

基于 H.264/AVC SP/SI 帧技术的 VCR 功能实现

吴业清, 王 嘉, 朱莹莹

(上海交通大学图像通信与信息处理研究所, 上海 200240)

摘要: 基于 H.264/AVC 的 SP/SI 帧能参照不同参考帧重构出相同图像帧的技术和多参考帧技术, 提出一种实现 H.264/AVC 视频 VCR 操作的方法。该方法能够消除双向码流中的漂移问题。实验结果显示, 与传统方法相比, 在相同播放速率下, 使用该方法可以降低约 40% 的网络带宽, 峰值信噪比的提高超过 1 dB。

关键词: H.264/AVC 标准; 数字 VCR; SP/SI 帧

Realization of VCR Functions Based on SP/SI Frame Technology of H.264/AVC

WU Ye-qing, WANG Jia, ZHU Ying-ying

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

【Abstract】 The main feature of SP/SI frame of H.264/AVC frames is that identical frames can be reconstructed even when different reference frames are used for prediction. This paper proposes a scheme to implement Video Cassette Recording(VCR) functions of H.264/AVC video by utilizing the SP/SI frame and multi-reference frame. It can eliminate the drifting problem of dual-bitstream technique. Experimental results show that the scheme can reduce the network bandwidth by about 40% under VCR operation compared with traditional one, and PSNR improves over 1 dB.

【Key words】 H.264/AVC standard; digital Video Cassette Recording(VCR); SP/SI frame

1 概述

在线多媒体内容的不断增多越来越要求能够支持对多媒体视频流有效快速的浏览。而实现用户对多媒体内容快速友好浏览的一个关键技术就是提供 VCR(Video Cassette Recording)功能。VCR 功能主要包括播放、倒播、停止(回到起始)、暂停、步进、步退、快进、快退和随机访问等。这些功能有助于用户更好地浏览多媒体内容, 并可以用于其他应用(如视频媒体的编辑)。

实现 VCR 操作面临着很多困难。最近一些相关研究实现了 MPEG 压缩视频流应用中的 VCR 功能。文献[1]提出了一种方法, 将输入的 I-P-B 比特流分为 I-P 帧和 B 帧 2 个部分, 然后用转码器将 I-P 帧转为反向帧顺序结构的 I-P 比特流。采用文献[2]的方法, 以原始 I-P 比特流前向运动矢量来估计新生成的 I-P 比特流的反向运动矢量, 可以降低转码过程的计算复杂度。对于 B 帧, 使用文献[3]建议的倒播运动矢量交换方案。然而这种转码过程要求很高的计算精度, 而且使用文献[2]的运动矢量近似方法也会导致漂移问题。可见上述方法都没有很好地解决由于实现 VCR 功能而引起的额外网络负载和计算复杂度问题。

文献[4]采用双向比特流技术实现 MPEG 视频流的 VCR 功能。这种方法的思路是在服务器端事先对原始视频序列进行反向编码以产生一个反向视频流, 解决视频播放中的倒播问题。虽然双向比特流方法可以很好地降低网络负载和解码复杂度, 但存在 3 个问题: (1)因为在执行 VCR 命令时反向(正向)的 P 帧只是用前向(反向)的 I 帧来近似, 所以会引起漂移问题, 虽然漂移问题可以通过额外的漂移比特流来补偿, 但会增加复杂度。(2)网络带宽和解码复杂度可进一步减小, 但仍传输了不少冗余帧。(3)这种方法在服务器端的存储容量为

正常情况下的 2 倍多。

文献[5]提出了在双向码流中使用 SP/SI 帧来实现 VCR 功能, 前向码流中的 SP/SI 帧为主 S 帧, 反向码流中的 SP/SI 帧为辅 S 帧, 在前向码流和反向码流 S 帧处形成 SI_{FB}/SP_{RB} 对和 SP_{FB}/SI_{RB} 对。使用这种方法可以消除一般双向码流的漂移问题, 但仍需要存储 2 个码流, 而且在进行 VCR 操作时, 需要在 2 个码流间不停切换, 操作比较复杂。根据 SP/SI 帧能参照不同参考帧重构出相同图像帧的性质, 本文采用 H.264/AVC 的 SP/SI 帧和多参考帧技术来实现 VCR 功能。

