



(12) 发明专利

(10) 授权公告号 CN 101371312 B

(45) 授权公告日 2015. 12. 02

(21) 申请号 200680052466. X

代理人 钱慰民

(22) 申请日 2006. 12. 08

(51) Int. Cl.

(30) 优先权数据

60/748, 437 2005. 12. 08 US

60/778, 760 2006. 03. 03 US

60/787, 031 2006. 03. 29 US

60/787, 043 2006. 03. 29 US

60/829, 618 2006. 10. 16 US

60/862, 510 2006. 10. 23 US

(85) PCT国际申请进入国家阶段日

2008. 08. 07

(86) PCT国际申请的申请数据

PCT/US2006/061815 2006. 12. 08

(87) PCT国际申请的公布数据

W02007/067990 EN 2007. 06. 14

(73) 专利权人 维德约股份有限公司

地址 美国得克萨斯州

(72) 发明人 S·希伯利 R·希万拉

A·埃尔法泽阿迪斯 J·伦诺克斯

R·萨西恩 M·萨克希纳

O·莎彼洛

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

H04N 19/172(2014. 01)

H04N 19/61(2014. 01)

H04N 19/37(2014. 01)

H04N 19/132(2014. 01)

H04N 19/166(2014. 01)

H04N 19/89(2014. 01)

H04N 19/31(2014. 01)

H04N 19/587(2014. 01)

H04N 19/573(2014. 01)

H04N 19/58(2014. 01)

H04N 19/577(2014. 01)

H04N 19/66(2014. 01)

H04N 19/67(2014. 01)

H04N 21/2662(2011. 01)

H04N 21/462(2011. 01)

H04N 21/63(2011. 01)

H04N 21/6375(2011. 01)

H04N 21/6583(2011. 01)

H04L 1/16(2006. 01)

H04L 1/18(2006. 01)

审查员 刘欣

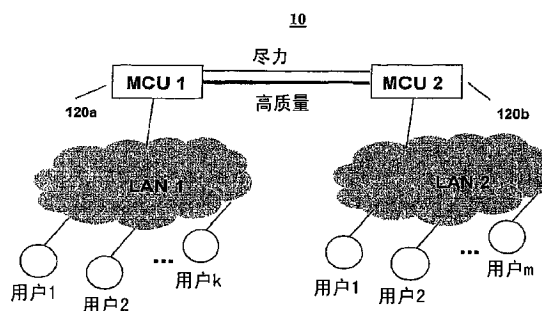
权利要求书3页 说明书26页 附图12页

(54) 发明名称

用于视频通信系统中的差错弹性和随机接入
的系统和方法

(57) 摘要

提供了一种用于视频通信系统中的差错弹性传输和随机接入的系统和方法。该视频通信系统基于可用在视频通信系统中的单层可缩放视频、或具有时间可缩放性的联播视频编码。视频信号传输中的一组视频帧或画面被指定使用安全或高可靠性链路或通过重传技术可靠或有保证地递送到接收器。被可靠递送的视频帧在差错发生之后或在随机接入期间用作接收器与所传送的视频信号重新同步的参考画面。



视频会议系统

1. 一种用于通信网络上传送端点与一个或多个接收端点或服务器之间的媒体通信的系统,所述系统包括:

编码器,它将所传送的媒体编码为具有包括最低时间层的多个不同层的线程化编码结构中的帧,

其中所传送的画面包括指示以下的元素:

对于最低时间级画面,标识所述画面的序号,以及

对于其它时间级画面,对按照解码次序最近的最低时间级画面的所述序号的引用,

以使得接收端点或服务器可通过检查在所述接收端点或服务器处是否已接收到与所引用画面的帧索引相对应的所述画面来检测最低时间级画面是否已丢失,

其中所述最低时间级画面被指定成将被可靠地递送到接收方目的地。

2. 如权利要求 1 所述的系统,其特征在于,所述数据元素另外指示与各个空间或质量层相关联的系列号,其中所述接收端点或服务器通过确定在所述接收端点或服务器处是否已接收到与所引用的系列号和序号相对应的所述帧来检测特定空间或质量层的最低时间级帧是否丢失。

3. 如权利要求 1 所述的系统,其特征在于,所述数据元素包括指示所述分组中是否存在最低时间级帧或其片段的标志。

4. 如权利要求 3 所述的系统,其特征在于,所述网络中的接收端点或服务器响应于所述接收端点或服务器检测到丢失的 R 帧或帧的一部分发送被格式化为 RTCP 反馈消息的否定确认消息,

其中所述反馈消息包括所述丢失帧的所述序号、指示所述丢失帧层的系列号、以及指示在由所述序号指示的画面之后的画面中哪些也丢失的位屏蔽。

5. 如权利要求 4 所述的系统,其特征在于,一旦接收到所述否定确认消息,所述网络中的所述传送端点或服务器就检查所述丢失帧是否已被新近的帧取代,并且其中如果未被代替,则所述传送端点或服务器相应地重传所述丢失分组,或者如果用包括已被取代的所述丢失帧的一定范围的帧的指示来取代所述丢失分组,则传送所述新近的帧。

6. 如权利要求 1 所述的系统,其特征在于,所述编码遵循 H. 264SVC,并且在针对 SVC 元素的 NAL 单元报头扩展中运送所述数据元素。

7. 如权利要求 6 所述的系统,其特征在于,所述数据元素包括所述针对 SVC 的 NAL 报头扩展中的附加字节,并且其中针对所述 SVC 的 NAL 报头扩展中的标志发信号通知所述附加字节的存在。

8. 一种用于媒体通信的系统,包括:

解码器,所述解码器将用提供两个或多个时间层的技术编码的经压缩数字视频解码,其中经压缩视频画面被构造成一个或多个分组,其中所述解码器被配置成接收:

包含至少一个指示以下的元素的数据元素的分组报头:

对于最低时间级画面,标识所述画面的序号,

对于其它时间级画面,对按照解码次序最近的最低时间级画面的序号的引用,

以使得接收器可通过检查与所引用的画面的帧索引相对应的所述画面在所述接收器处是否可用来检测最低时间级画面的丢失,其中所述最低时间级画面被指定成将被可靠地递送到接收方目的地。

9. 如权利要求 8 所述的系统,其特征在于,所述数据元素包括一组扩展比特和标志,所述比特和标志在被置位时指示存在所述一组扩展比特。

10. 如权利要求 8 所述的系统,其特征在于,一旦检测到丢失最低时间级画面,所述接收器生成指示所述丢失最低时间级画面的序号的被格式化为 RTCP 反馈消息的否定确认消息。

11. 一种用于媒体通信的系统,包括:

解码器,所述解码器将用提供两个或多个时间层的技术编码的经压缩数字视频解码,其中经压缩的视频画面被构造成一个或多个分组并使用 RTP 通过基于 IP 的网络接收,其中所述解码器被配置成接收:

包括以下的 RTP 报头扩展:

与每个层相关联的系列号,

与每个最低时间层画面相关联的序号,以及

用于指示分组是否包含画面或所述最低层时间画面的画面片段的标志,

其中所述序号被使用所述最低时间层画面作为参考的所有其它画面所引用,以使得可通过检查与所引用的系列号和序号相对应的所述画面在所述系统是否可用来立即检测出最低时间级画面的丢失,

其中所述最低时间级画面被指定成将被可靠地递送到接收方目的地。

12. 如权利要求 11 所述的系统,其特征在于,一旦检测到丢失最低时间层画面就发送被格式化为 RTCP 反馈消息的否定确认消息,并且所述反馈消息指示:所述丢失画面的所述序号、所述丢失画面所属的系列号、以及指示在由所述序号所指示的画面之后的画面中有哪些也丢失的位屏蔽。

13. 一种用于通信网络上传送端点与一个或多个接收端点或服务器之间的媒体通信的方法,其中所传送媒体被编码成具有包括最低时间层的多个不同层的线程化编码结构中的帧,所述方法包括提供指示以下的数据元素:

对于最低时间级画面,标识所述画面的序号,以及

对于其它时间级画面,对按照解码次序最近的最低时间级画面的所述序号的引用,

以使得接收端点或服务器可通过检查与所引用的画面帧索引相对应的所述画面在所述接收端点处是否可用来检测最低时间级画面是否丢失,

其中所述最低时间级画面被指定成将被可靠地递送到接收方目的地。

14. 如权利要求 13 所述的方法,其特征在于,所述数据元素另外指示与各个空间或质量层相关联的系列号,

其中所述接收端点或服务器通过确定在所述接收端点或服务器处是否已接收到与所引用的系列号和序号相对应的所述帧来检测特定空间或质量层的最低时间级帧是否丢失。

15. 如权利要求 13 所述的方法,其特征在于,所述网络中的接收端点或服务器响应于所述接收端点或服务器对丢失的 R 帧或帧的一部分的检测发送被格式化为 RTCP 反馈消息的否定确认消息,所述方法还包括:

在所述反馈消息中纳入所述丢失帧的所述序号以及指示在由所述序号指示的画面之后的画面中有哪些也丢失的位屏蔽。

16. 如权利要求 15 所述的方法,其特征在于,还包括:

一接收到所述否定确认消息,就在所述网络中的所述传送端点或服务器处核查所述丢失分组是否已被最近的帧取代;以及

如果未被代替则相应地重传所述丢失分组,或者如果用包括已被取代的所述丢失帧的一定范围的帧的指示来取代所述丢失分组,则重传所述最近的帧。

17. 如权利要求 13 所述的方法,其特征在于,所述编码遵循 H. 264SVC,并且在针对 SVC 元素的 NAL 单元报头扩展中运送所述数据元素。

18. 如权利要求 17 所述的方法,其特征在于,所述数据元素包括针对所述 SVC 的 NAL 报头扩展中的附加字节,并且其中针对所述 SVC 的 NAL 报头扩展中的标志发信号通知所述附加字节的存在。

19. 如权利要求 17 所述的方法,其特征在于,所述数据元素包括针对所述 SVC 的 NAL 报头扩展中与 FGS 编码有关的、未被最低质量层的画面使用的比特。

20. 一种用于将用提供两个或多个时间层的技术编码的经压缩数字视频解码的方法,所述方法包括:

在分组报头中接收用于指示以下的元素:

对于最低时间级画面,标识所述画面的序号,

对于其它时间级画面,对按照解码次序最近的最低时间级画面的所述序号的引用,

以使得接收器可通过检查与所引用的画面帧索引相对应的所述画面在所述接收器处是否可用来检测最低时间级画面的丢失,

其中所述最低时间级画面被指定成将被可靠地递送到接收方目的地。

21. 如权利要求 20 所述的方法,其特征在于,所述数据元素包括一组扩展比特和标志,所述比特和标志在被置位时指示存在所述一组扩展比特。

22. 一种用于将用提供两个或多个时间层的技术编码的经压缩数字视频解码的方法,其中经压缩视频画面被构造成一个或多个分组并使用 RTP 通过基于 IP 的网络被接收,所述方法包括:

接收包括以下的 RTP 报头扩展:

与每个层相关联的系列号,

与每个最低时间层画面相关联的序号,以及

用于指示分组是否包含画面或最低层时间画面的画面片段的标志,

其中所述序号被使用所述最低时间层画面作为参考的所有其它画面所引用,其中所述最低时间级画面被指定成将被可靠地递送到接收方目的地,以及

检查所接收到的画面中的所述 RTP 报头扩展以验证与所引用的系列号和序号相对应的所述画面的可用性,以使得最低时间级画面的丢失可被检测到。

23. 如权利要求 22 所述的方法,其特征在于,还包括:

一检测到丢失最低时间层画面就发送被格式化为 RTCP 反馈消息的否定确认消息,并且所述反馈消息指示:所述丢失画面的所述序号、所述丢失画面所属的系列号、以及指示在由所述序号所指示的画面之后的画面中有哪些也丢失的位屏蔽,由此传送系统可采取纠正动作。

用于视频通信系统中的差错弹性和随机接入的系统和方法

[0001] 相关申请的交叉引用

[0002] 本申请要求 2005 年 12 月 8 日提交的美国临时专利申请 S/N. 60/748, 437、2006 年 3 月 3 日提交的 S/N. 60/778, 760、2006 年 3 月 29 日提交的 S/N. 60/787, 031、2006 年 10 月 16 日提交的 S/N. 60/829, 618 以及 S/N. 60/862, 510 的权益。所有前述优先权申请特此通过全文引用结合于此。

技术领域

[0003] 本发明涉及视频数据通信系统。具体地, 本发明涉及在视频会议应用中提供差错弹性和随机访问能力的技术。

背景技术

[0004] 通过基于分组的现代通信网络(例如, 基于网际协议(IP)的网络)在发送者与接收者之间提供高质量数字视频通信至少由于此类网络上的数据传输通常是以尽力为基础执行的这个事实而成为技术上的难题。现代通信网络中的传输差错通常将其自身表现为分组丢失而非比特差错, 后者是早期通信系统中的特性。分组丢失常常是中间路由器中的拥塞的结果, 而非物理层差错的结果。

[0005] 当在数字视频通信系统中发生传输差错时, 确保接收器可从差错快速恢复并返回传入视频信号的无差错显示是很重要的。然而, 在传统数字视频通信系统中, 接收器的稳健性因传入数据被严重压缩以节约带宽这个事实而降低。此外, 用在通信系统中的视频压缩技术(例如, 最新编解码器 ITU-T H. 264 和 H. 263 或者 ISO MPEG-2 和 MPEG-4 编解码器)可在连续视频分组或帧之间形成十分强的时间依存性。具体地, 运动补偿预测(例如, 涉及 P 帧或 B 帧的使用)编解码器的使用形成其中所显示的帧依赖于先前帧的帧依存性链。依存性链可一直扩展到视频序列的开头。作为依存链的结果, 给定分组的丢失可影响接收器处多个连续分组的解码。由于给定分组的丢失而导致的差错传播仅以“内”(I)刷新点或者完全不使用任何时间预测的帧终止。

[0006] 数字视频通信系统中的差错弹性要求在传送信号中有至少某一级别的冗余度。然而, 这种要求与力图消除或最小化所传送信号中的冗余度的视频压缩技术的目的相反。

[0007] 在提供区分型服务的网络(例如, 基于区分型服务 IP 的网络、租用线路上的专用网等)上, 视频数据通信应用可采用网络特征来以无损或近似无损的方式向接收器递送视频信号数据的部分或全部。然而, 在未提供区分型服务的任意尽力型网络(诸如因特网)中, 数据通信应用必须依赖其自身的特征来实现差错弹性。在文本或字母数字数据通信中有用的已知技术(例如, 传输控制协议-TCP)不适于视频或音频通信, 这对由人类接口需求所导致的较低的端对端延迟有额外的约束。例如, TCP 技术可用于文本或字母数字数据输送中的差错弹性。TCP 保持重发数据直至确认所有数据被接收, 即使其包括若干秒的延迟。然而, TCP 不适用于现场或交互式视频会议应用中的视频数据输送, 因为无限制的端对端延迟对于参与者而言是无法接受的。

[0008] 相关的问题是随机接入。假定接收器加入现行视频信号的传输。典型示例是加入视频会话的用户,或者调谐至广播的用户。此类用户将必须在传入比特流中找到他/她能够开始解码或与编码器同步的点。然而,提供这种随机接入点对压缩效率具有相当大的影响。注意:根据定义,随机接入点是从该点开始任何差错传播终止的差错弹性特征(即,其是差错恢复点)。因此,由特定编码方案提供的随机接入支持越好,其可提供差错恢复支持越佳。逆过程可能并非总为真;其依赖于关于差错弹性技术已被设计成要解决的差错的持续时间和范围所作出的假设。对于差错弹性,某些状态信息可被假定为在接收器处发生差错时可用。

[0009] 视频通信系统中差错弹性的一方面涉及对压缩效率具有相当大的影响的随机接入(例如,当接收方加入现行视频信号的传输时)。例如,随机接入的实例是加入视频会议的用户,或者调谐至广播的用户。这样的用户将必须在传入比特流信号中寻找开始解码或要与编码器同步的合适的点。随机接入点是从该点开始任何传播差错终止的有效的差错弹性特征(或差错恢复点)。因此,提供良好的随机接入支持的特定编码方案通常具有提供较快速差错恢复的差错弹性技术。然而,逆过程依赖于关于差错弹性技术被设计要解决的差错的持续时间和范围的特定假设。差错弹性技术可假定某些状态信息在接收器处发生差错时可用。在此情形中,差错弹性技术不确保良好的随机接入支持。

[0010] 在用于数字电视系统(数字电缆 TV 或卫星 TV)的 MPEG-2 视频编解码器中,以周期间隔(通常为 0.5 秒)使用 I 画面来实现快速切换成流。然而,I 画面显著大于其 P 或 B 对应物(通常大 3 到 6 倍),并由此被避免使用,在低带宽和/或低延迟应用中尤其如此。

[0011] 在诸如视频会议的交互式应用中,请求帧内更新的概念通常被用于差错弹性。在操作中,更新涉及从接收方对发送方的、关于使得解码器被同步的内画面传输的请求。此操作的带宽开销是显著的。另外,当分组差错发生时,也存在此开销。如果分组丢失是由拥塞导致的,则使用内画面只会恶化拥塞问题。

[0012] 另一种用于差错稳健性的传统技术——在过去用于减轻由 IDCT 实现(例如,在 H. 261 标准中)中的失配而导致的漂移——是周期性地编码每个宏块帧内模式。H. 261 标准要求每 132 次强制帧内编码传送一宏块。

[0013] 随着强制要被编码为给定帧中帧内的宏块的百分比的增大,编码效率下降。相反,当此百分比比较低时,从分组丢失恢复的时间增加。强制帧内编码过程要求额外的精力以避免运动相关漂移,该漂移由于必须避免某些运动矢量值——即使它们是最有效的——而进一步限制了编码器的性能。

[0014] 除传统之外,单层编解码器、分层或可缩放编码是多媒体数据编码中的公知技术。可缩放编码用于生成以带宽效率方式共同代表给定媒体的两个或多个“经缩放”比特流。可缩放性可以多个维度——即时间、空间和质量(也称为 SNR “信噪比”可缩放性)——来提供。例如,视频信号可以 CIF 或 QCIF 分辨率且每秒 7.5、15 和 30 帧(fps)的帧速率下的不同层来可缩放地编码。取决于编解码器的结构,空间分辨率和帧速率的任何组合可从编解码器比特流获得。与不同层相对应的比特可作为单独的比特流传送(即,每层一个流)或它们可在一个或多个比特流中被复用在一起。为了便于本文的描述,与给定层相对应的经编码比特可称为该层的比特流,即使各个层被复用且在单个比特流中传送。专门设计成提供可缩放特征的编解码器包括例如 MPEG-2(也称为 ITU-T H. 262 的 ISO/IEC 13818-2)和

目前开发的 SVC(称为 ITU-T H. 264Annex G 或 MPEG-4 第 10 部分 SVC)。在共同转让国际专利申请 No. PCT/US06/028365——“SYSTEM AND METHOD FOR SCALABLE AND LOW-DELAY VIDEOCONFERENCING USING SCALABLE VIDEO CODING(用于使用可缩放视频编码的可缩放和低延迟视频会议的系统和方法)”——中描述了为视频通信专门设计的可缩放编码技术。注意,即使并非专门设计成可缩放的编解码器也可以时间维度呈现可缩放性的特性。例如,考虑 DVD 和数字 TV 环境中使用的非可缩放编解码器——MPEG-2Main Profile(主型)编解码器。此外,假定以 30fps 操作的编解码器以及使用 IBBPBBPBBPBBPBB(周期 $N = 15$ 帧)的 GOP 结构。通过连续消除 B 画面,接着消除 P 画面,得到总共三个时间分辨率——30fps(包括所有画面类型)、10fps(仅有 I 和 P)以及 2fps(仅有 I)——是可能的。连续消除过程导致可解码的比特流,因为 MPEG-2Main Profile 编解码器被设计成使得 P 画面的编码不依赖于 B 画面,并且类似的 I 画面的编码不依赖于其它 P 或 B 画面。在以下,具有时间可缩放特征的单层编解码器被认为是可缩放视频编码的特例,因此被包括在术语可缩放视频编码,除非另外明确指明。

