧对应同一个位置索引，那么则保留这些具有相同位置索引的最后一个音频帧，并删除其它音频帧。如果因为相邻音频帧的时间戳差距比较大而导致位置索引不连续，可以先确定所有空缺位置索引，然后为空缺位置索引对应设置静音帧，使得位置索引连续。

步骤 104，根据重采样后的音视频帧的时间戳生成位置索引，通过位置索引将音视频帧存储在第二存储空间。

在实施中，视频合成服务器在对音视频帧进行重采样之后，可以根据重采样的音视频帧的时间戳生成位置索引，然后通过位置索引将音视频帧存储在第二存储空间中，使得多个不同的数据流中具有相同位置索引的音频帧在逻辑上对应相同的物理时间，具有相同位置索引的视频帧在逻辑上也对应相同的物理

时间，再结合重采样后不同数据流的音视频帧具有相同的持续时间，从而方便后续合成时对于不同数据流的同步。此处，第二存储空间可以理解为预设的第二级缓存空间，用于存储已解码并经过重采样的数据。

可选的，对重采样后的音视频帧生成位置索引的方式可以具体如下：根据服务启动时间、第二存储空间的首个音视频帧的时间戳和对应的存储时间，确定目标数据流的时间戳偏移量；根据时间戳偏移量调整重采样后的音视频帧的时间戳；基于调整后的音视频帧的时间戳重新生成音视频帧的位置索引。

在实施中，视频合成服务器在启动视频合成服务时，可以记录当前时间 T_{start}（即服务启动时间），并在第二存储空间存入首个目标数据流的音视频帧时，记录首个音视频帧的时间戳 TS_{in} 和对应的存储时间 T_{curr}，进而可以计算目标数据流的时间戳偏移量 $Off= T_{curr}- T_{start} - TS_{in}$。之后，对于所有存入第二存储空间的音视频帧，视频合成服务器可以将音视频帧的时间戳加上时间戳偏移量 Off，作为该音视频帧新的时间戳，这样，将不同时候获取到的数据流或者具有不同起始时间戳的数据流的音视频帧的时间戳换算为统一的系统时间戳，便于后续的不同数据流之间的同步。接下来，视频合成服务器可以将调整后的音视频帧的时间戳除以该帧的持续时长，并向下取整，从而生成音视频帧的位置索引。

可选的，可以在第二存储空间中将目标数据流的音频数据和视频数据分别进行存储，相应的，步骤 104 的部分处理可以如下：将第二存储空间分为包括预设数目个缓存位置的音频存储空间和视频存储空间；根据音视频帧的位置索引将音视频数据的音频帧和视频帧分别存储至音频存储空间和视频存储空间。

在实施中，视频合成服务器可以将第二存储空间分为包括预设数目个缓存位置的音频存储空间和视频存储空间，每个缓存位置用于存储一帧音频帧或一帧视频帧，音频存储空间和视频存储空间的结构和属性完全一致。之后，视频合成服务器可以根据音视频帧的位置索引，将目标数据流的音频帧和视频帧分别存储至音频存储空间和视频存储空间。具体可以参考图 5，其中下方的矩形格子对应音频存储空间或视频存储空间中的缓存位置，矩形格子中的数值为缓存空间的缓存位置，上方的平行四边形代表音频帧或者视频帧，里面的序号为帧对应的位置索引。当存在 N 个缓存位置时，可以将每个音频帧或者视频帧的位置索引除以 N 后得到的余数作为存储的位置。可见，随着位置索引的增加，添加进第二存储空间的帧是循环覆盖的，并且位置索引相差 kN 的帧位于相同的缓

存位置。

步骤 105，按照位置索引周期性从第二存储空间中提取音视频帧，将音视频帧与其它数据流的音视频帧进行合成。

在实施中，视频合成服务器在第二存储空间中存储了目标数据流的音视频帧之后，可以基于一定的时间间隔，按照音视频帧的位置索引，周期性地从第二存储空间中提取音视频帧，并将提取出的音视频帧与其它数据流的音视频帧进行合成。由于不同数据流的视频帧都具有相同的持续时间，不同数据流的音频帧也都具有了相同的采样率、声道数、采样点数等，因此视频合成服务器一次性从多个不同数据流的第二存储空间中提取到的视频帧或者音频帧为逻辑上相同时间点所对应的帧数据，从而可以实现不同数据流之间的同步，并便于进行如视频画面合成、音频音量调节和混音等进一步的视频合成处理。

可选的，可以参考当前时间从第二存储空间中提取音视频帧， 相应的，步骤 105 的处理可以具体如下：基于预设的音视频输出参数确定音视频帧的标准帧时长；以标准帧时长为间隔，周期性根据当前时间和标准帧时长确定待提取索引；从第二存储空间中提取位置索引为待提取索引的音视频帧。

在实施中，视频合成服务器在从第二存储空间中提取音视频帧的过程中，可以先基于预设的音视频输出参数确定音视频帧的标准帧时长，然后以标准帧时长为间隔，周期性地根据当前时间和标准帧时长确定待提取索引，此处，待提取索引可以为当前时间除以标准帧时长。之后，视频合成服务器可以从第二存储空间中提取位置索引为待提取索引的音视频帧。

可选的，如果从第二存储空间提取不到某个音视频帧，则可以对应进行补帧处理，相应的处理可以如下：如果第二存储空间中不存在位置索引为待提取索引对应的音频帧，则创建静音帧；如果第二存储空间中不存在位置索引为待提取索引对应的视频帧，则复制最近提取的视频帧。

在实施中，当按照位置索引从第二存储空间中获取不到对应的帧时，视频合成服务器则可以进行补帧处理，对于视频帧，可以把最近一次从第二存储空间中提取的视频帧复制一份作为当前获取到的视频帧；对于音频帧，可以创建一个静音帧作为当前获取到的音频帧，其采样点数等于预设的采样点数。

