

H. 264 可分级扩展技术的介绍和分析

骆政屹 余松煜 宋 利 杨小康

(上海交通大学图像通信与信息处理研究所, 上海 200240)

摘 要 H. 264 可分级扩展是 ISO/IEC 组织和国际电信联盟最新的联合标准制订工作。目前, H. 264 扩展标准仍处于制订过程中, 但其主体框架已基本确定。为了使人们对正在制定的 H. 264 扩展标准有一概略了解, 该文首先对扩展标准的主体框架作了简要介绍, 并阐述了实现空域、时域和质量可分级的机制, 然后对其中的层间编码和精细质量可分级编码包含的关键技术进行了较为详细的分析, 并利用最新的参考软件 JSVM5.7 进行了性能以及复杂度的比较和测试, 结果表明, 该项技术在具有良好的支持分级能力的同时, 可保持较高的编码效率。

关键词 视频编码 可分级 H. 264

中图法分类号: TN919.81 文献标识码: A 文章编号: 1006-8961(2006)11-1578-06

Introduction and Analysis of the H. 264 Scalable Extension Techniques

LUO Zheng-yi, YU Song-yu, SONG Li, YANG Xiao-kang

(Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240)

Abstract The scalable extension of H. 264 is the newest work of developing joint standards for ISO/IEC and ITU. Now, this project is still in progress, and yet the main frame of the extension standard has been established. In order to make the ongoing standard known to people, in this paper a short introduction about the main frame of the H. 264 extension standard is provided first, with the mechanism of implementing spatial, temporal and quality scalability explained. Then the key techniques involved in inter-layer coding and fine grain quality scalable coding are analyzed in detail. Making use of the newest reference software JSVM5.7, we carry out some comparisons and tests on performance and complexity, which verify that those techniques keep quite good coding efficiency together with satisfactory scalable performance.

Keywords video coding, scalable, H. 264

1 引 言

近年来,随着编解码技术的发展,视频压缩与传输业务的应用领域不断扩大。对于不同接收条件的应用场合,传统的视频编解码技术需要为每种场合都提供一套压缩与传输系统。随着应用环境的增多,压缩与传输系统的重复性构建无疑耗费了巨大的资源。因此,视频业务迫切需要一种能够同时适用于各种应用环境的可分级编解码系统。

为了满足对可分级编解码的需求,某些传统的视频编码标准(如 MPEG-4)虽进行了适度的可分级

扩展,但实际使用的结果表明这些标准仍无法提供满意的可分级性能。基于高性能可分级编解码系统的需求,2004 年 10 月, MPEG 组织决定以压缩性能优异的 H. 264 标准为起点制订新一代的可分级编解码标准。2005 年 1 月, MPEG 组织和 VCEG 组织同意把可分级编解码标准作为现有 H. 264 标准的扩展和修正。

H. 264 标准是一种经典的混合式视频编解码标准,其中各种高效的编解码方法使 H. 264 标准具有良好的压缩效率。正在制订的 H. 264 扩展标准^[1]继承了 H. 264 标准的各项编解码技术,其显著特点是宏观上采用了分层式编解码结构。基于 H. 264

收稿日期:2006-04-30; 改回日期:2006-07-30

第一作者简介:骆政屹(1983 ~),男。2004 年获南京邮电学院学士学位,现为上海交通大学硕士研究生。主要研究方向为视频编解码。
E-mail: llzyynjupt@sina.com

扩展标准的主体框架结构,本文首先阐述了主要的可分级类型,接着对扩展标准中的关键编码技术作了详细地介绍和分析,并利用参考软件 JSVM5.7^[2-4]进行了性能测试。测试环境是 P4 3GHz CPU、512M 内存、Windows XP 操作系统,所用的编译工具是 Visual C++ 6.0。

2 分组结构

与 H. 264 标准一样,H. 264 扩展标准仍以 GOP (group of pictures)为单位进行编码。GOP 中每层一般包含一个关键帧(I 帧或 P 帧)和若干分级式 B 帧。GOP 的大小通常是 2 的指数次方。根据不同的分组方式,扩展标准支持固定式 GOP 和自适应 GOP 两种分组方式。其中,固定式 GOP 要求每次都以同样的 GOP 尺寸编码;而在自适应 GOP 下,则可以根据图像内容自适应地选择分组方式。

