



(12) 发明专利申请

(10) 申请公布号 CN 104980775 A

(43) 申请公布日 2015. 10. 14

(21) 申请号 201510144494. 5

H04N 21/858(2011. 01)

(22) 申请日 2015. 03. 30

H04N 21/433(2011. 01)

(30) 优先权数据

14305476. 5 2014. 04. 01 EP

(71) 申请人 汤姆逊许可公司

地址 法国伊西莱穆利欧市

(72) 发明人 吉勒斯·斯特劳布

尼古拉斯·勒·斯科阿尼克

克里斯托夫·诺依曼 史蒂芬·欧诺

(74) 专利代理机构 北京东方亿思知识产权代理

有限责任公司 11258

代理人 李晓冬

(51) Int. Cl.

H04N 21/262(2011. 01)

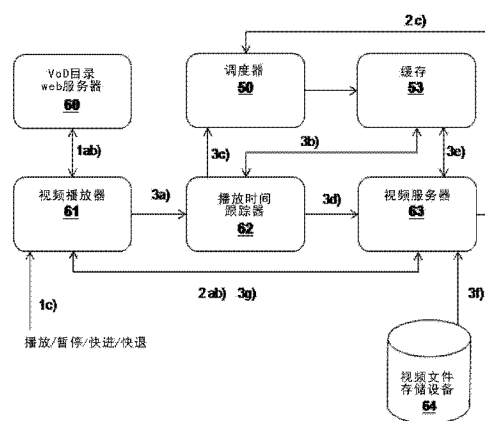
权利要求书3页 说明书12页 附图8页

(54) 发明名称

视频流式传输的方法、相应的设备及系统

(57) 摘要

本申请涉及视频流式传输的方法、相应的设备及系统。将视频以区块的形式提供给播放器。需要计算一些区块,因为它们被定位给特定用户。通过从相关联的队列弹出作业的云计算实例来完成对目标区块的计算。由于播放器使用多速模式,因此可能比最初预计的更早或更晚需要视频区块。为了将目标区块在正确的时间递送至播放器,播放时间跟踪器定期将超载区块请求插入到清单文件中。超载请求包括对未来需要的目标区块的引用以及这些区块的预期播出时间。对超载请求进行验证,并且如果确定未来需要的目标区块尚未被计算,则用于验证的消息被发送至调度器。该调度器对是否需要采取测量来确保及时递送目标视频区块进行验证。



1. 一种向视频播放器设备 (61) 进行视频流式传输的方法, 其中, 该方法包括:

从所述视频播放器设备接收 (71) 将视频流式传输至所述视频播放器设备的请求;

响应于将视频进行流式传输的所述请求, 将视频区块列表发送 (72) 至所述视频播放器设备, 所述视频区块列表包括用于播放所述视频的视频区块的资源定位符, 所述视频区块列表包括第一资源定位符类型的资源定位符和第二资源定位符类型的资源定位符, 所述第一资源定位符类型引用单个视频区块, 并且所述第二资源定位符类型包括: 引用第一视频区块的第一资源定位符、引用第二视频区块的第二资源定位符、以及所述第二视频区块的预期播出延时;

从所述视频播放器设备接收 (73) 对来自所述视频区块列表的视频区块进行流式传输的请求;

如果被包括在对视频区块进行流式传输的所述请求中的资源定位符为所述第一资源定位符类型, 则将所述单个视频区块发送 (75) 至所述视频播放器设备;

如果被包括在对视频区块进行流式传输的所述请求中的资源定位符为所述第二资源定位符类型, 则将所述第一视频区块发送 (76) 至所述视频播放器设备, 并且如果所述第二视频区块不可用, 则发送 (78) 对所述第二视频区块进行计算的请求, 对所述第二视频区块进行计算的所述请求包括所述预期播出延时。

2. 如权利要求 1 所述的方法, 其中, 对所述第二视频区块进行计算的所述请求被发送至用于对计算视频区块进行调度的调度器 (50), 所述调度器包括用于计算视频区块的多个作业队列 (51), 所述多个作业队列中的每个作业队列由特定分类的计算实例 (52) 进行服务, 并且每个计算实例将所计算的视频区块存储于缓存存储器 (54) 中, 并且所述调度器将用于计算视频区块的作业归属到根据所述相关联的预期播出延时而选择的特定分类的队列。

3. 如权利要求 2 所述的方法, 其中, 当接收到用于计算的所述请求时, 所述调度器对与对所述第二视频区块的所述引用相对应的作业是否被预计为在所述预期播出延时而返回计算结果进行验证, 如果否, 则将与对所述第二视频区块的所述引用相对应的所述作业的副本作业插入到分类紧急的队列中。

4. 如权利要求 2 所述的方法, 其中, 当接收到用于计算的所述请求时, 所述调度器对与对所述第二视频区块的所述引用相对应的作业是否被预计为在所述预期播出延时而返回计算结果进行验证。如果否, 则将与对所述第二视频区块的所述引用相对应的所述作业移入分类紧急的队列中。

5. 如权利要求 2-4 中任一权利要求所述的方法, 其中, 该方法还包括根据作业队列尺寸而添加或移除对每个作业队列进行服务的计算实例。

6. 如权利要求 2-4 中任一权利要求所述的方法, 其中, 作业队列分类至少包括尽力而为队列和紧急队列。

7. 如权利要求 2-4 中任一权利要求所述的方法, 其中, 所述计算包括利用将所述视频播放器设备的用户作为目标的内容对视频区块的图像帧中的像素区域进行覆盖。

8. 一种设备 (800, 62), 其中, 所述设备包括:

处理器 (801);

发送机 (802);

接收机 (803) ;

所述接收机被配置为 : 从视频播放器设备接收将视频流式传输至所述视频播放器设备的请求 ;

所述发送机被配置为 : 响应于将视频进行流式传输的所述请求, 将视频区块列表发送至所述视频播放器设备, 所述视频区块列表包括所请求的视频的视频区块的资源定位符, 所述视频区块列表包括第一资源定位符类型的资源定位符和第二资源定位符类型的资源定位符, 所述第一资源定位符类型引用单个视频区块, 并且所述第二资源定位符类型包括 : 引用第一视频区块的第一资源定位符、引用第二视频区块的第二资源定位符、以及所述第二视频区块的预期播出延时 ;

所述接收机还被配置为 : 从所述视频播放器设备接收对来自所述视频区块列表的视频区块进行流式传输的请求 ;

所述处理器被配置为确定被包括在对视频区块进行流式传输的所述请求中的资源定位符是否为所述第一资源定位符类型, 并且所述发送机还被配置为 : 如果被包括在对视频区块进行流式传输的所述请求中的所述资源定位符为所述第一资源定位符类型, 则将所述单个视频区块发送至所述视频播放器设备 ;

所述处理器还被配置为确定被包括在对视频区块进行流式传输的所述请求中的资源定位符是否为所述第二资源定位符类型, 所述发送机还被配置为如果被包括在对视频区块进行流式传输的所述请求中的所述资源定位符为所述第二资源定位符类型, 则将所述第一视频区块发送至所述视频播放器设备, 所述处理器还被配置为确定所述第二视频区块是否不可用, 所述发送机还被配置为如果所述第二视频区块不可用, 则发送对所述第二视频区块进行计算的请求, 对第二视频区块进行计算的所述请求包括对所述第二视频区块的所述引用和所述预期播出延时。

9. 一种系统, 所述系统包括根据权利要求 8 的设备 (800, 61)、调度器 (50) 和缓存存储器 (53), 其中, 所述调度器被配置为在计算实例 (52) 上对计算视频区块进行调度, 所述缓存存储器被配置为存储所计算的视频区块, 其中, 所述设备的所述发送机还被配置为将对所述第二视频区块进行计算的所述请求发送至所述调度器, 所述调度器使用用于存储多个作业队列 (51) 的存储器, 其中, 所述多个作业队列用于计算视频区块, 所述多个作业队列中的每个作业队列由特定分类的计算实例进行服务, 每个计算实例将可用的所计算的视频区块存储于所述缓存存储器中, 并且所述调度器将用于计算视频区块的作业归属到根据所述相关联的预期播出延时而选择的特定分类的队列。

10. 如权利要求 9 所述的系统, 其中, 所述调度器还被配置为 : 当接收到用于计算的所述请求时, 对与对所述第二视频区块的所述引用相对应的作业是否被预计为在所述预期播出延期内返回计算结果进行验证, 如果否, 则将与对所述第二视频区块的所述引用相对应的所述作业的副本作业插入到分类紧急的队列中。

11. 如权利要求 9 所述的系统, 其中, 所述调度器还被配置为 : 当接收到用于计算的所述请求时, 对与对所述第二视频区块的所述引用相对应的作业是否被预计为在所述预期播出延期内返回计算结果进行验证。如果否, 则将与对所述第二视频区块的所述引用相对应的所述作业移入分类紧急的队列中。

12. 如权利要求 9-11 中任一权利要求所述的系统, 其中, 所述调度器还被配置为 : 添加

或移除将每个作业队列作为作业队列尺寸的函数的计算实例。

13. 一种计算机程序产品,所述计算机程序产品能够从通信网络被下载和 / 或被记录在计算机可读的介质上和 / 或能够由处理器运行的介质上,所述计算机程序产品包括用于实现根据权利要求 1-4 中任一权利要求所述的方法的步骤的程序代码指令。

14. 一种非暂态计算机可读介质,所述非暂态计算机可读介质包括被记录在其上并且能够被处理器运行的计算机程序产品,所述计算机程序产品包括用于实现根据权利要求 1-4 中任一权利要求所述的方法的步骤的程序代码指令。