值得一提的是，当数据流网络波动导致第一存储空间内或者第二存储空间内的帧数据耗尽时，如果想在数据输入正常时该数据流的画面能够接上卡屏时

的画面继续播放，则可以在数据耗尽的一段时间后，重新生成上述的初始音频帧时间戳和初始视频帧时间戳和与之关联的开始解码的物理时间，以及第二存储空间的首个音视频帧的时间戳和对应的存储时间；而如果想要直接播放最新的数据流的画面，则无需进行任何操作。

本发明实施例中，获取目标数据流的音视频数据，将音视频数据存储在第一存储空间内；按照音视频数据的音视频帧的时间戳的顺序，从第一存储空间中读取并解码音视频数据的音视频帧；基于预设的音视频输出参数对解码后的音视频帧进行重采样处理；根据重采样后的音视频帧的时间戳生成位置索引，通过位置索引将音视频帧存储在第二存储空间；按照位置索引周期性从第二存储空间中提取音视频帧，将音视频帧与其它数据流的音视频帧进行合成。这样，视频合成服务器通过两级的数据缓存，平滑了网络波动的影响，同时引入重采样技术，并利用时间戳和位置索引对音视频数据进行存取，实现了多数据流的同步，可以更加方便有效地实现数据流合成。

此外，各数据流的拉流处理和合成处理之间，仅在第二存储空间处出现交集，没有其它额外的互相之间的依赖或者控制。比如需要拉取 4 路数据流，进行类似 4 个窗口的画面模式的合成，可以 4 路数据流分别独立地进行拉流、解码，以及放入各自的第二级缓存。各路数据流之间可以保证相对独立的输入和运行，不会因为其中一路出现拉流卡顿或者其它问题，导致其它数据流的输入出现问题。因此，本发明提供的方法同时能够在输入的各数据流之间、对数据流的拉流处理和合成处理之间进行解耦合，降低了拉流处理对合成处理的影响，即使某路数据流出现拉流问题或者数据流本身出现问题，也不会影响其它路的数据流，能够保证合成的画面和声音的稳定。

基于相同的技术构思，本发明实施例还提供了一种合成音视频数据流的装置，如图 6 所示，所述装置包括

第一存储模块 601，用于获取目标数据流的音视频数据，将所述音视频数据存储在第一存储空间内；

解码模块 602，用于按照所述音视频数据的音视频帧的时间戳的顺序，从所述第一存储空间中读取并解码所述音视频数据的音视频帧；

重采样模块 603，用于基于预设的音视频输出参数对解码后的音视频帧进行

重采样处理；

第二存储模块 604，用于根据重采样后的音视频帧的时间戳生成位置索引，通过所述位置索引将所述音视频帧存储在第二存储空间；

合成模块 605 ,用于按照所述位置索引周期性从所述第二存储空间中提取所述音视频帧，将所述音视频帧与其它数据流的音视频帧进行合成。

可选的，所述第一存储模块 601，用于：

将所述音视频数据包含的音频数据和视频数据分别存储在第一存储空间的音频存储空间和视频存储空间内。

可选的，所述第一存储模块 601 还用于：

如果所述目标数据流为直播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，删除所述第一存储空间中最早存储的音视频数据，继续存储所述音视频数据；

如果所述目标数据流为点播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，等待所述第一存储空间中存储的音视频数据被消耗后，继续存储所述音视频数据。

