

高清视频编码和编解码器 MEncoder 的研究

黄丽福, 陈 锋, 缪 文

(厦门大学 信息科学与技术学院 福建 厦门 361005)

【摘要】: ITU-T H.264 和 MPEG-4 Part 2(ASP)是当前最重要的两个高清视频编码标准。而 x264 和 Xvid 是这两个标准的最流行的具体实现之一。本文介绍了视频编解码中的一些基本知识,通过对 MEncoder 的选项介绍说明了影响视频编码的一些要素。最后利用 MEncoder 对 x264、Xvid 在不同压缩质量下编码的全高清视频,进行压缩率、编解码效率和压缩后视频质量等方面比较。

【关键词】: X264; H.264; Xvid; MPEG-4; MEncoder; 高清视频编码; 视频压缩

1、引言

随着网络带宽的增长、单位硬盘存储的价格下降以及对视频质量要求的提高,网络传输的视频中高清视频所占的比例越来越大。全高清视频(Full High-definition Video)是指 1080p 分辨率下的高清视频,这类视频对编解码器的要求是最高的,也是未来一段时间内的主流视频。根据 NTSC 制定的数字视频标准,对于通常 90 分钟长的全高清视频来说,要求的数据传输速率超过 1490Mbps,存储空间要求超过了 1TB。这显然远远超过了目前网络的带宽能力。所以高清视频的压缩标准显得更为重要。而经过压缩后的视频一般大小只有几十 GB(蓝光光碟容量)。视频编码的压缩要兼顾时间、压缩率和画面质量三个主要指标,所以不同的编解码器对视频的影响很大。

2、高清视频编码标准

当前流行的高清视频编码标准主要有 ITU-T H.264 [1]和 MPEG-4 Part 2 (ASP)[2],其中 H.264 也称为 MPEG-4 Part10 (AVC)。所以 H.264 和 MPEG-4 视频标准有一定的联系,但却是完全不同的标准。总体而言 H.264 标准强调实用性、高效性和可靠性,MPEG-4 视频部分则更灵活,支持近 20 个档次,每个档次都代表了一类特定的视频通信应用,而 H.264 只支持 3 个档次的的应用,第四个档次目前还在制定中。

2.1 H.264 和 MPEG-4 Part2

MPEG-4 Part2 是通常意义上的 MPEG-4 视频,由运动图像专家组(MPEG)于 1993 年开始起草,在 1999 年成为标准,标准号为 ISO/IEC 14496-2[2]。H.264 标准则是由国际电信联盟(ITU-T)的视频编码专家组(VCEG)发起的,H.264 标准制定的后期由 VCEG 和 MPEG 联合组成的专家组共同完成,所以在 2003 年被 ISO/IEC (MPEG-4 Part10)和 ITU-T(H.264 草案)同时支持即 ISO/IEC 14496-10[1]。

H.264 和 MPEG-4 Part2 只是提出了视频压缩的标准,具体由编解码器实现。Xvid [3]最初的名称为 XviD,由 openDviX 发展而来。Xvid 作为 MPEG-4 视频部分,也就是 MPEG-4 Part2 (ASP)标准的实现。x264 同 Xvid 一样也是开源的视频编解码器。不同的是 x264 实现的标准是 H.264/MPEG4 AVC。这两个开源编解码器 Xvid 和 x264,分别是当前 MPEG-4 Part2 ASP 和 H.264 标准中使用最广泛、性能最好的编解码器之一。

2.2 视频图像压缩的几个重要概念

运动补偿[4]在 MPEG-4 视频和 H.264 标准中,用来降低预测帧和当前帧之间差别的比特流。视频运动导致的图像变换一般都是以像素点为单位的移动,但是通过像素轨迹来计算复杂度太高,数量太大,所以一般采用基于宏块的移动来计算运动补偿。

宏块[4.5,6]把每个当前帧划分成许多长方形的块,这样可以有效的降低计算复杂度。从 16x16、8x8 到 4x4,越小的运动补偿分块,得到的残差图像(当前帧与预测帧之间相减的结果)精度越高[4]。但是越小的块划分,会使得搜索分块的算法复杂度上升,传输的运动矢量数目也越多,这些消耗的比特流可能会超过

