

基于压缩域的多路视频流混合方法^①

王中元^② 胡瑞敏^③ 常 军 钟 睿 韩 镇

(武汉大学国家多媒体软件工程技术研究中心 武汉 430072)

摘 要 针对视频会议应用中传统的像素域多路视频混合方法存在运算复杂度高、画面质量损伤的问题,提出了一种基于压缩域的替代方法,并给出了详细的码流映射算法步骤。该方法按照混合后多画面的空间位置关系,通过对输入的多路码流中宏块编码次序的重排和语法元素的映射,在码流级别将多路视频合成到同一画面中,并采取提前量化策略消除可能出现的二次量化失真,从而可兼具处理速度快和高保真的双重优点。以 H.263 为例验证了此方法的有效性。实验结果表明,与编解码器级联的方法相比,此方法的峰值信噪比(PSNR)平均提高 2dB,运算效率提高百倍以上。此研究工作有望为正在制定的国际视频编码标准 H.265 贡献一种视频混合解决方案。

关键词 视频混合, 压缩域, 转码, 视频会议

0 引 言

视频会议系统中的多点控制单元(multipoint control unit, MCU)承担多方会议的混音及画面拼接功能。音视频混合处理的运算效率决定了 MCU 的性能瓶颈,在目前普遍采用的基于像素空间的画面混合方式下,MCU 的运算开销随接入路数成线性增长,极大地限制了 MCU 的扩容潜力。研究高效率、高保真度的视频混合算法极具应用与标准价值。

视频混合器通过将接收的各通道画面拼接成一个大的画面,实现了终端能够同时观看到不同的会议场面的效果。早期的一种非标准化做法是开启多个逻辑通道^[1],MCU 将接收到的分会场的多路视频依然用多个逻辑通道转发到每个终端。这种方法要求终端绑定多个视频解码器,不符合视频会议标准 H.323 或会话发起协议(SIP)定义的通信规程。随着处理器能力的提升,开启多个逻辑通道的方式被弃用,转而寻求编解码器级联的方案^[2]。由于图像的编码和解码需要消耗大量的系统资源,因此对编解码器级联的视频混合方案的研究集中在如何降低二次编码过程的运算复杂度上。大量文献^[3-5]讨论了重用解码信息加速编码的各种策略,尤其是在运动矢量的合成和精化方面出现了不少实用化的技

术^[6,7]。但这类编解码器级联的视频混合应用框架存在两个不容忽视的固有缺陷,一是二次编码步骤降低了原有的画面质量,二是处理时间增加了媒体流的转发时延。文献[8]报道了一种分层媒体混合器的模型,可相应地缓解多 MCU 级联时的通信时延,而图像质量下降问题则无法回避。由于现有的 H.263、H.264 等视频编码协议给视频混合造成了极大的不便,所以国际电信联合会 ITU-T 在制定新一代视频编码标准 H.265 时,明确将“易于视频混合”作为需求目标之一^[9]。事实上,早在 1998 年发布的 H.263 的更新版本 H.263+ 中,视频编解码界已经注意到多路视频复合传输的需求,H.263+ 增加了连续存在多点(continuous presence multipoint, CPM)模式^[10]用于携带多子视频流,但极少被采用。受到压缩域视频转码(transcoding)启发,本文提出了一种在离散余弦变换(DCT)域实现多路视频画面混合的新的方法,在视频流的语法级别将多路视频流按照混合后的空间位置复合在一起,例如 4 个 QCIF 码流合并成 1 个 CIF 的码流,或 4 个 CIF 码流合并成 1 个 4CIF 的码流。该方法兼具计算复杂度低、处理速度快、延时小、不改变原有画面质量等特点。这一研究成果已经被应用到自主知识产权的 H.323 视频会议系统设计中,并申请了国家发明专利。

① 973 计划(2009CB320906),国家重大科技专项(2010ZX03004-003-03)和国家自然科学基金(61003184, 60970160, 61070080)资助项目。