[0015] 可缩放编解码器通常具有锥形比特流结构,其中组成比特流之一(称为“基层”)在恢复某种基本质量的原始媒体时是必需的。连同基层一起使用一个或多个剩余比特流(称为“增强层”)提升了所恢复媒体的质量。增强层中的数据丢失是容许的,但是基层中的数据丢失会导致所恢复媒体的显著畸变或完全失败。

[0016] 可缩放编解码器提出了类似于由单层编解码器针对差错弹性和随机接入所提出的那些挑战。然而,可缩放编解码器的编码结构具有单层视频编解码器中不存在的独特特性。此外,与单层编码不同,可缩放编码可能涉及从一个可缩放性层切换到另一个(例如,在 CIF 与 QCIF 分辨率之间来回切换)。

[0017] 联播是用于视频会议的编码解决方案,它不如可缩放视频编码复杂但具有后者的优点中的某些。在联播中,两个不同版本的源被编码(例如,以两种不同空间分辨率)并传送。每个版本是独立的,因为其解码并不依赖于另一版本的接收。类似于可缩放和单层编码,联播提出了类似的随机接入和稳健性问题。在下文中,联播被认为是可缩放编码(其中不执行层间预测)的特例,并且两者皆简称为可缩放视频编码技术,除非另外明确指明。

[0018] 现在针对改进对视频通信系统中经编码比特流的差错弹性和随机接入能力作出考虑。关注开发对由端对端延迟和系统所用的带宽具有最小影响的差错弹性和随机接入技术。合需的差错弹性和随机接入技术将可应用于可缩放和单层视频编码两者。

[0019] 发明概述

[0020] 本发明提供了用于在基于单层以及可缩放视频编码的视频通信系统中提升差错弹性并提供随机接入能力的系统和方法。

[0021] 在第一示例性实施例中,本发明提供了一种用于在不增加端对端延迟的情况下可靠地传送经编码的视频信号的最低或单个时间层的全部或部分,并在随后使用它来从分组丢失中恢复的机制。提供特定技术以通过 RTP 以及在使用 H. 264Annex G(SVC)NAL 单元时进行传送。

[0022] 在第二示例性实施例中,本发明提供了一种使用基于服务器的内帧来可靠地传送经编码的视频信号的最低或单个时间层,并随后在不会不利地影响其它接收器的情况下使用它来使特定接收器从分组丢失中恢复的机制。

[0023] 在第三示例性实施例中,本发明提供了一种通过以精心协调的方式使用内宏块编码使得从单层或可缩放视频编码的分组丢失中恢复成为可能的机制。

[0024] 在第四示例性实施例中,本发明提供了一种收集和聚集来自一个或多个接收方的反馈以便任选地选择画面参考帧以及分配内宏块的机制。

[0025] 在第五示例性实施例中,本发明提供了一种通过使用来自低分辨率空间层的信息来从高分辨率空间可缩放层的分组丢失中恢复的机制。

[0026] 此外,在第六示例性实施例中,本发明提供了一种在很少的延迟或没有延迟的情况下从低空间或质量分辨率切换到高空间或质量分辨率的机制。

[0027] 通过与速率-失真最优化量化器以及运动模式和矢量选择相耦合,这些实施例在或者独立或者相互组合的情况下允许构建具有高稳健性和较小带宽开销的十分有效的视频通信系统。

[0028] 本文的描述说明了如何使用这些技术来实现对给定视频系统的随机接入,以及藉由其接收器——无需对较低层进行完全解码——就可使用来自所述较低层的信息来有效地重构较高层的高空间分辨率数据的机制。本发明利用可缩放视频编码技术的特定属性来最小化对端对端延迟和带宽的影响。在诸如通过端对端要求苛刻(端对端的最大值为200ms)以及分组丢失率会很严重(即,较低的平均分组丢失率但有较长的突发)的IP网络的视频会议的通信应用中,本发明尤其有用。

[0029] 一旦适当地选择画面编码结构和传输模式,本发明的技术就使得在极小的带宽开销的情况下允许接近即时地进行层切换成为可能。

[0030] 附图简述

[0031] 图1是示出根据本发明原理的用于递送可缩放的经编码视频数据的示例性视频会议系统的框图;

[0032] 图2是示出根据本发明原理的与单层视频编码的使用相兼容的示例性最终用户终端的框图;

[0033] 图3是示出根据本发明原理的与可缩放或联播编码的使用相兼容的示例性最终用户终端的框图;

[0034] 图4是示出根据本发明原理的多点SVCS的内部交换结构的框图;

[0035] 图5是示出SVCS的操作原理的框图;

[0036] 图6是示出根据本发明原理的示例性视频编码器的结构的框图;

[0037] 图7是示出根据本发明原理的用于编码基层和时间增强层的视频编码器的示例性架构的框图;

[0038] 图8是示出根据本发明原理的用于空间增强层的视频编码器的示例性架构的框图;

[0039] 图9是示出根据本发明原理的示例性分层画面编码结构的框图;

[0040] 图10是示出根据本发明原理的另一示例性分层画面编码结构的框图;

[0041] 图11是示出根据本发明原理的包括时间和空间可缩放性的示例性画面编码结构的框图;

[0042] 图12是示出根据本发明原理的用于差错弹性视频通信的示例性分层画面编码结构的框图;

[0043] 图 13 是示出根据本发明原理的具有空间 / 质量可缩放性的用于差错弹性视频通信的示例性分层画面编码结构的框图；

[0044] 图 14 是示出根据本发明原理的用于使用肯定确认进行 LR 画面的可靠递送的通信协议的操作的时序图；

[0045] 图 15 是示出根据本发明原理的用于使用否定确认进行的 LR 画面的可靠递送的通信协议的操作的时序图；

[0046] 图 16 是示出根据本发明原理的、当 R 分组技术被用于通过 RTP 传输时传送终端的 LRP 发送模块的示例性架构的框图；

[0047] 图 17 是示出根据本发明原理的、当 R 分组技术被用于通过 RTP 传输时接收终端的 LPR 接收模块的示例性架构的框图；

[0048] 图 18 是示出根据本发明原理的、当 R 分组技术被用于通过 RTP 传输时服务器的 LRP 发送和接收模块的示例性架构的框图；

[0049] 图 19 示出根据本发明原理的用于 RTP 分组的命名 RTP 报头扩展的示例性结构；

[0050] 图 20 示出根据本发明原理的 RNACK 分组的反馈控制信息字段的示例性结构；

[0051] 图 21 示出现有技术系统中在发生分组丢失时 H. 264SVC 解码器如何达到非正确状态；

[0052] 图 22 示出目前针对现有技术系统定义的 H. 264SVC NAL 报头扩展；

[0053] 图 23 示出根据本发明原理的具有帧索引的经更改 H. 264SVC NAL 报头扩展定义；

[0054] 图 24 示出根据本发明原理的、帧索引位于报头的扩展中的经更改 H. 264SVC NAL 报头扩展定义；

[0055] 图 25 示出根据本发明原理的用于快速转发内恢复的示例性切片编码结构；

[0056] 图 26 示出根据本发明原理的快速转发内恢复可如何结合 SR(增强层)画面使用；

[0057] 贯穿附图中相同附图标记用于指代所例示的实施例的类似特征、元件、组件或部分,除非另外声明。此外,现在参照附图详细描述本发明是结合示例性实施例进行的。

[0058] 本发明的详细描述

[0059] 本发明提供了用于视频通信系统中差错弹性和随机接入的系统和方法。这些机制与可在视频通信系统中使用的具有时间可缩放性的可缩放编码技术以及单层和联播视频编码相兼容。

[0060] 这些系统和方法涉及在视频信号传输中指定一组视频帧或画面来向接收方可靠或有保证地递送。指定组视频帧的可靠递送可通过使用安全或高可靠性的链路、或通过重传技术来实现。可靠递送的视频帧被用作参考画面用于使接收器在差错发生或随机接入之后与所传送视频信号重新同步。

[0061] 在优选实施例中,示例性视频通信系统可以是在基于分组的网络上操作的多点视频会议系统 10。(参看例如图 1)。多点视频会议系统可包括任选网桥 120a 和 120b(例如,多点控制单元(MCU)或可缩放视频通信服务器(SVCS))以协调网络上的端点(例如,用户 1-k 和 1-m)之间的可缩放多层或单层视频通信。对于结合或不结合任选网桥 120a 和 120b 使用的点对点连接而言,示例性视频通信系统的操作是相同的且是有益的。

[0062] 在共同转让的国际专利申请 No. PCT/US06/28365——“SYSTEM AND METHOD FOR SCALABLE AND LOW-DELAY VIDEOCONFERENCING USING SCALABLE VIDEO CODING(用于使用

可缩放视频编码的可缩放和低延迟视频会议的系统和方法)”以及 No. PCT/US06/28366——“SYSTEM AND METHOD FOR A CONFERENCE SERVER ARCHITECTURE FOR LOW DELAY AND DISTRIBUTED CONFERENCING APPLICATIONS(用于低延迟和分布式会议应用的会议服务器架构的系统和方法)”中提供了基于可缩放视频编码的可缩放视频编码技术和视频会议系统的详细描述。此外,在 2005 年 12 月 22 日提交的美国临时专利申请 No. 60, 753, 343——“COMPOSITING SCALABLE VIDEO CONFERENCE SERVER(合成可缩放视频会议服务器)”中提供了基于可缩放视频编码的可缩放视频编码技术和视频会议系统的描述。所有前述国际和美国临时专利申请皆通过引用全部结合于此。

[0063] 图 1 示出了视频会议系统 10 的一般结构。视频会议系统 10 包括经由 LAN1 和 2 链接在网络 100 上的多个最终用户终端(例如,用户 1-k 和用户 1-m)和服务器 120a 和 120b。服务器可以是传统 MCU,或者可缩放视频编码服务器(SVCS)或合成可缩放视频编码服务器(CSVCS)。后者服务器具有与传统 MCU 相同的用途,但是具有显著减小的复杂度和改进的功能。(参看例如国际专利申请 No. PCT/US06/28366)以及 2005 年 12 月 22 日提交的美国临时专利申请 No. 60/753, 343。在本文的描述中,术语“服务器”可用于统指或者 SVCS 或者 CSVCS。

[0064] 图 2 示出被设计成与基于单层编码的视频会议系统(例如,系统 10)一起使用的最终用户终端 10 的架构。图 3 示出被设计成与基于多层编码的视频会议系统(例如,系统 10)一起使用的最终用户终端 140 的架构。终端 140 包括人机输入/输出设备(例如,相机 210A、话筒 210B、视频显示器 250C、扬声器 250D),以及耦合至输入和输出信号复用器和分用器单元(例如,分组 MUX 220A 和分组 DMUX 220B)的一个或多个网络接口控制器卡(NIC)230。NIC 230 可以是标准硬件组件,诸如以太网 LAN 适配器或任何其它合适的网络接口设备或者其组合。

[0065] 相机 210A 和话筒 210B 被设计成分别捕捉参与者的视频和音频信号以传输给其他会议参与者。相反,视频显示器 250C 和扬声器 250D 被设计成分别显示和回放从其他参与者接收到的视频和音频信号。视频显示器 250C 也被配置成任选地显示参与者/终端 140 自身的视频。相机 210A 和话筒 210B 输出分别经由模数转换器 210E 和 210F 耦合到视频和音频编码器 210G 和 210H。视频和音频编码器 210G 和 210H 被设计成压缩输入视频和音频数字信号以减小在电子通信网络上传输信号所需的带宽。输入视频信号可以是实况或预先录制和存储的视频信号。编码器压缩本地数字信号以使传输信号所需的带宽最小化。

[0066] 在本发明的示例性实施例中,音频信号可用本领域中已知的任何合适的技术(例如,G. 711、G. 729、G. 729EV、MPEG-1 等)来编码。在本发明的优选实施例中,可缩放音频编解码器 G. 729EV 被音频编码器 210G 用来编码音频信号。音频编码器 210G 的输出被发送到复用器 MUX 220A 以经由 NIC 230 通过网络 100 传输。

[0067] 分组 MUX 220A 可使用 RTP 协议执行传统复用。分组 MUX 220A 也可执行可由网络 100 提供的任何相关服务质量(QoS)处理。来自终端 140 的每个数据流在其自身的虚拟信道或 IP 技术中的“端口号”中传送。

[0068] 图 3 示出被配置成与其中使用可缩放或联播视频编码的视频会议系统一起使用的最终用户终端 140。在此情形中,视频编码器 210G 具有多个输出。图 3 示出例如标示为“基”和“增强”的两个层输出。终端 140 的输出(例如,单层输出(图 2)或多层输出(图

3)) 经由 LRP 处理模块 270A 连接到分组 MUX 220A。LRP 处理模块 270A (和模块 270B) 被设计成通过处理特定类型的帧 (例如, 图 12 和 13 中的“R”帧) 以及诸如视频序列报头数据的需要可靠传输的任何其它信息的传输来进行差错弹性通信 (“差错弹性 LRP 操作”)。如果视频编码器 210G 产生一个以上的增强层输出, 则增强层输出的每一个可以与图 3 中所示的相同的方式连接到 LRP 处理模块 270A。类似地, 在此情形中, 附加增强层将经由 LRP 处理模块 270B 提供给视频编码器 230A。或者, 增强层输出的一个或多个可直接连接到分组 MUX 220A, 而非经由 LRP 处理模块 270A。

[0069] 终端 140 还可配有一组视频和音频解码器对 230A 和 230B, 其中每一对对应于在视频会议中终端 140 处可看到或听到的每个参与者。应当理解, 尽管在图 2 和 3 中示出解码器 230A 和 230B 的若干实例, 但是使用单个解码器对 230A 和 230B 来连续处理来自多个参与者的信号是可能的。因此, 终端 140 可配有单个解码器对 230A 和 230B 或比参与者的数目少的解码器对 230A 和 230B。

[0070] 音频解码器 230B 的输出被连接到混音器 240, 后者又与驱动扬声器 250B 的数模转换器 (DA/C) 250A 连接。混音器将各个信号组合成用于回放的单个输出信号。如果到达的音频信号是预先混合的, 则可能不需要混音器 240。类似地, 视频解码器 230A 的输出可经由合成器 260 在视频显示器 250C 的帧缓冲器 250B 中被组合。合成器 260 被设计成将每个经解码的画面置于输出画面显示的适当区域。例如, 如果显示器被划分成四个较小的区域, 则合成器 260 从视频解码器 230A 的每一个处获得像素数据并将其放置在适当的帧缓冲器位置 (例如, 通过填充右下方的画面)。为了避免双重缓冲 (例如, 一次在解码器 230A 的输出以及一次在帧缓冲器 250B), 合成器 260 可被实现为驱动解码器 230A 的输出像素的放置的地址生成器。用于最优化各个视频输出对显示器 250C 的放置的其它技术也可用于类似效果。

[0071] 例如, 在 H. 264 标准规范中, 通过使用灵活宏块排序 (FMO) 方案将多个参与者的视图组合在单个经编码的画面中是可能的。在此方案中, 每个参与者占用经编码图像的一部分——包括其切片之一。在概念上, 单个解码器可用于解码所有参与者的信号。然而, 从实践角度看来, 接收器 / 终端将必须解码四个较小的独立编码的切片。因此, 图 2 和 3 中所示具有解码器 230A 的终端 140 可用在 H. 264 规范的应用中。注意: 用于转发切片的服务器是 CSVCS。

[0072] 在终端 140 中, 分用器 DMUX 220B 接收来自 NIC 320 的分组并经由接收 LRP 模块 270B 将它们重定向到适当的解码器单元 230A, 如图 2 和图 3 中所示。视频解码器 230A 的输入处的 LRP 模块 270B 终止接收终端的那一端处的差错弹性 LRP 操作 (图 12 和 13)。

[0073] MCU 或服务器控制块 280 协调服务器 (SVCS/CSVCS) 与最终用户终端之间的交互。在没有中间服务器的点对点通信系统中, 不需要服务器控制块。类似地, 在非会议应用中, 在接收最终用户终端处仅需要单个解码器。对于涉及所存储的视频 (例如, 预先录制的广播、预先编码的材料) 的应用中, 传送最终用户终端可不涉及音频和视频编码块或它们之前的所有终端块 (例如, 相机、话筒等) 的整体功能。具体地, 仅需要提供如下所说明的、与视频分组的可选择性传输有关的部分。

[0074] 应当理解, 终端 140 的各个组件可以是彼此互联的、物理上单独的软件和硬件设备或单元 (例如, 集成在个人计算机中), 或者可以是其任何组合。

[0075] 图 4 示出用于差错弹性处理应用中的示例性 SVCS 400 的结构。SVCS 400 的核心是确定来自每一个可能源的哪个分组被传送到哪个目的地和通过什么信道的交换机 410。(参见例如 PCT/US06/028366)。

[0076] 示例性 SVCS 400 的操作原理可参照图 5 来理解。在此示例中,传送终端或端点处的 SVC 编码器 510 产生除多个时间层(未在图中示出)之外的三个空间层。各个经编码的视频层在各个分组中自传送端点(SVC 编码器)传送到 SVCS 400。取决于网络条件或用户偏好,SVCS 400 决定哪个分组要被转发到所示三个接收方/解码器 520 的每一个。在图 5 所示示例中,SVCS 400 仅将第一和第二空间层转发到 SVC 解码器 520(0)、将所有三个空间层转发到 SVC 解码器 520(1)、以及仅将第一(基)层转发到 SVC 解码器 520(2)。

[0077] 再次参看图 4,除 PCT/US06/028366 中描述的交换机之外,SVCS 400 还包括分别设置在交换机的输入和输出处的 LRP 单元 470A 和 470B。SVCS 400 被配置成在其传入交换机连接处终止弹性 LRP 处理,以及在其传出交换机连接处发起弹性 LRP 处理。在使用 SVCS 400 的本发明的实现中,差错弹性 LRP 处理并非在网络上端对端地执行,而是仅在每个单独的连接段(例如,发送方到 SVCS、SVCS 到 SVCS 以及 SVCS 到接收方)上执行。然而,应当理解,本发明的差错弹性 LRP 处理可在使用或不使用 SVCS 的情况下以端对端方式在网络上执行。没有 LRP 单元 470A 和 470B 的 SVCS 400 可在使用 SVCS 的网络中用于端对端 LRP 处理。此外,SVCS 400 可配备有一个以上 NIC 230,如果 SVCS 400 通过不同网络连接用户,则通常是这种情形。

[0078] 图 6 示出可用于差错弹性视频通信系统中的示例性视频编码器 600 的架构。视频编码器 600 可例如是基于块的经运动补偿的转换编码器。H. 264/MPEG-4AVC 设计是视频编码器 600 的优选设计。然而,也可使用其它编解码器设计。例如,图 7 示出用于编码基于 SVC 设计的基层和时间增强层的示例性视频编码器 600' 的架构,而图 8 示出用于编码空间增强层的示例性视频编码器 600'' 的架构。(参见例如 PCT/US06/28365 和 PCT/US06/028366)。视频编码器 600' 和 600'' 包括可用于使用空间可缩放性降低系统中的输入分辨率(例如,从 CIF 到 CIF)的任选输入下采样器 640。