降低残差图像能量的比特流。

压缩遍数[7,8,6](Pass),即扫描视频数据的次数,两次以上的扫描(n-Pass)可以更好的利用前一次扫描的统计数据,提高预测的精确度以及优化整体比特流。Xvid 支持 1-Pass 压缩和 2-Pass 压缩,而 H.264 还支持 3-Pass,通常 2-Pass 的效果已经达到最优。

量化器[4.5,7](quantizer),用来降低标量或矢量的精度,即将一定范围内的信号映射到一个较小取值范围内。标量量化器输入信号的一个样本值映射到另一个量化输出值,而矢量量化器将输入信号的一组样本值映射成一组量化输出值。量化过程是一个有损过程,有效的降低了原始数据的精度从而实现节省比特流的目的。

PSNR 值[9],衡量视频画面主要客观标准。(参考本文 4.3)

视频帧[4]。视频由一个帧或多个帧(条带)来组织连续的视频图像编码。基本的帧类型包括 I 帧,也称为内部画面,由 DCT(离散余弦变换)压缩当前帧而来。P 帧,也称为预测帧,由相邻的前一帧(I 帧或者 P 帧)和当前帧来压缩当前帧的数据,只记录不同点。B 帧,也称为双向预测帧,根据相邻的前后帧和当前帧的差值,来压缩当前帧。

P 帧可以去除已经编码的帧间的冗余信息。而 B 帧不仅考虑当前帧之前的图像,也考虑不同帧之后的编码图像,可以更大程度的压缩图像质量。可以通过在 I 帧和 P 帧之间加入不同数量的 B 帧,提升画面质量或加快压缩速度。

3、视频编解码器 MEncoder

3.1 MEncoder 介绍

当前最流行的视频压缩和处理软件核心为 FFmpeg [10]和 MEncoder 两种。两者都开源且跨平台的视频压缩处理命令行程序,而且很多 FFmpeg 的开发者同时也是 Mplayer 的开发者,所以两者联系紧密。FFmpeg 支持从视频中截取单帧图片而 MEncoder 不支持,但是在视频压缩方面,两者的功能相差不大,主要差别在于 FFmpeg 支持的视频格式没有 MEncoder 完整。本文的测试只涉及到视频压缩,所以使用 MEncoder 作为测试工具。

3.2 MEncoder 的使用

压缩单个文件:mencoder [选项] 文件 [文件|URL|-] [-o 文件 | file://文件 | smb://[用户名:密码@]主机/文件路径]。压缩多个文件:mencoder [选项] 文件 1 [分选项] [文件 2] [分选项]。Encoder 基础选项[6]包括了输出文件名称(-o)。MEncoder 压缩转换选项主要包括了输出音频(-oac)和视频编解码器(-ovc),以及输出容器的格式(-of)。

3.2 与视频编码器相关的 Mencoder 主要选项[6]

在指定了-ovc(输出视频编码器)选项后,还可以指定一些与编码器相关的选项,分别是-xvidencopts 和-x264encopts。对 Xvid 的-xvidencopts 和 X264 的-x264encopts 都有几个重要的子选项。(具体子选项中的 PSNR 和速度提升或损失[6])。

在分析完重要的编码相关的参数后,可以看下一些典型的图像质量对参数的设置:

描述	编码选项	压缩帧数	PSNR 损失
超高	chroma_opt:vhq=4:bvhq=1:quant_type=mpeg	16fps	0dB
高	vhq=2:bvhq=1:chroma_opt:quant_type=mpeg	18fps	-0.1dB
快	turbo:vhq=0	28fps	-0.69dB
实	turbo:nochroma_mc:ntrellis:max_bframe_s=0:vhq=0	38fps	-1.48dB

表 3-1 xvid 选项

描述	编码选项	压缩帧数	PSNR 损失
超高	subq=6:partitions=all:8x8dct:me=umh:frameref=5:bframes=3:b_pyramid:weight_b	6fps	0dB
高	subq=5:8x8dct:frameref=2:bframes=3:b_pyramid:weight_b	13fps	-0.89dB
快	subq=4:bframes=2:b_pyramid:weight_b	17fps	-1.48dB

表 3-2 x264 选项

4. 视频质量及其测试方法

4.1 视频质量的度量

测试方法主要包括了主观测试和客观测试两种方法。客观方式对比主观方式有很大的好处在于容易量化的方式来展示压缩的效果,但是目前缺乏非常完善的衡量参数使得客观测试的结果和主观感觉完全相同。有时压缩的效果在客观测试中表现良好,但是主观观测却难以接受。