视频流式传输的方法、相应的设备及系统

技术领域

[0001] 本公开涉及云计算领域,并且具体涉及在使用云计算的目标视频环境中维护用于提供视频区块(chunk)的最优计算成本的领域。

背景技术

[0002] 长久以来,将广告插入到去往消费者设备的内容广播成为内容和服务提供商的收入来源。通过双向数据通信网络将媒体文件流式传输至消费者设备已司空见惯,并且点播视频流式传输也越来越成功。TV 点播(TVoD)或TV重播、时移和视频点播(VoD)技术使得能够将内容传输至单个用户。该新的内容消费模型为从所谓的目标广告中产生收入创建了机会。目标广告是将一组消费者或单个消费者作为目标的广告。目标确定是基于人口统计信息或基于消费者简档(profile)的。广告通常是在所请求的视频的开始处被插入的附加的视频序列,并且附加的广告视频可以在观看期间或在所请求的视频的结尾处被插入。然而,消费者倾向于跳过所插入的广告间歇,这意味着内容提供商会有收入损失。解决方案包括:当插入的广告被播出时,阻塞视频多速(trick)模式,从而使得所插入的广告不能被跳过。对于服务提供商或消费者而言,这都不是一个令人满意的解决方案。因此,替代的解决方案已被开发,该替代的解决方案允许直接修改内容本身,而不是插入广告间歇,内容视频帧被修改为包括广告,从而消除了插入不可跳过的广告间歇的需要。主要发明人为McAlister的文件WO 02/37828A2描述了在将视频流式传输给用户的同时利用目标广告来代替视频帧中的广告放映区域。广告不再集中于广告间歇,而是成为视频帧的一部分,并且从而可以不再被消费者绕过。该技术被称为“覆盖”,该技术要求在视频帧中包括广告放映区域。当将视频流式传输至消费者时,广告放映区域被识别,并且目标广告代替广告放映区域。然而,众所周知,VoD服务器处理各种各样的请求,这些请求依赖于内容流行度和时隙变化。在流式传输期间执行的该代替操作要求较高的计算能力。在该操作环境中,在进行VoD流式传输期间用目标广告代替广告放映区域的解决方案的可扩展性仍成问题。

[0003] 因此,存在提供可扩展的技术方案的需要,该可扩展的技术方案保证在将视频流式传输至用户的同时,目标广告可以在预期的播出时间并且以最小的成本被覆盖在视频内容的视频帧的像素区域中。

发明内容

[0004] 所公开的方法、设备和系统解决背景部分中所讨论的缺点中的至少一些缺点。

[0005] 为此,本公开在第一个实施例中包括视频流式传输的方法。该方法包括如下步骤:从视频播放器设备接收将视频流式传输至该视频播放器设备的请求。响应于对视频进行流式传输的请求,该设备向视频播放器设备发送视频区块列表,该视频区块列表包括用于播放视频的视频区块的资源定位符。该视频区块列表包括第一资源定位符类型的资源定位符和第二资源定位符类型的资源定位符。第一资源定位符类型引用单个视频区块,第二资源定位符类型包括:引用第一视频区块的第一资源定位符、引用第二视频区块的第二资源定

位符、以及第二视频区块的预期的播出延时。该方法还包括如下步骤：从视频播放器设备接收对来自视频区块列表的视频区块进行流式传输的请求。如果被包括在对视频区块进行流式传输的请求中的资源定位符为第一资源定位符类型，则该设备将单个视频区块发送至视频播放器设备。然而，如果被包括在对视频区块进行流式传输的请求中的资源定位符为第二资源定位符类型，则该设备向视频播放器设备发送第一视频区块，并且如果第二视频区块不可用，则该设备发送对第二视频区块进行计算的请求，对第二视频区块进行计算的所述请求包括预期的播出延时。

[0006] 根据视频流式传输的方法的第二个实施例，对第二视频区块进行计算的请求被发送至用于对计算视频区块进行调度的调度器。该调度器包括用于计算视频区块的多个作业队列，多个作业队列中的每个作业队列由特定分类的计算实例进行服务，并且每个计算实例将所计算的视频区块存储于缓存存储器中。调度器将用于计算视频区块的作业归属到根据相关联的预期的播出延时而选择的特定分类的队列。

[0007] 根据视频流式传输的方法的第三个实施例，当接收到用于计算的请求时，调度器对与对第二视频区块的引用相对应的作业是否预计在预期的播出延时而返回计算结果进行验证。如果否，则调度器将与对第二视频区块的引用相对应的作业的副本作业插入到分类紧急的队列中。

[0008] 根据视频流式传输的方法的第四个实施例，当接收到用于计算的请求时，调度器对与对第二视频区块的引用相对应的作业是否预计在预期的播出延时而返回计算结果进行验证。如果否，则调度器将与对第二视频区块的引用相对应的作业移入分类紧急的队列中。

[0009] 根据视频流式传输的方法的第五个实施例，该方法还包括作为以队列尺寸为变量的函数而添加或移除对每个队列进行服务的计算实例。第五个实施例可以与第二个至第四个实施例中的任意实施例进行组合以提供特定的有利的实施例。

[0010] 根据视频流式传输的方法的第六个实施例，作业队列分类至少包括尽力而为 (best effort) 队列和紧急队列。第六个实施例可以与第二个至第五个实施例中的任意实施例进行组合以提供特定的有利的实施例。

[0011] 根据视频流式传输的方法的第七个实施例，所述计算包括利用将视频播放器设备的用户作为目标的内容对视频区块的图像帧中的像素区域进行覆盖。第六个实施例可以与之前的实施例中的任意实施例进行组合以提供特定的有利的实施例。

[0012] 本公开还涉及设备。所公开的设备包括处理器、发送机和接收机。接收机被配置为从视频播放器设备接收将视频流式传输至该视频播放器设备的请求。发送机被配置为：响应于对视频进行流式传输的请求，向视频播放器设备发送视频区块列表，该视频区块列表包括所请求的视频的视频区块的资源定位符。该视频区块列表包括第一资源定位符类型的资源定位符和第二资源定位符类型的资源定位符。第一资源定位符类型引用单个视频区块。第二资源定位符类型包括：引用第一视频区块的第一资源定位符、引用第二视频区块的第二资源定位符、以及第二视频区块的预期的播出延时。接收机还被配置为：从视频播放器设备接收对来自视频区块列表的视频区块进行流式传输的请求。处理器被配置为确定被包括在对视频区块进行流式传输的请求中的资源定位符是否为第一资源定位符类型，并且发送机还被配置为：如果被包括在对视频区块进行流式传输的请求中的资源定位符为第一资

源定位符类型,则向视频播放器设备发送单个视频区块。处理器还被配置为确定被包括在对视频区块进行流式传输的请求中的资源定位符是否第二资源定位符类型。发送机还被配置为:如果被包括在对视频区块进行流式传输的请求中的资源定位符为第二资源定位符类型,则向视频播放器设备发送第一视频区块,并且处理器还被配置为确定第二视频区块是否不可用。发送机还被配置为:如果第二视频区块不可用,则发送对第二视频区块进行计算的请求,对第二视频区块进行计算的所述请求包括对第二视频区块的所述引用和预期的播出延时。

[0013] 本公开还包括一种系统,该系统包括上述设备。该系统还包括调度器和缓存存储器,其中,该调度器被配置为在计算实例上对计算视频区块进行调度,该缓存存储器被配置为存储所计算的视频区块。该设备的发送机还被配置为将对第二视频区块进行计算的请求发送至调度器。调度器使用存储器以存储用于计算视频区块的多个作业队列,多个作业队列中的每个作业队列由特定分类的计算实例来服务,每个计算实例将可用的所计算的视频区块存储于缓存存储器中,并且调度器将用于计算视频区块的作业归属到根据相关联的预期的播出延时而选择的特定分类的队列。

[0014] 根据该系统的变体实施例,调度器还被配置为:当接收到用于计算的请求时,对与对第二视频区块的引用相对应的作业是否预计在预期的播出延期内返回计算结果进行验证,如果否,则将与对第二视频区块的引用相对应的作业的副本作业插入到分类紧急的队列中。

[0015] 根据该系统的变体实施例,调度器还被配置为:当接收到用于计算的请求时,对与对第二视频区块的引用相对应的作业是否预计在预期的播出延期内返回计算结果进行验证,如果否,则将与对第二视频区块的引用相对应的作业移入分类紧急的队列中。

[0016] 根据该系统的变体实施例,调度器还被配置为:添加或移除将每个作业队列作为作业队列尺寸的函数的计算实例。

[0017] 本公开还包括可从通信网络下载和/或被记录在计算机可读的介质上和/或可由处理器执行的计算机程序产品,该计算机程序产品包括用于根据第一个至第七个实施例中的任何实施例实现该方法的步骤的代码指令。

[0018] 本公开还包括非暂态计算机可读介质,该非暂态计算机可读介质包括被记录在其上并且能够被处理器运行的计算机程序产品,包括用于根据第一个至第七个实施例中的任何实施例实现该方法的步骤的代码指令。

附图说明

[0019] 当阅读以下的说明书和附图时会发现本公开的实施例的其他特征和优点。

[0020] 图 1 示出了目标内容覆盖的处理;

[0021] 图 2 示出了用于确定可作为目标的视频区块的视频预处理;

[0022] 图 3 示出了清单(manifest)文件生成和使用;

[0023] 图 4 示出了云计算装置的使用,用于使用亚马逊网络服务(Amazon AWS)来提供包括目标广告的视频流作为视频区块的图像帧的像素区域中的覆盖;

[0024] 图 5 图解地示出了用于目标视频区块计算的示例云计算环境;

[0025] 图 6 根据本发明的实施例示出了播放时间跟踪器和其他模块之间的关联;

[0026] 图 7 是示出根据本发明的方法的具体实施例的流程图；

[0027] 图 8 是用于实现本发明的方法的实施例的示例设备。

具体实施方式

[0028] 随后的描述提供了示例性实施例。这些实施例不旨在限制本公开的范围、适用性或配置,而是向本领域技术人员提供使他们能够实现这些实施例的说明。应当理解,在保持在本公开的范围之内的同时,对示例性优选实施例的功能和布置可以做出各种更改。