② 男,1972 年生,博士,副教授;研究方向:视频编解码;E-mail: wzy_hope@163.com

③ 通讯作者,E-mail: hrm1964@public.wh.hb.cn

(收稿日期:2010-05-17)

利^[11]。

1 码流语法映射

本文算法的关键是如何恰当利用多路视频头信息的相关性来重构混合后视频流的头信息。依照 H.263 码流自身的层次性结构,多路视频头信息的相关性可以分为图像层相关性、块组层相关性及宏块层相关性三个层次。相应地,三层结构的编码参数的映射关系如图 1 所示。下面给出每层的映射算法。

1.1 图像层相关性

图像层相关性的映射要点是混合后视频流的图像层与混合前多路视频流的图像层同时相关,具体映射方法列于表 1。

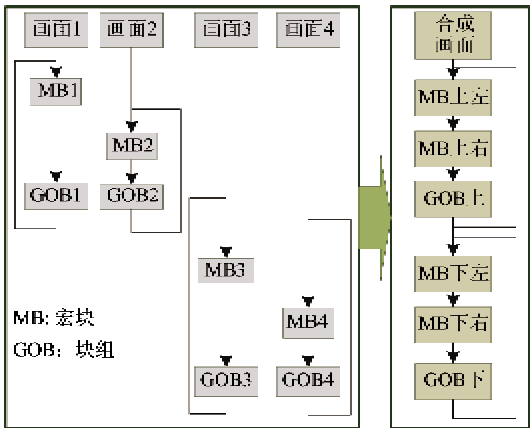


图 1 码流映射示意图

表 1 图像层相关性映射

相关性	映射依据	映射规则
时域参考 (TR) 值相关性	混合前多路视频流的时域参考值 TR 相互独立,而混合后则要求多个 TR 值统一为一个 TR 值	采用 N 路视频流 TR 值的均值作为新的 TR 值,其中 N 为当前未结束的视频流数目,若某一路视频流已结束,则 N 相应减 1
图像类型相关性	混合前多路视频流的 I 帧间隔不同,因此会出现 I 帧、P 帧混合为一帧的情况	规定当前多路视频流均为 I 帧,混合后也为 I 帧;否则,混合后为 P 帧,混合前的 I 帧混合后都转为 P 帧的 INTRA 宏块
图像量化步长 (PQUANT) 值相关性	混合后视频流的 PQUANT 值只与第一路 (左上角) 视频流相关	直接将第一路视频流的 PQUANT 值作为混合后的 PQUANT 值

1.2 块组层相关性

考虑到混合后视频流的块组层只与混合前同一水平方向的两路视频流的块组层相关,与其它视频流无关。因此可以规定:只要当同一水平方向上左

边的视频流存在 GOB 层,则混合后在此水平方向上存在 GOB 层,否则不存在 GOB 层。块组层相关性的映射规则总结见表 2。

表 2 块组层相关性映射

相关性	映射依据	映射规则
块组帧标记 (GFID) 值相关性	因为 GFID 的变化与帧类型 (PTYPE) 值的变化是同步的,而决定混合后视频流 PTYPE 值变化的是图像类型,故 GFID 值的相关性其实就是图像类型相关性	直接根据混合后视频流 I/P 帧类型确定混合后的 GFID 值
块组量化步长 (GQUANT) 值相关性	只有当同一水平方向上左边的视频流存在 GOB 层,混合后在此水平方向上才会存在 GOB 层	混合后的 GQUANT 值直接等于左边视频流的 GQUANT 值

1.3 宏块层相关性

混合后视频流的宏块只与其混合后的相邻宏块相关,因为只有在四路视频交界处和 GOB 层变化处混合前后的相邻宏块会发生变化,所以宏块层重构的关键在视频交界处和 GOB 层变化处宏块的处理。宏块层重构面临转换的语法元素最多,包含宏块头信息、运动矢量及 DCT 系数。本节只讨论前两种, DCT 系数的修正放在下一节论述。