4.2 主观测试

主观测试主要是通过类似图灵测试的方式,随机给出或直接告知压缩后的视频和原始视频的播放顺序,让受测者给定评分。但是由于压缩是有损的,而人主观的知识和经验对判断会有较大的影响,所以需要进行大范围、反复的测试才能得到一个比较客观的结果。这方面采用的标准[11]主要有 ITU-R BT.500-9、ITU-R BT.500-11、ITU-T P.800、ITU-T P.910[12]等。莫斯科国立大学图像实验室的 MSU Perceptual Video Quality tool[11],实现了上述标准中的一些测试方法 [11],包括:DSCQS,SCACJ,MSUCQE等。

4.3 客观测试[7,8]

客观测试是视频压缩和视频处理判断的主要方式。测试的主要指标有 MSE[13]、PMSE[13]、PSNR[14]、SSIM[15]等。这四种度量都是利用压缩前后图像的损失来评价视频图像的质量。PSNR 是目前衡量视频图像压缩质量的一个重要客观参考量,但是 PSNR 的缺陷和其他客观测试参考量一样,都是数值上的计算比较,而人眼的视觉会侧重画面的一部分,而不是全部,客观值不能反映这种重点区域的优劣。针对特定类型的视频,像真人特写等画面较多且要求较高的,可以结合一些图像识别上的方法,如人脸识别等对重点区域进行评价,但是目前还没有关于这方面的具体方法。SSIM 方法虽然可以在一定程度上克服传统方法的不足,但是计算较其他方法更为复杂。对高清影片而言,画面的整体压缩损失不大,影响视频压缩质量的主要是细节上表现,所以 PSNR 值作为一个全局的客观标准还是可以衡量影片的整体压缩质量。

5、MEncoder 压缩测试

测试视频采用长度为 20m42s,码率 40Mbps,MPEG-2 编码 1080p 的蓝光视频片段,主要为飞行战斗画面和人物叙事画面,可以有效检测视频编码在压缩动态和静态画面时的优劣。采用表 3-1、表 3-2 中提及的 Xvid 和 x264 选项对视频进行压缩,设定压缩方式为 2-pass,视频码率 4000Kbps,音频为拷贝,容器为 avi。下面是测试结果及分析:

原始视频大小为 4.3GB,压缩后的视频大小见表 5-1。由于

为了测试高清视频的压缩效率,设定的码率比较高,所以 x264 的 fast 选项压缩和 Xvid 的 realtime 选项压缩后的视频大小同其他两个选项没有太大差别。而 Xvid 和 x264 之间的差别较大,因为 x264 比 Xvid 在运动补偿、熵编码等运算中采用的宏块更小、编码方式更复杂,从而提高了 x264 的压缩率。这种高压率率是以解码时 CPU 的占用率提高为代价的,在 Intel 酷睿 2 E4500 2.2GHz 的机器上播放压缩后的最高画质的视频,x264 占用率为 39%~45%,而 Xvid 为 20%~28%。

编码和编码方式	Fast / Realtime	High	Very High
X264	687MB	688MB	688MB
Xvid	2.1GB	2.3GB	2.3GB

表 5-1 压缩后视频文件大小

视频的 PSNR 值是衡量视频质量的客观评价方式中最重要的一种,本文利用 pnmppsnr 和 Mplayer 结合计算指定帧的 PSNR 值。图 5-1 给了 500 帧两种视频编码在不同的压缩方式下压缩后计算得到的 PSNR 值。可以看到对同一帧画面,Xvid 和 x264 的处理都比较接近,这是因为对高清视频图像而言,视频的 PSNR 值损失较小,造成视频压缩损失的主要因素除了视频编码外,与视频画面静态或者动态有关,较大变化的连续帧有可能在压缩中损失更多的 PSNR 值。从图 5-2 可以看到 Xvid 比 x264 的平均 PSNR 值要低 0.6db 左右,也就是说 x264 压缩后的画面质量较 Xvid 高,但这个幅度差别不是很大。Xvid 比 x264 整体画质较低是由于 x264 在压缩时采用了精度更高的宏块大小,虽然提高了压缩算法的复杂度,但对压缩后的视频帧质量效果更好。不同的视频压缩方式对画面影响不大,甚至因为高清视频对码率要求较高的情况下,Xvid 的 realtime 反而效果比其他两种压缩方式要好一些。