分组方式以及分组大小与最终的编码性能密切相关,因为 GOP 尺寸越大,B 帧所占的比例越多。由于 H. 264 扩展标准是采用分层式的编码结构,因此各层增加的 B 帧会使编码延迟大大增加,但由此获得的增益是总体码率的下降。

3 可分级类型

H. 264 扩展标准支持多种可分级类型,其中空域可分级、时域可分级和质量可分级是可分级模式的典型代表。

3.1 空域可分级

与传统 H. 264 标准不同,在每个 GOP 内,扩展标准允许编码同一视频源的多层图像,且层与层之间可以有不同的空间分辨率。例如,如图 1 所示,第 1 层到第 3 层可以分别编码视频源 QCIF、CIF 和 4CIF 格式的图像。这样,编码后便能得到一个具有 3 种空间分辨率的码流。在码流发送阶段,就可以向不同的接收方发送不同比例的码流。其中接收到全部码流的客户端可以显示 4CIF 格式的图像,而接收部分码流的客户端则可显示 CIF 或 QCIF 格式的图像。通过这种由不同比例的码流解码出不同空间分辨率的图像的方式,H. 264 扩展标准就达到了空域上可分级的效果。

3.2 时域可分级

H. 264 扩展标准是利用多个编码层来实现空域

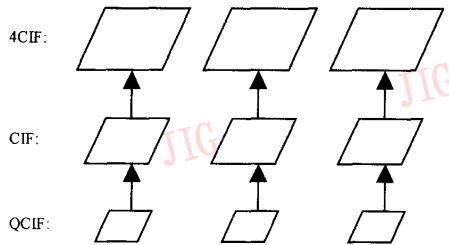


图 1 空域可分级示例

Fig. 1 Example of spatial scalability

的可分级。具体在每一层内,编码按照逐次分级的方式进行,即第 1 次编码关键帧图像(I 帧或 P 帧),下面依次编码不同级别的 B 帧图像。例如,当 GOP 尺寸为 8 时,可按照第 8 帧、第 4 帧(第 1 级 B 帧)、第 2、6 帧(第 2 级 B 帧)、第 1、3、5、7 帧(第 3 级 B 帧)的次序逐级编码。第 8 帧图像(P 帧)选取上一个 GOP 的关键帧为参考图像,其余各个 B 帧图像一般均以上一轮编码的重建图像为参考帧,详细的编码流程如图 2 所示。

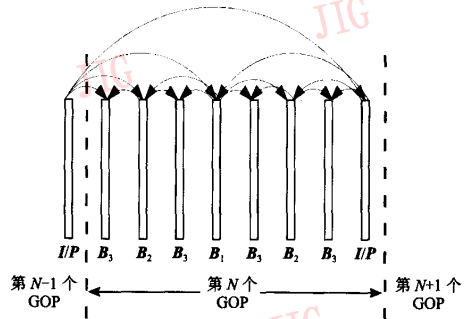


图 2 时域可分级示例

Fig. 2 Example of temporal scalability

编码码流发送后,能接收到全部码流的客户就可以显示 GOP 中的全部图像;接收到中等码流的客户能显示 GOP 中的第 2、4、6、8 帧图像;而接收到最低码流的客户则只能显示 GOP 中的第 8 帧图像。这样,通过层内的逐级编码,H. 264 扩展标准就实现了时域上的可分级。

表 1 和表 2 是空域和时域可分级的实验结果。测试时,GOP 尺寸固定为 8,设置参考软件为两层编码结构,第 1、2 层的图像格式分别为 QCIF 和 CIF。共编码 100 帧图像,帧率为 30Hz,两层的量化参数都为 36。其编码采用 32pixels 范围内的快速运动估