编码指示位 (COD) 值相关性。I 帧的 INTRA 块转为 P 帧 INTRA 块时,会增加 1 比特的 COD 位。此外,当量化步长差值 (DQUANT) 值因 $[-2, +2]$ 范围的限制而必须进行二次量化时,以前的非零系数有可能全部被重新量化为零系数,致使宏块不再存在非 INTRA DC 系数,此时 COD 值有可能由 0 变为 1,故在本算法中,会对二次量化后的系数重新统计,以确定新的 COD 值。

宏块类型及色度的编码块样式(MCBPC)值相关性。可以从三个方面考虑,第一点,I 帧转变为 P 帧,MCBPC 的变长码表会改变;第二点,当需要进行二次量化时,宏块级量化步长(QUANT)的差值可能由零变为非零,也可能由非零变为零,这种情况下宏块类型会在 INTER 与 INTER + Q, INTRA 与 INTRA - Q 之间改变,从而改变 MCBPC 值;第三点,二次量化可能会使色度块的系数量化为零,使某一色度块不存在非 INTRA DC 系数,从而改变 MCBPC 值。亮度的编码块样式(CBPY)值相关性与 MCBPC 相似,二次量化也会改变亮度块的非零 INTRA DC 系数,从而改变 CBPY 值。

运动矢量数据(MVD)相关性。H. 263 标准的宏块的运动矢量采用了差分编码技术,差分编码值是当前宏块的运动矢量和“预测因子”之差。预测因子取自三个相邻宏块的运动矢量(左、上、上右)的中值,因此,MVD 与这三个相邻宏块相关。此外,当 GOB 头非空时,处于 GOB 顶端的宏块的候选预测因子 MV2(上)、MV3(上右)均置为 MV1(左),因此 MVD 还与 GOB 层相关。由于多路视频混合后不仅 GOB 层会发生改变,图像交界处的三个相邻宏块也会发生改变,故先根据混合前的 GOB 层和相邻宏块重构出运动矢量值,再由混合后的 GOB 层和相邻宏块求取新的预测因子,以得到新的运动矢量差值。

2 二次量化失真的消除

2.1 问题的提出

压缩域操作码流引起的重量化误差漂移问题很早就被注意到,Tang 等人^[12]在最近的研究中对转码中的重量化与插值误差的补偿给出了一个比较有说服力的结果,但这并不能完全套用到处理码流合成的误差上,原因在于二者产生的根源不同。2010 年 4 月,Cock 等人^[13]用长文报道了对 H. 264 转码中重量化误差在时间和空间方位上漂移(drift)和累积(propagation)的定量分析,提出的漂移补偿(drift compensation)思路对克服 H. 264 闭环转码的质量损失有作用。在 H. 263 中,宏块层码率控制不允许相邻宏块的量化步长值发生陡变,DQUANT 被限制在 $[-2, +2]$ 之间^[10]。在多路视频混合时,原先两路图像的宏块会在视频交界处和换行处相邻(如图 2 所示),这两路图像原本是相互独立的,其质量可能会有较大的差异,例如一路码率高而另一路码率低,此时相邻宏块间的 DQUANT 就会比较大,无法

保证它们恰好处于 $[-2, +2]$ 之间。因此,为确保组合后相邻宏块间的 DQUANT 的幅度不超过 2,就必须根据 $[-2, +2]$ 限制重新计算量化步长。