[0079] 图 6 还示出可使用视频编码器 600 实现的编码过程。编码器 600 中的增强参考控制(ENC REF CONTROL)620 用于创建“线程化”编码结构。(参见例如 PCT/US06/28365 和 PCT/US06/028366)。基于标准块的经运动补偿的编解码器基于 I、P 和 B 画面的常规结构。例如,在诸如 IBBPBBP 的画面序列(按显示次序)中,‘P’帧是从前一 P 帧或 I 帧预测的,而 B 画面是使用前一以及后一 P 或 I 帧两者来预测的。尽管连续的 I 或 P 画面之间的 B 画面的数目如同 I 画面出现的速率一样可变化,但是例如 P 画面不可能用作在时间上比最近的 P 画面早的另一 P 画面的参考。H. 264 是个例外,因为编码器和解码器保持两个参考画面列表。选择哪些画面用作参考以及哪些参考用于要被编码的特定画面是可能的。图 6 中的帧缓冲器块 610 代表存储参考画面列表的存储器,而编码器侧的增强参考控制 620 决定对当前画面使用哪些参考画面。

[0080] 参照示出示例性分层画面编码结构 900 的图 9 可更好地理解增强参考控制 520 的操作。为了实现多个时间分辨率,用于视频通信系统中的编解码器可生成多个单独的画面“线程”。给定级的线程被定义为使用或者来自相同线程的画面、或者来自较低级线程的画面来进行运动补偿的画面序列。线程的使用允许实现时间可缩放性,因为可消去任何数目

的顶级线程而不影响其余线程的解码过程。

[0081] 在本发明的优选实施例中,使用具有一组三个线程的编码结构(例如,图9中的结构900)。在图9中,画面标志中的字母‘L’指示任何可缩放层。L之后的数字(0、1和2)标识时间层,例如,其中“0”对应于最低、或最粗糙的时间层,而“2”对应于最高或最细的时间层。图9中所示的箭头指示预测的方向、源和目标。在多数应用中,将仅使用P画面,因为使用B画面增加编码延迟达捕捉和编码用于B画面的参考画面所占用的时间。然而,在并非延迟敏感的应用中,可能除L0画面之外,画面的一部分或全部可以是B画面。类似的,L0画面可以是形成传统画面组(GOP)的I画面。

[0082] 继续参看图9,层L0仅是相隔四个画面的常规P画面序列。层L1具有与L0相同的帧速率,但是仅允许根据前一L0帧进行预测。层L2帧是根据最近的L0或L1帧预测的。L0提供完全时间分辨率的四分之一(1:4),L1是L0帧速率的两倍(1:2),而L2是L0+L1帧速率的两倍(1:1)。

[0083] 可在设计成适应本发明的特定实现的不同带宽/可缩放性要求的编码结构中类似地构造比以上讨论的三个层L0、L1和L2更多或更少的层。图10示出在仅有两个层L0和L1的线程化编码结构1000中转换一系列传统预测帧IPPP...的示例。此外,图11示出用于空间可缩放性的线程化编码结构1100的示例。编码结构1100包括由字母‘S’指示的用于增强层的线程。应当注意,增强层的帧可具有与基层的帧不同的线程结构。

[0084] 用于编码时间层的视频编码器600’(图7)可被扩充以编码空间和/或质量增强层。(参见例如PCT/US06/28365和PCT/US06/028366)。图8示出用于空间增强层的示例性编码器600”。除基层信息也可用于编码器600”之外,编码器600”的结构和功能类似于基层编解码器600’。此信息可包括运动矢量数据、宏块模式数据、经编码的预测差错数据或经重构的像素数据。编码器600”可重用此数据中的部分或全部以针对增强层S作出编码决定。该数据必须被缩放成增强层的目标分辨率(如果基层为QCIF且增强层为CIF,则缩放2的倍数)。尽管空间可缩放性通常需要维护两个编码环路,但是通过将基层中用于增强层编码的数据仅限于可从当前画面的基层中编码的信息计算出的那些值来执行单环路解码是可能的(例如,在H.264SVC草案标准中)。例如,如果基层宏块是帧间编码(inter-code)的,则增强层不能使用该宏块的经重构像素作为预测的基础。然而,它可使用其运动矢量和预测差错值,因为它们可通过只解码当前基层画面中包含的信息来获得。由于解码器的复杂度被显著降低,因此单环路解码是可取的。

[0085] 质量或SNR可缩放性增强层编解码器可以与空间可缩放性编解码器相同的方式来构造。对于质量可缩放性,与在较高分辨率版本的输入上构建增强层不同,编解码器以相同空间分辨率编码残余预测差错。关于空间分辨率,在单或双环路编码配置中,基层的所有宏块数据可在增强层被重用。出于简便起见,本文的描述一般针对使用空间可缩放性的技术。然而,应当理解,相同的技术可应用于质量可缩放性。

[0086] 通过引用结合于此的国际专利申请PCT/US06/28365[SVC编码]描述了线程编码结构(例如,编码结构900)在其针对存在传输差错的稳健性方面所具有的不同优点。在基于运动补偿预测的传统技术水平的视频编解码器中,时间依存性是固有的。给定画面处的任何分组丢失不仅影响该特定画面的质量,而且或者直接或者间接地影响以给定画面作为参考的所有将来画面。这是因为解码器对将来的预测可构建的参考帧将与编码器处所用的

参考帧不同。随后的差异或漂移会对传统技术水平的视频编解码器造成极大的影响。

[0087] 相反,图 9 中所示的线程结构创建三个完备线程或依存链。关于 L2 画面发生的分组丢失将仅影响 L2 画面;L0 和 L1 画面仍可被解码和显示。类似地,关于 L1 画面发生的分组丢失将仅影响 L1 和 L2 画面;L0 画面仍可被解码和显示。此外,线程结构可被创建成包括 S 画面的线程或依存链(例如,图 11)。图 11 中所示的示例性 S 分组线程结构 1100 具有与图 9 中所示的 L 画面线程结构 900 相类似的特性。S2 画面处的分组丢失仅影响特定画面,而 S1 画面处的丢失还将影响随后的 S2 画面。在任一情形中,一旦将下一 S0 画面解码,漂移就将终止。

[0088] 再次参看图 9,在画面质量方面,L0 画面处发生的分组丢失可能是灾难性的,因为将会影响所有画面类型。如前所述,对此问题的传统解决方案是周期性地编码 L0 画面作为内画面或 I 画面。然而,实现这种解决方案的带宽开销会是显著的,因为 I 画面通常比 P 画面大 3-6 倍。此外,导致需要使用 I 画面的分组丢失通常是网络拥塞的结果。试图通过网络发送 I 画面来补救分组丢失只会恶化拥塞问题。

[0089] 比使用 I 画面传输来补救分组丢失更好的技术是将 L0 的特定百分比的内宏块编码为在任何给定画面中内部。这种技术有助于将比特率负荷分摊到多个画面上而不是将负荷集中到单个画面中。已被编码为在给定画面中内部的宏块在同一周期中无需再次被强制编码为在内部。在有限数目的画面之后,接收器/解码器将具有接收到的、关于画面中的所有宏块位置的内部信息。在使用这种技术时,在编码器处必须注意不将失真预测引入到已经由运动预测编码为内部的区域中(即,“安全”对“非安全”帧区域)。因此,在编码器处,在给定周期中出于稳健性目的已将宏块编码为内部之后,对同一帧区域的将来时间预测可仅从同一周期中已被编码为内部的位置开始进行。在给定 L0 画面中以帧内模式编码宏块的约 10-15%可实现良好的权衡。结果,在约十个 L0 帧(即,40 个画面、或每秒 30 帧下的 1.3 秒)之后,解码器将与编码器在 L0 层处重新同步。应当注意,当解码器在内部刷新周期开始之后立即加入流时,它必须等待下一周期开始以及等待直至下一周期完成,以便进行同步(即,总共接近两个周期的延迟)。由于画面编码结构(例如,结构 900)的层依存性,随后的 L1 和 L2 画面也将被准确解码,只要它们的数据被准确接收到。因此,如果基层 L0 和某些增强层画面以其递送被保证的方式传送,则在分组丢失的情形中没有严重后果的情况下其余层可基于尽力来传送。这种有保证传输可使用诸如区分型和 FEC 等已知技术来执行。在本文的描述中,也可参照提供这种区分型服务质量的作为两个实际或虚拟信道的高可靠性信道(HRC)和低可靠性信道(LRC)(图 1)。(参见例如 PCT/US06/28365 和 PCT/US06/28366)。在使用可缩放视频编码结构(图 11 中的结构 1100)的视频通信系统中,层 L0-L2 和 S0 可例如在 HRC 上可靠传送,而 S1 和 S2 可在 LRC 上传送。尽管 S1 或 S2 分组的丢失可导致有限的漂移,但是仍期望能够尽可能多地避免信息丢失。

[0090] 内宏块编码技术的一个缺点是在某些差错状况下,实现充分的 I 块所需的 L0 帧之一可能丢失,由此阻碍了过程的收敛。这种技术的另一个缺点是无论信道的状况如何都有编码效率惩罚。换言之,强制内宏块将产生带宽开销,即使通信中完全没有分组丢失也是如此。

[0091] 本发明的差错弹性技术通过采用 L0 层的子集或整个 L0 层的可靠传输来克服用于补偿分组丢失的传统技术的前述限制。差错弹性或可靠性通过重发来确保。本发明的差错

弹性技术不仅出于显示目的而被设计成恢复丢失的画面,还被设计成创建正确的参考画面以便于解码依赖于丢失分组中所包含的那些画面(整体或部分)的将来画面。在本发明的系统实现中,L0画面的可靠传输可通过LRP模块(例如,图2中的模块270A和270B以及图4中的模块470A和470B)根据适当的保护协议(例如,图14中的协议1400)在发送和接收方之间使用肯定或否定确认来执行。

[0092] 图12示出示例性画面编码结构1200,其中L0基层和L1-L2时间增强层与至少一个可靠传送的基层画面相耦合以进行差错弹性视频通信。在编码结构1200中,除标示为L0-L2画面的常规基层和增强层之外,还有称为LR(‘R’代表可靠)的新的画面类型。注意:在图12中所示的编码结构1200中,层LR和L0-L2可等效于分别被标示为L0-L3,因为LR画面总是经编码的视频信号的最低时间层。根据用于差错弹性视频通信的本发明,可以是P画面的LR画面被指定成将被可靠地递送到接收方目的地。

[0093] 本发明的差错弹性技术的操作可通过考虑其中L0画面之一因分组丢失而被破坏或丢失的示例来理解。如前所述,在传统通信系统中,丢失L0画面对所有后续L0-L2画面的影响是严重的。使用本发明的画面编码结构1200,在丢失L0画面之后的下一“可靠递送”LR画面提供了重新同步点,接收方/编码器可在该点之后继续无失真地解码和显示。

[0094] 在图12中所示的编码结构1200中,LR画面之间的时间距离例如为12帧。LR画面的可靠递送利用了具有极长时间距离(6帧或更多)的P画面约为I画面的大小的一半,并且可靠递送并非旨在确保相关画面的及时显示而是旨在创建合适的参考画面以备将来使用这个事实。因此,LR画面的递送可通过在连续LR画面之间的一段时间内极轻微地增加系统中的带宽来实现。

[0095] 编码结构1200可使用现有的H.264AVC标准——在该标准下LR画面可例如被存储在解码器处作为长期参考画面并可用MMCO命令来替换——来实现。

[0096] 图13示出示例性画面编码结构1300,其中LR画面的概念被应用于增强层画面(或者空间或者质量可缩放性)。这里,要被可靠传送的画面被标示为SR,而对于LR画面,它们组成空间或质量增强层的最低时间层。

[0097] 注意:尽管出于例示的目的LR画面的概念在此通常被描述成应用于经编码的视频信号的最低时间层,但此概念也可根据本发明原理被扩展或应用到附加层。此扩展应用将导致以可靠方式传送附加层。例如,参看图12,除LR画面之外,L0画面也可被纳入可靠(重)传送机制中。类似地,可包括任何空间/质量增强层的画面(来自最低或附加时间层)。此外,视频序列报头或其它数据可被当作或视为与系统中的LR等效,以使得它们(报头或其它数据)被可靠地传送。在下文中,出于描述简便的目的,假定只有LR画面被可靠传送,除非另外明确指明。然而,应当容易理解,附加层或数据可以完全相同的方法来可靠传送。

[0098] 当没有分组丢失时,用于LR帧的可靠递送的带宽开销为零或可忽略是合乎需要的。这意味着动态闭环算法应当被用于可靠递送机制。在例如LR帧被主动重发多次的情况下,使用开环算法也是可能的。

[0099] 图14示出用于LR帧的可靠递送的优选机制或协议1400。协议1400采用基于肯定确认(ACK)消息的机制向发送方(例如,发送器、SVCS1或SVCS2)指示特定LR画面已被预期接收方(例如,SVCS1、SVCS2或接收器)接收到。参看图14中的时间轴,如果在指定时

间间隔内（例如，一次往返行程的时间（RTT））未接收到确认，则发送器处的定时器发起给定 LR 画面的重传。除针对 LR 画面使用常规、周期或静态结构定义之外，还可采用动态结构。在此情形中，在系统操作中动态地定义 LR 画面。在发送器从所有接收器接收到对传送流中特定帧的接收的肯定确认之后，视频通信系统随后可将此帧指定为 LR 帧并使用它作为新的锚定或同步点。换言之，在所有接收器已确认它们皆已正确接收到特定画面之后，发送编码器将使用该特定画面作为 LR 画面。如果特定 LR 画面变得陈旧，则发送器可放弃该特定 LR 画面，并且在任何时候尝试建立与较新画面的重新同步点。如果使用否定确认（NACK）消息取代肯定 ACK 消息，则协议 1200 的操作是类似的。在此情形中，发送器一旦接收到 NACK 就立即重发给定画面。

[0100] 当通信系统中存在 SVCS 时，它可任选地充当 ACK 消息的聚集点。在此情形中，SVCS 可仅向发送器发送单个归纳确认消息以指示所有预期上游接收器皆已接收到 LR 画面。这种特征有助于使通信系统的不同组件之间的控制消息通信量最小化。或者，SVCS 可充当 ACK 消息的终止点（‘ACK 终止模式’）。在此模式中，SVCS 立即确认接收到的 LR 画面并对其进行高速缓存。在此情形中，发送器不希望来自 SVCS 的上游的其他接收器的其它确认。然后，‘终止模式’SVCS 按需执行对下游 SVCS 或接收器的重发以确保可靠递送，并在所有接收器皆已确认接收之后从其高速缓冲器中移除 LR 画面。可利用此模式来将特定接收器 / 端点与有问题的连接隔离开，从而使其它端点之间的通信不受到影响。注意：在 ACK 终止模式中，不再可能在发送器处动态地将画面定义为 LR 画面，因此在此情形中周期性或静态模式是适当的。

[0101] 示例性协议 1200（使用肯定确认，但不用 ACK 聚集或终止）的操作的细节可参照图 14 来理解。附图示出例如通过两个单独的 SVCS 单元 1 和 2 通信的发送器和接收器。应当理解，协议 1200 的操作通常与其中不使用 SVCS 的系统（例如，具有发送器与接收器之间的直接连接的系统）以及其中使用一个或多个 SVCS 的系统中的相同。

[0102] 参看图 14，发送器在时刻 t_0 传送作为 LR 状态的候选的 L0 帧。该帧可在一个或多个传输层分组中输送。在此出于描述方便起见，可假定使用单个分组。此外，如果使用帧分段——在此情形中重传会影响丢失的特定片段，但不一定是整个帧——则操作是相同的。

[0103] 在 SVCS1 处期望在给定时刻 t_1-t_0 内接收到包含 LR 帧（LR）的分组。此时，发送器期望 SVCS1 生成针对该帧的肯定确认消息（ACK）。如果在系统的往返行程时间（RTT）内没有接收到这种 ACK，则发送器假定分组已丢失并在时间 t_2 重传 LR 帧。现在假定在 SVCS1 处接收到帧。针对发送器的 ACK 可由 SVCS1 生成，该 SVCS1 还将帧转发到 SVCS2。与发送器相类似，SVCS1 还将经历多次帧重传直至 SVCS2 确认其接收。图 14 示出在时间 t_6 由 SVCS2 接收到由 SVCS1 发送的 LR 帧。然后，SVCS2 将持续向接收器传送该帧直至其接收到来自接收器的 ACK（例如，ACK 1410）（例如，在时间 t_8 ）。当最终用户接收器（并非中间 SVCS）接收到 LR 帧时，它通知初始发送器其现在具有可用作供将来画面的编码用的参考画面的此新的正确接收到的帧。此 ACK14010 经由 SVCS 传播到达发送器（例如，在时间 t_{10} ）。在特定视频通信会话中的所有接收器确认对新的 LR 帧的正确接收之后，发送器可随后使用所传送的帧作为参考画面。

[0104] 如前所述，在 H. 264 视频编码标准中，通过将候选传送画面标记为长期参考画面来便于使用所传送的帧作为参考画面。类似的标记技术可与其它编码方案一起使用。候选

传送画面未被用作参考画面直至已从所有接收器收集到肯定 ACK。注意：在 LR 协议 1400 运行的整个时间内，发送器持续传送经编码的视频。换言之，由于协议要求可能的重传，因此不引发附加的端对端延迟。LR 处理机制的目标是创建用于将来画面的编码的可靠参考画面。实际上，LR 画面的初始传送被破坏且在特定接收器处无法适当的播放是可能的。发送器（或 SVCS）将持续传送画面直至其被特定接收器正确地接收到，而接收器将持续尝试解码和回放发送器将继续传送的后续视频帧。

[0105] 图 15 示出使用否定确认 (NACK) 的协议 1500 的操作。与使用 ACK 的协议的操作的不同之处在于现在接收端点或 SVCS 具有检测何时 LR 画面未被接收到且已丢失的任务。RTP 或 H. 264 传输中丢失检测用的具体技术将在下文中描述（例如，参照图 16-24）。这些技术使得一接收到任何后续画面就能够检测到丢失。在协议 1500 的操作中，当接收端点或 SVCS 检测到 LR 画面已丢失时，它就向传送端点或 SVCS 发送 NACK 消息。然后，传送端点或 SVCS 从其高速缓冲器获得丢失的画面，并重传或者丢失的帧或者将使得接收器能够重新同步其解码器的更近的 LR 画面。

[0106] 继续参看图 15，假定使用图 12 的画面编码结构（四个时间层，即 LR 和 L0-L2），以及发送器和接收器通过 SVCS 通信。此外，假定由发送器在时间 t_0 传送的 LR 画面丢失，而后续的画面——L0 画面——被成功地传送到 SVCS。一接收到 L0 画面，SVCS 就检测作为参考的 LR 画面已丢失，并传送 NACK，后者由发送器在时间 t_R 接收到。同时，发送器也已在时间 t_2 传送 L1 帧。一旦在时间 t_R 接收到 NACK，发送器就向 SVCS 重传最近的 LR 画面。发送器继续以适当的时间间隔传送初始画面流，例如，在时间 t_3 传送 L2 画面而在时间 t_4 传送 L1 画面。注意：无论所需的 LR 画面是否已丢失，SVCS 都立即向下游接收器转发已从接收器处成功接收到的任何画面。假定针对接收器的所有传输都成功，则当在接收器处接收到重传的 LR 画面时，接收器将具有用以解码在较早时间 t_3 和 t_4 接收到的 L0 和 L1 画面所必需的所有信息。尽管可能太迟而不能显示这些画面，但是接收器（例如，在解码画面而非显示它们的“恢复模式”中）可将其解码以便具有用于正确解码在时间 t_5 到达的 L2 画面的正确参考画面。如果接收器具有足够的 CPU 能力，则这种解码可比实时更快地完成。在时间 t_5 ，接收器可随后开始常规地解码和显示没有差错的传入视频信号，并且没有因丢失而导致任何延迟。应当注意，如果所选的接收器改为在 L2 之前显示 LR、L0 和 L1 画面，则通信会话的常规（没有丢失的情况下）端对端延迟将增加 SVCS 恢复所丢失的 LR 画面所用的时间量。在交互式通信中，附加延迟是不合需要的，并且消除该附加延迟是本发明的优点之一。