表 1 Tempete 空域和时域可分级测试结果

Tab.1 Test of spatial and temporal scalability for tempete

		码 率 (kbps)			
时域 (Hz)		3.75	7.5	15	30
空域	QCIF	38.1438	50.5320	65.3520	80.4096
	CIF	181.0338	228.4536	279.4680	335.1912

表 2 Mobile 空域和时域可分级测试结果

Tab.2 Test of spatial and temporal scalability for mobile

		码 率 (kbps)			
时域 (Hz)		3.75	7.5	15	30
空域	QCIF	61.8808	78.3192	93.7080	107.5776
	CIF	295.2485	359.1960	424.3512	492.3864

计,基层模式设为带 SEI (supplemental enhancement information) 信息的 H.264 兼容模式,层间预测处于自适应状态,而其余编码选项均取默认设置。

由表 1 和表 2 可见,随着空域和时域分辨率的增加,码率逐渐增大。这样在码流发送阶段,就可以为不同接收条件的客户端提取和发送不同类型的码流,然后接收方就可根据所收到的码流显示出不同空间和时间分辨率的图像。

3.3 质量可分级

除了空域和时域可分级之外,H.264 扩展标准还支持质量(信噪比)可分级。质量可分级包括粗粒可分级 CGS (coarse grain scalability) 和细粒可分级 FGS (fine grain scalability) 两种。CGS 与空域可分级类似,只是编码时各层输入的是相同分辨率的图像,其中信噪比的变化可通过各层间不同的量化参数实现。由于仅仅是通过各层的量化参数来调节信噪比,因此可分级的粒度一般比较大,其在码流提取阶段也只有有限的几种码率可供选择。表 3 和表 4 是 CGS 的测试结果。测试时,GOP 尺寸固定为 16,3 层图像均设为 CIF 格式,共编码 100 帧图像,帧率为 30Hz。第 1 至第 3 层的量化参数分别为 36、

表 3 Tempete 质量粗粒可分级测试结果

Tab.3 Test of coarse grain quality scalability for tempete

层级	量化参数	码率 (kbps)	Y 分量峰值信噪比 (dB)	U 分量峰值信噪比 (dB)	V 分量峰值信噪比 (dB)
1	36	300.2088	31.6870	36.6309	38.9589
2	27	1174.5120	36.2696	39.9796	41.8023
3	18	4220.1000	43.0234	44.6947	45.6238

表 4 Mobile 质量粗粒可分级测试结果

Tab.4 Test of coarse grain quality scalability for mobile

层级	量化参数	码率 (kbps)	Y 分量峰值信噪比 (dB)	U 分量峰值信噪比 (dB)	V 分量峰值信噪比 (dB)
1	36	419.7960	30.5740	35.3707	34.6351
2	27	1677.5400	35.4428	38.6733	38.2618
3	18	5587.9464	42.7895	44.0843	44.0728

27 和 18,其余参数与之前的测试相同。

随着量化参数的降低,码率和信噪比逐渐升高,因此发送时可根据需要发送几种不同信噪比的码流。

质量的细粒可分级有一套完整的理论体系,第 5 节将单独介绍 H.264 扩展标准中的质量精细粒度可分级技术。

除了以上 3 种可分级类型外,H.264 扩展标准还在酝酿其他可分级方式,如计算复杂度可分级、感兴趣区可分级等。需要特别指明的是,各种分级方式之间并不是孤立存在的,应用时可以相互组合,以满足对分级方式的多样化需求。

4 层间编码

如前所述,H.264 扩展标准仍然把图像分为 I、P、B 帧。对于层内的每一幅图像,扩展标准允许其可以进行 H.264 标准的运动估计和模式判决。考虑到一般会有多个编码层,各层图像又都来自于同一视频源,因此,除了传统的编码方式以外,各层间的相关性使得 H.264 扩展标准具有一系列层间编码方式。层间编码总体上遵循在相邻两层间由低层向高层预测的原则。