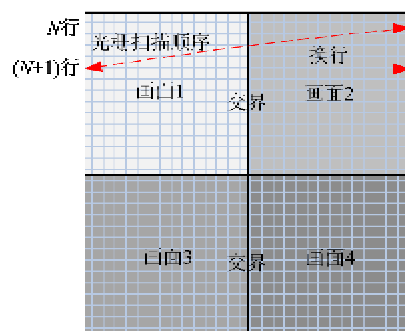


图 2 混合图像宏块的位置关系示意

一旦量化步长被改变,就要对 DCT 系数相应地采取二次量化,即先按第一次的量化步长进行反量化,再用算得的新量化步长再次量化。这一过程不仅影响了 DQUANT、COD、MCBPC、CBPY 的数值,产生的二次量化误差进而还会影响到相邻位置之后的宏块的量化。以合并后的大图像的角度看,当操作以光栅扫描顺序从低码率子图区域向高码率子图区域转移时,势必出现第一次量化步长较小而第二次量化步长较大的情况,重量化在更低的精度下执行,量化失真加剧而改变了原有图像质量。尤其是当画面交界和换行处宏块量化器相差较大时,连续多个宏块都将被执行二次量化,从而导致原先的高码率图像质量显著下降。由此可见,必须采取措施消除二次量化失真对画面质量的负面影响。

进一步分析可知,不仅两幅子图像相邻位置之前的 DCT 系数无需施加二次量化,对于由高码率子图向低码率子图过渡的情况,原先低码率视频的质量亦不会受损伤。这正是下文在研究应对策略时仅针对低码率向高码率过渡这一种情形的缘故。

2.2 提前量化策略

避免二次量化误差最直接的方法是对每一行都设置 GOB 层,因为 GOB 层的 GQUANT 位允许重新设置量化步长,从而避开了 DQUANT 的约束。但是设置 GOB 层只能使组合位置中左边的图像(即图 2 中的画面 1、3)不受二次量化误差影响,而不能避免右边图像(即图 2 中的画面 2、4)的质量下降。文献[14]提出了一种借助 H. 263 + 选项 Annex K 的思路,但添加的高级选项可能不被低版本解码器理解,缺少普适性。

总之,包括设置 GOB 层在内的原有的方法只能

在由低码率视频变为高码率视频的交界处和换行处才可以检测到量化步长的陡变,之后才能按 $-2/\text{MB}$ 的速率将低码率视频的大量化步长缓慢变化到高码率视频的小量化步长上,以致高码率视频质量的下降难以完全避免。为此,本文进一步提出称之为提前量化的方法来彻底解决这个问题。

提前量化方法的基本思想并非是在低码率视频到达高码率视频的交界处和换行处才开始降低步长,而是提前按 $-2/\text{MB}$ 的速率降低低码率视频的大量化步长,使得到达交界处和换行处时,低码率视频量化步长已经能够比较平滑地迅速过渡到高码率视频的量化步长。这样就可以根本上避免高码率视频主观质量的下降,对低码率视频而言,由于第一次量化步长已经很大,第二次的小量化步长量化对其主观质量和码率基本上都不会产生明显影响。算法的具体步骤如下:

(1) 分别计算低码率视频与高码率视频第一次交界处或换行处的前后 N 个宏块 (N 为合并前视频的一行宏块的数目,如 QCIF 则 N 为 11) 的量化步长的均值 $\overline{QP}_l, \overline{QP}_h$ 。

(2) 预测以下一次低码率视频与高码率视频交界处或换行处为基准需要提前量化的距离(以宏块为单位) $L = ((\overline{QP}_l - 1) - (\overline{QP}_h + 1))/2 - 1$ 。

(3) 根据提前量化后交界处或换行处的实际量

化差值 ΔQP 对提前量化距离进行修正:当 $\Delta QP > 2$ 时,增大提前量化的距离,增量 $\Delta L = \Delta QP/2 - 1$; 当 $\Delta QP < -2$ 时,减小提前量化的距离,减量 $\Delta L = |\Delta QP|/2 - 1$ 。