[0107] 使用 RTCP 或其它反馈机制，可使用例如上述的肯定和否定确认技术通知发送器特定接收器正经历丢失分组。反馈可与针对每个单独分组的各个 ACK/NACK 消息一样详细。使用反馈使得编码器能够计算（确切或大致地）解码器的状态，并相应地进行动作。此反馈被生成并由可靠性和随机接入控制 (RFC) 模块 530（图 6）来收集。RRC 模块可随后指令编码器使用内宏块，或适当地增加其频率以在需要时进一步帮助同步过程。

[0108] 当使用肯定确认时并且为了使已经历丢失分组的接收器能够重新同步到经编码的比特流，发送器可使用最近的 LR 画面作为参考画面来选择编码当前帧。在知道此 LR 画面已被可靠接收到的情况下，发送器可使用 LR 画面作为参考来将当前画面编码为 P 画面。在接收器正确地接收到当前画面之后，从此时开始它可在参考画面缓冲器方面与编码器同步。换言之，将消除解码器中出现的任何漂移。

[0109] 类似地,当使用否定确认时,解码器可通过解码给定画面的所有必要参考画面来与比特流重新同步,即使它们太迟到达而不能被显示时也是如此。如果解码器可比实时更快地(换言之,所占用的解码时间少于画面之间的时间)解码,则其最终可与接收到的比特流同步。通过在同步点开始显示,解码器可继续常规解码和显示操作而没有任何附加的端对端延迟被添加到通信会话。

[0110] 用于重新同步接收器的这些技术在涉及例如 5-10 个参与者的中型到大型视频会议中具有明显的优势。在此类会议中,使用 I 帧实现已经历分组丢失的接收器的重新同步会对所有参与者施加相当可观的带宽惩罚。实际上,最弱链路(即,具有最多差错的链路)上的参与者将影响具有最强链路的参与者的质量。通过使用 LR 画面,消除了内画面的使用。尽管基于 LR 画面的 P 画面也具有带宽开销,只要帧之间的时间距离不是太大,开销就会显著小于 I 画面的。用于重新同步的 LRP 技术也适于诸如往返行程延迟、服务器的分布等的系统参数。系统越好,在接收器处准确接收到时 LR 画面将建立得更快,从而导致对基于 LR 画面的更好预测,这又将导致更小的开销。

[0111] 注意,当使用反馈时,可能不必先验决定 LR 帧的结构。实际上,LR 帧的结构可通过收集和整理来自所有接收器的反馈来统计和动态地建立。被确认为被所有接收器接收到的帧可被自动视为 LR 帧。

[0112] 在某些情形中,LR 画面的缺点在于到视频会议的单个较差的连接仍会使针对所涉及的所有参与者的质量下降。在此类情形中,中间 SVCS 可扮演发送器代理的角色,并在其余参与者未受影响地继续会议的同时持续重发所需的数据。例如,在转发 SVCS 到毗连 SVCS 或所连接的端点的连接使得实现来自其对等设备的肯定确认的时间大于预先配置的值的情况下,转发 SVCS 可被配置成就像端点确实发回肯定确认(包括发回适当的 ACK)那样对待该端点。这种配置限制了有问题的端点或 SVCS 连接对整体系统的影响。从此时开始,转发 SVCS 将仅向其有问题的对等设备传送 LR 帧,因为它是最终与解码过程重新同步所需的最小信息。如果较新的 LR 帧自发送器到达转发 SVCS,则它们将继续被重传到有问题的 SVCS 或端点,由此给予有问题的 SVCS 或端点与发送器比特流同步的其它机会。由于没有其它帧(除 LR 之外的)在此链路上传送,因此不会因这种重传导致附加拥塞。实际上,如果此类经高速缓冲和重传的 LR 帧的数目超过特定预定义数目(例如,2-3),则转发 SVCS 可考虑要终止特定的有问题的 SVCS 或端点连接。所终止的 SVCS 或端点随后将必须使用其可用的任何合适的随机进入机制来重新加入视频会议会话。

[0113] 在连接或链路中断是暂时的情况中,接收端点可以其正确次序解码重传的 LR 帧并重新加入会话。期望由于 LR 帧的数目远小于帧的总数,因此 CPU 负载将不会是个问题且接收端点可跟得上解码过程。

[0114] 应当理解,图 14 中所示的协议 1400 是示例性的,并且其可易于针对其它系统性能的改进进行更改。例如,在经更改的协议 1400 中,一直传回到发送器的确认(例如,图 14 中所示的 ACK[RCVR] 消息)不必从接收端点发起,而是可仅从最靠近链中端点的最后 SVCS 发起。连接到端点的最后 SVCS 可首先发回 ACK[RCVR],并随后如上所述地继续向端点可靠传送或重传 LR 帧。协议 1400 的这种更改避免在发回 ACK[RCVR] 之前必须等待预先配置的时间。

[0115] 如可为本领域技术人员所显见的,根据本发明的原理,可用其它合适的传输层机

制替代用于实现 LR 帧的可靠传输的 ARQ 协议（例如，协议 1400）。用于 LR 帧的合适传输的合适传输层机制包括诸如主动重传、以及诸如使用交织的 Reed-Solomon（理德 - 所罗门）码更复杂的 FEC（前向纠错）技术和混合式 FEC-ARQ 技术（参见例如 Rubenstein 等人在 2001 年 3 月的计算机通信杂志上发表）。

[0116] 在本发明的实现中重点考虑接收器（例如，接收端点或 SVCS）如何在最小延迟的情况下检测到 LR 画面已丢失。本发明包括基于画面编号和画面编号引用的技术。此技术通过向 LR 画面指派与这些 LR 画面分组一起被运送的序号来操作。接收器维护其已接收到的 LR 画面的编号的列表。另一方面，非 LR 画面包含最近 LR 画面在解码次序中的序号。此序号引用允许接收器甚至在接收后一 LR 画面之前检测丢失的 LR 画面。当接收器接收到 LR 画面时，其可通过将其画面编号与其所维护的画面编号列表（接收到的画面的编号应当比前一画面的编号大一，或者如果计数重新开始则为 0）作比较来检测是否已丢失先前 LR 画面中的一个或多个。当接收器接收到非 LR 画面时，它进行测试以查看所引用的 LR 画面的编号是否存在于其编号列表中。如果不存在，则假定为丢失，并且可发起纠正动作（例如，NACK 消息被传回发送器）。

[0117] LR 画面可使用标志或其它信令手段（例如，根据其分组报头或分组有效载荷参数导出）等来标识，或者它们的存在被暗示（例如，通过它们在经编码的视频序列中的次序）。作为 LR 画面编号的使用的例示，假定两个画面 LR 和 L0 的序列以此次序被传送。接收器的编号列表最初为空。此外，假定对 LR 画面指派序号 0。LR 画面将在其分组中指示编号 0 的情况下被传送。L0 画面也在包括编号 0 作为对其所依赖的 LR 画面——最近 LR 画面——的参考的情况下被传送。如果 LR 画面丢失，则接收器将接收到包含对具有编号 0 的 LR 画面的引用的帧 L0。由于此编号在其列表中不存在（列表仍为空），因此接收器检测到具有编号 0 的 LR 画面已丢失。它可随后请求重发丢失的 LR 画面。

[0118] 注意，使用 LR 画面编号技术检测丢失的 LR 画面可在接收端点以及中间 SVCS 两者处执行。例如在 LRP（接收）模块 270B（图 2 和 3）或模块 470B（图 4）处执行此操作。

[0119] 本文描述了 LR 画面编号技术的两个不同实施例。一个实施例（在下文中称为‘R 分组’技术）适用于系统使用 RTP 协议进行传输的情况。另一实施例适用于系统使用 H. 264 Annex G (SVC) 草案标准的情况。

[0120] 对于 R 分组技术，假定 RTP 协议（在 UDP 和 IP 上）被用于两个终端之间——可能经由一个或多个中间服务器——的通信。注意：媒体传送终端可执行实时编码，或者可从本地或其它存储（RAM、硬盘、存储区网、文档服务器等）访问媒体数据。类似地，接收终端可执行实时解码，以及它可将接收到的数据存储在本地或其它存储中以备将来回放，或者两者。对于在此的描述，假定（并非限制）实时编码和解码正在进行。

[0121] 图 16 示出传送终端的 LRP 发送模块（例如，图 2 中的模块 270A）的架构。LRP 发送模块包括具有用于可能要求重传的分组的本地存储（例如，缓冲器 1605）的分组处理器（R 分组控制器 1610）。R 分组控制器 1610 对 R 分组作标记，并且还对 RNACK 作出响应。R 分组控制器被连接到实现 RTP/UDP/IP 协议栈的复用器 MUX 1620 和分用器 DMUX 1630。尽管 MUX 1620 和 DMUX 1630 在图 16 中被示为单独的实体，但它们可被组合在同一单元中。MUX 1620 和 DMUX 1630 被连接到提供物理层接口的一个或多个网络接口控制器（NIC）。在优选实施例中，NIC 是以太网适配器，但可使用任何其它 NIC，这对于本领域技术人员是显而易

见的。

[0122] 类似地,图 17 示出接收终端的 LRP 接收模块(例如,图 2 中的模块 270B)的示例性架构。在此,R 分组控制器(例如,控制器 1610')负责分组丢失检测和适当 NACK 消息的生成。此外,图 18 示出服务器的 LRP 发送和接收模块(例如,图 4 中的模块 420A 和 420B)——可与接收终端的组件以及背靠背(back-to-back)连接的传送终端的那些组件相同——的结构。

[0123] 在优选实施例中,传送终端根据 RTP 规范对媒体数据进行分组。注意:尽管针对 RTP 定义了不同的分组(称为“有效载荷”)格式,但它们可共享同一公共报头。本发明针对 RTP 分组引入命名报头扩展机制(参见 2006 年 2 月 D. Singer 的 draft-ietf-avt-rtp-hdext-01(进行中)——“A general mechanism for RTP Header Extensions(RTP 报头扩展的通用机制)”)以使得 R 分组可被适当处理。

[0124] 根据本发明,在包含 R 分组的 RTP 会话中,用命名报头扩展机制来对各个分组作标记。R 分组报头扩展元素标识 R 分组自身以及先前发送的 R 分组。例如,此报头扩展元素具有名称“com.layeredmedia.avtr-packet/200606”。每个 R 分组包括——且每个非 R 分组应当包括——此类形式的报头扩展元素。

[0125] 图 19 示出发明性命名报头扩展的示例性数据字段格式,其中这些字段定义如下。

[0126] ID :4 比特

[0127] 针对此报头扩展元素协商的本地标识符,如例如 D. Singer 在 2006 年 12 月的 draft-ietf-avt-rtp-hdext-01(进行中)——“A general mechanism for RTP Header Extensions(RTP 报头扩展的通用机制)”中所定义的。

[0128] 长度(len) :4 比特

[0129] 此报头扩展元素的数据字节的长度减一,不计入报头字节(ID 和 len)。如果存在第二个字(取代范围),则这将具有值 6,如果不存在,则为 2。因此,其值必须或者为 2 或者为 6。

[0130] R :1 比特

[0131] 指示包含此报头扩展元素的分组是系列 SER 中具有 R 序号 RSEQ 的 R 分组的比特。如果此比特未被置位,则报头扩展元素替代地指示媒体流在系列 SER 中的最近 R 分组具有 R 序号 RSEQ。如果此比特未被置位,则取代范围应当不存在(即, len 字段应当为 2)并且在存在的情况下必须被忽略。

[0132] 保留,必须为零(MBZ) :3 比特

[0133] 保留比特。这些必须在传送时被设为零,并且在接收时被忽略。

[0134] 系列 ID(SER) :4 比特

[0135] R 分组系列的标识符由此报头扩展元素来描述。如果媒体编码器仅描述单个 R 分组系列,则此应当具有值 0。例如,使用图 13 中所示的可缩放视频画面编码结构,L 分组(空间增强基层,所有线程)可比方将 SER 设为零,而 S 分组(空间增强层,所有线程)可将 SER 设为 1。

[0136] R 分组序号(RSEQ) :16 比特

[0137] 指示此 R 分组在系列 SER 中的编号的无符号序号。在给定系列中每发送一个 R 分组,此值递增 $1 \pmod{2^{16}}$ 。各个序列的 RSEQ 值是独立的。

[0138] 取代范围的起始 (SUPERSEDE_START) :16 比特

[0139] 被此 R 分组取代的最早 R 分组 (包括此分组) 的 R 序号, 通过模 2^{16} 计算出。(由于此值使用模运算, 因此值 RSEQ+1 可用于 SUPERSEDE_START 以指示在取代范围末端之前的所有 R 分组已被取代。) 此字段是任选的, 并且仅在 len = 6 时存在。

[0140] 取代范围的末端 (SUPERSEDE_END) :16 比特

[0141] 被此 R 分组取代最后 R 分组 (包括此分组) 的 R 序号, 通过模 2^{16} 计算出。此值必须位于闭合范围 [SUPERSEDE_START..RSEQ] 模 2^{16} 中。此字段是任选的, 并且仅在 len = 6 时存在。

[0142] RTP 分组可包含多个 R 分组标记元素, 只要这些元素的每一个都具有一个不同的 SER 值。然而, RTP 分组不能包含一个以上的、R 比特被置位的这种报头扩展元素, 即, R 分组不可属于一个以上的系列。

[0143] 媒体流中使用 R 分组的所有 RTP 分组应当包括针对所有活跃系列的标记元素。

[0144] 当此报头扩展元素的第二个字存在时, 它指示此 R 分组取代某些先前接收到的 R 分组, 这意味着这些分组不再是重构流状态所必需的。第二个字必须仅在其 R 比特被置位的报头扩展元素中出现。

[0145] R 分组可仅取代系列中由元素的 SER 字段标识的 R 分组。R 分组不能取代其它系列中的分组。

[0146] 对于取代元素而言, 具有 SUPERSEDE_END = RSEQ 是有效的。这指示 R 分组取代其自身, 即, 此 R 分组直接变成与流状态无关。实际上, 这样做的最通常原因是结束一系列; 这可通过发送具有取代范围 (SUPERSEDE_START, SUPERSEDE_END) = (RSEQ+1, RSEQ) 的空分组 (例如, RTP No-op 分组, 参见 F. Andreassen 在 2005 年 5 月的 draft-ietf-avt-rtp-no-op-00 (进行中) 的“A No-op Payload Format for RTP (RTP 的 No-op 有效载荷格式)”), 以使得该系列不再包含任何未经取代的分组。

[0147] 在系列中发送的第一 R 分组应当在取代范围为 (SUPERSEDE_START, SUPERSEDE_END) = (RSEQ+1, RSEQ-I) 下发送, 以明确该范围内没有其它 R 分组存在。

[0148] R 分组可冗余地包括要被取代的分组范围内已被取代的分组。

[0149] R 分组的丢失由接收器来检测, 并且由该接收器用 RTCP 反馈消息向发送器进行指示。R 分组否定确认 (NACK) 消息是由 PT = RTPFB 且 FMT = 4 (作为示例) 标识的 RTCP 反馈消息 (例如, 参见 J. Ott 等人在 2006 年 7 月的 RFC4585 的“Extended RTP Profile for RTCP-based Feedback (RTP/AVPF) (基于 RTCP 反馈的扩展 RTP 概况 (RTP/AVPF))”)。根据本发明, 也可选择其它值。FCI 字段必须包含至少一个且可包含一个以上的 RNACK。

[0150] RNACK 分组用于指示一个或多个 R 分组的丢失。丢失分组经由分组序号、系列标识符和位屏蔽来标识。

[0151] RNACK 消息的结构和语义类似于 AVPF 通用 NACK 消息的那些。

[0152] 图 20 示出了 RNACK 反馈控制信息 (FCI) 的示例性语法, 其中各个字段定义如下:

[0153] R 分组序号 (RSEQ) :16 比特

[0154] RSEQ 字段指示接收器尚未接收到的 RSEQ 值。

[0155] 系列 ID (SER) :4 比特

[0156] 哪个 R 分组序列正被此报头扩展元素描述为丢失的标识符。

[0157] 后续丢失 R 分组的位屏蔽 (BLR) :12 比特

[0158] BLR 允许报告紧随由 RSEQ 指示的 RTP 分组之后的 12 个 R 分组中的任一个的丢失。指示 BLR 的最低有效位为比特 1, 且其最高有效位为比特 12, 如果接收器尚未接收到系列 $SER \pmod{2^{16}}$ 中的 R 分组号 (RSEQ+i), 则位屏蔽的比特 i 被设为比特 1, 并且指示此分组丢失; 否则比特 i 被设为 0。注意: 发送器不能假定接收器已接收到 R 分组, 因为其位屏蔽被设为 0。例如, 如果序列中对应于 RSEQ 的分组和后续的 R 分组已丢失, 则 BLR 的最低有效位可被设为 1。然而, 发送器不能仅因为比特 BLR 的比特 2 到 15 为 0 就推断已接收到分组 RSEQ+2 到 RSEQ+16; 所有发送器都知道接收器此时尚未将它们当作丢失来报告。

[0159] 当接收器检测到其尚未接收到非替代的 R 分组时, 它遵照 RTCP (参见 J. Ott 和 S. Wenger 在 2004 年 8 月的 draft-ietf-avt-rtcp-feedback-1 1 (进行中) - “Extended RTP Profile for RTCP-based Feedback (RTP/AVPF) (基于 RTCP 反馈的扩展 RTP 概况 (RTP/AVPF))”) 尽可能快地发送 RNACK 消息。在多点情况中, 这包括监听来自其它接收器的 RNACK 分组但不针对已被报告的丢失 R 分组发送 RNACK。

[0160] 当发送器接收到 RNACK 分组时, 其检查该分组是否已被取代。如果其尚未被取代, 则发送器重传被发送的 RNACK 所针对的分组 (使用例如 RTP 重传有效载荷, 参见 J. Rey 等人在 2006 年 7 月的 RFC 4588 的 “RTP Retransmission Payload Format (RTP 重传有效载荷格式)”)。如果分组已被取代, 则它重传最近的分组, 后者的 R 分组元素指示包括所请求分组的取代分组范围。

[0161] 取代重传先前已发送的分组, 发送器还可选择生成和发送取代 RNACK 中所请求分组的新的 R 分组。

[0162] 在某一时间段之后, 如果接收器尚未接收到或者被发送的 RNACK 所针对的 R 分组或者取代该分组的 R 分组的重传, 则它应当重传 RNACK 消息。接收器不能以超过 AVPF 所准许的频率发送 RNACK 消息。它在可能的情况下应当执行对到发送器的往返行程时间的估计, 并且不应当以超过每往返行程时间一次的频率发送 RNACK 消息。(如果接收器也充当 RTP 发送器, 且发送器正发送关于接收器的流的 RTCP 接收报告, 则可根据发送器报告的 LSR 和 DLSR 字段推断出往返行程时间。) 如果往返行程时间不可用, 则接收器不应当比设置时间段更频繁地发送 RNACK 消息。可能值为 100 毫秒, 尽管根据应当环境其它值也可能是合适的, 这对本领域技术人员是显而易见的。