可选的，所述第一存储模块 601 还用于：

根据预设的最大播放延迟和最大网络延迟确定所述第一存储空间的大小；

在获取目标数据流的音视频数据过程中，根据检测到的播放延迟需求和/或实时网络延迟调整所述第一存储空间的大小。

可选的，所述解码模块 602，还用于：

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的音视频帧的时间戳时长；

如果所述时间戳时长大于预设播放速率与读取并解码的所述音视频数据所用的物理时长的乘积，则在当前周期暂停从所述第一存储空间读取所述音视频数据的音视频帧。

可选的，所述解码模块 602，用于：

记录从所述第一存储空间中读取并解码的所述音视频数据的初始音频帧时间戳和初始视频帧时间戳；

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

将所述最新音频帧时间戳和所述最新视频帧时间戳的较小值与所述初始音频帧时间戳和所述初始视频帧时间戳的较小值之差确定为所述音视频数据的音视频帧的时间戳时长。

可选的，所述解码模块 602，用于：

检测已解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

如果所述最新音频帧时间戳大于等于所述最新视频帧时间戳，则按照所述视频帧的时间戳的顺序，从所述视频存储空间中读取并解码所述音视频的视频帧；

如果所述最新音频帧时间戳小于所述最新视频帧时间戳，则按照所述音频帧的时间戳的顺序，从所述音频存储空间中读取并解码所述音视频数据的音频帧。

可选的，所述重采样模块 603，用于：

基于预设的标准视频帧率和解码后的视频帧的时间戳确定所述视频帧对应的位置索引；

根据所述位置索引对所述视频帧的时间戳进行更新。

可选的，所述重采样模块 603，还用于：

如果存在多个视频帧对应同一个位置索引，则保留所述多个视频帧中最后进行重采样处理的视频帧，并删除所述多个视频帧中的其它视频帧；

如果所述视频帧对应的位置索引不连续，则确定所有空缺位置索引，并将每个空缺位置索引的相邻位置索引对应的视频帧复制为所述空缺位置索引对应的视频帧。

可选的，所述重采样模块 603，用于：

基于预设的音频采样率和声道数对解码后的音频帧进行转换；

按照预设的采样点数对所述音频帧进行拆分重组，并将重组后的音频帧的第一个采样点的时间戳确定为所述音频帧的时间戳。

可选的，所述重采样模块 603，还用于：

根据预设的音频采样率、采样点数和音频帧的时间戳确定所述音频帧对应的位置索引；

如果存在多个音频帧对应同一个位置索引，则保留所述多个音频帧中最后进行重采样处理的音频帧，并删除所述多个音频帧中的其它音频帧；

如果所述音频帧对应的位置索引不连续，则确定所有空缺位置索引，并为每个空缺位置索引对应设置静音帧。

可选的，所述第二存储模块 604，用于：

根据服务启动时间、所述第二存储空间的首个音视频帧的时间戳和对应的存储时间，确定所述目标数据流的时间戳偏移量；

根据所述时间戳偏移量调整重采样后的音视频帧的时间戳；

基于调整后的音视频帧的时间戳重新生成所述音视频帧的位置索引。

可选的，所述第二存储模块 604，用于：

将所述第二存储空间分为包括预设数目个缓存位置的音频存储空间和视频存储空间；

根据所述音视频帧的位置索引将所述音视频数据的音频帧和视频帧分别存储至所述音频存储空间和所述视频存储空间。

可选的，所述合成模块 605，用于：

基于预设的音视频输出参数确定音视频帧的标准帧时长；

以所述标准帧时长为间隔，周期性根据当前时间和所述标准帧时长确定待提取索引；

从所述第二存储空间中提取所述位置索引为所述待提取索引的音视频帧。

可选的，所述合成模块 605，还用于：

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的音频帧，则创建静音帧；

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的视频帧，则复制最近提取的视频帧。

本发明实施例中，获取目标数据流的音视频数据，将音视频数据存储在第一存储空间内；按照音视频数据的音视频帧的时间戳的顺序，从第一存储空间中读取并解码音视频数据的音视频帧；基于预设的音视频输出参数对解码后的音视频帧进行重采样处理；根据重采样后的音视频帧的时间戳生成位置索引，通过位置索引将音视频帧存储在第二存储空间；按照位置索引周期性从第二存储空间中提取音视频帧，将音视频帧与其它数据流的音视频帧进行合成。这样，视频合成服务器通过两级的数据缓存，平滑了网络波动的影响，同时引入重采样技术，并利用时间戳和位置索引对音视频数据进行存取，实现了多数据流的

同步，可以更加方便有效地实现数据流合成。本发明提供的方法同时能够在输入的各数据流之间、对数据流的拉流处理和合成处理之间进行解耦合，降低了拉流处理对合成处理的影响，即使某路数据流出现拉流问题或者数据流本身出现问题，也不会影响其它路的数据流，能够保证合成的画面和声音的稳定。

需要说明的是：上述实施例提供的合成音视频数据流的装置在合成音视频数据流时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的合成音视频数据流的装置与合成音视频数据流的方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

图 7 是本发明实施例提供的视频合成服务器的结构示意图。该视频合成服务器 700 可因配置或性能不同而产生比较大的差异，可以包括一个或一个以上中央处理器 722（例如，一个或一个以上处理器）和存储器 732，一个或一个以上存储应用程序 772 或数据 777 的存储介质 730（例如一个或一个以上海量存储设备）。其中，存储器 732 和存储介质 730 可以是短暂存储或持久存储。存储在存储介质 730 的程序可以包括一个或一个以上模块（图示没标出），每个模块可以包括对视频合成服务器中的一系列指令操作。更进一步地，中央处理器 722 可以设置为与存储介质 730 通信，在视频合成服务器 700 上执行存储介质 730 中的一系列指令操作。

视频合成服务器 700 还可以包括一个或一个以上电源 726，一个或一个以上有线或无线网络接口 750，一个或一个以上输入输出接口 758，一个或一个以上键盘 756，和/或，一个或一个以上操作系统 771，例如 Windows Server，Mac OS X，Unix，Linux，FreeBSD 等等。

视频合成服务器 700 可以包括有存储器，以及一个或者一个以上的程序，其中一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行上述合成音视频数据流的指令。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

5

以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权 利 要 求 书

1、一种合成音视频数据流的方法，其特征在于，所述方法包括：

获取目标数据流的音视频数据，将所述音视频数据存储在第一存储空间内；

按照所述音视频数据的音视频帧的时间戳的顺序，从所述第一存储空间中读取并解码所述音视频数据的音视频帧；

基于预设的音视频输出参数对解码后的音视频帧进行重采样处理；

根据重采样后的音视频帧的时间戳生成位置索引，通过所述位置索引将所述音视频帧存储在第二存储空间；

按照所述位置索引周期性从所述第二存储空间中提取所述音视频帧，将所述音视频帧与其它数据流的音视频帧进行合成。

2、根据权利要求 1 所述的方法，其特征在于，所述将所述音视频数据存储在第一存储空间内，包括：

将所述音视频数据包含的音频数据和视频数据分别存储在第一存储空间的音频存储空间和视频存储空间内。

3、根据权利要求 1 所述的方法，其特征在于，所述方法还包括：

如果所述目标数据流为直播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，删除所述第一存储空间中最早存储的音视频数据，继续存储所述音视频数据；

如果所述目标数据流为点播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，等待所述第一存储空间中存储的音视频数据被消耗后，继续存储所述音视频数据。

