

一种改进的多视点视频编码预测结构

韩宝英¹, 郑江滨¹, 钟 升^{1,2}

(1 西北工业大学 计算机学院, 陕西 西安 710129; 2 西安微电子技术研究所, 陕西 西安 710054)

摘 要: 多视点视频编码(Multi-view Video Coding, MVC)标准采用预测关系固定的编码结构,不能根据视频序列调整预测关系,导致无法获得更好的编码效率。本文针对这一缺陷提出一种确定预测结构的方法,首先对图像进行下采样获得粗粒度图像,然后利用粗粒度图像间的灰度方差获得灰度相近的视点,最后对所选取的灰度相近的视点运用全局视差确定 I-视点。经实验验证,本文的方法可以提高编码效率并减少编码中的缓存容量。

关键词: 多视点视频编码; 预测结构; 灰度均值; 灰度方差

中图分类号: TP919.3

文献标识码: A

文章编号: 1000—7180(2012)05—0062—05

An Improved Multi-view Video Coding Prediction Structure

HAN Bao-ying¹, ZHENG Jiang-bin¹, ZHONG Sheng^{1,2}

(1 Department of Computer, Northwestern Polytechnical University, Xi'an 710129, China;

2 Xi'an Microelectronics Technology Institute, Xi'an 710054, China)

Abstract: An improved Multi-view Video Coding prediction structure is proposed to adapt to different videos. Firstly, coarse-grained picture was got with down-sampling; then, the views with similar gray using Gray-scale Variance are fought; finally, I-view is chosen by calculating global parallax between each of the views with similar gray. Experiments results show that the proposed method can improve coding efficiency and decrease cache for coding.

Key words: Multi-view Video Coding; prediction structure; mean gray; gray-scale variance

1 引言

传统的二维视频缺少场景的三维深度信息而缺乏立体感。多视点视频是利用空间位置不同的多路摄像机同步拍摄同一场景得到的一组视频序列,增加了场景的深度信息,使得场景具有真实性。因此,多视点视频在任意视点视频、3D 电视等下一代多媒体业务中存在广阔前景。

多视点视频由多路视频构成,视频数据量随视点数目增加而增大。巨大的视频数据量给多视点视频的存储和传输带来极大挑战。ITU-T 和 MPEG 组成联合视频组(Joint Video Team, JVT)提出多视点视频编码(Multi-view Video Coding, MVC),MVC 的主要任务是在不引起视频序列明显失真情

况下,尽可能去除视点间冗余和时间冗余,以减小视频序列数据量。MVC 研究内容包括预测结构、预测工具和高层语法 3 个方面。其中,预测结构主要是设计视频序列中视点间和图像间的预测关系,合理的预测结构能够很好的处理预测关系,提高编码效率。

当前主要的预测结构有:文献[1]提出的最小生成树结构,它将待编码图像看成二维图结构,并用图的最小生成树作为预测结构。文献[2]提出的 GoGOP 结构是二维视频编码图像组(Group Of Picture, GOP)结构的扩展,它将视频序列分成不同的图像组,图像组的组内和组外分别采用不同的预测方式进行编码。文献[3]提出的二维 Hierarchical B 结构^[3]以优良的性能及较好的适应性,被 JVT 采纳进参考模型 JMVC^[4]。

收稿日期: 2011—09—20; **修回日期:** 2011—11—04

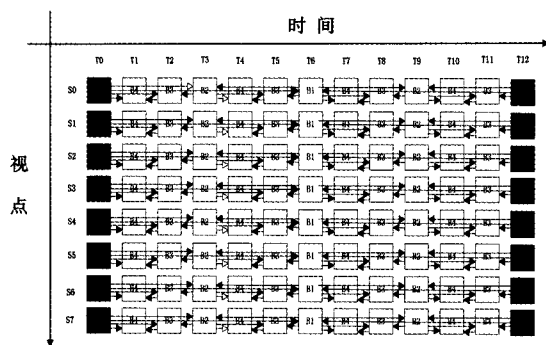
基金项目: 国家自然科学基金项目(60970069); 西北工业大学研究生创业种子基金(Z2001134)