[0029] 在随后的描述中,术语“覆盖”表达了将内容覆盖在视频帧中的原始内容上的动作。这导致了这样的视频帧,即,原来被包括在该视频帧中的像素中的至少一些像素被覆盖图像的像素代替。诸如透明度、颜色设置和对比度之类的特征可以被用来将覆盖图像适应到原始的视频帧内容。

[0030] 在随后的描述中,视频内容的一系列连续的图像帧被称为“视频区块”。包括用于利用目标广告进行覆盖的可覆盖的像素区域的视频内容的一系列连续的图像帧被称为“可作为目标的视频区块”。目标内容被覆盖的一系列连续的图像帧被称为“目标视频区块”。在将视频内容流式传输至消费者设备的过程中,通过将把特定消费者作为目标的广告覆盖到所确定的可作为目标的视频区块中,视频内容被适应到特定用户,因而所获得的目标视频区块作为视频内容的一部分被流式传输至该消费者。

[0031] 在随后的描述中,术语“计算”被用来指示用于从可作为目标的视频区块获得目标视频区块的处理操作。

[0032] 在随后的描述中,术语“播出延时 (play out delay)”被用来指示从给定的时刻起计数(例如,从给定的视频区块起计数)的延时,在该给定的时刻内,预计视频播放器需要播出另一视频区块。替代地,当预计视频播放器需要播出另一视频区块时,预期的播出“延时”是预期的播出“时间”,该预期的播出“时间”例如与参考时钟(例如, NTP(网络时间协议))的时间、多个设备之间共享的共同时间参考、或者未来时间或时钟参考的壁钟 (wall clock) 时间相对应。

[0033] 原则上,云计算可以提供所需的扩展性,因为其允许潜在无限的点播计算资源。利用诸如由 Amazon AWS 提供的云计算平台,对计算实例定价依赖于计算速度并且依赖于对在给定的时刻准备好计算结果的保证。理想地,对目标视频区块进行计算可以通过云计算平台所提供的各种类型的计算实例以优化计算成本的方式被派遣。例如,对于靠近点播视频开始处的可作为目标的视频区块,需要相对较高成本、相对较高速度的云计算实例;而对于离开始处二十分钟的可作为目标的视频区块,相对较低成本、相对较低速度的云计算实例就够了。然而,如下文所解释的,各种考虑因素扰乱了该理想场景:

[0034] (i) 消费者不是强制地以线性方式来观看点播视频,他/她可能使用多速模式。根据消费者的多速模式动作,因而可能需要比预计的更早或更晚来播出视频区块。

[0035] (ii) 对云计算实例进行定价不稳定,因为其通过竞标被确定。

[0036] (iii) 相对低价的云计算实例(例如,亚马逊“spot”)不能保证在给定的延时之内给出结果。

[0037] (iv) 甚至标准的云计算实例也可能在作业完成之前发生故障。

[0038] 图 1 示出了目标内容覆盖在流式传输的视频内容中的处理。图像帧 10 表示未改

变的视频的原始的图像帧。图像帧 11 表示改变后的视频图像帧,在该改变后的视频图像帧中覆盖有目标内容 111(此处是在桥栏杆中)。

[0039] 图 2 示出了用于确定可作为目标的视频区块的视频预处理。项目 20 表示视频序列。项目 21、22 和 23 表示在视频预处理的过程中被确定的视频图像帧,这些视频图像帧包括像素区域,目标内容可以被重叠在这些像素区域中。可以采用不同的技术来识别出这些像素区域,例如,通过对就改变而言相对稳定的连续帧中的像素区域进行检测。

[0040] 图 3 示出根据本发明的一个或多个实施例的清单文件生成和使用。根据用于该示例的场景,存在两个消费者 33 和 34。当这些消费者请求视频时,每个消费者接收清单文件。消费者 33 接收清单文件 32。消费者 34 接收清单文件 31。由清单生成功能 30 生成清单文件。清单文件包括指向所请求的视频的视频区块的统一资源定位符(URL)。因此,清单文件是给定的视频内容的视频区块部分的列表。对于消费者 33,清单文件包括 URL1 和 URL3,其中,URL1 指向通用视频区块 35,该区块对于所有的消费者而言是相同的,URL3 具体将消费者 33 作为目标,因为其指向目标视频区块 36a,目标视频区块 36a 包括被覆盖在与出现在视频区块的图像上的桥栏杆相对应的像素区域上的广告。消费者 34 接收包括 URL1 和 URL2 的清单文件。URL1 与用于消费者 33 的是相同的。URL2 指向通用视频区块 36,通用视频区块 36 对于消费者 33 和消费者 34 而言是相同的。这些消费者的消费者设备(未示出)中的每个消费者设备中的视频播放器将读取清单文件并且获取所引用(reference)的 URL,以接收并播放视频区块。

[0041] 图 4 示出了云计算装置的使用,用于使用亚马逊网络服务(Amazon AWS)提供包括目标广告的视频流作为视频区块的图像帧的像素区域中的覆盖,其中,本发明的一个或多个实施例可以被实现。

[0042] ● 计算实例(例如,“EC2”(弹性计算云;EC2 微型(micro)401、EC2 插播(spot)402、EC2 大型(large)407))运行计算任务(例如,确定目标、覆盖、用户简档建立、清单文件生成)。EC2 是提供大小可变的计算能力并且为不同种类的操作系统和不同种类的“实例”配置提供虚拟计算环境的 web 服务。典型的实例配置是“EC2 标准”或“EC2 micro”。“EC2 micro”实例很好地适于周期性地要求附加的计算周期的 web 站点和较低吞吐量应用。有多种方式获得 AWS 中的资源。第一种方式被称为“点播”,该方式将以给定的价格保证资源可用。第二种方式被称为“spot”,该方式允许以较便宜的价格获得资源,但没有可用性的保证。“EC2 spot”实例允许通过竞标机制来获得 EC2 计算能力的价格。这些实例可以显著降低时间灵活、容忍中断的任务的计算成本。对于相同的 EC2 实例类型,这些价格通常显著低于点播价格。

[0043] ● 存储实例(例如,“S3”(简单存储服务;404、405、406))存储诸如通用视频区块和可作为目标的视频区块之类的数据。“S3”提供了能够被用来在任意时间存储和取回任意数量的数据的简单 web 服务接口。存储空间价格取决于所期望的可靠性,例如,具有高可靠性的标准存储和用于存储非关键、可复制的数据的降低冗余存储。

[0044] ● “云端(CloudFront)”(403)计划数据递送。云端是用于内容递送的 web 服务,并且云端与其他 AWS 服务相集成来利用低延迟且高数据传送速度将内容分发给端用户,并且能够被用于对内容进行流式传输。

[0045] 在该图中,元素 400 描绘了用户设备(例如,机顶盒、PC、平板电脑或移动电话)。

可靠的 S3 404 被用于存储通用视频区块和可作为目标的视频区块。降低可靠的 S3(405) 被用来存储能够容易地被重新计算的目标视频区块。降低可靠的 S3(405) 被用作缓存, 以将目标视频区块保存在存储器中达一段时间。可靠的 S3 406 被用于存储: 可作为目标的视频区块、广告或覆盖内容。EC2 spot 实例 402 被用来对目标视频区块进行预计算。通过 EC2 spot 实例进行的该计算(被称为“批量”计算)例如在清单生成时被触发。点播 EC2 large 实例(407) 被用来实现“即时(on the fly)”或“实时”覆盖。按照如下计算目标视频区块: 从可靠的 S3(406) 取回可作为目标的视频区块, 对该可作为目标的视频区块进行解码, 对覆盖内容进行选择(例如广告, 被作为以用户偏好为变量的函数而选择), 将覆盖内容覆盖到可作为目标的视频区块的图像帧的像素区域上, 并且将这些图像帧重新编码, 从而获得目标视频区块。取决于先前提到的对目标视频区块进行的“即时”或“批量”计算, 对可作为目标的视频区块进行的解码、对覆盖内容进行的选择、覆盖以及重新编码相应地在 EC2 spot 示例(402) 或在 EC2 large 实例(407) 中进行。当然, 所描述的这种变体只是若干个可能的策略中的一个。其他策略可以包括根据不同的参数(例如, 延时、任务大小以及计算实例成本)针对“即时”计算或“批量”计算中的任一个来使用不同的 EC2 示例(例如, micro、medium 或 large), 以使得对这些实例的使用被优化来提供具有较好服务质量的划算的解决方案。然后在“批量”计算的情形中, 目标视频区块被存储于被用作缓存的降低可靠的 S3(405) 中, 或者在“即时”计算的情形中, 目标视频区块被直接从 EC2 large 实例 407 进行服务并且可选地被存储于降低可靠的 S3 405 中。如果时间允许, 则出于计算成本的原因, 对目标视频区块进行批量计算更为可取。因此, 开始对目标视频区块进行批量计算的较好的时刻是当清单被生成时。然而, 如果用户快进到尚未被计算的可作为目标的视频区块, 则需要更多成本的“即时”计算。

[0046] 因此, 根据可作为目标的视频区块在视频播放时间轴上的位置, 将根据优化成本的策略来选择云计算实例。一旦清单文件被生成, 策略可以被确定。然而, 如果用户使用多速模式(例如, 快进), 则计算策略必须被再次访问。如果对目标视频区块进行的计算以 spot 实例开始, 则当用户快进视频时不能保证该目标视频区块按时可用。因此, 存在在视频区块播出之前的一段时间再次访问计算策略的需要, 以确保在突发事件的情形中视频区块的可用性。可能的解决方案可以是在 VoD 服务器中跟踪用户播放时间。但这要求对每个 VoD 用户的播放时间进行监控。这样的解决方案会损害可扩展性。并且, 通常, 基于云的解决方案要求无状态设计, 以便减轻云服务器之间的负载均衡。因此期望提供允许充分提前再次访问视频区块计算策略的无状态解决方案, 以便能够在预期的播出时间内提供区块。本发明的实施例提供了用于针对视频区块的递送进行播放时间跟踪的方法和设备。