[0163] 上述 RNACK 机制也可作为肯定确认 ‘RACK’ 消息来应用。在此情形中, 接收器向发送器指示哪些分组已被正确接收。根据本发明的原理, 在对分组报头的语义进行适当修改的情况下, 与 RNACK 消息相同的设计可被用作这些 ‘RACK’ 消息。RACK 消息可具有有效载荷特定解释, 例如, 它们可对应于切片或整个帧。在此类情形中, RACK 消息必须确认相关切片或帧所涉及的所有各个分组。

[0164] 在同一系统中组合使用 RACK 和 RNACK 也是可能的。

[0165] R 分组技术具有若干优点。首先, 它使得发送器能够将所生成 RTP 流中的分组子集指示为高优先级 (R) 分组。

[0166] 它还使得接收器能够确定其何时已丢失 R 分组、何时接收到流的任何分组, 并且无需考虑经编码的流的依存性结构如何。

[0167] 它还使得接收器能够向发送器指示其在何时已丢失 R 分组。这可通过否定确认已

被标识为丢失的任何分组来实现。任选地,接收到的 R 分组可由接收器进行肯定确认。

[0168] 另外,它使得不管已丢失多少其它非 R 分组,接收器都能够确定其自接收到的最后分组起未丢失任何 R 分组。

[0169] 又一个优点在于,它使得发送器能够——或者以编解码器知晓的方式(例如,H. 264 切片)或者以编解码器未知晓的方式(例如,RFC 3984 分段单元)——将帧划分成任何数目的 R 分组。

[0170] 另一个优点在于,它使得发送器能够声明 R 分组取代先前的 R 分组,即,先前的某些 R 分组不再是建立流状态所必需的。这包括能够声明给定分组之前的所有 R 分组已被取代,以及一定范围的 R 分组被取代两者。

[0171] 最后,另一个优点在于,它允许编码器以允许 R 分组状态从 FEC 流恢复的方式对其媒体流——或者对所有分组或者选择性地仅对于 R 分组——应用前向纠错(FEC)(参见例如 A. Li 在 2006 年 3 月的 draft-ietf-avt-ulp-17(进行中)——“RTP Payload Format for Generic Forward Error Correction(通用前向纠错的 RTP 有效载荷格式)”。

[0172] 允许接收器以最小延迟检测 LR 画面(包括 SR 画面)已丢失的第二示例性检测技术可应用于基于 H. 264Annex G(SVC) 草案标准的系统。在此情形中,H. 264Annex G(SVC)NAL 单元被用作进行传输的基础。目前 H. 264SVC 编解码器的设计不运送足以允许接收器确定是否已接收到流的全部最低时间层(R)——即 H. 264SVC 术语中的“关键画面”的信息。例如,参看图 21,帧 0 和帧 3 都是自身存储在长期参考缓冲器中的位置 0 的关键帧。帧 4 参考长期参考缓冲器中的位置 0。如果帧 3 完全丢失,则帧 4 不能被正确解码。然而,接收器在 H. 264Annex G(SVC) 草案标准下无法了解这些;接收器如同它能够使用帧 0 作为帧 4 的参考那样来进行操作,因而显示不正确的图像。

[0173] 用于使得解码器能够检测帧丢失的机制要将连续帧号或索引指派给关键图片,并且使非关键帧通过参考最近的关键帧的帧索引来指示该最近的关键帧。通过检查关键帧索引,流接收器可确定其是否确实接收到当前帧之前的流的全部关键画面。对于以 H. 264SVC 语法提供帧索引信息,存在多种可能性。以下参照图 23 和 24 描述了两种替换性实施例。

[0174] 图 22 示出 SVC NAL 报头扩展的结构,如在目前的 H. 264Annex G 草案(参见例如 2006 年 7 月在 Klagenfurt 由 J. Reichel、D. Santa Cruz 和 F. Ziliani 修改的联合视频组的文献 JVT-T201 的 T. Wiegand、G. Sullivan、J. Reichel、H. Schwarz、M. Wien 等人的“Joint Draft 7, Rev. 2: Scalable Video Coding(联合草案 7,修订版 2:可缩放视频编码)”,2006 年 7 月在 Klagenfurrt 的联合视频组的文档 JVT-T0 83(经修改)的“On High Level Syntax(关于高级语法)”,这两篇文献皆通过引用全部结合于此)中定义的。图 22 示出 3 字节报头的结构,以及各个字段的名称及其比特长度。dependency_id(依存性_ID)(D)、temporal_level(时间_级)(T)以及 quality_level(质量_级)(Q)字段分别指示空间/粗粒度质量、时间和细粒度质量维度中的点。换言之,它们指示 NAL 的有效载荷在由可缩放编码器提供的一组分辨率中的位置。注意,此方案中的基层通过 $D = Q = T = 0$ 来标识。

[0175] 此外,注意:当 $T = Q = 0$ 时,fragmented_flag(分段_标志)、last__fragment_flag(最后_片段_标志)、fragment_order(片段_次序)字段没有使用,因为它们仅对 FGS 经编码的数据相关($Q > 0$)。这些字段提供总共 4 比特。如果后缀 reserved_zero_two_bits(保留_零_两_比特)被包括在内,则总共为 6 比特。类似地,当 $T > 0$ 但 $Q =$

0 时,总共 4 比特的字段 fragmented_flag、last_fragment_flag、fragment_order 不被使用。如果添加后缀保留比特,则总共为 6 比特。通过注意条件 $T = Q = 0$ 对应于关键画面,而 $T > 0$ 且 $Q =$ 对应于非关键画面,可看到存在可用于引入帧编号的若干比特。可用的比特数目受非关键帧比特的限制。

[0176] 图 23 示出根据用于以 H. 264SVC 语法提供帧索引信息的示例性技术的经更改的 SVC NAL 扩展报头的结构。注意:报头的长度不变;然而,某些比特取决于 T 和 Q 字段的值而被不同地解释。在 $T = 0$ 且 $Q = 0$ 的情况下,F、LF、FO 和 R2 字段被解释为指定了指派给当前接入单元的关键画面帧索引的 FI 字段(key_picture_frame_idx(关键_画面_帧_索引))。在 $T > 0$ 且 $Q = 0$ 的情况下,F、LF、FO 和 R2 字段被解释为指定按照解码次序相对于当前接入单元的最近关键帧的 key_pic_frame_idx 的 LFI 字段(last_key_picture_frame_idx(最后_关键_画面_帧_索引))。

[0177] 对非关键帧使用 6 比特允许表示 64 个连续帧号。使用在每秒 30 帧下低至 4 个的关键画面时段,每个帧号周期为 8.4 秒。最小周期时间为 4.2 秒,对应于 2 个的关键画面时段。明显地,较长的时间提供更好的稳健性,因为参考画面与到达的画面之间帧号重复的机会得以减小。

[0178] 用于以 H. 264SVC 语法提供帧索引信息的技术的第二实施例通过使用保留比特之一作为扩展标志——当它被置位时发信号通知报头中存在附加比特或字节——来允许较长长度的帧索引。图 24 示出此实施例的示例性 SVC NAL 报头扩展结构,其中最初 3 字节报头的最后比特现在用作扩展标志(EF,extension_flag(扩展_标志))。当 EF 标志被置位时,报头中存在附加字节。取决于 T 字段(temporal_level(时间_级))的值,此附加字节被解释为 FI 或 LFI 字段。

[0179] 在两个实施例(3 字节或 4 字节 SVC NAL 报头扩展)中,FI 字段值递增并满足以下约束:

[0180] 如果当前画面是 IDR 画面,则 FI 值应当等于 0;以及

[0181] 否则,即,如果当前画面不是 IDR 画面,则令 PrevTL0FrameIdx 等于按照解码次序 T 等于 0 的最近画面的 FI 值。当前画面的 FI 的值应当等于: $(\text{PrevTL0FrameIdx} + 1) \% 256$ 。数字 256 表示 FI 字段的动态范围(最大值+1),并且应当针对不同的 FL 字段长度调节成值 2^L (以比特计的 FI 长度)。

[0182] 在 RTP 传输上下文和 H. 264SVC NAL 传输上下文中,根据本发明的用于指示 R 画面索引值并在非 R 画面中引用它的替换性机制对于本领域技术人员将是显而易见的。

[0183] 现在关注在视频通信系统中使用 LR 画面进行可靠传输和随机接入的替换性实施例(参看图 1)。在本发明的替换性实施例中,SVCS 单元可被配置成通过解码所有 LR 画面并将最近的 LR 画面保留在缓冲器中来便于 LR 画面的可靠传输。当接收器经历分组丢失时,它可从 SVCS 请求最近 LR 画面的副本。此画面现在可在 SVCS 处被编码成高质量内画面并被传送到接收器。此经编码的画面被称为内 LR 画面。尽管带宽开销可能很高,但是这仅影响特定 SVCS 与经历分组丢失的接收器之间的链路。内 LR 画面随后可被接收器用作应当被包含在其参考画面缓冲器中的实际参考画面的良好近似。为了改进这种近似,帧内编码较佳地应当具有极高质量。由 H. 264 支持的 SI/SP 技术还可用来提供对用于与比特流同步所需的参考帧的准确再现。在此情形中,SI 和 SP 画面两者必须由编码器生成。SI 画面由未

接收到 SP 画面的接收器使用。通过构造,使用 SI/SP 画面机制没有漂移。注意:尽管 SI/SP 机制当前仅被 H. 264AVC 支持,但是完全可对 SVC 类型(可缩放)编码应用相同的方法。SI 画面可被 SVCS 进行高速缓存,并且仅提供给新参与者。

[0184] 在最接近接收最终用户的 SVCS 不具有用于持续解码 LR 画面(或者在 LR 画面不存在的情况下的 L0 画面)的计算能力的情形中,该任务可被指派给在传输路径的较早级处的 SVCS。在极端情形中,指派(以及最终用户的相关联请求)可在发送器自身处进行。

[0185] 注意:常规解码画面与在使用内 LR 画面之后解码的那些画面之间的匹配并不一定精确(除非使用 SI/SP 帧)。然而,结合内宏块,在传输期间可能存在的视觉伪像被极大地减少的同时,视频通信系统可逐渐恢复同步。这种技术的益处在于其将差错处理完全局部化在经历分组丢失的链路上。结果,其他参与者在其视频信号的质量上完全不遭受惩罚。

[0186] 以上差错弹性技术也可用于提供对经编码视频信号的随机接入。例如,在图 1 中所示的视频会议示例中,当最终用户 3 加入最终用户 1 与 2 之间的现行视频会议时,最终用户 3 将开始从两个最终用户 1 和 2 处接收经编码的视频流。为了能够适当地解码这些流,最终用户 3 处的视频解码器必须与最终用户 1 和 2 处的解码器同步。这要求将最终用户 3 处的参考画面与最终用户 1 和 2 处的参考画面对齐。

[0187] 如前所述,由于使用内画面会对系统带宽产生较大影响——对于较大会议的媒体尤其如此——因此使用内画面没有吸引力。内宏块的替换性技术可用于在较小的时段内实现同步。

[0188] 在本发明的实施例中,基于服务器的内 LR 画面被直接用于随机接入。当参与者第一次加入会议时,它立即请求这种内 LR 画面,并随后进入差错恢复模式(就像丢失了分组那样)。通过同时使用内宏块,解码器将快速与编码器同步,但是在处于差错恢复模式期间,视觉伪像将被最小化。注意:发送编码器通过会话的信令机制了解新的用户何时加入通信会话,并且可由此发起使用内宏块或按需增加其频率。这例如通过图 6 中所示的 RRC 模块 630 来实现。因此,与内宏块相关联的编码效率的降低仅被限于新的用户加入会话的期间。

[0189] 由基于服务器的内画面所导致的计算复杂度并非十分高。假定每三个 L0 帧中有一个是 LR 帧,则仅需要解码 8% 的帧。将仅需要对很少部分的帧进行编码。实际上,仅集中在随机接入问题上(例如,参与者改变分辨率、或预订会话),则仅需要对 10% 或更少的帧进行编码。还可通过限制每个被处理的流生成 I 帧的频率来将编码限于任何期望值。例如,假定 8% 的帧被解码且 2% 被编码(对应于每 48 帧的随机进入),则与必须解码和编码整个流的译码 MCU/服务器的传统实现相比,总的复杂度小于 3.5% ($8\% \times 25\% + 2\% \times 75\% = 3.5\%$,假定编码复杂度是解码复杂度的 3 倍)。类似于传统译码 MCU,基于服务器的内 LR 画面技术可将内帧请求(例如,对于差错恢复、随机接入以及画面大小的改变)与传送器隔离开,并由此限制了这种帧内请求对其它参与者端点的影响。

[0190] 如前所述,如果服务器不具有用于基于服务器的内画面处理的 CPU 能力,或者如果服务器未在会议会话中预订所请求的流,则内画面请求可传播到下一 SVCS(即,更靠近特定视频流的传送器)。如果系统中没有服务器具有合适的内画面处理功能,则内画面请求甚至可传播到发送器/传送器本身。

[0191] 以基于服务器的 LR 画面为基础的视频会议保留基于可缩放视频和联播的视频会议的优点。这些优点包括因无需抖动缓冲器而实现的最小服务器延迟(即使具有 LR 画面)、

改进的差错弹性、以及比传统 MCU 的复杂度小一个量级的复杂度。

[0192] 以上所述的 LR 和基于服务器的内 LR 画面技术也可直接应用于空间可缩放性和 SNR 或质量可缩放性。LR 画面和基于服务器的内 LR 画面的概念可应用于空间或质量层的任一个。例如,图 13 示出具有三个时间层和两个空间或质量层的示例性画面编码结构 1300。除差错弹性和随机接入之外,空间可缩放性和 SNR 可缩放性需要考虑层切换。例如,当正在以 CIF 分辨率观看参与者的最终用户决定切换至 QCIF 或反之,可能导致层切换的需要。对于差错弹性和随机接入而言,层切换是类似的,但并不相同。不同分辨率(空间或质量)之间的相关性可被有益地用于创建有效的层切换机制。

[0193] 应当注意:如当前在 H. 264SVC 标准化努力中所研究的,在空间可缩放性方面,在单环路中操作接收器是可能的。如果在高分辨率下执行的预测不使用要求以低分辨率应用运动补偿的任何低分辨率信息,则单环路操作可能的。换言之,预测可使用内宏块、运动矢量、预测模式、经解码的预测差错值,但不使用低分辨率的实际经解码的像素。虽然从计算的观点来看,单环路解码使得可缩放解码器较为简单,但是其使得从低到高或从高到低的分辨率切换成为一个复杂的问题。单环路解码的替换方案是其中以两个或多个接收分辨率解码接收到的信号的多环路解码。多环路解码显著地增大了解码复杂度,因为它类似于同时操作多个解码器(每个经解码的分辨率一个)。

[0194] 在许多视频会议应用中,分辨率之间的频率切换是必需的。例如,考虑其中有 5 个人参加并且说话者在较大窗口中呈现而其他参与者则在较小的窗口呈现的中型会议中的动态布局。通过使用两个分辨率下的 LR 画面,解码器可维护解码环路,后者使得两个分辨率下的参考画面缓冲器的内容相近似——准确而言是在 LR 时间点处。当从一个分辨率切换到另一个时,LR 画面可用作解码成另一分辨率的起始点。假定这些 LR 画面是每 4 个 L0 画面中的一个,则转换在 0.4 秒内发生,同时计算开销小于单环路解码的 10% (准确地为 1/12)。当解码器仅“预订”LR 画面时,SVCS 可向解码器传送分割成较小片的 LR 帧。这些较小的片可分散在 LR 周期上的所有帧之间,以在给定链路上维护平滑的比特率。或者,SVCS 可跨时间地分散于来自多个流的不同 LR 帧上。

[0195] 两个分辨率下的内宏块也可用来便于层切换。假定端点希望从低分辨率转到高分辨率。其将持续解码低分辨率信号并以高分辨率显示它(上采样),同时将在“差错恢复”模式中开始解码高分辨率信号但不显示它。当接收器确信其高分辨率解码环路与编码器充分同步时,它可将显示切换到经解码的高分辨率画面且任选地停止解码低分辨率环路。相反,当从高分辨率转到低分辨率时,接收器可使用高分辨率画面作为低分辨率编码环路的良好参考画面,并以低分辨率在常规差错恢复模式(并且进行显示)继续。通过这种方法,端点将避免必须持续接收高分辨率数据。

[0196] 使用内宏块的一个可能的缺点是其在切换或进入时间与施加于流的当前接收器上开销的量之间造就一种权衡。切换越快速,则对于当前接收器而言,将有越多的开销。以上 [0066] 所述方法或在服务器上生成内帧的是用于有效地避开这种权衡的一种可能的方法,但它在服务器上的确需要附加媒体处理。本发明之下的其它方法如下:

[0197] 方法(a),其中内宏块被包括在 LR/SR 帧中(以使得低速率切换或进入将可能具有极低的开销),同时 SVCS 高速缓存 LR/SR 帧。当新的接收器进入流时,SVCS 仅向其提供这些帧以使得接收器可比实时更快地将它们解码(通常为 1 : 8)并且缩短进入时间。

[0198] 方法 (b), 其中作为方法 (a) 的补充, SVCS 移除高速缓存的 LR/SR 画面中存在的、对于接收器而言因后继 I 宏块而变得冗余的帧间宏块。这在 LR/SR 帧由编码器以切片形式来配备的情况下可更容易实现, 以使此操作将仅需要删去此类冗余的中间切片。在以下描述中, 这两种方法 (a) 和 (b) 都称为“内宏块快速转发”。

[0199] 图 25 示出内宏块快速转发的操作。附图示出在三个连续时刻 $t = i$ 到 $i+2$ 处各自被编码成三个单独切片的 LR 画面 2500 (LR i 到 $i+2$)。在各个时刻, 三个切片之一被编码为帧内 (A)。当进行组合时, 这三个画面一起向解码器提供每个宏块的至少一个内版本。为了用于创建参考画面, 除内切片 A 之外, 解码器还必须接收图中所示的阴影切片 (B)。这些阴影切片是使用宏块数据从同一位置处的先前切片预测的。在实现快速转发内恢复时, 服务器需要高速缓存提供这种内切片编码的任何连续 LR 画面。一旦有来自接收器的请求, 服务器就仅需传送内切片以及图 25 中所指示的阴影切片 B。图 25 中所示的非阴影切片 (C) 无需被传送。

[0200] 注意: 并非所有 LR 画面都必须提供此类内切片编码。例如, 假定 LR 画面的传输模式诸如为 LRI LRI LRI LR LR LR, 其中 ‘I’ 上标指示存在内切片, 则服务器不仅必须高速缓存内切片及其在 LRI 画面中的依存切片, 还必须高速缓存后续 LR 画面中的依存切片。

[0201] 此技术可被扩展到高分辨率同步。例如, 在如上所述同步到基层之后, 接收器可在开始时显示经上采样的基层信息。同时, 它可在增强 (S) 层 (通过 SRI 画面) 中发起相同的过程。注意: 这些画面在 SVCS 处无需被高速缓存, 而是接收器一被添加到会话就指示编码器开始生成它们。由于恢复点是根据高速缓存的基层来确定的, 因此这不会增加同步时间。这将仅影响由接收器看到的初始视频质量。图 26 示出使用由三个切片构成的 LR 画面的示例的这种高分辨率同步过程。