本文在 JMVC 基础上综合考虑编码效率与计算复杂度因素,给出一种根据视频序列确定 I-视点位置的方法。

2 JMVC 的几种预测结构

2.1 Simulcast 预测结构

JMVC 最简单的预测结构是每个视点的时域预测结构相同,不同视点之间不相互参考预测。图 1 中 $T_0 \cdots T_{12}$ 表示时间序列, $S_0 \cdots S_7$ 表示各个视点,可见图中 $S_0 \cdots S_7$ 各视点间独立进行编/解码,各个视点的预测只有时间序列 $T_0 \cdots T_{12}$ 方向上 I0 图像、 $B_1 \cdots B_4$ 各图像间的运动补偿预测。Simulcast 结构实现简单,视点之间的并行性和随机接入性较好,但由于此结构没有利用视点之间相关性,编码效率不高。



$T_0 \cdots T_{12}$ 为时间序列, $S_1 \cdots S_7$ 为各个视点; 黑色 I0 为各个视点中的 I 帧, 白色 $B_1 \cdots B_4$ 为 B 帧, 下标为所在分层

图 1 Simulcast 结构

2.2 Hierarchical B 结构

2.2.1 一维 Hierarchical B 结构

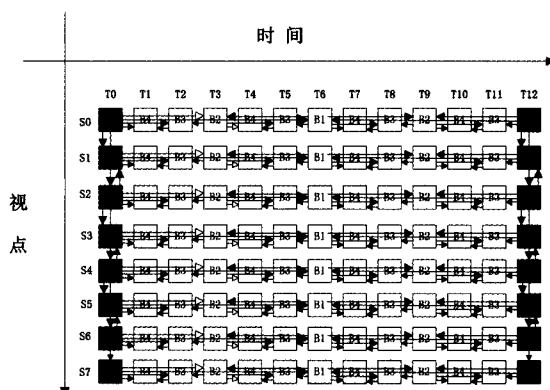
一维 Hierarchical B 结构的视频图像组 (Group Of Picture, GOP) 由关键图像 (Anchor Picture) 和非关键图像 (Non_anchor Picture) 像组成。关键图像如果是 P 图像, 只能参考上一个 GOP 的关键图像, 如图 1 各视点中的 I0 图像。非关键图像分布在关键图像之间, 并和当前的关键图像构成一个 GOP。一个 GOP 中除了关键图像都是非关键图像, 非关键图像都为 B 图像, 采用双向预测, 如图 1 各视点中 $B_1 \cdots B_4$ 各图像。

Hierarchical B 结构增加了 B 图像数目, 编码图像和参考图像相关性得到加强, 编码效率较 IBPP 和 IPPP 有较大提高。

2.2.2 二维 Hierarchical B 结构

视点/时间二维 Hierarchical B 结构是一维 Hierarchical B 结构的扩展, 视点内的预测结构都是

GOP 长度为 12 的一维 Hierarchical B 结构; 视点间采用 IBPBP 结构去除视频序列的视点间冗余。如图 2 所示 S_0 是 I-视点, 该视点不参考其他视点进行编码。其他视点分成 B 视点和 P 视点, B 视点参考 2 个视点, P 视点只参考一个视点, 图中 S_1, S_3, S_5 各视点为 B 视点, S_2, S_4, S_6, S_7 各视点为 P 视点。



$T_0 \cdots T_{12}$ 为时间序列, $S_1 \cdots S_7$ 为各个视点; 黑色 I0, P, B1 视点间预测中的 I 帧, P 帧; 白色 $B_1 \cdots B_4$ 为各视点中 B 帧, 下标为所在分层

图 2 二维 Hierarchical B 结构

Hierarchical B 结构与 Simulcast 结构相比, 增加了视点间预测, 充分利用视点间相关性, 因此可以得到更大的编码增益。

2.3 自适应结构

Simulcast 结构没有视点间预测, 导致编码效率不高; 二维 Hierarchical B 结构引入视点间预测, 但视点间参考关系固定, 无法进一步提高编码效率, 最近有学者开始对视点间的预测结构设计进行优化研究^[5-6], 其中, 针对不同视频序列自适应调整预测结构是一个重要研究方面, 这是由于多视点视频序列在拍摄过程中, 摄像机的空间位置关系确定, 视频序列中视点间相关性确定, 根据视频特性合理调整各个视点间的参考关系, 可以提高视差估计的准确性, 进而提高编码效率。Park 等人通过计算所有视点间的全局视差确定 I-视点的位置^[7], 但当视点较多时此方法计算量巨大。针对这一缺陷, 本文提出一种简化计算的 I-视点自适应预测结构。