图 5-1 不同编码压缩方式的 PSNR 值

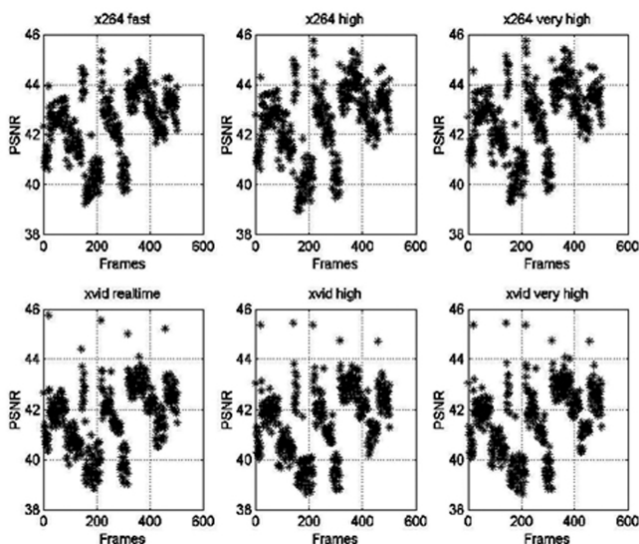
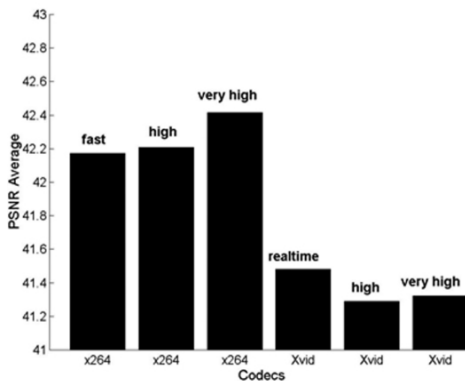


图 5-2 不同编码压缩方式的平均 PSNR 值



(下转第 38 页)

破,核心网设备需同时支持两种协议,大量现网设备需要改造,实现难度大,近期内不可行。

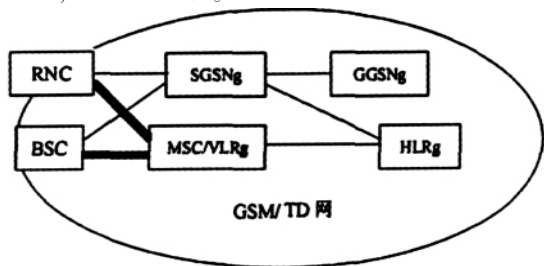


图 2 TD、GSM 混合组网

方案	独立组网	混合组网
优点	网络改造工程量小,实现简单。	共用核心网设备,双网负荷可自动平衡。
缺点	双网负荷很难平衡。	需设备支持,实现难度大,目前不可行。

表 2 两种组网方案对比

2.3 业务策略

方案一:TD 优先方式

只要有 TD-SCDMA 网络覆盖,支持 TD 系统的 UE 全部附着在 TD-SCDMA 小区,所有业务在 TD 网络上进行。

这种方案适合 TD 建网初期,可使用户尽量感受到 TD-SCDMA 网络的覆盖。这样可以减少 UE 附着在 GSM 网络的机会,使在双网并存的区域视频电话总能顺利建立与使用。

方案二:按用户习惯

根据用户以往使用两种网络的频率大小的策略:该策略主要是根据用户以往使用 GSM 和 TD-SCDMA 网络的频率来决定接入的网络。若用户以往登陆使用 GSM 网络的次数远多于使用

TD-SCDMA 网络的次数,UE 自动接入到 GSM 网络中信号最强的小区,否则,UE 将接入到 TD-SCDMA 网络中信号最强的小区。该策略考虑了用户习惯和用户类型,具有一定的智能性,因为高端用户往往驻留在 TD-SCDMA 网络上,而低端用户则使用 GSM 网络更多一些。

这种方案适合双模手动切换手机较多的网络。

方案三:按业务分类

根据用户发起的业务选择驻留网络,CS 话音业务全部选择 GSM 网络,即在 GSM 系统发起的话音留在 GSM 系统中,直到电话在 GSM 网络中终止。TD 系统发起的话音业务尽快切换至 GSM 网络。在 GSM 网络服务小区不发生拥塞的情况下,系统将禁止切换到 TD-SCDMA 网络,即使 TD-SCDMA 网络的质量较高。同时 GSM 网络承载全部 WCDMA 漫游用户。