2 基于 SP/SI 帧的 VCR 实现方案

2.1 H.264/AVC SP/SI 帧技术

H.264/AVC 为了顺应视频流的带宽自适应性和抗误码性能的要求, 定义了 2 种新的帧类型: SP 帧和 SI 帧, SP 帧和 SI 帧技术参见文献[6-7]。SP 帧通过参照不同的参考帧重构出相同的图像帧。充分利用这个特性, SP 帧可以取代 I 帧广泛应用于流间切换、拼接、随机接入、快进快退以及错误恢复等, 同时大大降低码率的开销。与 SP 帧相对应, SI 帧则是基于帧内预测的编码技术。尽管 SP 帧的编码效率略低于 P 帧, 但远高于 I 帧。除了主 SP 帧外, 还有一个相应的辅 SP 帧, 它和主 SP 帧有着相同的重构值。

在此举一个码流切换的例子来说明主 SP 帧和辅 SP 帧的工作原理。如图 1 所示, 码流 A 和码流 B 分别是同一视频序列采用了不同信源编码参数编码得到的 2 个视频流。当正常传送时, 根据当前网络的有效带宽, 将比特流 A 或者比特流 B 传

作者简介: 吴业清(1980-), 男, 硕士, 主研方向: 视频压缩编解码, 视频传输; 王 嘉, 副教授; 朱莹莹, 硕士

收稿日期: 2007-11-01 **E-mail:** alban@sjtu.edu.cn

给用户。在时刻 n 需要从比特流A转换到比特流B时，发送辅 $SP_{12,n}$ 帧代替主 $SP_{2,n}$ 帧。当解码 $SP_{12,n}$ 时，解码器在时刻 n 能够获得和解码 $SP_{2,n}$ 一样的参考值，因此，能够在 $n+1$ 时刻无缝解码比特流B。因为辅 $SP_{12,n}$ 帧和主 $SP_{2,n}$ 帧能够得到相同的重构值，所以在流间切换时，编码器和解码器之间不会引入不匹配。

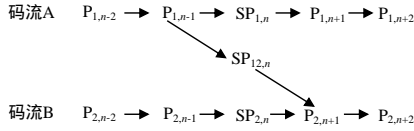


图1 SP帧的码流切换

使用SP/SI帧会有性能损失，在信源编码部分插入SP/SI帧会增加编码复杂度。而且SP/SI帧的编码过程增加了额外的量化过程，因此，SP帧的编码效率不会超过P帧的编码效率。

2.2 基于SP/SI帧的VCR功能实现原理

使用SP/SI帧编码技术实现VCR功能的原理如图2所示，为了说明方便，本文不使用B帧。

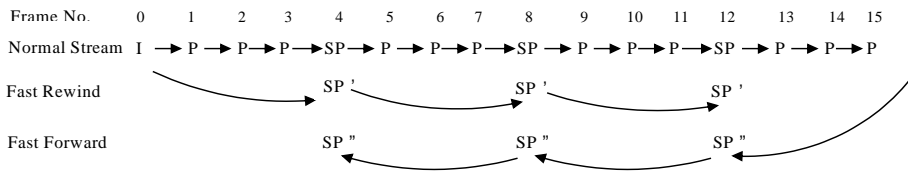


图2 本文方案的编码结构

利用H.264/AVC SP/SI技术对原始视频序列进行编码，生成包含主SP帧的常规流，这部分码流在正常播放时传输。为了实现快进、快退、倒播和随机访问功能，根据主SP帧生成相应的辅SP'帧(用于快进)和辅SP''帧(用于倒播和快退)，辅SP'帧以其最左边的SP'帧或I帧作为参考帧和对应的主SP帧生成，在进行快进操作时才需要传输这部分码流以代替传输主SP帧。同样，辅SP''帧是根据将其最右边的SP''帧或I帧作为参考帧和对应的主SP帧生成，在进行快退和倒播时才需要传输这部分码流以代替传输主SP帧。码流之间、帧与帧间的参考特性参见图2。