[0047] 当用户请求给定的视频时, 针对该视频中的可作为目标的视频区块做出如下决定: 根据用户简档, 利用将该特定用户作为目标的内容来对这些可作为目标的视频区域进行覆盖。清单文件被生成, 并且被传输至用户。对于该视频中在缓存区域中尚不可用的每个目标视频区块, 用来计算目标视频区块的作业被生成并且被投递到作业队列。作业可以被认为“目标区块计算请求”。每个作业包括预期播出时间参数或预期播出延时参数, 该预期播出时间参数或预期播出延时参数指定可以被用来计算目标视频区块的延时限制或时间限制。预期播出时间是对于给定的实际的播放位置和播放速度, 当目标视频区块被认为是视频播放器所需的时的预期的时间。预期的播出时间可以包括附加的留白(margin) 延

时,以考虑用于将目标视频区块从存储设备传输至视频播放器的存储延时和传输延时。不同类型的云实例由诸如每时间单位的计算速度和计算成本之类的参数来表征。计算速度可以被表达为对实时视频播出速度的比率,例如,1 意味着计算持续时间等于视频区块播出持续时间,2 意味着计算持续时间是播出持续时间的两倍,等等。当然,这取决于实例特征,例如,对于 AWS EC2,取决于 ECU(EC2 计算单元)的数目。根据云计算实例关于计算速度和可靠性的不同特征来定义不同的作业队列。

[0048] 图 5 图解地示出了用于目标视频区块计算的示例云计算环境。所描绘的是:调度器 50、作业队列 51、实例池 52、扩缩器 53 以及缓存 54。队列中的作业包括与目标视频区块的视频区块标识符和该视频区块的预期播出时间有关的相关联的信息。调度器接收用于计算目标视频区块的请求并且生成作业。调度器根据预期播出时间来决定将作业放置在哪个队列中。例如以 FIFO 顺序(先进先出)向实例池给予来自队列的作业。如果针对队列中的作业的目标视频区块已经存在于缓存中(即,目标视频区块被称为“可用”),则该作业从该队列中被移除,并且下一作业被获取。如果该目标视频区块尚未在缓存中,则开始该作业。当视频区块已被计算时,它们被存储于缓存中(即,计算结果(也就是目标视频区块)被存储于存储器区域;缓存是存储器区域的示例,存储器区域的其他示例为数据库、易失性或非易失性存储器、外部存储设备或内部存储设备)。经计算的视频区块被从缓存递送至消费者设备上的视频播放器(未示出)。在图 5 的示例实施例中,示出了四种类型的队列:“尽力而为”队列 Q1、针对紧急作业的“紧急”队列 Q2、针对“正常”作业的“正常”队列 Q3、以及针对“可选”作业的“可选”队列 Q4。每个特定队列中的作业被归属到特定的实例池中;用于“尽力而为”队列 Q1 的“Spot XLarge”实例池,用于“紧急”队列 Q2 的“XLarge”实例池,用于“正常”队列 Q3 的“XLarge”实例池,用于“可选”队列 Q4 的“Spot Micro”实例池。扩缩器 53 根据队列 Q1 和 Q3 中的作业流从“Spot XLarge”和“XLarge”实例池添加或移除实例。可以有若干个扩缩策略用于开始/停止为特定队列进行服务的给定池的特定实例。对于给定的队列,存在唯一的扩缩策略。扩缩策略决定何时从实例池开始新的实例,例如,通过确定队列中存在的作业数目或通过确定实例池中当前实例的状态。扩缩策略还决定给定的实例何时被停止以及何时被从实例池中移除。对于给定的队列,存在两种极端的情形:

[0049] (i) 如果该策略是优化成本,则尝试避免具有处于空闲状态(即,不被用于计算目标视频区块)的实例;假设实例是按小时收费的,则该策略是在开始新的实例之前等待该队列被相对填满。在该极限情形中,用于计算视频区块的延时将相对较高。

[0050] (ii) 在另一极端情形中,该策略是最小化延时,尝试具有尽可能小的队列,承担如下风险:一些实例在某一时间可能不对区块进行计算(当队列为空时,即,这些实例被称为“空闲”)。

[0051] 开始和停止实例是通过观察队列中的作业数目、当前运行实例的数目及其类型来完成的。因而,队列由为其服务的实例池的特征及其扩缩策略进行定义。spot 实例与其他实例的不同之处在于没有实例将被授予的保证;然而,如果被授予,则与其他类型的实例相比,其价格将相对较低。spot 实例可能在任何时候被关闭,甚至在被授予之后;如果其在一个小时之内被关闭,则不对其进行收费。调度器 50 接收针对目标视频区块的请求(即,“作业”),并且针对每个请求,调度器 50 接收视频区块的标识符和播出之前的时间。调度器根据预期播出时间将作业归属到正确的队列。如果预期播出时间高于阈值,则该作业可以

被投递到尽力而为队列 Q1。来自实例池 52 的实例从其所归属的队列中获取作业并且运行这些作业。然而,由于 spot 实例可能在任意时间被终止,并且可能在它们有时间交付其作业结果之前被终止,因此需要用于识别未处理的作业并且按照需要重新创建这些作业的机制。如果作业的预期播出时间低于阈值,则调度器仍根据播出之前的时间,将该作业归属到正常队列 Q3 或紧急队列。根据最简形式的变体实施例,只存在两个队列,一个用于“慢”作业,一个用于“快”作业。对于更加一般的情形,支持多于一的任意数目的队列。根据变体实施例,定义了优先级队列,在该优先级队列中,例如根据预期播出时间来对作业进行排序。扩缩器的角色是通过开始和停止池内的实例来将队列的长度保持在阈值以下。当队列较大并且仍在增长时,扩缩器添加实例。当队列较小并且在缩减时,可以停止实例。扩缩器确保存在足够的用于处理紧急队列 Q2 的作业的空闲快速实例。在视频区块到期之前不能被处理的任何视频区块可以简单地被丢弃,因此在该情形中,后续将没有目标视频区块被提供。

[0052] 图 6 根据本发明的实施例示出播放时间跟踪器与其他模块之间的关联。该图示出:VoD 目录 web 服务器 60、调度器 50、缓存 53、视频播放器 61、播放时间跟踪器 62、视频服务器 63 以及视频文件存储设备 64。

[0053] 调度器 50 一旦将作业投递到合适的队列中,则在投递该作业与需要相关的目标视频区块的时刻之间可能发生一定数目的事件:

[0054] ●如果该作业在由 spot 实例服务的队列中,则不保证在预期播出时间内分配实例以计算区块;

[0055] ●当前正在执行作业(即,计算区块)的实例可能在完成该任务之前就失效,并且将不计算视频区块;

[0056] ●如果用户执行多速模式,则预期播出时间可能改变,导致作业队列归属不再合适。

[0057] 因此,由调度器采用的把作业布置到不同队列中的决定被定期地再次访问,以便最大化如下可能性:所请求的目标视频区块在正确的时间可用并且利用最低的可能成本进行计算。为了利用无状态解决方案来实现该目标,参数被插入到允许播放时间跟踪的视频区块 URL 中。例如,在被发送至视频播放器的清单文件中,对于每 10 个视频区块或每 20 个视频区块,视频区块的 URL 通过附加信息(例如,预计很快会被请求的视频区块的标识符)和播出时间或播出延时被“超载(overload)”。示例清单文件看起来如下:

[0058] # 视频“m”的清单文件

[0059] http://server/m/chunk0.ts

[0060] http://server/m/chunk1.ts

[0061] http://server/m/chunk2.ts

[0062] http://pttracker/? chunk = m/chunk6.ts&duein = 20&dest = http://server/m/chunk3.ts

[0063] http://server/m/chunk4.ts

[0064] http://server/m/chunk5.ts

[0065] http://server/m/chunk6.ts

[0066] ...

[0067] 以下动作应用于该图中的功能部件之间的箭头:

[0068] 1a:消费者在 VoD 目录中搜索内容；

[0069] 1b:视频播放器 61 获得到达所选内容的 URL；

[0070] 1c:消费者按压播放；

[0071] 2a:视频播放器 61 从视频服务器 63 请求内容；

[0072] 2c:视频服务器 63 请求调度器 50 调度作业以提供所请求的内容的目标视频区块；

[0073] 3a:视频播放器 61 从清单文件请求 URL；

[0074] 3b:播放时间跟踪器 62 在缓存 53 中验证视频区块 6 的可用性；

[0075] 3c:如果视频区块 6 不在缓存 53 中,则播放时间跟踪器 62 请求调度器 50 针对视频区块 6 发起紧急作业；

[0076] 3d:播放时间跟踪器 62 将针对视频区块 3 的请求转发给视频服务器 63；

[0077] 3e:视频服务器 63 从缓存 53 获取视频区块；

[0078] 3f:视频服务器 63 从视频文件存储设备 64 获取视频区块；

[0079] 3g:视频服务器 63 将视频区块 3 提供给视频播放器 61。

[0080] 操作视频播放器 61 的消费者从 VoD 目录 web 服务器 60 选择视频内容(动作 1a)。消费者一旦选择了视频(例如,视频“m”),则视频播放器获得到达该内容的 URL(动作 1b),并且消费者利用所获得的 URL 通过发布播放命令开始播放该视频(动作 1c)。该播放命令被发送到视频服务器 63(动作 2a)。作为响应,视频服务器将清单文件或视频区块列表(例如,上面的视频区块列表)发送给视频播放器(动作 2b)。当视频被播放时,视频播放器 61 发送对与清单文件中的条目相对应的视频区块的 http get 请求(动作 3a)。针对非超载的 URL 的请求被转发至视频服务器(例如,针对视频区块 0、1、2、4、5、6)(动作 3d)。因此,针对超载的视频区块(例如,针对视频区块 3)的请求首先被发送至播放时间跟踪器(动作 3a),播放时间跟踪器提取出播放时间信息,然后该请求被转发至视频服务器(动作 3d)。无论来自缓存(如果视频区块在缓存中)(动作 3e)还是来自视频文件存储设备(如果视频区块不在缓存中)(动作 3f),所有被请求的视频区块由视频服务器来服务,并且将视频区块发送至视频播放器(动作 3g)。由播放时间跟踪器提取出的播放时间信息(例如,关于视频区块 3)包括时间值和视频区块标识符(在该情形中标识视频区块 6)。时间值指示在播放器将对由所嵌入的视频区块标识符标识的视频区块进行请求之前剩余的时间。播放时间跟踪器对目标视频区块是否存在于缓存 53 中进行检查(动作 3b)。如果该目标视频区块在缓存中,则不采取动作。如果其不在缓存中,则播放时间跟踪器将请求传送给调度器(动作 3c)以对目标视频区块进行紧急计算。被发送至调度器的请求包括视频区块标识符(例如,视频区块 URL)和上面所讨论的时间值。从播放时间跟踪器接收到请求后,调度器基于预期播出时间(例如,针对视频区块 6 为 20s)决定做什么:如果用于对紧急请求的视频区块进行计算的作业正由可靠的实例进行处理,并且预计该作业将在预期播出时间内完成,或者如果用于计算视频区块的作业在紧急队列 Q2 或可选队列 Q4 中,则不需要采取动作。如果用于计算视频区块的作业在尽力而为队列 Q1 中或正常队列 Q3 中,则可能需要将该作业移到紧急队列 Q4 中。因此,根据预期播出时间,调度器可以对作业的队列归属进行更改。根据优选实施例,调度器不对作业的队列归属进行更改,而是创建副本作业,将该副本作业投递到紧急队列;并且在开始在实例上执行来自队列的作业之前,相应的实例对相关的目标

视频区块是否已经处于缓存中进行验证。如果处于,则不需要执行该作业,将该作业从该队列移除,并且该实例从该队列获取下一作业。该优选实施例具有如下优点:在除了作业被获取以由实例进行处理之外的其他时候,不需要从队列移除作业,这简化了队列管理。根据变体实施例,调度器定期对队列中的作业是否被预计在与其相关联的预期播出时间内完成进行验证,并且自主地采取动作来改善完成它们的可能性,例如,通过按照需要将副本作业投递到紧急队列。调度器还可以对正由实例执行的作业是否进展太慢以至于预计不能在预期播出时间内完成进行验证,然后创建副本作业并将其投递到紧急队列。在接收到来自播放时间跟踪器的请求时,对于该请求中所指定的视频区块,对预计作业是否会在预期播出时间内完成进行附加验证。

[0081] 根据替代实施例,清单文件中的 URL 被签名以提供改善的安全性。

[0082] 根据替代实施例,使用了两个队列,第一个队列由 spot 实例池进行服务,第二个队列由用于紧急计算视频区块的按需实例进行服务。第一个队列被配置为在播出之前的预定时间段(例如,20s)得到计算结果。第二个队列被配置为在小于预定时间段的时间段内(例如,20s、10s 或 5s 内)得到计算结果。

[0083] 根据替代实施例,URL 定期被超载到清单文件中,例如,如果视频区块的典型持续时间为 10s,则每隔 6 个视频区块 URL 被超载到清单文件中。超载 URL 的频率是一种折中:如果太高,则播放时间跟踪器将需要许多 CPU 时间(例如,在若干个 EC2 实例上运行以支持负载)。当太低时,错过多速模式动作的风险太高,该风险将导致错过的目标视频区块。最后,播放时间跟踪器的 CPU 负载意味着计算成本。可能有若干个变体:

[0084] (i) 使用固定的 URL 超载频率,该固定的 URL 超载频率是计算成本与反应度之间的折中:每 n 个 URL 中将一个 URL 超载到清单文件中。n 的值将取决于平均视频区块持续时间。

[0085] (ii) 仅当预计在预定播出时间或播出延期内(例如,在 100s 或 200s 之内)需要目标视频区块时才进行 URL 超载。

[0086] (iii) 同 (ii),但针对清单文件中的若干个 URL。

[0087] (iv) 在同一个超载的 URL 中针对若干个目标视频区块进行 URL 超载,并且在超载的 URL 中指定每个目标视频区块的时间值。

[0088] 可以将上述变体进行组合来形成特定有利的变体实施例。

[0089] 根据又一实施例,根据由清单文件所去往的视频播放器所选择的播放速度对 URL 超载周期性进行适配。例如,如果所选择的播放速度为快进 16x,则 URL 超载的周期性是播放速度为正常播放(1x)的频率的 16 倍。尽管在内容(由于缩短的显示时间,该内容以比正常播放速度更快的速度来观看)中提供可作为目标的视频区块可能被看作是资源的浪费,但该变体有利地允许在视频播放器切回到正常的播放速度时准备好提供目标视频区块。在该图中,设备之间的连接被描绘为逻辑连接,但本领域技术人员将容易地理解,所有设备可以被连接到一个或多个网络(例如,图 8 的网络 805),通过这一个或多个网络可以使得所描绘的逻辑连接发生。

[0090] 图 7 是根据本发明的特定实施例的流程图。在步骤 70 中,对用于该方法的变量和参数进行初始化。在步骤 71 中,从视频播放器接收对视频进行流式传输的流式传输请求。在步骤 72 中,视频区块列表被发送至视频播放器,该视频区块列表包括所请求的视频的视

频区块的资源定位符。在步骤 73 中,从视频播放器接收对来自视频区块列表的视频区块进行流式传输的请求。在步骤 74 中,确定被包括在对视频区块进行流式传输的所述请求中的资源定位符是否为第一类型。如果资源定位符为第一类型,则在步骤 75 中,将该资源定位符中所引用的视频区块发送至视频播放器,并且该处理继续步骤 71。如果资源定位符是第二类型,则在步骤 76 中,将该资源定位符中所引用的第一视频区块发送至视频播放器,并且在步骤 77 中确定该资源定位符中所引用的第二视频区块是否可用(例如,在存储设备中(例如,在缓存区域中))。如果第二视频区块可用,则该处理继续步骤 71。如果第二视频区块不可用,则在步骤 78 中,将对该资源定位符中所引用的第二视频区块进行计算的请求发送至调度器,以计算目标视频区块,该请求包括对第二视频区块的引用和预期的播出时间或延时,并且该处理继续步骤 71。

[0091] 图 8 是适于实现本公开的方法的示例设备 800。设备 800 例如实现图 6 的播放时间跟踪器 62。该设备包括处理单元 801、发送机接口 802、接收机接口 803,它们通过内部数据和通信总线 804 进行互连。接收机接口和发送机接口被连接到网络 805,网络 805 允许该设备与其他设备(例如,视频播放器、调度器、缓存存储器、以及视频服务器)进行通信。接收机 803 从视频播放器接收将视频进行流式传输的请求。响应于该请求,发送机 802 向视频播放器发送视频区块列表,该视频区块列表包括所请求的视频的视频区块的资源定位符,该列表包括至少两种类型的资源定位符,第一种类型包括单个资源定位符,该资源定位符引用第一资源区块,而第二种类型除了包括引用第一视频区块的资源定位符,还包括引用第二视频区块的资源定位符和第二视频区块的预期播出时间或延时。接收机然后从视频播放器接收对来自视频区块列表的视频区块进行流式传输的请求。处理单元 801 确定被包括在对视频区块进行流式传输的请求中的资源定位符是否为第一类型。如果处理单元已经确定被包括在对视频区块进行流式传输的请求中的资源定位符为第一类型,则发送机将单个资源定位符中所引用的第一视频区块发送至视频播放器。如果处理单元已经确定被包括在该请求中的资源定位符为第二类型,则发送机将多个资源定位符中所引用的第一视频区块发送至视频播放器,并且处理单元确定该多个资源定位符中所引用的第二视频区块是否可用。如果处理单元确定第二视频区块不可用,则发送机发送对第二视频区块进行计算的请求,对第二视频区块进行计算的该请求包括对第二视频区块的引用和预期播出时间或延时。

[0092] 除了图 8 中所示的设备架构,其他设备架构是可能的并且与本公开兼容。本领域技术人员可以理解,本原理的各方面可以被实施为系统、方法或计算机可读介质。因此,本原理的各方面可以采取完全硬件实施例的形式、完全软件实施例(包括固件、常驻软件、微代码,等等)的形式或硬件方面和软件方面进行组合的实施例的形式,它们在本文中一般都可以被定义为“电路”、“模块”或“系统”。而且,本原理的各方面可以采用计算机可读存储介质的形式。可以使用一个或多个计算机可读存储介质的任意组合。

[0093] 因此,例如,本领域技术人员将理解,本文所呈现的框图表示实施本发明原理的示例性系统部件和/或电路的概念性视图。类似地,应当理解,任意流程图、流程框图、状态转换图、伪代码等表示可以在计算机可读存储介质中被充分表示并且由计算机或处理器来执行的各种处理,不管这样的计算机或处理器是否被明确示出。