4.1 层间帧内纹理预测

因为各层图像是同一视频源的不同分辨率版本,所以,用低层重建图像来预测高层图像可以降低层间的相关性和提高压缩效率。遵从参考软件,不妨称这种层间预测编码方式为 IntraBL 模式。这种 IntraBL 模式可以被高层图像采用,具体方法是先确定是否有对应于当前宏块的基层宏块。如果存在,则将此重建的宏块经适当插值后(若两层图像的空间分辨率相同,则无须插值)作为当前层宏块的预测。下面只需对当前层宏块的预测差值进行编码即可。

另外,可否使用 IntraBL 模式还与解码环路的设置有关。解码环路一般可分为完全单环、除关键帧(I、P 帧)以外单环和普式多环 3 种。其中完全单环是指任何帧的宏块,只有在其基层宏块采用帧内编码时才可应用 IntraBL 模式;除关键帧(I、P 帧)以外单环是指 B 帧的宏块有使用 IntraBL 模式的上述限制,而对 I、P 帧则没有特殊要求;普式多环下任何帧都没有使用 IntraBL 模式的上述限制。做出这种分类主要是考虑到解码器的运算复杂度^[5]。在完全单环下,由于不存在多重预测环路,因此解码器的运算量最低;而在普式多环下,多重预测环路使得解码器的运算复杂度最高。目前,H. 264 扩展标准草案中仅限于完全单环下使用 IntraBL 模式,不过参考软件 JSVM5.7 仍保留有上述 3 种使用 IntraBL 模式的方式。

4.2 层间运动预测

当对应的基层宏块采用帧间编码模式时,其运动信息可用于预测当前层宏块。层间运动预测一般包括基层模式(base layer mode)和 1/4 像素修正模式(quarter pel refinement mode)两种。

在基层模式下,当前宏块的运动矢量、参考帧索引、宏块分割方式等信息完全由相应的基层宏块得到。如果两层图像具有不同的空间分辨率,则基层宏块的运动矢量、分割方式等需经适当地插值放大后再赋予当前宏块。1/4 像素修正模式只有在两层图像具有不同的空间分辨率时才可使用。在 1/4 像素修正模式下,当前宏块的运动矢量、参考帧索引、宏块分割方式等信息也是从基层宏块推得。不过与基层模式不同的是,其得到的运动矢量需要在当前层 1/4 像素范围内做修正。

层间运动预测之后,当前层宏块就可以利用得到的信息直接进行运动补偿,进而对预测残差编码。由于各层图像均来源于同一视频源,因此恰当运用层间运动预测可以减少大量不必要的层内运动估计和模式判决。从算法优化的角度来说,这也是一个减少编码器复杂度的着力点。

4.3 层间残差预测

当相应的基层宏块和当前层宏块都有预测残差时,则可以通过用基层残差预测当前层残差的方式来提高编码效率。若两层图像具有不同的空间分辨率,则用以预测的基层残差同样要经过适当地插值放大。

层间运动预测和层间残差预测在是否使用上可以相互组合,由此能够得到 4 种不同的层间编码

方式。

4.4 修正 IntraBL 模式^[6]

修正 IntraBL 模式一般适用于高层 B 帧宏块,当对应的基层宏块是帧间编码宏块时,则可以选择此种编码方式。在修正 IntraBL 模式下,当前层宏块仍由基层宏块得到运动信息。但在计算最终的预测信号时,则要结合基层宏块的预测残差进行平滑处理。

表 5 和表 6 是参考软件在不同层间预测状态下的测试结果。测试时,GOP 尺寸固定为 16,两层图像分别为 QCIF 和 CIF 格式。共编码 100 帧图像,帧率为 30Hz。两层的量化参数均设为 27,其余参数与之前的测试相同。

表 5 Tempete 层间预测测试结果

Tab.5 Test of inter-layer prediction for tempete

层间预测 状态	编码时间 (s)	总码率 (kbps)	高层 Y 分量峰值 信噪比 (dB)	高层 U 分量峰值 信噪比 (dB)	高层 V 分量峰值 信噪比 (dB)
始终关闭	195.625	1309.8192	36.2980	39.9417	41.7990
始终打开	59.875	1477.4880	35.9503	39.9543	41.7271
自适应	391.969	1226.1720	36.2927	39.9662	41.7948