(4) 将修正后的提前量化距离作为下一次的提前量化距离。

(5) 重复修正、量化,直到一帧图像结束。

3 实验结果与分析

针对本文方法,我们同编解码器级联方法进行了实验比较。实验输入的 4 路 QCIF 格式码流依次为 Container, Foreman, Akiyo, Silent, 其中每路码流均为 100 帧,由 H.263 的 TMN8 编码器编码产生,每路码率 96kbps,帧率 25fps。图 3 和图 4 分别示出了两种方法的主观视频质量对比和峰值信噪比 (PSNR) 对比,可以看出,本文方法的视觉效果明显要好,PSNR 平均提高 2dB。至于运算时间,本文方法只有视频流的解复用和复用的操作,以及少量的二次量化的计算,而这些运算相对于复杂的编解码过程而言所占比例微不足道,实验结果统计数据也证实了这一点,在普通笔记本电脑上混合 100 帧 4 路 QCIF 的运算时间相差达百倍以上。



图3 混合视频的主观效果对比

针对本文提出的提前量化方法,我们也进行了比较实验。实验输入的 4 路 QCIF 码流依次为 Foreman, Silent, Silent, Silent, 每路码流均为 100 帧,由 TMN8 编码器编码,帧率 25fps,其中 Foreman 码率 48kbps,三路 Silent 码率均为 96kbps。图 5 是用提

前量化方法改进前后的主观视频质量对比,可以很明显的看到,虽然输入的三路 Silent 完全一样,但是由于低码率视频 Foreman 导致了严重的二次量化误差,使得右上角的 Silent 质量严重下降,使用提前量化方法改进后,右上角的 Silent 质量有了明显改善。

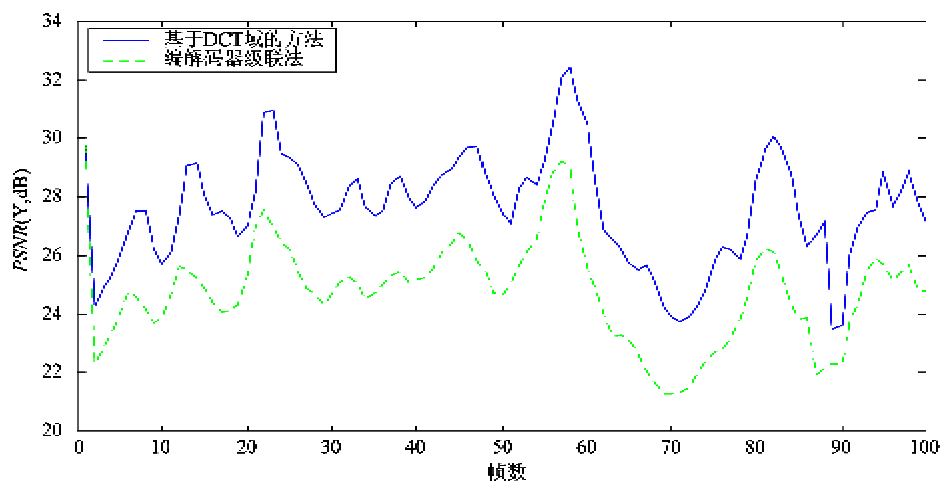


图 4 基于 DCT 域方法 vs 编解码器级联方法



图 5 提前量化方法的主观效果

4 结 论

文章提出了在 DCT 域压缩码流的语法和语义级别实现多路视频混合的方法,相比于传统的基于像素域的方法,此方法运算复杂度低,且处理过程没有引入新的失真。以 H. 263 标准为例,对此方法进行了算法实现和仿真验证,对于基于块的其它编码标准,如 MPEG-4、AVS 或者 H. 264,本文方法同样适用。值得指出的是,由于实际通信环境中视频传输面临的问题更复杂多变,诸如送到画面混合器的多路图像尺寸、帧率不一样,要求混合器输出多种码率等等问题,仅仅通过比特流语法映射是不够的。下一步我们希望借鉴视频转码中的空间分辨率转码、时间分辨率转码以及比特率缩减转码的成果,来尝试解决这些问题。研究成果有望为正在制定的国际视频编码标准 H. 265 贡献一种视频混合解决方案。