3 改进的预测结构

预测结构包括视点间的预测和 GOP 内图像间的预测两个方面, 视点间的预测中确定 I-视点的位置具有重要意义, 本文在下采样的基础上利用图像间灰度方差选出候选视点, 再通过候选视点间的全局视差确定 I-视点, 避免对所有视点计算全局视差

减少冗余计算;时域方向上,GOP 内 B1 层中图像数目影响编解码所需缓存容量,可通过 GOP 内 B1 层中图像数目控制编解码缓存的容量。

3.1 视点间的预测

3.1.1 选择候选视点

(1) 视点图像下采样

为简化计算对原始图像进行下采样,由于图像具有灰度连续性,因此下采样并不会严重改变图像间的灰度相关性.图像连续 Q 采样度个像素的均值作为新图像的像素值,如式(1)所示可得到图像的粗粒度图像.这里将采样度 Q 设置为 2 或 4.

$$I(i,j) = \frac{1}{Q \times Q} \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} |P(i,j)| \quad (1)$$

式中 $I(i,j)$ 和 $P(i,j)$ 代表粗粒度图像 I 与原始图像 P 在点 (i,j) 处的像素值。

(2) 图像相关性

图像间灰度值方差体现两幅图像相关程度,通过计算视点与其它视点的平均灰度值方差,选取 N 个视点中 $N/2$ 个视点作为 I -视点候选视点。

$$D(I_u, I_v) = E[(I_u - I_v)^2] - [E(I_u - I_v)]^2 \quad (2)$$

$$DAG(v) = \frac{1}{N} \sum_{i=0}^{N-1} |D(I_v, I_i)| \quad (3)$$

式(2)中, $D(I_u, I_v)$ 为视点 u 中图像 I_u 与视点 v 中图像 I_v 的灰度值方差,反映图像 I_u 与 I_v 的相关性程度.式(3)得出视点 v 与其他视点的平均灰度值方差,如表 1 所示.选取视点 1、视点 2、视点 3、视点 4 作为 I -视点候选视点。

表 1 下采样度 4 和 16 时

视点	$q=4$	$q=16$
0	1674	1109
1	1648	1005
2	1628	983
3	1680	1035
4	1648	1081
5	1919	1228
6	1963	1270
7	2202	1442

3.1.2 确定 I -视点位置

灰度值方差不能精确反映图像间相关性,为准确定 I -视点,计算候选视点图像间全局视差,与其

他视点的平均全局视差值最小的视点为 I -视点。

$$MAD(e) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |I_u(i,j) - I_v(i,j)| \quad (4)$$

$$G(u,v) = \min_{0 \leq e \leq m} \{MAD(e)\} \quad (5)$$

$$MAGD(v) = \frac{1}{N} \sum_{n=0}^{N/2-1} |G(v,n)| \quad (6)$$

$$MAGD_I = \min_{0 \leq v \leq N-1} \{MAGD(v)\} \quad (7)$$

式(4)中 $MAD(e)$ 表示水平偏移 e 像素后,视点 u 与视点 v 中每个像素的绝对误差.式(5)最小化 $MAD(e)$,得到视点 u 和视点 v 的全局视差用 $G(u,v)$ 表示.式(6)中, $MAGD(v)$ 表示视点 v 与其他视点的平均全局视差.根据式(7)与其他视点平均全局视差最小的视点为 I -视点,如表 2 所示,视点 3 被选为 I -视点。

表 2 候选视点 $MAGD(v)$ 值

视点	$q=4$	$q=16$
1	42.3	29.2
2	37.4	20.8
3	32.5	19.9
4	40.9	28.5

3.2 时域图像间的预测

MVC 参考标准在时域上的预测采用一维 Hierarchical B 结构.如图 3 所示 GOP 内图像分为多个层次, B 的下标 1、2、3 代表该 GOP 相应的时域分层.增加 B1 层图像的数目可减小编码图像与参考图像间的距离,进而可减小所需缓存容量。