[0094] 计算机可读存储介质可以采用计算机可读程序产品的形式,该计算机可读程序产

品被实施于一个或多个计算机可读介质中,并且具有被实施于其上的可由计算机执行的计算机可读程序代码。鉴于在其中存储信息的固有能力和从其中提供信息检索的固有能力,本文所使用的计算机可读存储介质被认为是非暂态存储介质。计算机可读存储介质可以是(例如而非限制)电子的、磁的、光的、电磁的、红外的或半导体的系统、装置、或设备、或前述任意适当组合。应当理解,尽管下面提供了本原理可以被应用于的计算机可读存储介质的更多的具体示例,但本领域技术人员将理解它们只是说明性而非穷尽列举:便携式计算机磁盘、硬盘、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或闪存)、便携式压缩盘只读存储器(CD-ROM)、光储存设备、磁存储设备、或前述任意合适的组合。

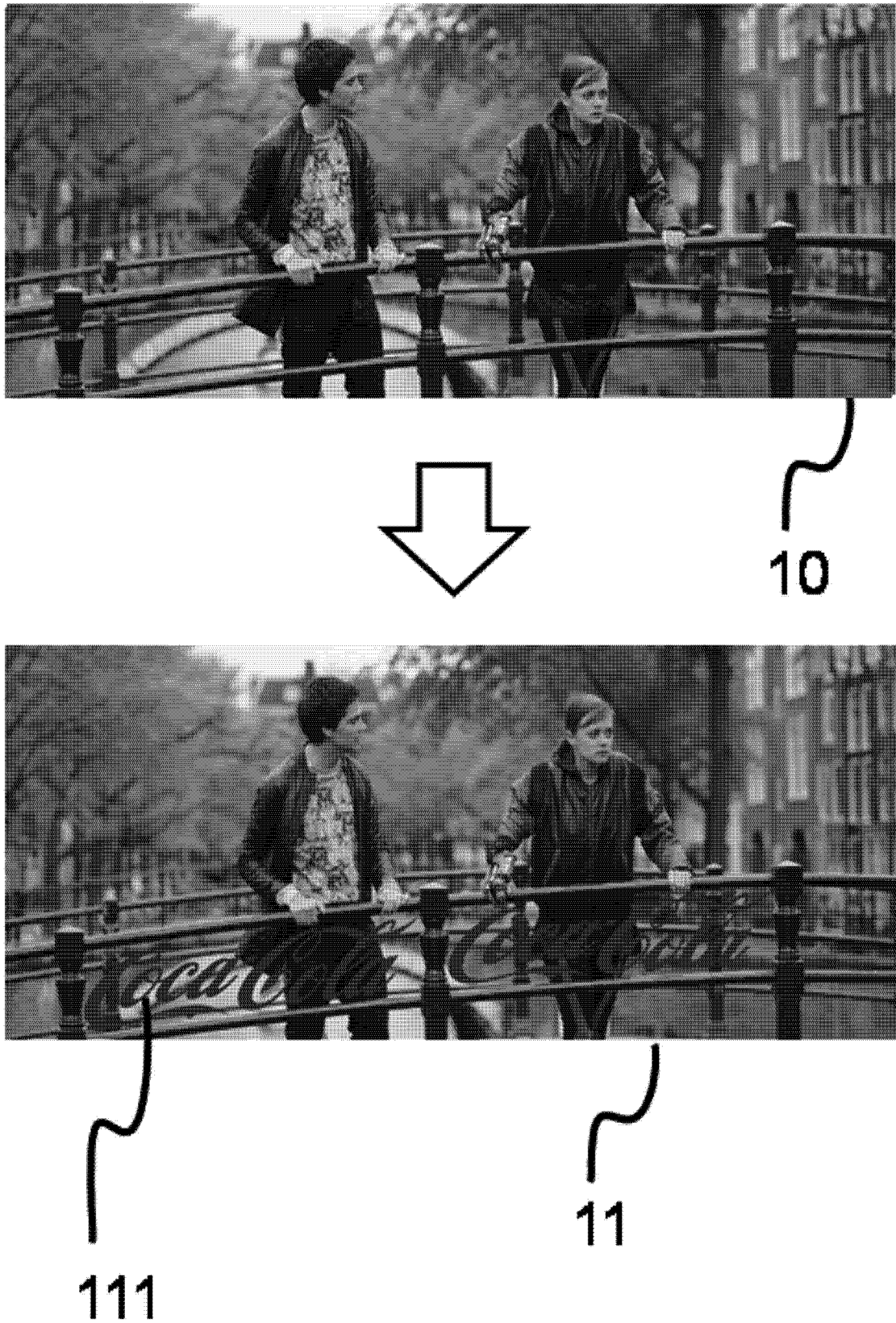


图 1

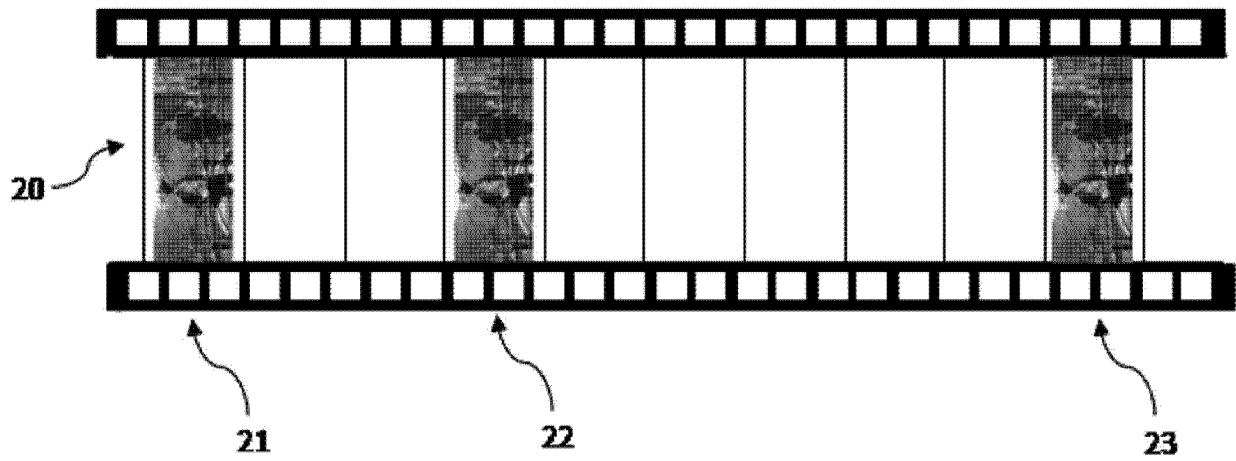


图 2

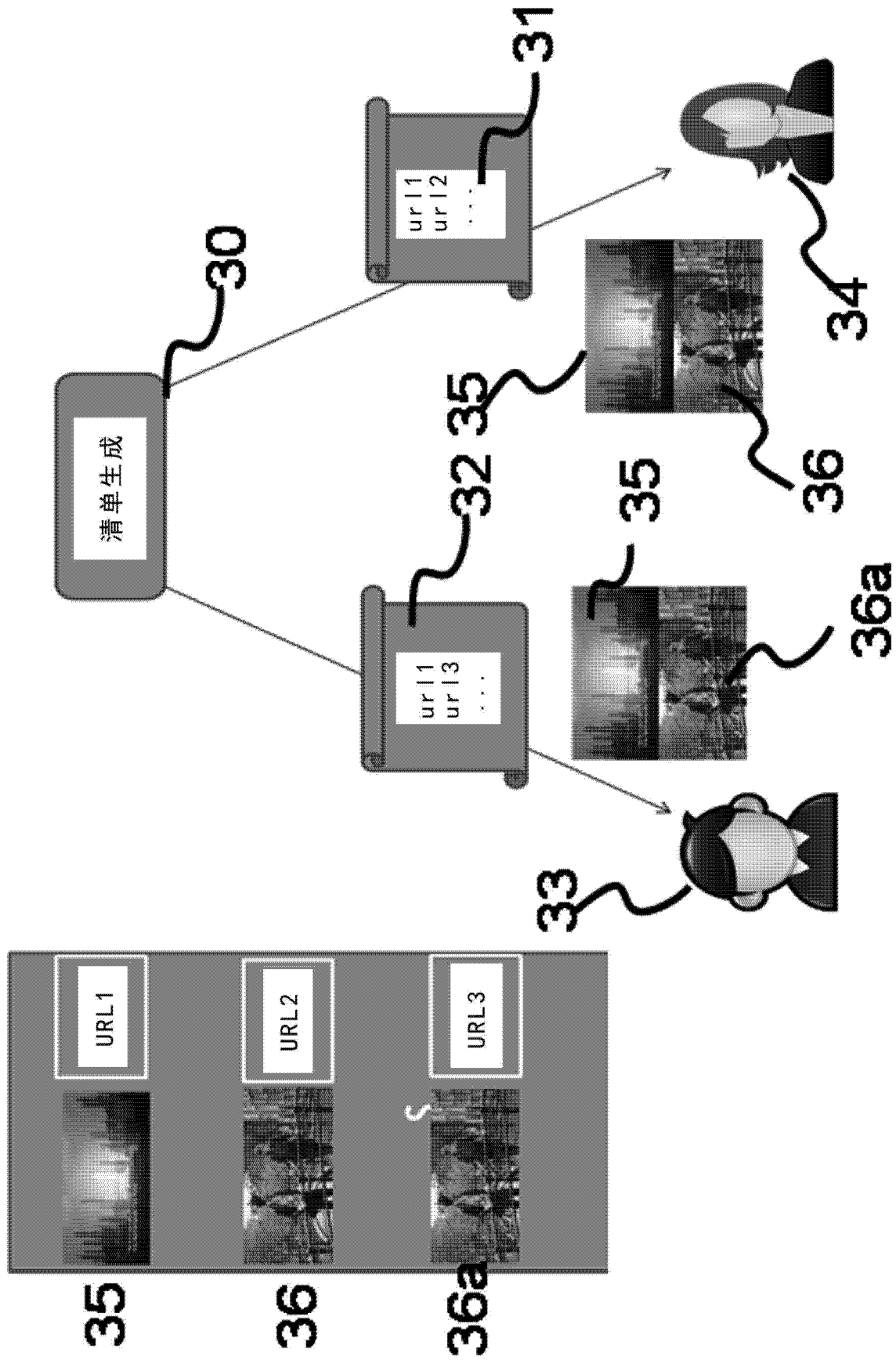