现通过2个例子来说明利用SP/SI帧实现VCR功能的原理。假设从帧 P_2 开始要以4倍速快进，到帧 P_{13} 后以正常速度播放，利用本文方法只须传输 $P_2, P_3, SP'_4, SP'_8, SP'_{12}, P_{13}$ 共6帧。而采用传统方法则须传输 $P_2 \sim P_{13}$ 间所有的12帧，采用本文方法所须传送的帧明显减少，当播放速度更快时，优势更加明显。

考虑另一个例子，当以正常速度播放到 P_{15} 时要求倒播。因为当解码和显示帧 P_{15} 后，它的参考帧 P_{14} 就被丢弃了，如果按照传统方法解码显示帧 P_{14} ，先需要解码帧 P_{14} 之前的所有参考帧直到之前最近的I帧，即需要传送帧 P_{14} 之前最近的I帧和帧 P_{14} 之间的所有帧，在图2中需要传送 $I_0 \sim P_{14}$ 间共15帧，而采用本文方法只须传送 $I_{16}, SP''_{12}, P_{13}, P_{14}$ 共4帧，效果改善明显。

如果快进和快退的速率等于16时，在图2中只须传送I帧。当速率大于16，可以丢弃某些I帧来实现快进和快退。

2.3 SP/SP'/SP''的编码原理

主SP帧的编解码原理与P帧相似，本文不再详细介绍，而是主要介绍辅SP帧的编解码原理。辅SP帧的编解码框图如图3所示。

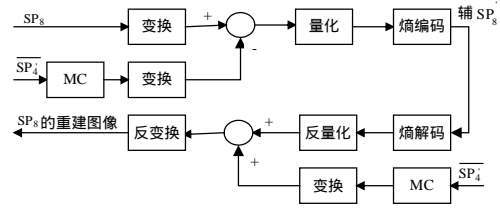


图3 辅SP帧的编解码框图

假设主 SP_8 帧的重建图像为 $I_c(x, y)$ ，它由重建系数 l_{rec} 做反量化和反变化而得。假设另一预测帧为 $\overline{SP'_4}$ (帧 SP'_4 的重建图像)，若要得到与主 SP_8 帧具有相同的重建图像 $I_c(x, y)$ 的辅 SP'_8 帧，就要找到新的预测残差系数 $l_{rec,2}$ ，使得使用 $\overline{SP'_4}$ 也能准确地重建图像 $I_c(x, y)$ 。对 $\overline{SP'_4}$ 进行变化和量化，量化后的系数记为 $l_{pred,2}$ ，则辅 SP'_8 帧的预测残差系数 $l_{err,2}$ 可由下式计算得到： $l_{err,2} = l_{rec} - l_{pred,2}$ 。然后把计算得到的 $l_{err,2}$ 进行无损的熵编码。解码端生成 $\overline{SP'_4}$ 后对其进行变换和量化，得到 $l_{pred,2}$ ，其值与编码器相同。 $l_{pred,2}$ 与接收到的预测残差系数 $l_{err,2}$ 相加，和为 l_{rec} 。对 l_{rec} 进行反量化和反变换就可得到重建图像 $I_c(x, y)$ ，其与主 SP_8 帧的重建图像完全相同，这样 P_9 就可将重建图像作为参考帧进行解码了。