4、根据权利要求 1 所述的方法，其特征在于，所述方法还包括：

根据预设的最大播放延迟和最大网络延迟确定所述第一存储空间的大小；

在获取目标数据流的音视频数据过程中，根据检测到的播放延迟需求和/或实时网络延迟调整所述第一存储空间的大小。

5、根据权利要求 1 所述的方法，其特征在于，所述方法还包括：

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的音视频帧的时间戳时长；

如果所述时间戳时长大于预设播放速率与读取并解码的所述音视频数据所用的物理时长的乘积，则在当前周期暂停从所述第一存储空间读取所述音视频数据的音视频帧。

6、根据权利要求 5 所述的方法，其特征在于，所述周期性检测从所述第一存储空间中读取并解码的所述音视频数据的音视频帧的时间戳时长，包括：

记录从所述第一存储空间中读取并解码的所述音视频数据的初始音频帧时间戳和初始视频帧时间戳；

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

将所述最新音频帧时间戳和所述最新视频帧时间戳的较小值与所述初始音频帧时间戳和所述初始视频帧时间戳的较小值之差确定为所述音视频数据的音视频帧的时间戳时长。

7、根据权利要求 2 所述的方法，其特征在于，所述按照所述音视频数据的音视频帧的时间戳的顺序，从所述第一存储空间中读取并解码所述音视频数据的音视频帧，包括：

检测已解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

如果所述最新音频帧时间戳大于等于所述最新视频帧时间戳，则按照所述视频帧的时间戳的顺序，从所述视频存储空间中读取并解码所述音视频的视频帧；

如果所述最新音频帧时间戳小于所述最新视频帧时间戳，则按照所述音频帧的时间戳的顺序，从所述音频存储空间中读取并解码所述音视频数据的音频帧。

8、根据权利要求 1 所述的方法，其特征在于，所述基于预设的音视频输出参数对解码后的音视频帧进行重采样处理，包括：

基于预设的标准视频帧率和解码后的视频帧的时间戳确定所述视频帧对应的位置索引；

根据所述位置索引对所述视频帧的时间戳进行更新。

9、根据权利要求 8 所述的方法，其特征在于，所述方法还包括：

如果存在多个视频帧对应同一个位置索引，则保留所述多个视频帧中最后进行重采样处理的视频帧，并删除所述多个视频帧中的其它视频帧；

如果所述视频帧对应的位置索引不连续，则确定所有空缺位置索引，并将每个空缺位置索引的相邻位置索引对应的视频帧复制为所述空缺位置索引对应的视频帧。

10、根据权利要求 1 所述的方法，其特征在于，所述基于预设的音视频输出参数对解码后的音视频帧进行重采样处理，包括：

基于预设的音频采样率和声道数对解码后的音频帧进行转换；

按照预设的采样点数对所述音频帧进行拆分重组，并将重组后的音频帧的第一个采样点的时间戳确定为所述音频帧的时间戳。

11、根据权利要求 10 所述的方法，其特征在于，所述方法还包括：

根据预设的音频采样率、采样点数和音频帧的时间戳确定所述音频帧对应的位置索引；

如果存在多个音频帧对应同一个位置索引，则保留所述多个音频帧中最后进行重采样处理的音频帧，并删除所述多个音频帧中的其它音频帧；

如果所述音频帧对应的位置索引不连续，则确定所有空缺位置索引，并为每个空缺位置索引对应设置静音帧。

12、根据权利要求 1 所述的方法，其特征在于，所述根据重采样后的音视频帧的时间戳生成位置索引，包括：

根据服务启动时间、所述第二存储空间的首个音视频帧的时间戳和对应的存储时间，确定所述目标数据流的时间戳偏移量；

根据所述时间戳偏移量调整重采样后的音视频帧的时间戳；

基于调整后的音视频帧的时间戳重新生成所述音视频帧的位置索引。

13、根据权利要求 12 所述的方法，其特征在于，所述通过所述位置索引将所述音视频帧存储在第二存储空间，包括：

将所述第二存储空间分为包括预设数目个缓存位置的音频存储空间和视频存储空间；

根据所述音视频帧的位置索引将所述音视频数据的音频帧和视频帧分别存储至所述音频存储空间和所述视频存储空间。

14、根据权利要求 1 所述的方法，其特征在于，所述按照所述位置索引周期性从所述第二存储空间中提取所述音视频帧，包括：

基于预设的音视频输出参数确定音视频帧的标准帧时长；

以所述标准帧时长为间隔，周期性根据当前时间和所述标准帧时长确定待提取索引；

从所述第二存储空间中提取所述位置索引为所述待提取索引的音视频帧。

15、根据权利要求 14 所述的方法，其特征在于，所述方法还包括：

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的音频帧，则创建静音帧；

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的视频帧，则复制最近提取的视频帧。

16、一种合成音视频数据流的装置，其特征在于，所述装置包括：

第一存储模块，用于获取目标数据流的音视频数据，将所述音视频数据存储在第一存储空间内；

解码模块，用于按照所述音视频数据的音视频帧的时间戳的顺序，从所述第一存储空间中读取并解码所述音视频数据的音视频帧；

重采样模块，用于基于预设的音视频输出参数对解码后的音视频帧进行重采样处理；

第二存储模块，用于根据重采样后的音视频帧的时间戳生成位置索引，通

过所述位置索引将所述音视频帧存储在第二存储空间；

合成模块，用于按照所述位置索引周期性从所述第二存储空间中提取所述音视频帧，将所述音视频帧与其它数据流的音视频帧进行合成。

17、根据权利要求 16 所述的装置，其特征在于，所述第一存储模块，用于：

将所述音视频数据包含的音频数据和视频数据分别存储在第一存储空间的音频存储空间和视频存储空间内。

18、根据权利要求 16 所述的装置，其特征在于，所述第一存储模块还用于：

如果所述目标数据流为直播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，删除所述第一存储空间中最早存储的音视频数据，继续存储所述音视频数据；

如果所述目标数据流为点播流，则在所述第一存储空间存储所述音视频数据的过程中，当所述第一存储空间已满时，等待所述第一存储空间中存储的音视频数据被消耗后，继续存储所述音视频数据。