CS 数据业务和 PS 业务选择 TD 网络,即使 GPRS 优先附着在 TD 网络,CS 数据业务优先使用 TD 网络。

这种方案需要双模自动切换手机支持。

方案	TD 优先方式	按用户习惯	按业务分类
优点	适合 TD 建网初期快速发展用户。	适合双模手动切换终端,参数设置简单。	充分发挥 TD 网络在数据业务上的优势。
缺点	TD 网络负荷增长可能过快,影响网络安全。	完全按照用户习惯,驻留 GSM 网络时,某些业务受限,造成双网负荷不均衡。	需要终端支持,需要网络设备支持,可能需要网络改造,参数设置难度大,系统处理负荷大。

表 3 三种业务方案对比

三、小结

TD-SCDMA 系统的网络建设必然是阶段式渐进的,因而 TD-SCDMA/GSM 双模混合组网是否成功也直接决定了 TD 系统的发展。混合组网是 TD 一个重要的课题,还需要我们更多的关注。

(上接第 36 页)

6、结论

本文介绍了高清视频压缩的一些概念并讨论了当前流行的两种高清视频编码 Xvid 和 x264。通过对 MEncoder 使用方法和具体选项的分析,比较两种视频编码中重要压缩方法和影响因素。还对视频质量的主观测试方法和客观测试方法做了讨论,给出了几种流行的视频评价方式。最后对两种编码六种不同的压缩选项进行了测试,并利用视频帧的 PSNR 值对两种视频编码压缩质量做了比较。由于 Xvid 的视频比较发布较早,所以较 x264 而言存在不小的差距,特别是视频压缩率上 x264 是 Xvid 的三倍,但是 CPU 占用率比较有优势。由于高清视频要求的视频帧率比较高,所以不同的压缩方式对视频帧的画质影响不是非常明显。

参考文献:

[1] ISO/IEC 14496-10 and ITU-T Rec. H.264, Advanced Video Coding [S], 2003

[2] ISO/IEC 14496-2, Coding of Audio-Visual Objects Part 2:Visual [S], 2001

[3] 王清亮,李中福,刘玉珊,郑黎, Xvid 视频编码技术 [J], 河南职业技术学院学报,2004,第 32 卷第 1 期

[4] Iain E.G.Richardson. H.264 and MPEG-4 Video Compression [M]. England:Wiley, 2003:[9-25,85-133,159-195].

[5] 刘峰. 视频图像编码技术及国际标准 [M]. 北京:北京邮电大学出版社, 2005

[6] MPlayer team, Encoding with MEncoder[S], (2009) [2009.6]

<http://www.mplayerhq.hu/DOCS/HTML/en/index.html>

[7] Dmitriy Vatolinx, 264 Codec Strong and Weak Points [S], (2008.11) [2009.5], <http://compression.ru/video/cod>
ec_comparison/pdf/x264_options_analysis_08.pdf

[8] Dmitriy Vatolinx, MPEG-4 AVC/H.264 Video Codecs Comparison [S], (2008.11),[2009.5], <http://compression.ru/vid>
eo/codec_comparison/pdf/x264_options_analysis_08.pdf

[9] Thomos, N., Boulgouris, N. V., Strintzis, M. G.,Optimized Transmission of JPEG2000 Streams Over Wireless Channels, IEEE Transactions on Image Processing[C], 2006.1,[15-16].

[10] FFmpeg, FFmpeg Documentation [S], (2009) [2009.6] <http://ffmpeg.org/ffmpeg-doc.html#SEC1>

[11] Dmitriy Vatolinx, MSU Subjective Comparison of Modern Video Codecs[S],(2006.1), [2009.6], <http://compression.ru/video/> codec_compar-
ison/subjective_codecs_comparison_en.html

[12] ITU,ITU-T P.910[S].(2008.4) [2009.6]. <http://www.videoc>
rity.com/PDF/T-REC-P.910-199909-!!! PDF-E%5B1%5D.pdf.

[13] George Casella, E.L. Lehmann, Theory of Point Estimation [M], Springer, 1999

[14] Xiangjun, L., Jianfei, ROBUST TRANSMISSION OF JPEG2000 ENCODED IMAGES OVER PACKET LOSS CHANNELS [C],ICME 2007,[947-950]

[15] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity [C], IEEE Transactions on Image Processing, 2004.4, vol.13, no.4, [600-612]