[0202] 参看图 26, SVCS 高速缓存 LRI 画面的整个周期 2610 以及后续 LR 画面 (2610')。当客户端加入 (例如, 在 A 点) 时, SVCS 尽可能快地向接收器传送所有高速缓存的 LR 画面。一解码所有这些画面, 接收器现在就处于同步 (例如, 在 B 点) 并可开始 LR 流的常规解码。它还可显示经解码的、上采样到高分辨率的画面。同时, 在 A 点处, 通知编码器来生成 SRI 画面 2620。这些画面开始在 C 点到达接收器。一接收到整个周期的 SRI 画面 (例如, 在 D 点), 接收器就从显示上采样的基层画面切换到显示经解码的全分辨率画面。尽管 LR 恢复是通过比实时更快速地解码来实现的, 但是 SR 恢复是通过实时解码来实现的。在此示例中, 接收器能够在 B 点产生显示输出 (虽然以较低的质量)。应当理解, 根据本发明的原理, 可对 SR 恢复使用不同的定时或速率。例如, 在带宽允许的情况下, SR 恢复可连同 LR 恢复一起被快速转发。此外, 内宏块可一直存在于 SR 画面中, 而非仅在可适于较大会议或与频繁分辨率改变相关联的会议时按需发起。最后, 如果已在接收器中解码 LR 画面, 则仅快速转发 SR 级所需的信息可被提供给解码器。

[0203] 可在正确时间指示解码器开始使用如在 H. 264 规范中定义的恢复点 SEI 消息来显示画面。参数 `recovery_frame_cnt` (恢复_帧_计数) 和 `exact_match_flag` (准确_匹配_标志) 可用于指示完成恢复的帧号, 以及是否与编码器相匹配。

[0204] 在内宏块被减小以使得刷新需要较大数目的 LR/SR 帧的情况中, 快速转发方法将要求发送大量 LR/SR 帧, 从而导致比质量相当的一个 I 帧更大的总带宽利用。此外, 在许多视频切换技术 (例如, 语音激活切换) 中, 许多接收器将需要切换到较低或较高分辨率下的

相同画面。在这些情形中,方法(a)可增加执行R帧的解码以及向切换或进入接收器发送常规内帧的服务器(方法(c))。这种经增加的方法(a)提供了在当前预订流的端点处维持较小开销的同时降低与基于服务器的内帧方法相关联的计算开销与减小进行切换时的带宽开销以及切换时间本身之间的良好权衡。

[0205] 在另一方法(d)中,取决于系统中的约束,快速方法可仅用于缩短对同步的等待时间而非完全消除这种等待时间。例如,如果系统中的进入端点是带宽有限的,则可能并不较快地向它发送提前同步所需的所有LR/SR画面。相反,为了更快速地同步,可向进入端点发送或提供较小的储备(backlog)。

[0206] 以上所描述的各种技术和方法在实践时可被组合或更改。例如,快速转发方法可仅应用于LR级(最低空间/质量分辨率)帧,这些帧随后可被解码并上采样以用作后继增强层帧的参考。实际上,可在之后用于传送增强层帧的带宽和用于将它们解码的CPU可在同步期间使用以更快速地传送和解码LR帧。

[0207] 在编码器并非带宽有限的情形中,编码器在周期性基础上生成I帧或切片。编码器可操作成使得就在I切片或画面之前的帧将作为紧接其后的帧的参考。SVCS可高速缓存这种帧内信息,并且阻止将其转发到当前接收此流的端点,由此避免任何开销。对于新的参与者,SVCS将提供此I画面以及任何后续的R帧,以使得新的参与者可赶上实时。如果从编码器到SVCS可用其它带宽,则有可能传送所有LR帧,并且添加I切片或画面作为附加冗余画面。冗余画面将在SVCS处被高速缓存,而常规LR画面被转发到接收者。高速缓存的I切片或画面可如前所述地用于帮助接收器同步到特定流,同时未对当前参与者造成任何带宽开销。

[0208] 以上所述的方法也可用于在一个到多个流送应用的、要求低延迟和某些交互性度量的上下文中,并在本发明下被要求权利。

[0209] 前述切换技术的潜在的缺点是它在从低分辨率切换到高分辨率时要求双解码环路。一种替换性切换技术仅要求单环路解码结构。在要实现从低分辨率到高分辨率的切换时,解码器切换到由以较低分辨率解码的参考画面初始化的高分辨率解码环路。从此时开始,解码并显示高分辨率画面,并且最终经由内宏块与传送器同步。

[0210] 当进行单环路解码时,视频编码器仅以参与者所请求的大小编码画面是可能的。这在以多个分辨率进行编码时是优点,例如,极低分辨率的编码可用于差错隐藏目的。

[0211] 此外,根据本发明,空间和/或SNR可缩放性可用于差错隐藏。例如,假定单环路CIF/QCIF编码。如果在高分辨率上发生差错,对于差错隐藏,解码器可上采样QCIF分辨率的内宏块,并使用可用运动矢量、模式和在CIF层编码的预测差错。如果双环路解码是可能的,或在工作中可基于差错的检测进行,则解码器还可使用上采样的经解码的QCIF图像作为将来的帧的参考或用于显示目的。使用在CIF层处使用的宏块和/或消除对被破坏画面的依存性的时间结构,视频通信系统将快速地从损坏中恢复。

[0212] 图13中所示的相同LR方案也可用于稳健性目的。当在增强层发生分组丢失时,低分辨率LR帧可提供恢复点。经解码的帧可用作高分辨率参考画面缓冲器的估计,或者可替代高分辨率帧显示直至高分辨率解码环路恢复。当与内宏块组合时,这可以是有效的差错弹性技术。此外,可在计算负载与切换速度之间进行权衡。例如,通过更多地解码低分辨率层(例如,所有L0画面),存在用于恢复高分辨率层的更多和更好的数据。对增强层信号

使用 LR 帧也是可能的。

[0213] 如在图 13 的画面编码结构中,当存在一个以上的空间或质量分辨率时,快速转发恢复和隐藏可同时发生。例如,当解码器不接收所要求的 SR 画面时,它可使用隐藏来解码后续 SR 和 S0-S2 画面。当缺少的 SR 画面通过重发变得可用时,解码器可随后重新解码自 SR 丢失起接收到的插入 SR 画面,并且可能已被隐藏地显示,以使其为后续 SR 画面产生正确的参考画面。注意:如果 SR 重传足够快,且重传的 SR 早于在丢失的 SR 之后的 SR 画面到达,则在允许其对必须解码并接着显示的画面生成正确的参考画面的情况下,解码器也可解码已被隐藏地显示的 S0 和 S1 画面的任一个或全部。如果画面以切片形式来构造,则根据本发明的原理,本文所述的隐藏和快速转发恢复技术皆可独立地应用于切片中的每一个。

[0214] 在空间可缩放性方面,带宽效率在跨时间与跨空间分辨率之间存在有益的相互作用。例如,在单环解码中,基层处的内宏块对改进较高空间层的编码效率是有益的。此外,试验已表明,编码的质量越高(即,QP 值越小),则运动估计的有效性越低。LR 帧的典型大小是 L0 帧的两倍,但是大小差异随着质量的提高而减小。因此,对于越高分辨率和/或画面质量,可令所有 L0 帧来将 LR 帧用作参考而没有显著的编码效率惩罚。由于保证 LR 帧将被可靠接收到,因此使用它们提供了更大程度的差错弹性解决方案而在带宽中没有过度的惩罚。

[0215] 针对视频通信系统在 LR 画面与内宏块的使用之间的选择取决于所遇到的特定网络状况、参与者的数目和若干其它因素。为了使视频通信系统的效率最佳化,联合考虑这些技术的每一种在解码过程中的效果是很重要的。理想地,如果编码器完全知晓解码器的状态——包括丢失分组,则最大化将来的帧的质量是可能的。这可在编码器与所有解码器之间维护紧密反馈环路的情况下实现。这通过 RRC 模块 630(图 6)来表示。可在所有级下提供反馈,例如从各个宏块、切片、画面或整个层提供。

[0216] RRC 模块 630 可被配置成根据模式选择、运动矢量选择等连同参考画面选择(常规或 LR 参考)以及强制内宏块编码过程的统计一起来协调编码器的决定。此外,RRC 模块 630 可被配置成维护与帧中可用于运动补偿预测的安全对非安全部分有关的状态信息。这些决定可以与编码器联合的方式作出。可为编码器所用的反馈越是详细,可作出的决定越好。

[0217] 如果编码器知道解码器处所采用的差错隐藏策略,则假定使用反馈编码器将能够计算解码器的准确状态,即使存在分组差错也如此。如果实际分组丢失信息不可用,则编码器仍可使用统计技术来估计分组丢失的概率性影响以及在执行速率-失真最优化时计入分组丢失。例如,较高的丢失率将导致较大百分比的帧内经编码的宏块。

[0218] 类似地,可将诸如新用户加入会议的操作引入编码器的最优化过程中。在此情形中,向新用户随机接入点的需要转化为编码器处极高百分比的内宏块。使用可缩放编码,在层切换中观测到相同的现象。

[0219] 出于系统效率起见,由 RRC 630 管理的反馈信息无需直接到达特定编码器。作为替换方案,中间 SVCS 可过滤反馈消息并向编码器呈递合并结果。系统中的中间节点可对反馈消息采取动作。例如,考虑 NACK 消息的情形。NACK 可从最近的中间节点(SVCS)触发重传。NACK 可一直传播至源,在那里它被用于跟踪解码器的状态。此信息可例如引发编码器将参考画面索引切换成指向 LR 画面(或知道其自身已被适当地接收且当前在编码器的缓

冲器中可用的画面)。NACK/ACK 消息传送的概念直接导致用于运动补偿预测是安全或非安全的画面或画面区的概念,后者又自然地导致 LR 画面的概念。具有固定周期性结构的 LR 帧允许用 NACK 分发,并且类似地紧密 NACK/ACK 反馈的使用实现 LR 画面的完全动态选择。

[0220] 作为 NACK/ACK 反馈消息所隐含的“推入 (push)”方法的替换方案的是“拉出 (pull)”架构。在拉出架构中,LR 画面无需被确认,相反地被缓冲在每个中间 SVCS 处,并且当端点或其它下游服务器确定它们已错失 LR 分组时基于请求(例如,类似于对新的 I 帧的请求)来重传。

[0221] 在这种拉出架构的变体中,所有 L0 分组(或已用于给定应用的可缩放编码方案的其它最低时间级)被缓冲在每个中间 SVCS 处并基于请求来重传。此变体可使得端点在等待错失的 L0 分组的同时如果不具有解码已到达的所有 L0 分组的 CPU 带宽,则总是处于力图迎头赶上的模式中。然而,拉出架构的这种变体的优点在于,不存在仅出于差错弹性的目的而引入的略大 LR 帧的附加开销。

[0222] 可靠性分组(不管 LR 还是 L0)之间的间隔应当根据最弱参与者(端点或另一服务器)的 CPU 和带宽约束来确定。过于频繁到达的可靠性分组会在恢复期间超出端点的容量。视频通信系统可被配置成向发送器发信号通知参与者的恢复能力,以使得可靠性分组之间的间隔可尽可能地与最弱参与者所能处理的一样小,但是不比它小。

[0223] 构成编码器决定作出过程的一部分的是宏块编码类型 (mb_type) 的选择。此决定计入与给定以上考虑的情况下的帧间编码相关联的失真和速率。与(受约束的)帧内编码相关联的失真和速率在无需考虑多个解码器的情况下被计算出。取决于对成本函数的选择,必须对每个空间分辨率和 mb_type 计算一个或多个失真值。

[0224] 当对解码器状态或成本函数的建模不准确时,内宏块类型可替代或附加地选择以符合随机模式。适当数量的内宏块可根据对信道差错概率以及隐藏能量的量的估计来确定。

[0225] 虽然已描述了被视为是本发明的优选实施例的那些实施例,但是本领域技术人员应当认识到,可作出其它或进一步的改变和更改而不背离本发明的精神,并且其旨在要求保护落在本发明的真实范围内的所有这些改变和更改。

[0226] 应当理解,本发明的系统和方法可用任何硬件和软件的组合来实现。用于实现和操作前述系统和方法的软件(即,指令)可被设置在计算机可读介质上,这些计算机可读介质可包括但不限于:固件、存储器、存储设备、微控制器、微处理器、集成电路、ASICs、可在线下载的媒体以及其它可用的介质。

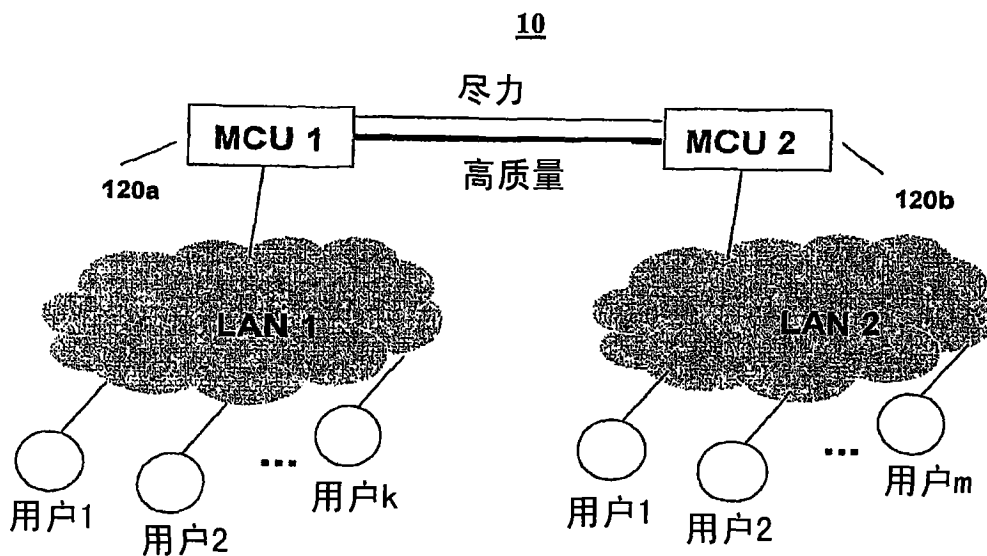


图 1

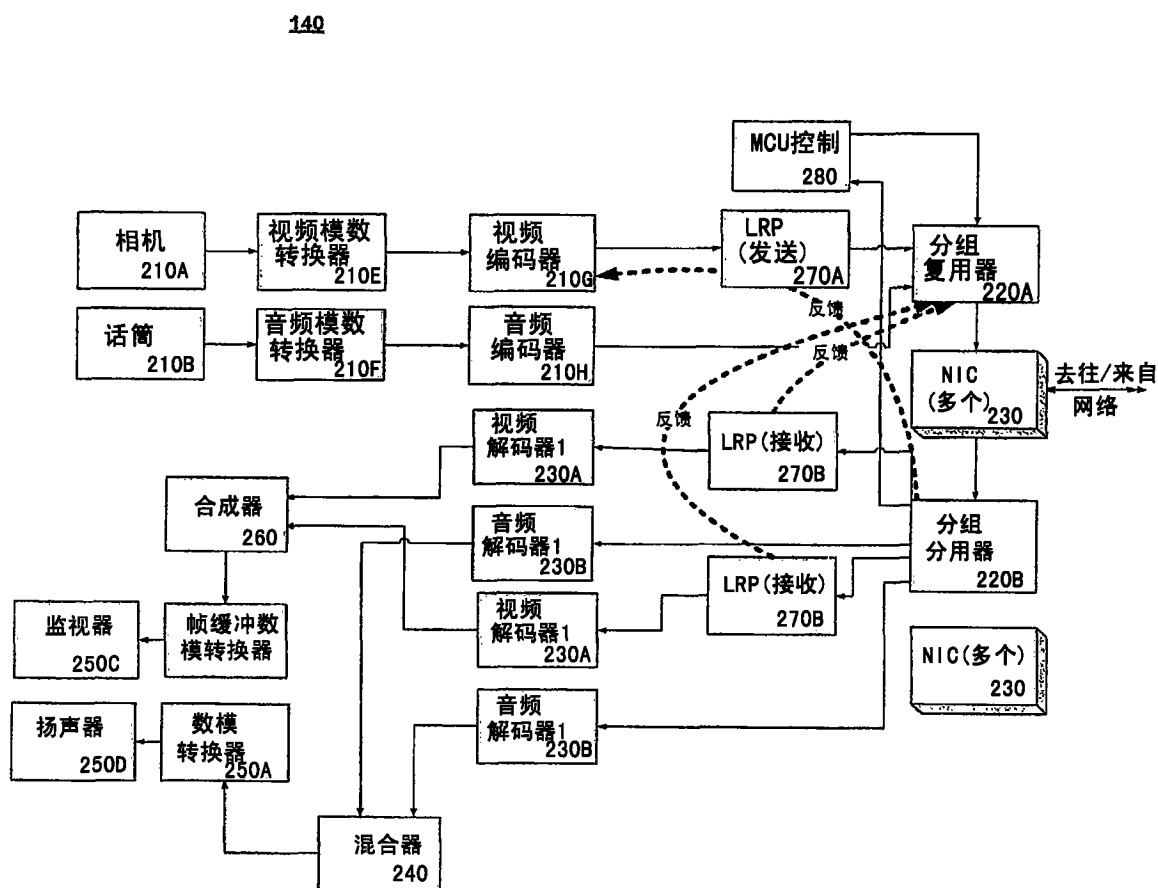
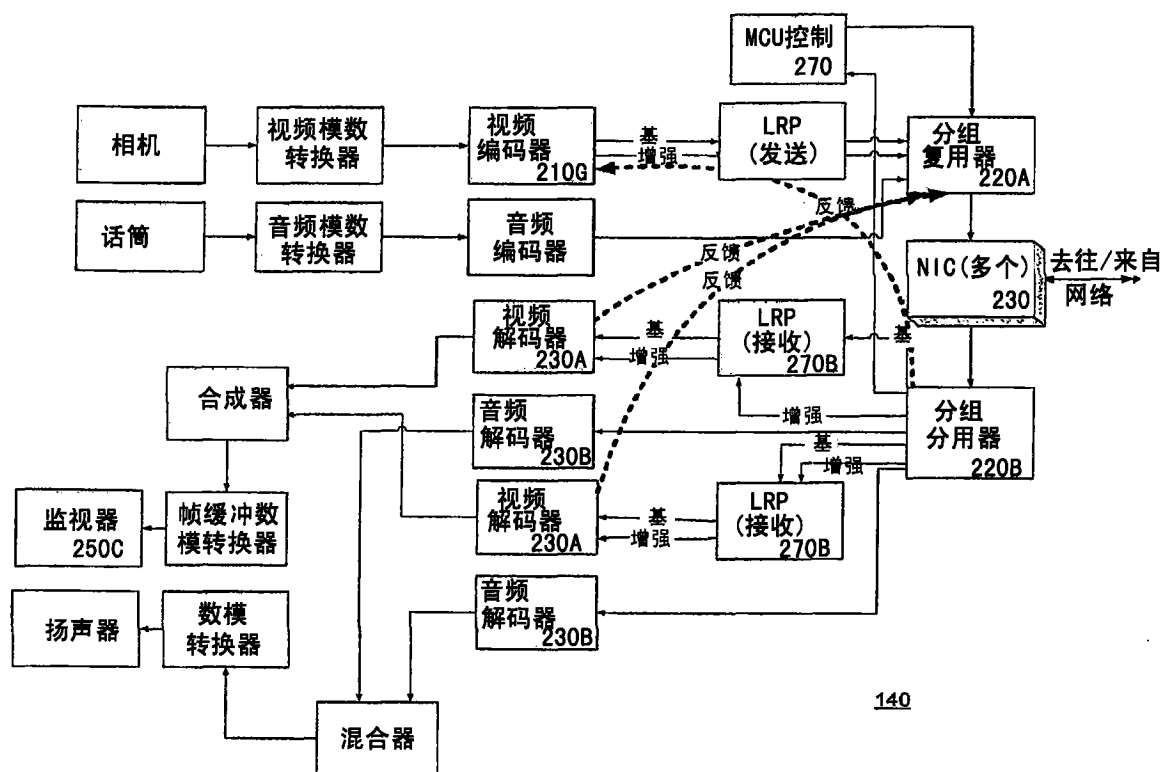
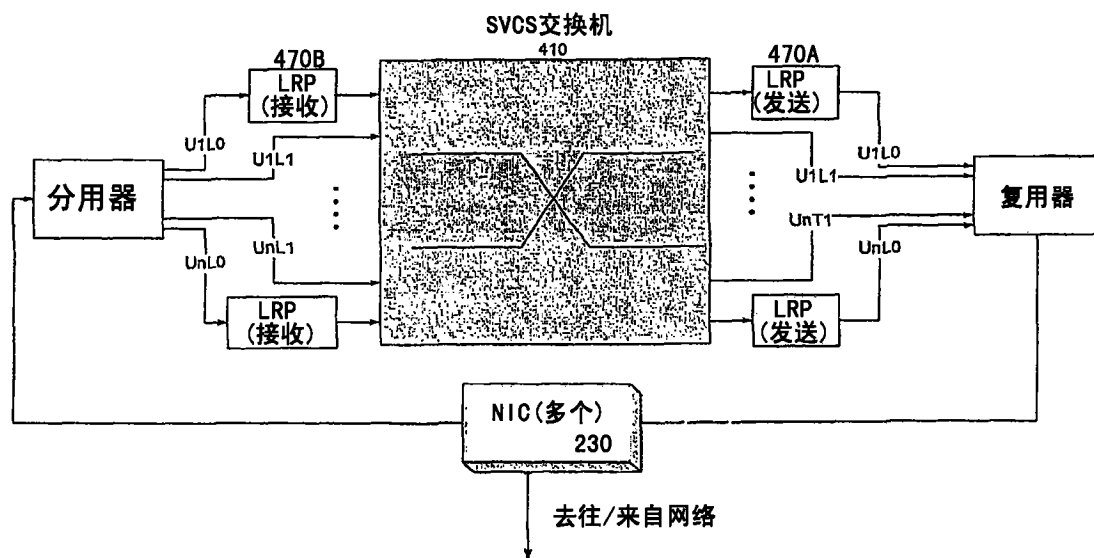