19、根据权利要求 16 所述的装置，其特征在于，所述第一存储模块还用于：

根据预设的最大播放延迟和最大网络延迟确定所述第一存储空间的大小；

在获取目标数据流的音视频数据过程中，根据检测到的播放延迟需求和/或实时网络延迟调整所述第一存储空间的大小。

20、根据权利要求 16 所述的装置，其特征在于，所述解码模块，还用于：

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的音视频帧的时间戳时长；

如果所述时间戳时长大于预设播放速率与读取并解码的所述音视频数据所用的物理时长的乘积，则在当前周期暂停从所述第一存储空间读取所述音视频数据的音视频帧。

21、根据权利要求 20 所述的装置，其特征在于，所述解码模块，用于：

记录从所述第一存储空间中读取并解码的所述音视频数据的初始音频帧时

间戳和初始视频帧时间戳；

周期性检测从所述第一存储空间中读取并解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

将所述最新音频帧时间戳和所述最新视频帧时间戳的较小值与所述初始音频帧时间戳和所述初始视频帧时间戳的较小值之差确定为所述音视频数据的音视频帧的时间戳时长。

22、根据权利要求 17 所述的装置，其特征在于，所述解码模块，用于：

检测已解码的所述音视频数据的最新音频帧时间戳和最新视频帧时间戳；

如果所述最新音频帧时间戳大于等于所述最新视频帧时间戳，则按照所述视频帧的时间戳的顺序，从所述视频存储空间中读取并解码所述音视频的视频帧；

如果所述最新音频帧时间戳小于所述最新视频帧时间戳，则按照所述音频帧的时间戳的顺序，从所述音频存储空间中读取并解码所述音视频数据的音频帧。

23、根据权利要求 16 所述的装置，其特征在于，所述重采样模块，用于：

基于预设的标准视频帧率和解码后的视频帧的时间戳确定所述视频帧对应的位置索引；

根据所述位置索引对所述视频帧的时间戳进行更新。

24、根据权利要求 23 所述的装置，其特征在于，所述重采样模块，还用于：

如果存在多个视频帧对应同一个位置索引，则保留所述多个视频帧中最后进行重采样处理的视频帧，并删除所述多个视频帧中的其它视频帧；

如果所述视频帧对应的位置索引不连续，则确定所有空缺位置索引，并将每个空缺位置索引的相邻位置索引对应的视频帧复制为所述空缺位置索引对应的视频帧。

25、根据权利要求 16 所述的装置，其特征在于，所述重采样模块，用于：

基于预设的音频采样率和声道数对解码后的音频帧进行转换；

按照预设的采样点数对所述音频帧进行拆分重组，并将重组后的音频帧的第一个采样点的时间戳确定为所述音频帧的时间戳。

26、根据权利要求 25 所述的装置，其特征在于，所述重采样模块，还用于：

根据预设的音频采样率、采样点数和音频帧的时间戳确定所述音频帧对应的位置索引；

如果存在多个音频帧对应同一个位置索引，则保留所述多个音频帧中最后进行重采样处理的音频帧，并删除所述多个音频帧中的其它音频帧；

如果所述音频帧对应的位置索引不连续，则确定所有空缺位置索引，并为每个空缺位置索引对应设置静音帧。

27、根据权利要求 16 所述的装置，其特征在于，所述第二存储模块，用于：

根据服务启动时间、所述第二存储空间的首个音视频帧的时间戳和对应的存储时间，确定所述目标数据流的时间戳偏移量；

根据所述时间戳偏移量调整重采样后的音视频帧的时间戳；

基于调整后的音视频帧的时间戳重新生成所述音视频帧的位置索引。

28、根据权利要求 27 所述的装置，其特征在于，所述第二存储模块，用于：

将所述第二存储空间分为包括预设数目个缓存位置的音频存储空间和视频存储空间；

根据所述音视频帧的位置索引将所述音视频数据的音频帧和视频帧分别存储至所述音频存储空间和所述视频存储空间。

29、根据权利要求 16 所述的装置，其特征在于，所述合成模块，用于：

基于预设的音视频输出参数确定音视频帧的标准帧时长；

以所述标准帧时长为间隔，周期性根据当前时间和所述标准帧时长确定待提取索引；

从所述第二存储空间中提取所述位置索引为所述待提取索引的音视频帧。

30、根据权利要求 29 所述的装置，其特征在于，所述合成模块，还用于：

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的音频帧，则创建静音帧；

如果所述第二存储空间中不存在所述位置索引为所述待提取索引对应的视频帧，则复制最近提取的视频帧。

5

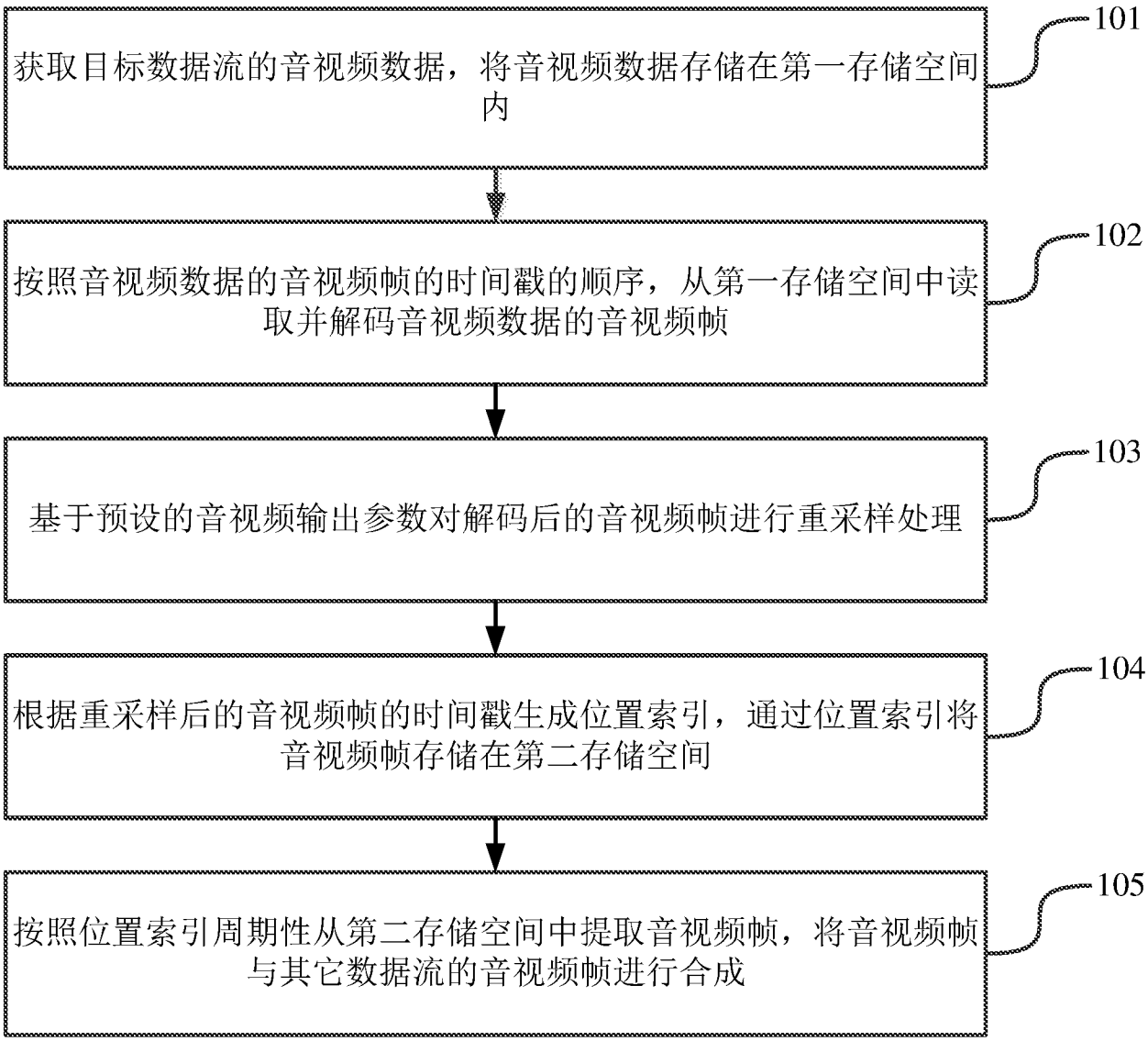

获取目标数据流的音视频数据，将音视频数据存储在第一存储空间内
101
按照音视频数据的音视频帧的时间戳的顺序，从第一存储空间中读取并解码音视频数据的音视频帧
102
基于预设的音视频输出参数对解码后的音视频帧进行重采样处理
103
根据重采样后的音视频帧的时间戳生成位置索引，通过位置索引将音视频帧存储在第二存储空间
104
按照位置索引周期性从第二存储空间中提取音视频帧，将音视频帧与其它数据流的音视频帧进行合成
105