表 6 Mobile 层间预测测试结果

Tab.6 Test of inter-layer prediction for mobile

层间预测 状态	编码时间 (s)	总码率 (kbps)	高层 Y 分量峰值 信噪比 (dB)	高层 U 分量峰值 信噪比 (dB)	高层 V 分量峰值 信噪比 (dB)
始终关闭	200.718	1899.4464	35.4680	38.6376	38.2673
始终打开	60.938	2244.1176	35.2860	38.7191	38.3781
自适应	434.922	1842.1200	35.4908	38.6879	38.3068

由表 5 和表 6 可见,在保持一定压缩效率的前提下,层间预测可以显著提高图像的编码速度。另外,测得的编码时间也表明,各层的运动估计和模式判决是造成 H. 264 扩展标准编码复杂度高的主要原因。如要提高编码速度,恰当运用层间预测将是优化工作的重点所在。

5 精细质量可分级

相比于质量的粗粒可分级,细粒可分级能更好地满足某些应用需求,因此,一直是近年来的研究热点。FGS 编码的显著特点是可以对生成的码流进行

更加自由的切割,以获取满足信噪比和码率要求的码流。为避免歧义,把层间编码中的每层称为基本质量层,而把附加的精细质量可分级层称为 FGS 层。

5.1 I 帧

相比于 P、B 帧,I 帧的 FGS 编码较为简单。编码时只需把原图像与基本质量层或之前 FGS 层的相关数据相减,然后对得到的编码误差进行细量化编码即可。

5.2 P、B 帧

由于 FGS 层码流可被任意切割,因此当 FGS 编码开关被打开时,P 帧的基本质量层一般仍选用参考帧的基本质量层做运动补偿。P 帧的 FGS 层编码一般需要一个基本质量层参考和一个 FGS 层参考。在 AR-FGS(coding with adaptive reference FGS)^[7]方式下,FGS 层参考图像一般都是参考帧的最高 FGS 层重建;在普通 FGS 下,根据当前编码的 FGS 层级,FGS 层参考图像一般是参考帧中相同 FGS 层的重建值。在得到两幅参考图像后,就可以对二者进行适当加权调整,以获得最佳的预测图像。

打开 FGS 编码开关时,B 帧的基本质量层要依据单、双闭环来选取参考图像。在单闭环下,B 帧的基本质量层以参考帧的最高 FGS 层为参考图像;而在双闭环下,则只有参考帧的基本质量层作为 B 帧基本质量层的参考。与 P 帧的 FGS 编码不同,B 帧的 FGS 层一般都选用参考帧的最高 FGS 层进行预测。

P、B 帧的 FGS 层由预测值算得预测残差后,要先与基本质量层或之前 FGS 层的相关数据相减,然后再进行细量化编码。

I、P、B 帧 FGS 层的量化参数由低到高逐层递减。H.264 扩展标准中 FGS 编码时的系数扫描方式是一大亮点。对 FGS 层的变换系数进行自适应算术编码时,根据之前几层的对应位置是否有编码过的非零值,扩展标准把变换系数分为重要性(significant)系数和细化(refinement)系数,进而采用类似于位平面的方式分别对两类系数进行周期模式、矢量模式^[8]等编码。

一般情况下,FGS 层的运动和模式信息均从基本质量层获得,虽然 H.264 扩展标准也允许 FGS 层重新进行模式判决和预测,不过,由此来获得编码增益往往要以计算复杂度的急剧增加为代价。

表 7 和表 8 是利用参考软件对精细质量可分级进行测试的结果。测试时,GOP 尺寸固定为 16,共

编码 100 帧图像,帧率为 30Hz。采用单层编码结构,图像为 CIF 格式,量化参数为 36。设定编码 3 个 FGS 层,各层间量化参数相差 6,FGS 层直接采用基本质量层的运动和模式信息。