图 2



最终用户终端 (可缩放或联播编码)

图 3



SVCS

图 4

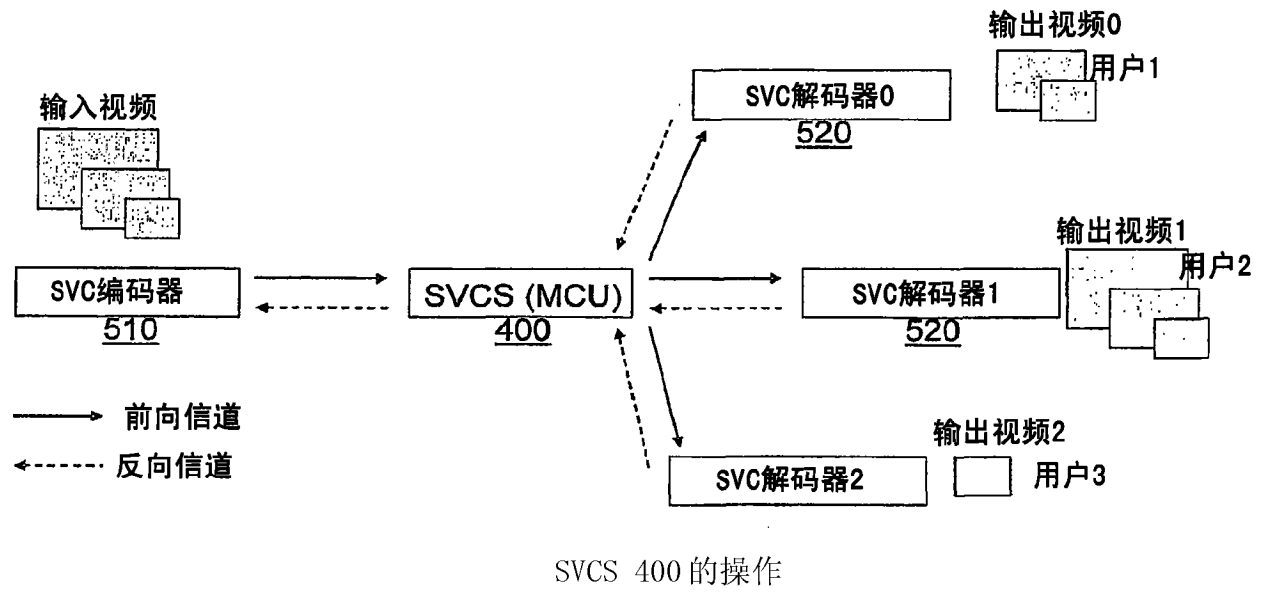


图 5

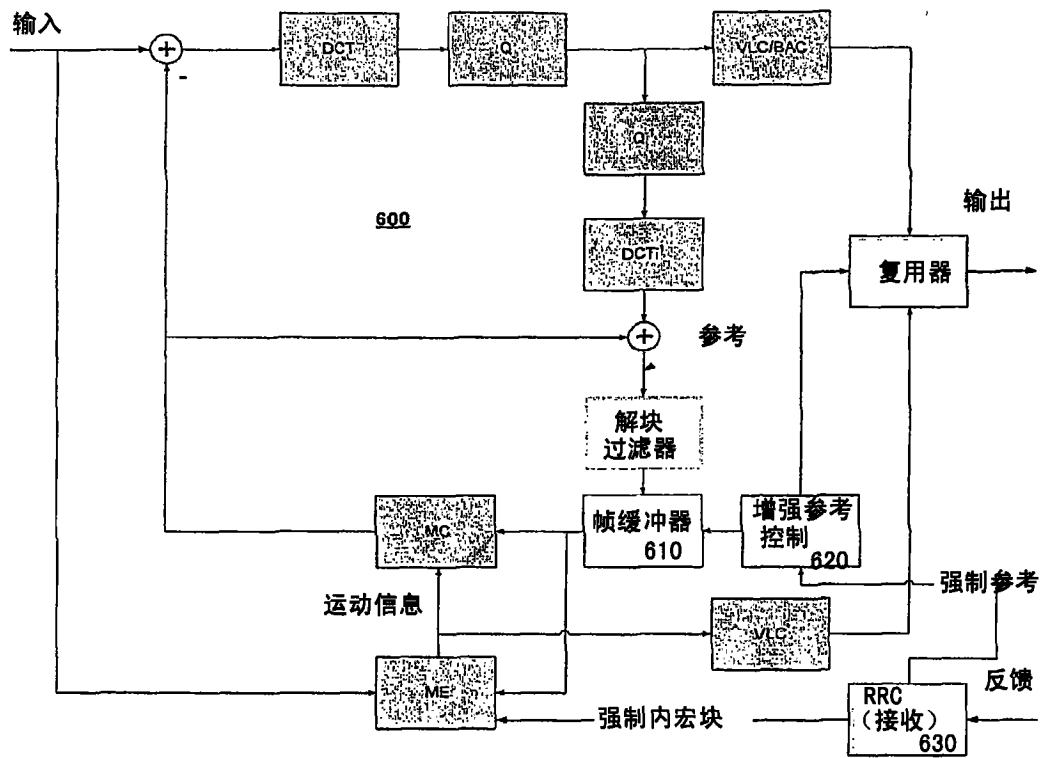


图 6

视频编码器（基和时间增强层）

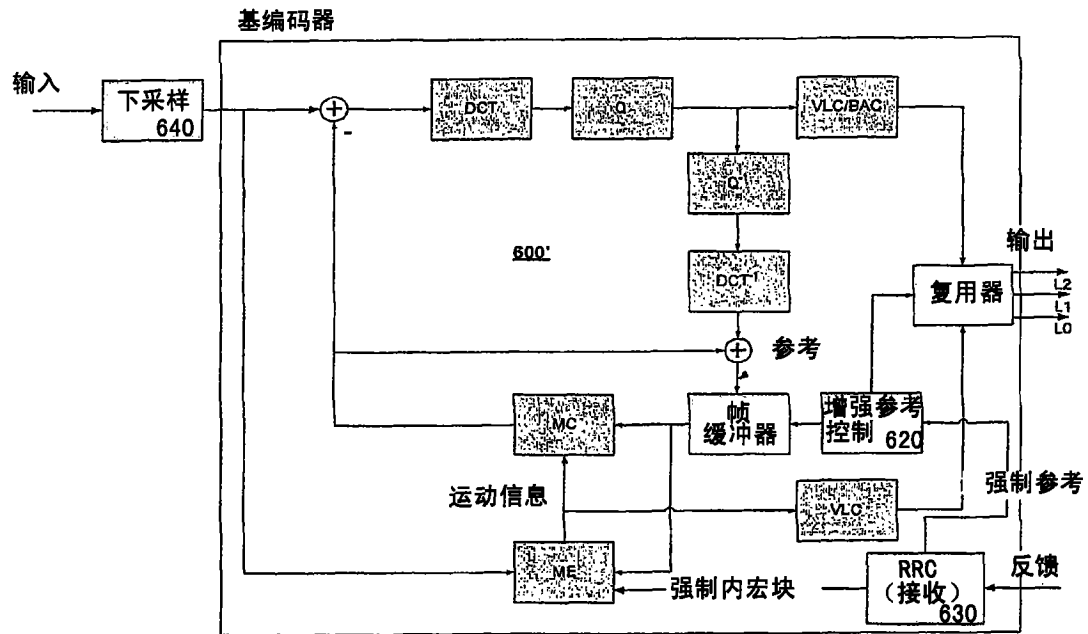


图 7

空间可缩放性增强层编解码器

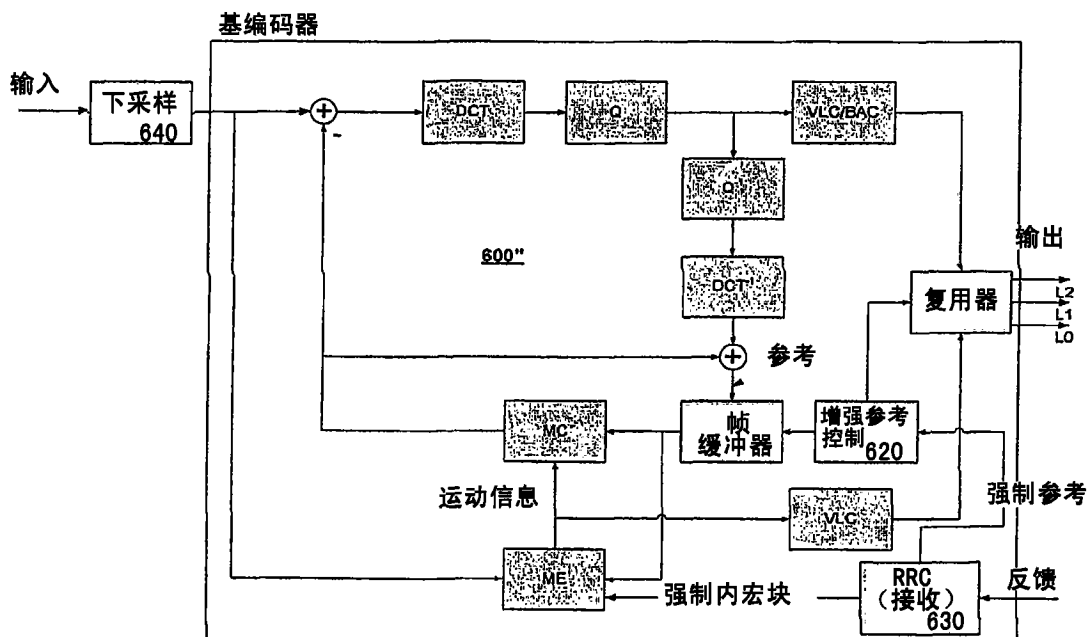
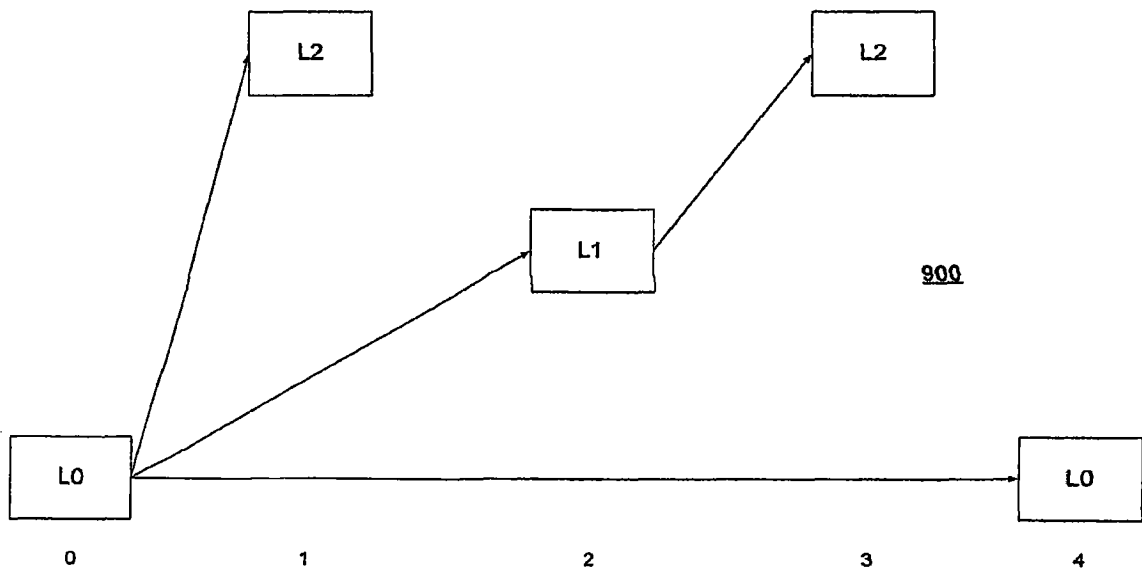
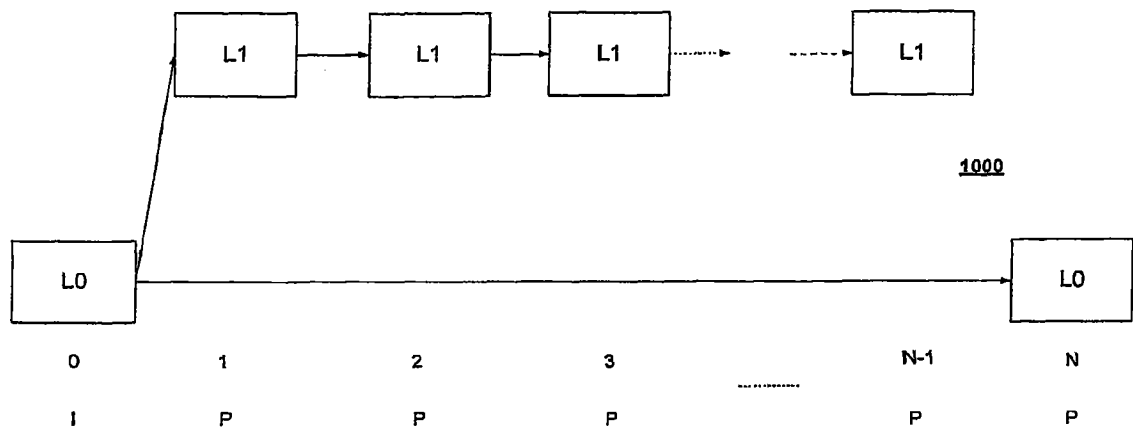