图 1

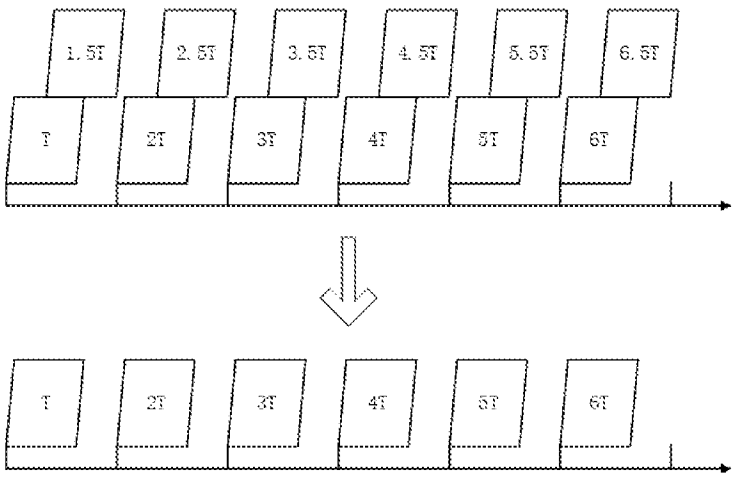

1.5T
2.5T
3.5T
4.5T
5.5T
6.5T
1T
2T
3T
4T
5T
6T
1T
2T
3T
4T
5T
6T

图 2

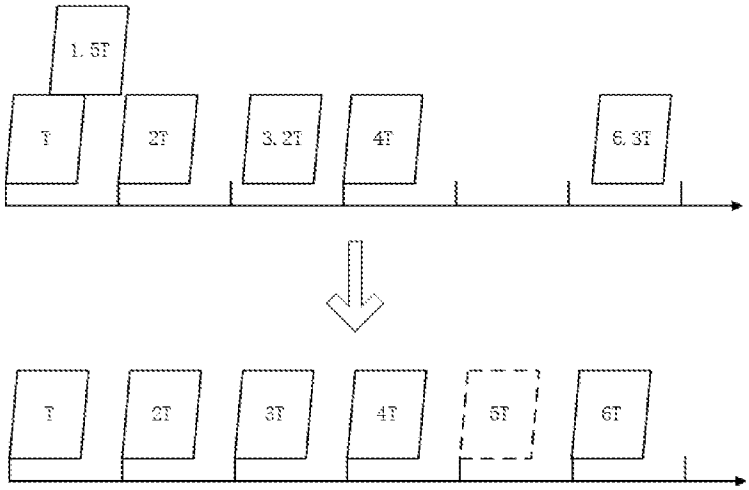

1.5T
T
2T
3.2T
4T
6.3T
T
2T
3T
4T
5T
6T

图 3

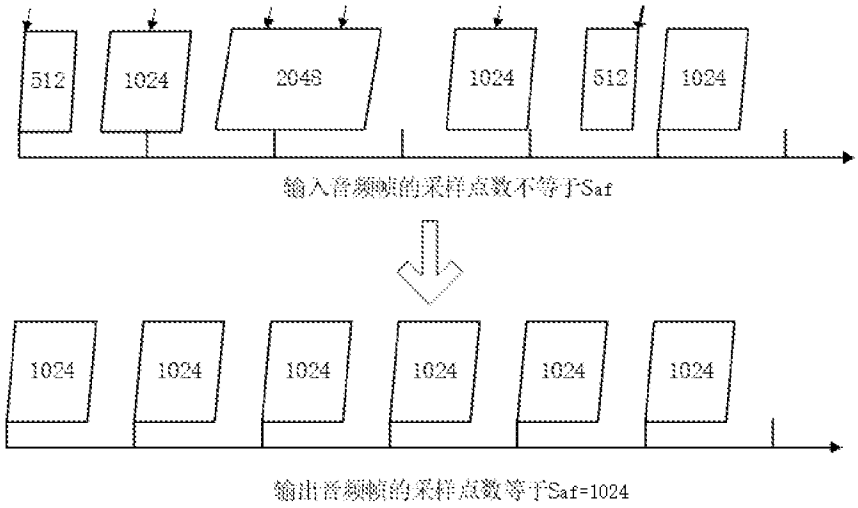

512
1024
2048
1024
512
1024
输入音频帧的采样点数不等于Saf
1024
1024
1024
1024
1024
1024
输出音频帧的采样点数等于Saf=1024

图 4

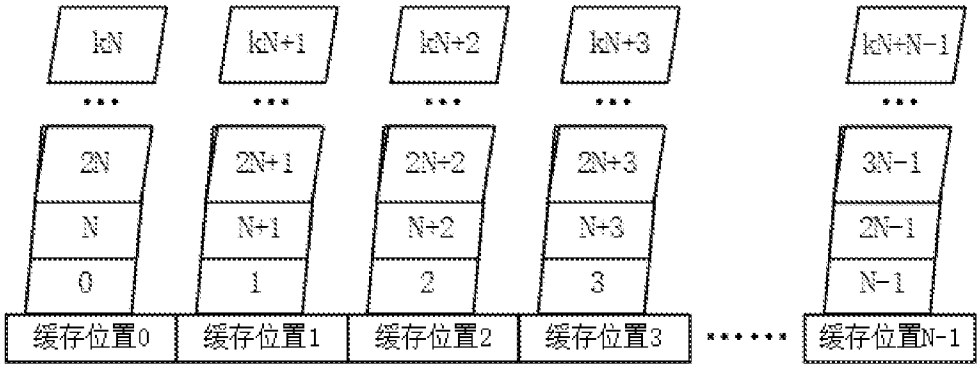

kN
kN+1
kN+2
kN+3
kN+N-1
2N
N
0
2N+1
N+1
1
2N+2
N+2
2
2N+3
N+3
3
3N-1
2N-1
N-1
缓存位置0
缓存位置1
缓存位置2
缓存位置3
缓存位置N-1