图 3

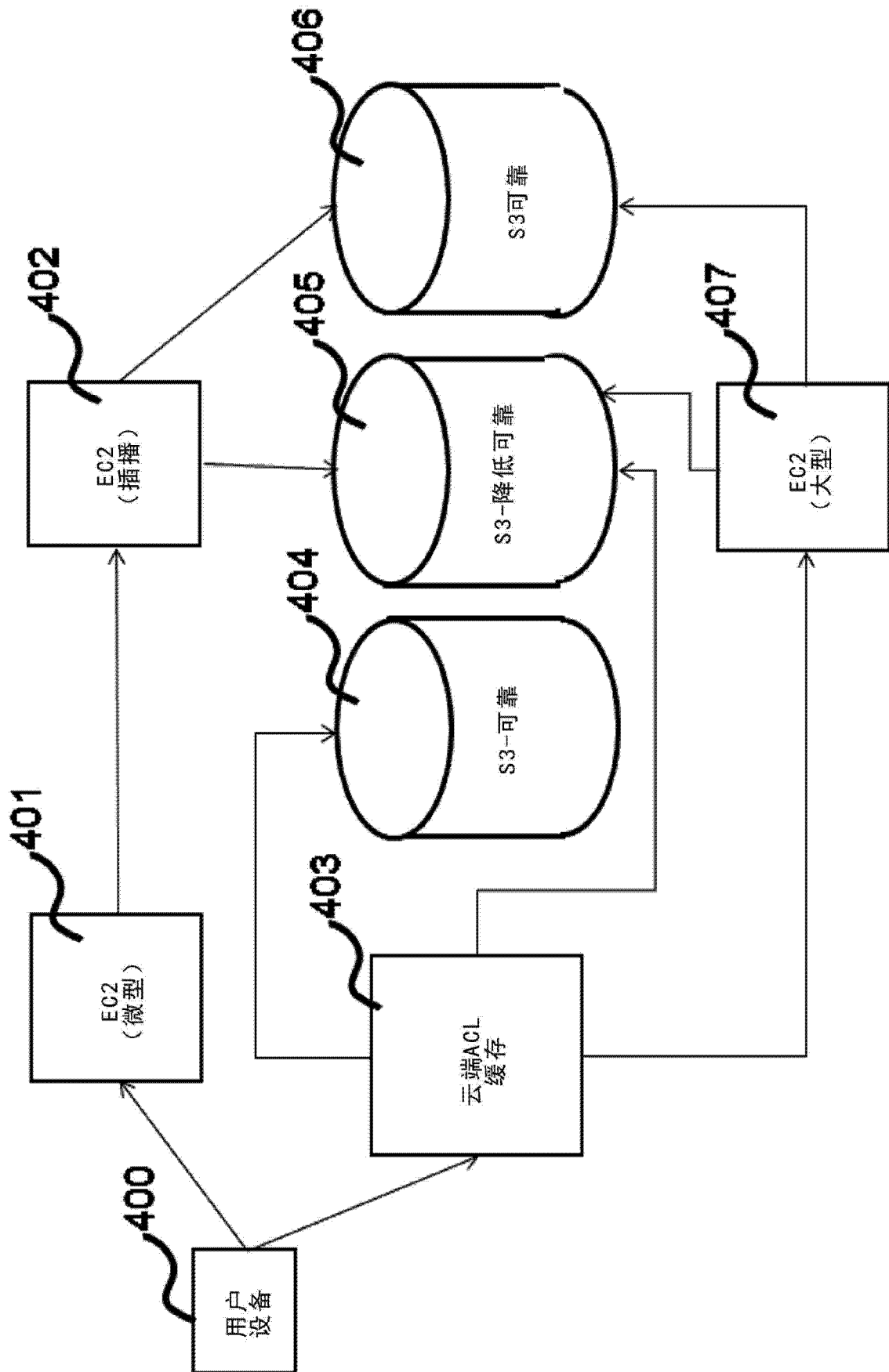


图 4

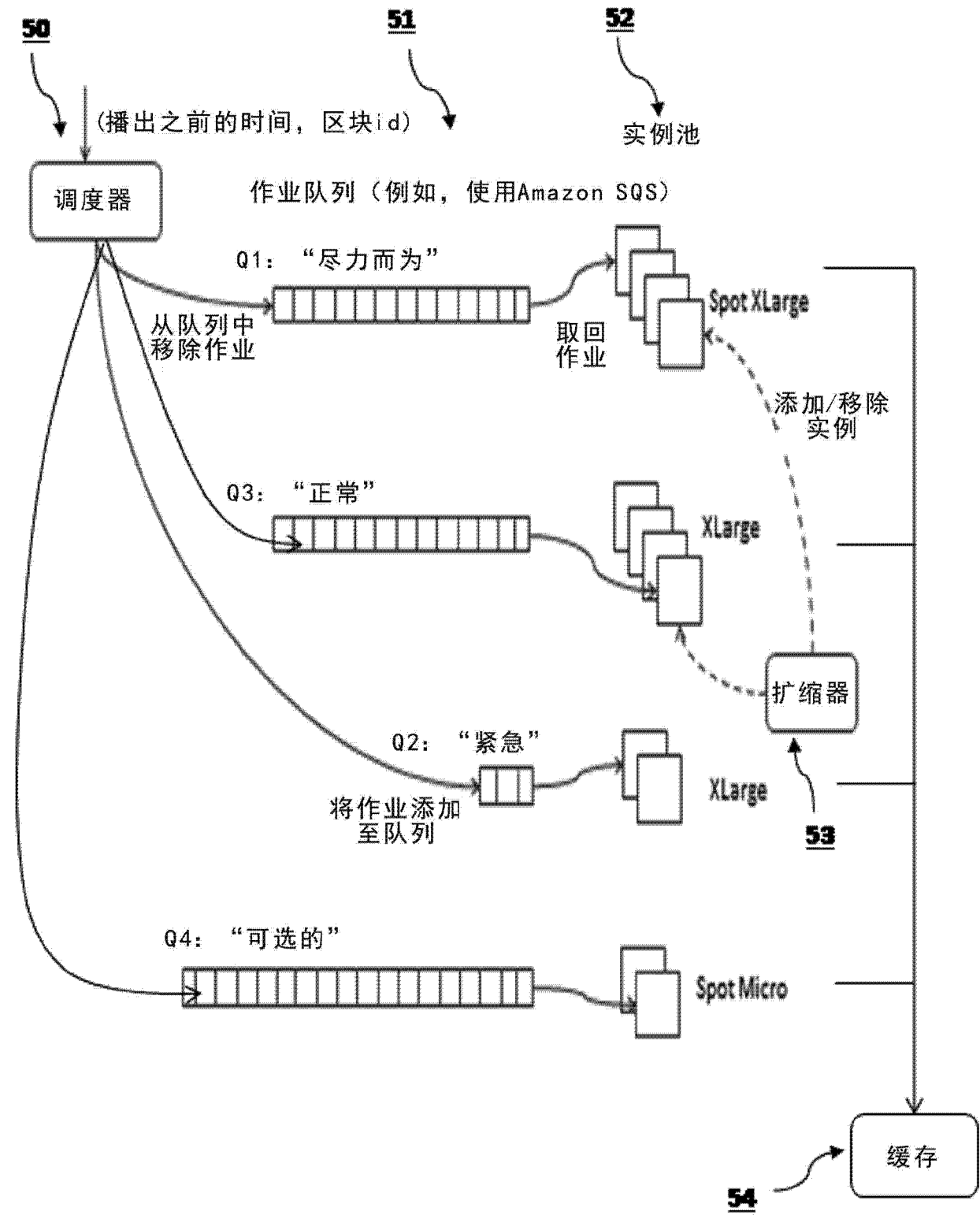


图 5

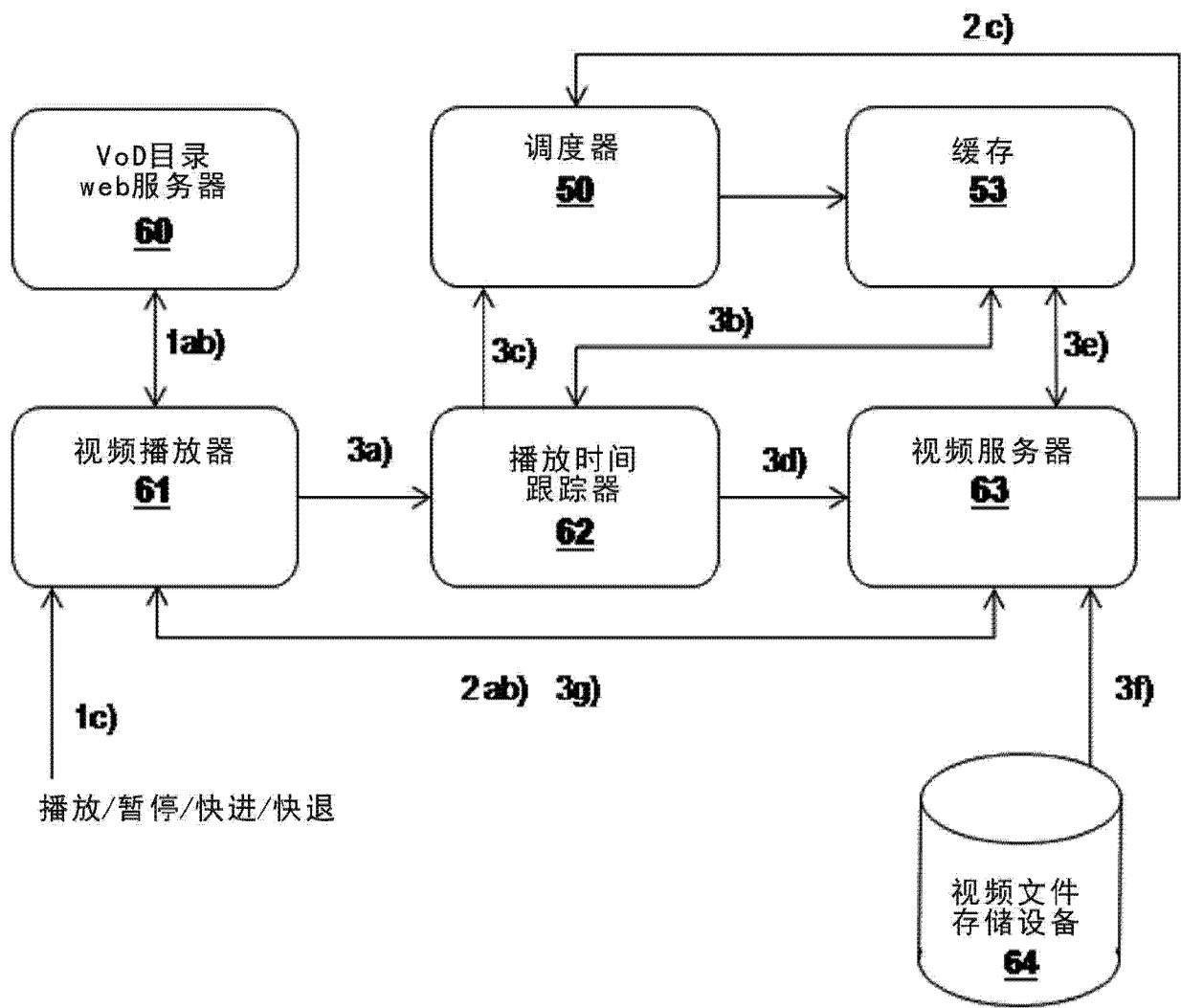


图 6

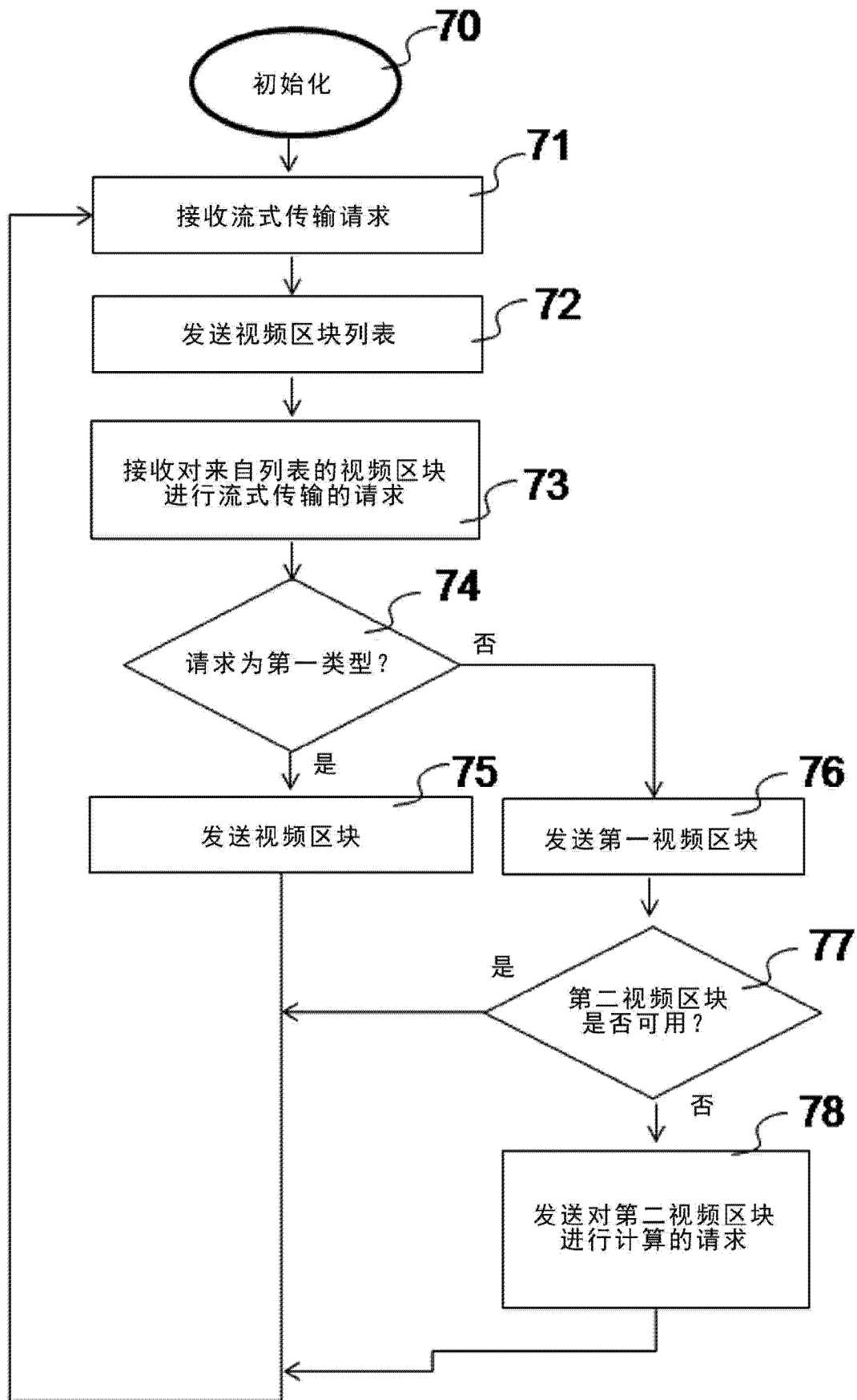


图 7

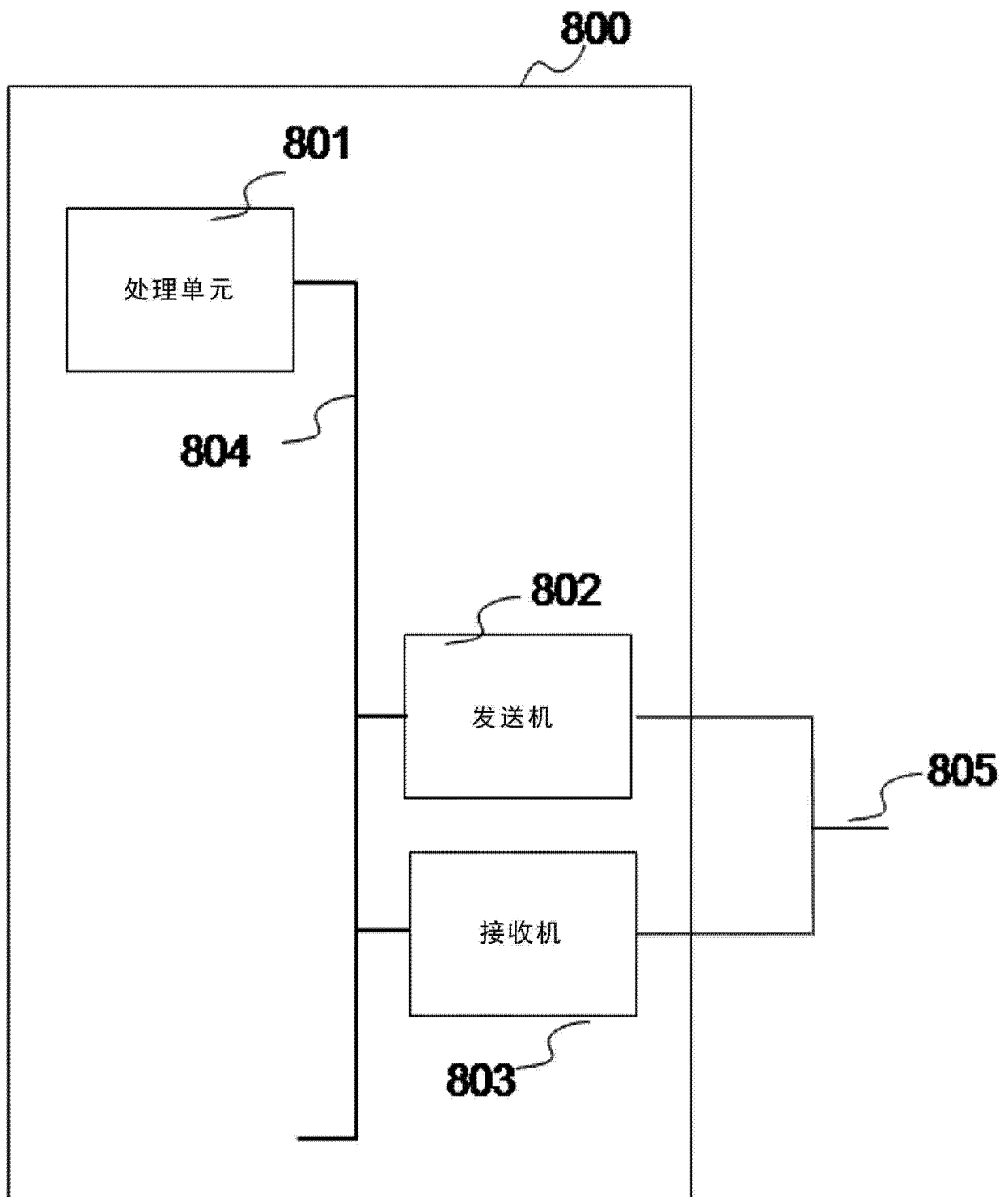


图 8