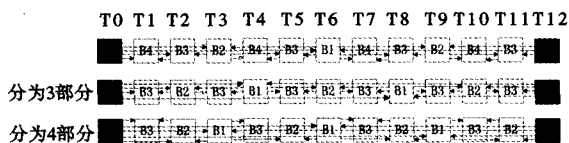


图 3 GOP 内 B1 层图像数目不同示意图

4 实验结果

JVT 对 MVC 算法性能确定了 15 项基本要求,其中包括:编码效率、计算复杂度、所需缓存区容量等^[8-9].本文采用 JMVC 8.3.1 作为实验平台,对提出的方法进行验证,测试序列采用 MERL 提供的 ballroom,测试环境所用参数见表 3,实验结果采用客观衡量标准-峰值信噪比(PNSR):

$$\text{PSNR}_Y[\text{dB}] = 10 \cdot \log_{10} \frac{255^2}{\text{MSE}} \tag{8}$$

$$\text{MSE} = \frac{1}{N_1 N_2} \sum_{n \in R} [I(n) - I(n+d)]^2 \tag{9}$$

其中 R 指 $N_1 \times N_2$ 块大小, $I(n)$ 为当前图像位于 n 点的像素值, $I(n+r)$ 为参考图像在 $n+r$ 点处的像素值. 测试参数如表 3 所示.

表 3 测试参数

参数名	数值
帧率	25
GOP 大小	12
搜索范围	64 pixel
BasisQP	32 37 42
搜索方式	fastsearch
编码帧数	13
参考帧数	2

视点位置不同导致编码效率不同,如图 4 所示. ballroom 视频序列中将视点 3 作为 I-视点具有较好的率失真性能,这与前文所得结果相符. 可见,本文方法根据视频序列特性,自适应选择 I-视点的位置,编码效率比固定预测结构要好,实验结果表明平均效率可提高 0.08~0.36 dB.

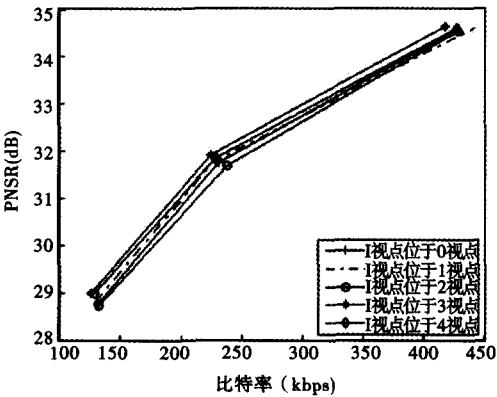


图 4 比特率-率失真曲线所需缓存容量

在有 N 个视点的视频序列中, Park 的方法共需计算 C_N^2 次全局误差, 当视点较多时, 这种选取方法计算量过大. 以 ballroom 视频序列, 需要计算 $C_8^2=28$ 个视点的全局误差, 本文提出的方法只需计算 $C_4^2=6$ 个视点的全局误差, 大大简化了计算复杂度.

解码所需缓存容量是衡量编码结果的一个重要标准. 表 4 所示为采用 Hierarchical B 结构解码一

个 GOP 各图像, 分别所需的缓存容量. 图 5 为时域上 GOP 内 B1 层图像数为 2 帧、3 帧、4 帧的率失真曲线. 由表 4 和图 5 可见, 增加 B1 层的图像数目解码所需参考帧数减少, 且率失真差别不明显.

表 4 解码各帧所需参考帧数

帧号	$N=2$	$N=3$	$N=4$
T0	0	0	0
T1	5	4	4
T2	4	3	3
T3	3	4	2
T4	5	2	5
T5	4	5	4
T6	2	4	2
T7	5	5	5
T8	4	2	4
T9	3	4	2
T10	5	3	4
T11	4	4	3
总数	44	40	38