图 5

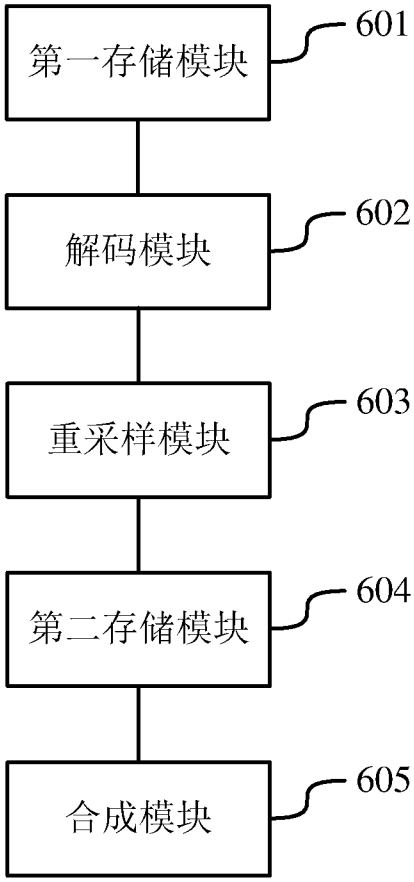


图 6

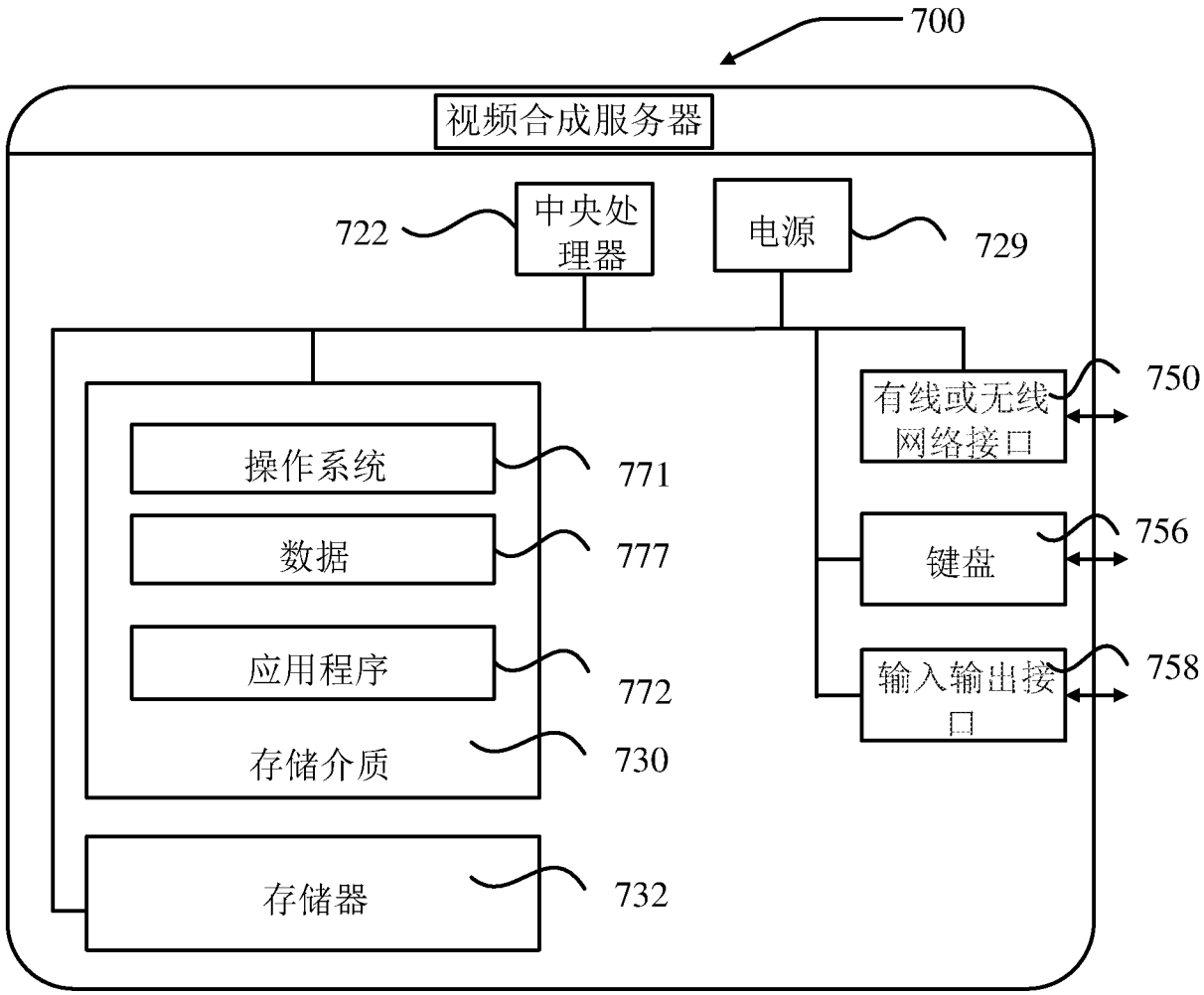

700
视频合成服务器
722
中央处理器
电源
729
750
有线或无线网络接口
操作系统
771
数据
777
键盘
756
应用程序
772
输入输出接口
758
存储介质
730
存储器
732

图 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/091206

A. CLASSIFICATION OF SUBJECT MATTER

H04N 21/433(2011.01)i; H04N 21/4402(2011.01)i; H04N 21/8547(2011.01)i; H04N 21/439(2011.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N,G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

VEN; CNABS; CNTXT; CNKI: 重采样, 上采样, 下采样, 升采样, 降采样, 二次, 采样, 取样, 合成, 同步, 拼接, 合并, 混合, 多路, 多通道, 缓存, 缓冲, 存储, 视频, 时间戳, 位置, 索引, 指针, 更新, re-sample, resample, sample, up-sample, down-sample, second, syhthsize, compose, synchronize, multiple, channel, buffer, storage, video, time stamp, location, position, index, pointer

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 101472090 A (CHINA DIGITAL VIDEO (BEIJING) LIMITED) 01 July 2009 (2009-07-01) description, pp. 4-9, and figure 7	1-30
Y	CN 1407831 A (KONINKLIJKE PHILIPS ELECTRONICS N.V.) 02 April 2003 (2003-04-02) description, p. 2, line 26 to p. 4, line 6, and figure 2	1-30
Y	CN 107613409 A (COMBA TELECOM SYSTEMS (CHINA) CO., LTD. ET AL.) 19 January 2018 (2018-01-19) description, paragraphs 0123-0132	1-30
Y	WO 2008084179 A1 (NDS LTD. ET AL.) 17 July 2008 (2008-07-17) abstract, and description, p. 10, line 24 to p. 13, line 30	4, 19

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 January 2019	**28 February 2019**

Name and mailing address of the ISA/CN	Authorized officer
State Intellectual Property Office of the P. R. China (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/091206

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 101472090 A	01 July 2009	CN 101472090 B	28 December 2011