在编码过程中，P帧的量化参数 Q_p 大于SP帧的量化参数 Q_{sp} ，因此，SP帧比P帧大。而且， Q_{sp} 作为辅SP帧的量化参数包含在环路中，从而影响到主SP帧的编码比特数大小。当 Q_{sp} 取值较大，即量化精度降低时，会增大帧缓存中用来作为预测参考的帧的误差，降低预测值的精度，导致主SP帧所需编码比特数增大；若 Q_{sp} 取值减小，主SP帧的编码所需比特数减小，但会使辅SP帧的编码比特数增大，因此，应该在两者之间进行折中。

3 仿真分析

实验以H.264/AVC的参考软件JM 10.1为平台，结合文献[8]所提供的SP/SI帧编解码器进行(该SP/SI帧编解码器原理参见文献[9])。该编解码由Perl语言编写，需要在Linux将其编译整合到JM 10.1中去，由此在VS.NET平台上进行编译和运行，实现SP/SI帧的编解码。本文选择的帧率为每秒30帧，使用extended profile。

表1给出了各种实现方法在前向播放和反向播放切换时平均PSNR结果，其中，F/R指Forward->Rewind；R/F指Rewind->Forward。实验结果显示，本文方法除了Football序列外，其他3个序列的PSNR都有提高。

表1 前向播放和反向播放间切换时3种方法的平均PSNR dB

序列	格式	传统双向码流		SP/SI实现的双向码流		本文方法	
		F/B	R/F	F/B	R/F	F/B	R/F
Salesman	352 × 288	33.997	34.161	34.441	34.660	35.603	35.499
Foreman	352 × 288	34.475	34.686	35.733	35.784	35.892	35.881
Football	352 × 240	33.107	33.212	33.850	33.916	32.539	32.603
Carphone	176 × 144	34.218	34.583	35.536	35.642	36.284	36.386

表2比较了以4倍速快进时本文方法和传统方法所须传送的比特率，以及以4倍速快退所需的比特率。由于反向播

放和反向快退时帧所处位置不同,用传统方法所需传送的帧数就不同,无法进行统计,因此只给出本方法4倍速快退时所需传送的比特率。

表2 4倍速快进和快退时2种方法所需的比特率 Kb/s

序列	格式	4倍快进		4倍快退
		传统方法	本文方法	本文方法
Salesman	352 × 288	1 485.16	991.57	1 073.84
Foreman	352 × 288	1 878.16	1 102.58	1 197.20
Football	352 × 240	8 734.20	3 290.09	3 463.84
Carphone	176 × 144	495.84	293.01	310.90

图4给出了同一视频序列在不同量化参数下进行4倍速快进时本文方法与传统方法所需传送比特数的比较。

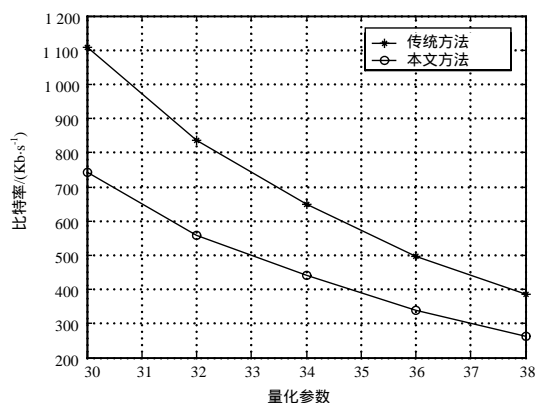


图4 4倍速快进时所需传送比特数的比较

4 结束语

本文提出了使用 SP/SI 帧和多参考帧技术来实现 H.264/AVC 的 VCR 功能。它不会像双向码流实现 VCR 功能

(上接第 232 页)



图6 改进后的 CamShift 跟踪结果

6 实验结果分析

实验数据采用羽毛球比赛视频,每帧的图像大小是 320 × 240,实验环境是 P4 CPU 2.4 GHz, 256 MB 内存。为了更好地分析算法效率,本文采用文献[5]提出的公式:查准率 = 跟踪到正确运动员区域 / 跟踪到的所有运动员区域,得到的结果如表 1 所示。