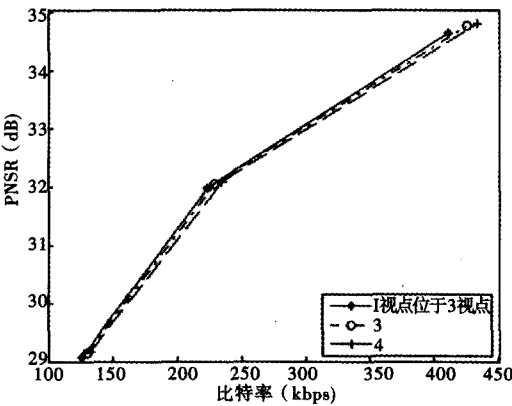


图 5 比特率-率失真曲线

5 结束语

预测结构是决定压缩效率、可伸缩性、随机访问、平行处理、所需缓存等多项性能的关键因素. 本文综合考虑编码效率、计算复杂度和所需缓存等因素, 在 MVC 参考标准预测结构的基础上提出了一种新的预测结构的方法, 此方法可以根据不同视频序列自适应调整 I-视点的位置, 进而改变预测结构, 在简化计算复杂度的同时提高了编码效率, 降低了解码所需的缓存容量.

参考文献:

- [1] Chen Chiaping, Chen Chusong. Image set compression through minimal-cost prediction structures[J]. Image Processing, 2004(2): 1289-1292
- [2] Imata H K. Multi-view video coding using reference picture selection for free-viewpoint video communication[J]. Picture Coding Symposium, 2004: 15-17.
- [3] Merkle P, Muller K, Smolic A, et al. Efficient compression of multi-view video exploiting inter-view dependencies based on H. 264/MPEG4-AVC[C]// IEEE international Conference on Multimedia and Expo. Toronto, 2006: 1717-1720.
- [4] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG. Joint multi-view video model [R], Geneva, 2009.
- [5] Huaiyi Pan. Development of multi-view video coding using hierarchical B pictures[J]. Congress on Image and Signal Processing, 2008(1): 497-503.
- [6] Liu Hai. Inter-view scalability and prediction multi-view video coding scheme[C]// International Forum on Information Technology and Applications, 2010(1): 129-133.
- [7] Park P K, Oh K J, Ho Y S. Efficient view-temporal prediction structures for multi-view video coding[J]. Electronics Letters, 2008, 44(2): 102-103.
- [8] 韦耿, 吴乐南, 朱斌. 移动环境视频编码功率-码率-失真模型研究[J]. 微电子学与计算机, 2011, 28(3): 13-16.
- [9] 纪涛, 张太镒. 分布式视频编码中一种基于混合模型的边信息生成算法[J]. 微电子学与计算机, 2010, 27(9): 22-25.

作者简介:

韩宝英 男, (1985—), 硕士研究生. 研究方向为多视点视频编码.

郑江滨 男, (1971—), 博士, 教授, 博士生导师. 研究方向为计算机视觉、模式识别、图像处理、多媒体嵌入式处理系统.

(上接第 61 页)

6 结束语

本文提出了一种基于 RTP 实时传输协议和前向纠错机制的方法, 对实时视频流进行传输. RTP 协议和 FEC 机制的结合, 能够较好的解决实时视频无线传输中的丢包问题, 提高了系统的服务质量. 该方案目前已经应用于基于 Linux 平台的视频采集和实时传输系统中, 并取得了良好的效果.

参考文献:

- [1] Schulzrinne H, Casner S, Frederick R, et al. RFC3550 RTP: a transport protocol for real-time application[S]. IETF, 2003.
- [2] 景慧燕, 唐存琛, 马玉利. 基于 RTP 的 MPEG-4 视频监控系统的的设计[J]. 计算机工程与设计, 2006, 27(8): 1439-1441.
- [3] 程念科, 胡瑞敏, 王中元. 基于 RS 编码的视频差错恢复方法[J]. 计算机应用研究, 2006, 23(4): 66-67.
- [4] 杨宗凯, 彭杰, 余江. 实时视频通信中的自适应前向纠错方案设计[J]. 计算机工程与科学, 2007, 29(8): 43-45.
- [5] 陈敏. 网络实时视频传输研究[D]. 广州: 华南理工大学, 2004.
- [6] 方新勇, 马瑞芳, 刘萍芬, 等. 基于 RTP 的流媒体传输系统的设计与实现[J]. 微电子学与计算机, 2007, 24(11): 183-185.
- [7] 葛镜. 基于 RTP 协议的视频流媒体实时传输[D]. 武汉: 华中师范大学, 2007.

作者简介:

王 峰 男, (1989—), 硕士研究生. 研究方向为高速实时信号处理