CN 1407831 A	02 April 2003	KR 20030013286 A	14 February 2003
		US 2003034982 A1	20 February 2003
		JP 2003076982 A	14 March 2003
		EP 1282078 A1	05 February 2003
CN 107613409 A	19 January 2018	None	
WO 2008084179 A1	17 July 2008	None	

国际检索报告

国际申请号
PCT/CN2018/091206

A. 主题的分类

H04N 21/433(2011.01)i; H04N 21/4402(2011.01)i; H04N 21/8547(2011.01)i; H04N 21/439(2011.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

H04N,G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

VEN;CNABS;CNTXT;CNKI:重采样，上采样，下采样，升采样，降采样，二次，采样，取样，合成，同步，拼接，合并，混合，多路，多通道，缓存，缓冲，存储，视频，时间戳，位置，索引，指针，更新，re-sample, resample, sample,up-sample, down-sample, second, syhthsize, compose, synchronize, multiple, channel, buffer, storage, video, time stamp, location, position, index, pointer

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	CN 101472090 A (新奥特北京视频技术有限公司) 2009年 7月 1日 (2009-07-01) 说明书第4-9页，图7	1-30
Y	CN 1407831 A (皇家菲利浦电子有限公司) 2003年 4月 2日 (2003-04-02) 说明书第2页第26行-第4页第6行，图2	1-30
Y	CN 107613409 A (京信通信系统中国有限公司等) 2018年 1月 19日 (2018-01-19) 说明书第0123-0132段	1-30
Y	WO 2008084179 A1 (NDS LTD等) 2008年 7月 17日 (2008-07-17) 说明书摘要、说明书第10页第24行-第13页第30行	4，19

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2019年 1月 31日	2019年 2月 28日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 高巍巍 电话号码 86-（010）-62089149

表 PCT/ISA/210 (第2页) (2015年1月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/091206

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	101472090	A	2009年 7月 1日	CN	101472090	B	2011年 12月 28日
CN	1407831	A	2003年 4月 2日	KR	20030013286	A	2003年 2月 14日
				US	2003034982	A1	2003年 2月 20日
				JP	2003076982	A	2003年 3月 14日
				EP	1282078	A1	2003年 2月 5日
CN	107613409	A	2018年 1月 19日		无		
WO	2008084179	A1	2008年 7月 17日		无		