表 7 Tempete 精细质量可分级测试结果

Tab.7 Test of fine grain quality scalability for tempete

层级	码率 (kbps)	Y 分量峰 值信噪比 (dB)	U 分量峰 值信噪比 (dB)	V 分量峰 值信噪比 (dB)
基本质量层	294.8520	31.4301	36.5446	38.8879
FGS 第 1 层	670.3560	33.8880	38.5480	40.4251
FGS 第 2 层	1756.7496	37.6670	41.0675	42.7198
FGS 第 3 层	3999.2232	42.9326	44.5710	45.4901

表 8 Mobile 精细质量可分级测试结果

Tab.8 Test of fine grain quality scalability for mobile

层级	码率 (kbps)	Y 分量峰 值信噪比 (dB)	U 分量峰 值信噪比 (dB)	V 分量峰 值信噪比 (dB)
基本质量层	413.1432	30.2609	35.3455	34.6375
FGS 第 1 层	938.9016	32.8196	37.4204	36.6857
FGS 第 2 层	2406.6696	37.0020	39.8064	39.5089
FGS 第 3 层	5215.3512	42.7174	43.9622	43.9504

可被自由切割是 FGS 码流的重要特性。利用测试中生成的码流和参考软件的码流提取工具,就可以实现更加细化的信噪比可分级。表 9 和表 10 是对应的测试结果。

表 9 Tempete 码流提取测试结果

Tab.9 Test of bitstream extraction for tempete

提取码率 (kbps)	图像格式	Y 分量峰 值信噪比 (dB)	U 分量峰 值信噪比 (dB)	V 分量峰 值信噪比 (dB)
600.8088	CIF	33.3991	38.1715	40.2528
1200.8568	CIF	35.2937	39.8178	41.7491
2400.9096	CIF	38.6672	41.4989	43.2515

表 10 Mobile 码流提取测试结果

Tab.10 Test of bitstream extraction for mobile

提取码率 (kbps)	图像格式	Y 分量峰 值信噪比 (dB)	U 分量峰 值信噪比 (dB)	V 分量峰 值信噪比 (dB)
800.8104	CIF	32.0405	36.7971	36.0014
1600.8648	CIF	34.1726	38.4515	37.8025
3200.9160	CIF	38.0927	40.1929	39.8873

6 结 论

H.264 扩展标准是为了满足可分级要求而对原有标准的扩展。相比于之前的视频可分级方法,H.264 扩展标准无论在性能还是灵活性上都有显著的优势。目前,H.264 扩展标准仍处于制订过程中,但其主体框架已基本确立。本文基于最新的参考软件,以具体实例对 H.264 扩展标准作了分析和介绍,同时揭示了各项可分级技术的特点。

参考文献 (References)

- 1 ITU-T and ISO/IEC JTC1. Scalable video coding-joint draft 6 [S]. JVT-S201, JVT, ITU-T/ISO, April, 2006.
- 2 ITU-T and ISO/IEC JTC1. JSVM 5.7 Software [S/OL]. cvs://garconient.rwthiaachen.de:/cvs./jvt, June 2006.
- 3 Reichel Julien, Schwarz Heiko, Wien Mathias. Joint scalable video model JSVM-6[S]. JVT-S202, JVT, ITU-T/ISO, April 2006.
- 4 Schwarz Heiko, Wien Mathias, Vieron Jerome. JSVM software manual[S/OL]. cvs://garconient.rwthiaachen.de:/cvs./jvt, May 2006.
- 5 Schwarz Heiko, Hinz Tobias, Marpe Detlev, *et al.* Constrained inter-layer prediction for single-loop decoding in spatial scalability [A]. In: Proceedings of the IEEE International Conference on Image Processing (ICIP) [C], Genoa, Italy, 2005.
- 6 Han Woo-jin. Modified intraBL design using smoothed reference [R]. JVT-R091, JVT, ITU-T/ISO, January 2006.
- 7 Wang Xiang-lin, Karczewicz Marta. CE7 report, multiple FGS layer coding for low-delay applications[R]. JVT-R077, JVT, ITU-T/ISO, January 2006.
- 8 Isabelle Amonou, Nathalie Cammas, Sylvain Kervadec, *et al.* Complexity reduction of FGS passes[R]. JVT-R069, JVT, ITU-T/ISO, January 2006.