参考文献

- [1] 边学工,胡瑞敏,陈军等. 基于分层排队网络模型的 MCU 性能预测及优化研究. 计算机学报, 2004, 27 (2): 209-215
- [2] 汪陈伍,朱志祥. 一种在像素域内实现多画面合成的方法. 西安邮电学院学报, 2005, 10 (4): 127-130
- [3] 杨高波,余圣发,张兆杨. 压缩域 H. 264 视频转换编码及其关键技术分析. 通信学报, 2006, 27(10): 124-130
- [4] 褚晶辉,俞斯乐,鲁照华. 视频转换编码及其实现技术的研究. 电子学报, 2004, 32(10): 1678-1683
- [5] Atkins C B. Blocked recursive image composition. In: Proceedings of the ACM Multimedia 2008, Vancouver, Canada, 2008. 821-824
- [6] Xu X Z, Wu Z M, He Y. A refined motion estimation strategy for adaptive interpolation filter. *Signal Processing: Image Communication*, 2010, 25(2): 121-129
- [7] Chen M J, Chu M C, Lo S Y. Motion vector composition algorithm for spatial scalability in compressed video.

- IEEE Transactions on Consumer Electronics*, 2001, 47(3): 319-325
- [8] Rangan P V, Vin H M, Ramanathan S. Communication architectures and algorithms for media mixing in multimedia conferences. *IEEE Transactions on Networking*, 1993, 1(1): 20-30
- [9] Question 6/16-Visual coding. [http://www. itu. int/ITU-T/studygroups/com16/sg16-q6. html](http://www.itu.int/ITU-T/studygroups/com16/sg16-q6.html):ITU,2008
- [10] H.263: Video coding for low bit rate communication. [http://www. itu. int/rec/T-REC-H. 263-199802-S](http://www. itu. int/rec/T-REC-H.263-199802-S): ITU, 2008
- [11] 胡瑞敏, 王中元, 韩镇等. 基于 DCT 空间的多路画面混合方法. 中国专利, ZL200310111331.4. 2006-03-01
- [12] Tang Q, Nasiopoulos P, Ward R K. Compensation of requantization and interpolation errors in MPEG-2 to H. 264 transcoding. *IEEE Transactions on Circuits and Systems for Video Technology*, March 2008, 18(3): 314-325
- [13] Cock J D, Notebaert S, Lambert P, et al. Requantization transcoding for H. 264/AVC video coding. *Signal Processing: Image Communication*, 2010, 25(4): 235-254
- [14] Yoo K Y, Seo K D. Syntax-based mixing method of H. 263 coded video bitstreams. In: *Proceedings of International Conference on Consumer Electronics, Las Vegas, USA, 2005*. 403-404

A DCT-based multipoint video composition scheme

Wang Zhongyuan, Hu Ruimin, Chang Jun, Zhong Rui, Han Zhen

(Wuhan University National Multimedia Software Engineering Research Center, Wuhan 430072)

Abstract

Aiming at the problems of high computational complexity and picture quality degradation of traditional pixel domain video mixing methods useful for multipoint conferences, this paper proposes a multipoint video composition scheme based on the compressed domain of discrete cosine transform (DCT), and describes the details of the bitstream mapping algorithm. With the rearrangement of the macroblock coding order and the mapping of the syntax element, the scheme combines multiple channel video frames together into the unique picture on the syntax layer according to the spatial position relation of the composition stream, and then a pre-quantization policy is particularly presented to remove requantization errors. To verify the availability, the details of the algorithm are integrated into the H.263 codec. The experimental results revealed that compared with the cascaded method, the average peak signal to noise ratio of the proposed method (PSNR) was improved almost 2dB and the operational efficiency increased a hundredfold. It is possible that this research can provide a video mixing solution for the international video coding standard H.265 which is under development.

Key words: video composition, DCT domain, transcoding, video conference