表1 实验结果

实验视频	统计帧数	传统 CamShift 的查准率/(%)	改进后 CamShift 的查准率/(%)
目标对象与背景颜色不相似	80	92	94
目标对象与背景颜色相似	80	62	86

实验结果表明,利用本文改进后的 CamShift 算法对于跟踪与背景颜色很相似的运动目标具有更强的鲁棒性。该跟踪算法具有较好的实时性,实验过程中处理速度达到 16 帧/s。

那样产生参考不匹配和漂移问题,而且无须存储前向和反向双向码流,因此,本文的方法能够明显减少所需传送的帧,是一种很有效的方法。

参考文献

- [1] Chen Ming-Syan, Kandlur D D. Downloading and Stream Conversion: Supporting Interactive Playout of Videos in a Client Station[C]//Proc. of the 2nd Int'l Conf. on Multimedia Computing and System. Washington, USA: IEEE Press, 1995.
- [2] Wee S J. Reversing Motion Vector Fields[C]//Proc. of IEEE Int'l Conf. on Image Processing. [S. l.]: IEEE Press, 1998.
- [3] Wee S J, Vasudev B. Compressed-domain Reverse Play of MPEG Video Streams[C]//Proc. of International Symposium on Voice, Video, and Data Communications. Boston, USA: [s. n.], 1998.
- [4] Lin Chiwen, Zhou Jian, You Jeongnam, et al. MPEG Video Streaming with VCR Functionality[J]. IEEE Trans. on Circuits and Systems for Video Technology, 2001, 11(3): 415-425.
- [5] Tak Piu, Chan Yui-Lam, Siu Wan-Chi. Adopting SP/SI-Frames in Dual-Bitstream Video Stream with VCR Support[C]//Proceedings of ICASSP'06. [S. l.]: IEEE Press, 2006.
- [6] Sun Xiaoyan, Li Shipeng, Wu Feng, et al. The Improved SP Frame Coding Technique for the JVT Standard[C]//Proc. of International Conference on Image Processing. Barcelona, Spain: [s. n.], 2003.
- [7] 毕厚杰. 新一代视频压缩编码标准-H.264/AVC[M]. 北京: 人民邮电出版社, 2005.
- [8] H.264 SP Frame Codec[Z]. (2005-09-20). <http://www.stanford.edu/~esetton/H2642.htm>.
- [9] Setton E, Girod B. Video Streaming with SP and SI Frames[C]//Proceedings of VCIP'05. Beijing, China: [s. n.], 2005.

7 结束语

运动员的检测和跟踪是对体育视频进行分析和处理的关键步骤。本文采用的自适应高斯混合模型能够较好地处理动态背景中光照的缓慢变化,但是当场景中光照发生突变时,比如突然打开或关闭照明灯,用该模型检测出的前景中会出现大量的“伪前景”,即光照发生突变的背景区域。在后续的工作中将进一步研究场景中光照突变的检测和“伪前景”的消除方法。

参考文献

- [1] Stauffer C, Grimson W E L. Adaptive Background Mixture Models for Real-time Tracking[C]//Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Los Alamitos, USA: IEEE Computer Society Press, 1999.
- [2] Bradski G R, Clara S. Computer Vision Face Tracking for Use in a Perceptual User Interface[J]. Intelligence Technology Journal, 1998, (2): 1-15.
- [3] 刘扬, 黄庆明, 高文. 自适应高斯混合模型球场地检测算法及其在体育视频分析中的应用[J]. 计算机研究与发展, 2006, 43(7): 1207-1215.
- [4] Comaniciu D, Meer P. Robust Analysis of Feature Spaces: Color Image Segmentation[C]//Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 1997.
- [5] Yilmaz A, Javed O, Shah M. Object Tracking: A Survey[J]. ACM Computing Surveys, 2006, 38(4): 1-45.