图 8



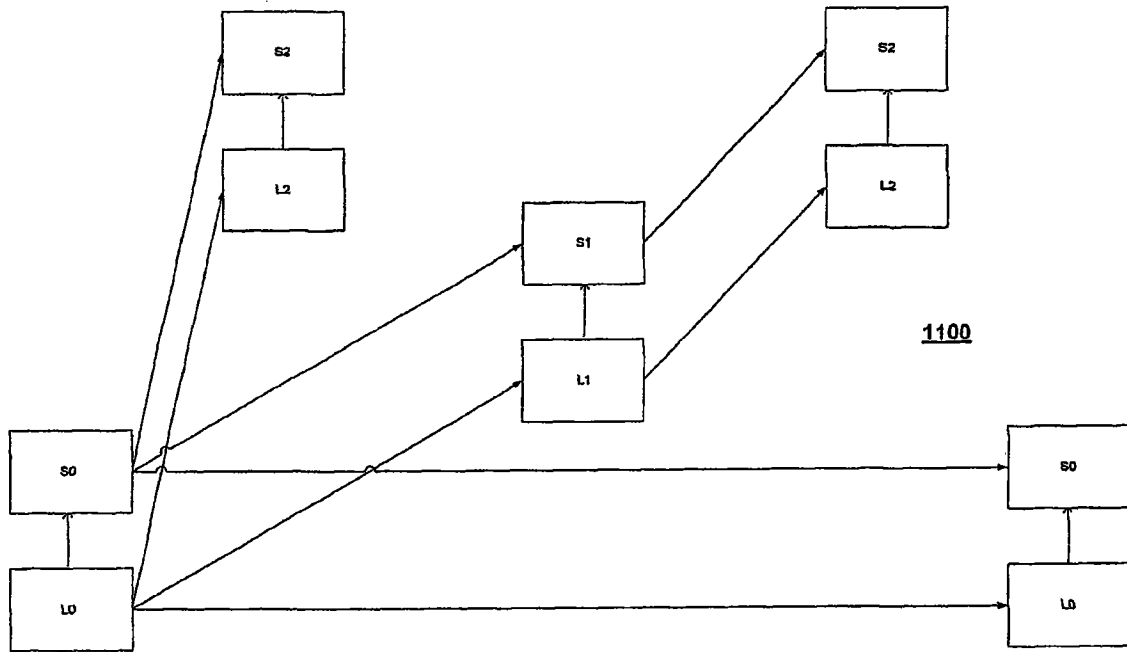
画面编码结构

图 9



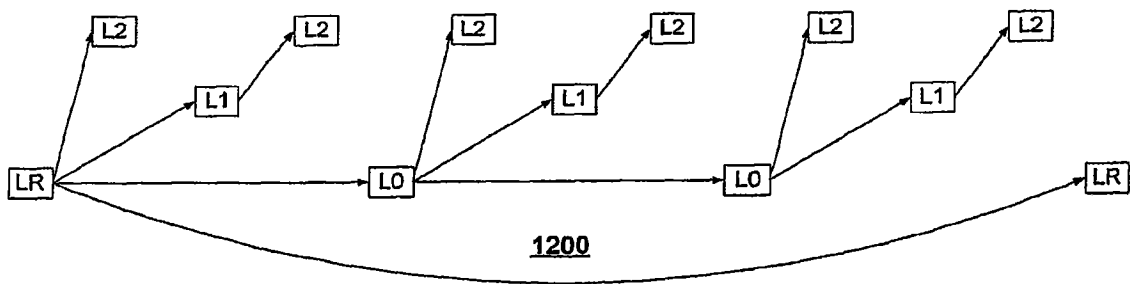
替换性画面编码结构示例

图 10



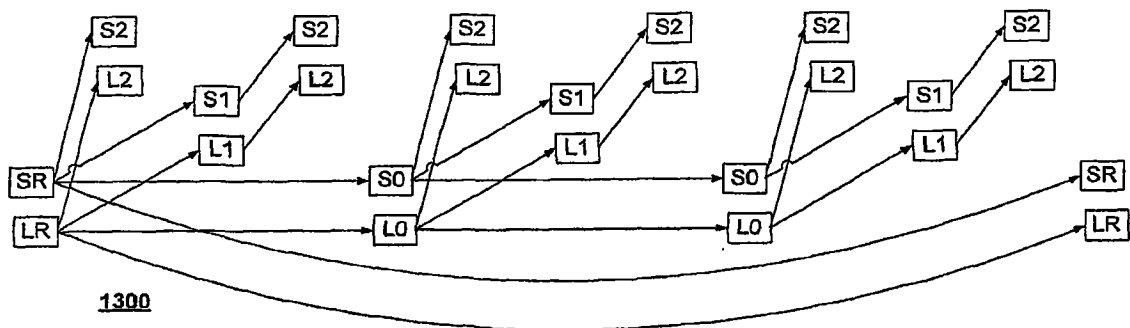
用于空间可缩放性的画面编码结构

图 11



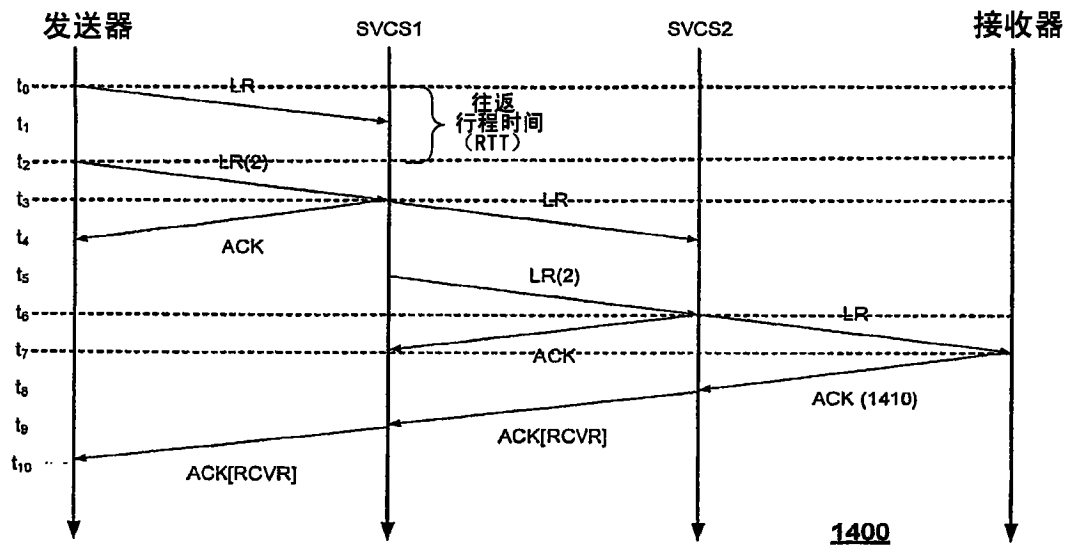
具有LR画面的画面编码结构

图 12



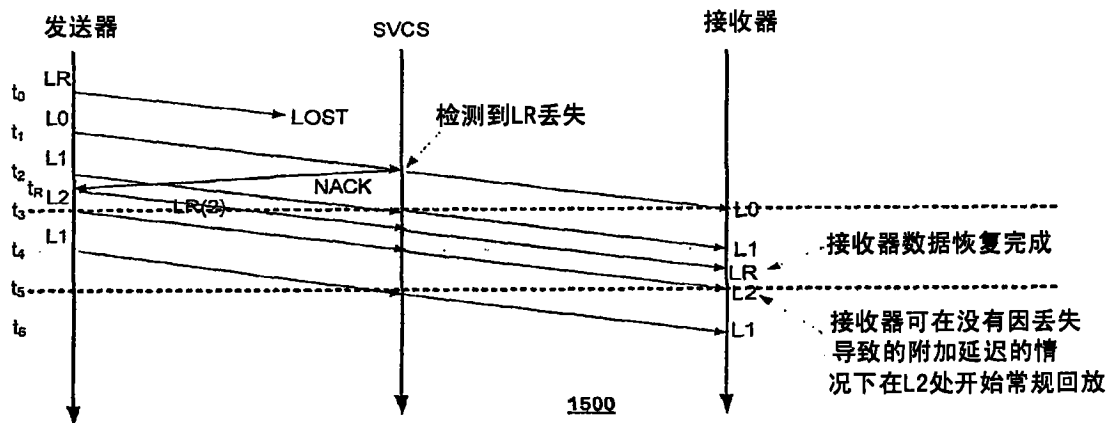
用于空间可缩放性的具有SR画面的画面编码结构

图 13



使用肯定确认的LR保护 (LRP) 协议

图 14



使用否定的LR保护 (LRP) 协议

图 15

使用RTP的具有LRP的传送终端

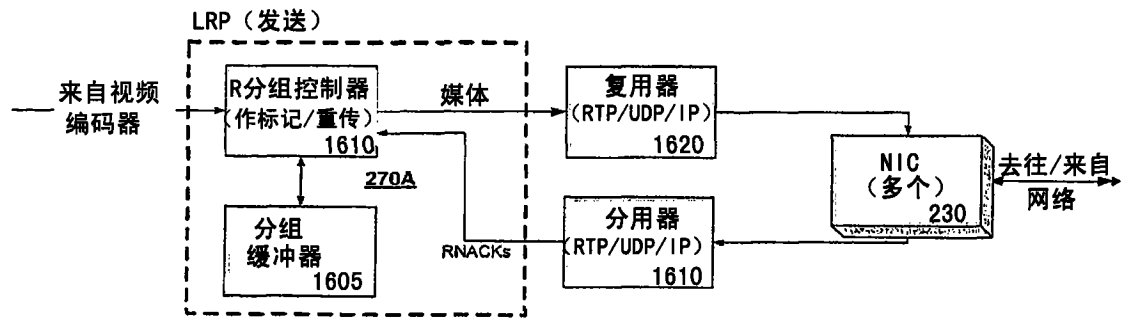
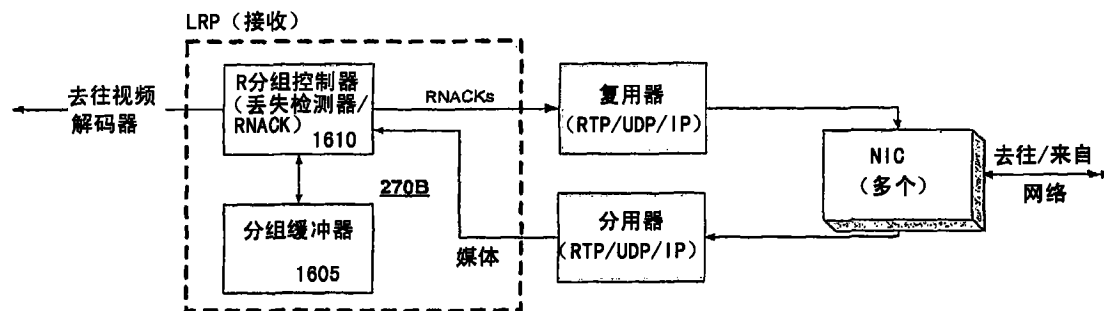
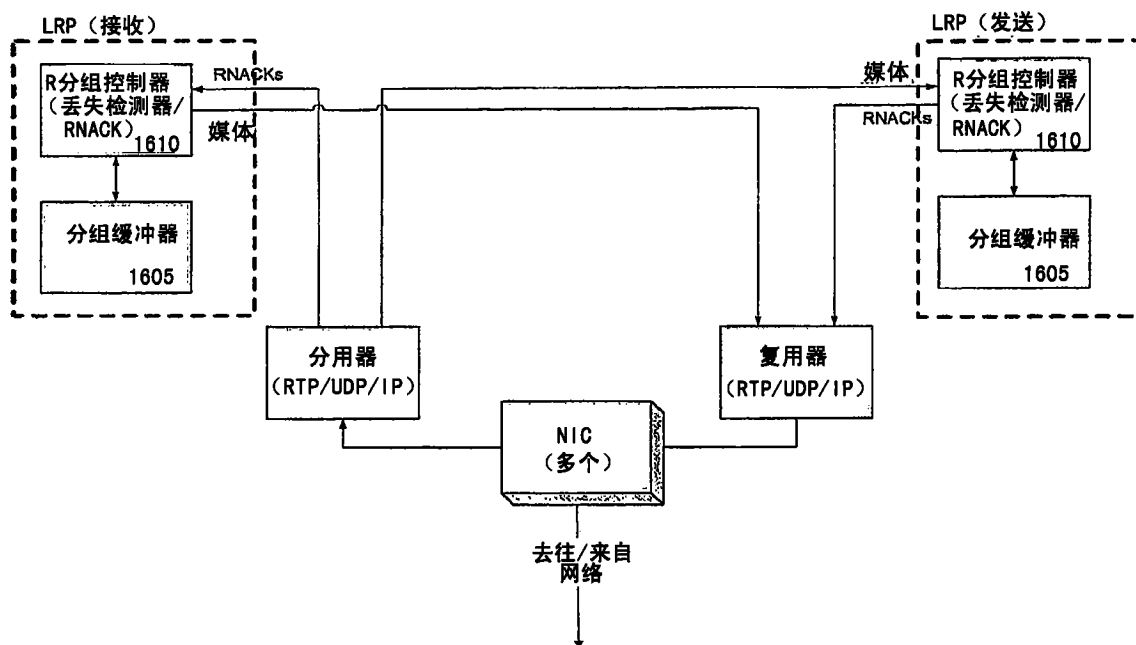


图 16



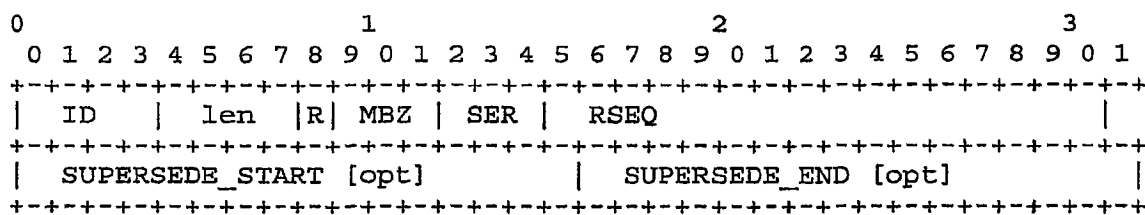
使用RTP的具有LRP的接收终端

图 17



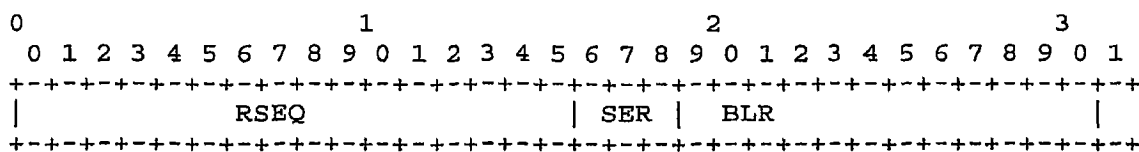
使用RTP的具有LRP的服务器

图 18



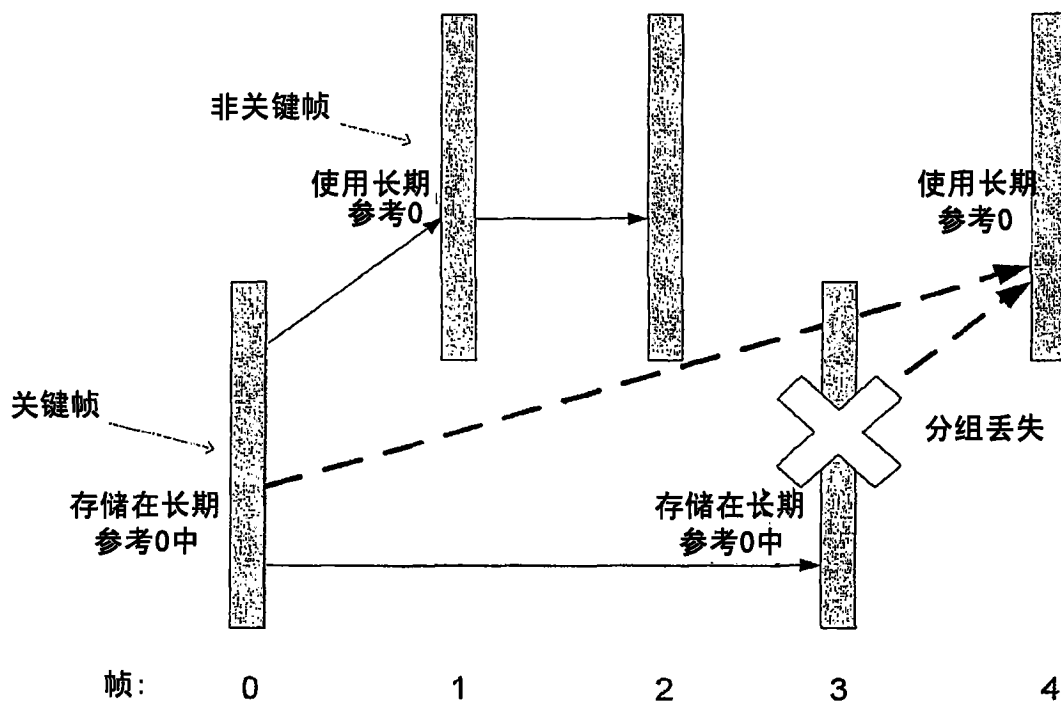
用于RTP分组的RTP报头扩展

图 19



RNACK反馈控制信息 (FCI) 字段

图 20



当分组丢失发生时纠正H.264 SVC中的编码器状态的示例(现有技术)

图 21

3 字节 SVC NAL 报头扩展 (现有技术)

字节0								字节1								字节2							
0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
R1	PID						DF	T				D				Q	L	B	F	LF	FO	R2	

R1:保留_零_比特 (1)

PID:优先级_id(6)

DF:可丢弃_标志 (1)

T:时间_级 (3)

D:依存性_id(3)

Q:质量_级 (2)

L:层_基_标志 (1)

B:使用_基_预测_标志 (1)

F:分段_标志 (1)

LF:最后_段_标志 (1)

FO:段_次序 (2)

R2:保留_零_两_比特 (2)

图 22

具有帧索引的 3 字节 SVC NAL 报头扩展

字节0								字节1								字节2													
0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7						
R1	PID						DF	T				D				Q		L	B	F	LF	FO	R2						
								0												0		FI							
								>0												0		LFI							

字段描述 (括号中为比特长度)

R1:保留_零_比特 (1)

PID:优先级_id(6)

DF:可丢弃_标志 (1)

T:时间_级 (3)

D:依存性_id(3)

Q:质量_级 (2)

L:层_基_标志 (1)

B:使用_基_预测_标志 (1)

F:分段_标志 (1)

LF:最后_段_标志 (1)

FO:段_次序 (2)

R2:保留_零_两_比特 (2)

FI:关键_画面_帧_索引 (6)

LFI:最后_关键_画面_帧_索引 (6)

图 23

具有附加比特中的帧索引的 SVC NAL 报头扩展

字节0								字节1							字节2								
0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
R1	PID						DF	T			D		Q		L	B	F	LF	FO	R2	EF=0		
								0													1		
								>0													1		

字节3							
0	1	2	3	4	5	6	7
<不存在>							
FI							
LFI							

字段描述（括号中为比特长度）

- R1 :保留 _ 零 _ 比特 (1)

PID :优先级 _id(6)

DF :可丢弃 _ 标志 (1)

T :时间 _ 级 (3)

D :依存性 _id(3)

Q :质量 _ 级 (2)
- L :层 _ 基 _ 标志 (1)

B :使用 _ 基 _ 预测 _ 标志 (1)

F :分段 _ 标志 (1)

LF :最后 _ 段 _ 标志 (1)

FO :段 _ 次序 (2)

R2 :保留 _ 零 _ _ 比特 2(1)

EF :扩展 _ 标志 (1)

FI :关键画面帧索引 (8)

LFI :最后关键画面帧索引 (8)

图 24

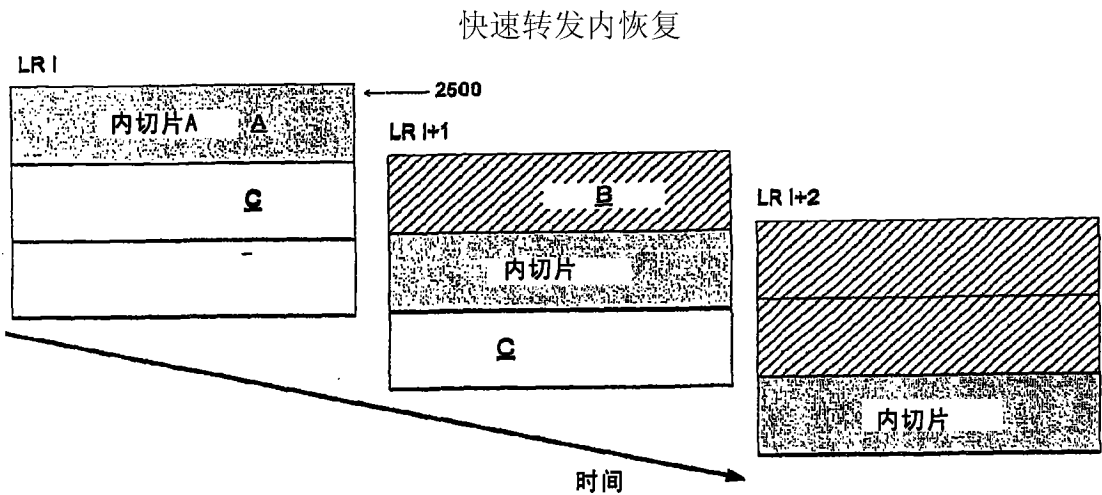


图 25

用于SR画面的增强的快速转发内恢复

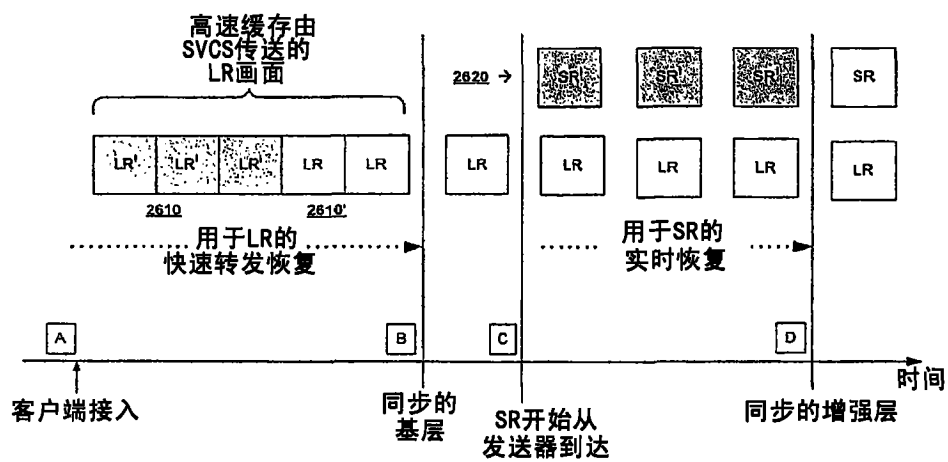


图 26