

一种新的视点间预测结构

严涛^① 安平^{①②} 沈礼权^{①②} 张兆扬^{①②}

^①(上海大学通信与信息工程学院 上海 200072)

^②(新型显示技术及应用集成教育部重点实验室 上海 200072)

摘要: 多视点视频编码除应具有较高的编码效率外, 还应该包括后向兼容性、时间随机访问和视点可分级性等, 这些都主要取决于所采用的预测结构。目前所提供的多视点视频编码(Joint Multi-view Video Coding, JMVC)采用固定的视点间预测结构, 难以适应复杂情况的多视点视频编码。该文综合考虑编码效率和用户随机访问等因素, 根据多视点视频相关性分析自适应调整视点间预测结构, 以获得较好的编码综合性能。试验结果表明, 与 JMVC 相比, 该文的方法在提高编码效率的同时, 有较好的随机访问性能。

关键词: 多视点视频编码; 随机访问; 相关性分析

中图分类号: TN919.81

文献标识码: A

文章编号: 1009-5896(2009)07-1610-05

A New Inter-view Prediction Structure

Yan Tao^① An Ping^{①②} Shen Li-quan^{①②} Zhang Zhao-yang^{①②}

^①(School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China)

^②(Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai 200072, China)

Abstract: Multi-view Video Coding (MVC) demands high compression rates, temporal random access, spatial random access, as well as view scalability, which based on inter-view prediction structure. Joint Multi-view Video Coding (JMVC) proposed the fixed view-temporal prediction structure for the standardization MVC, which is difficult to cope with various characteristics of multi-view videos. Considering coding efficiency and random access on, the multi-view coding scheme based on correlation analysis is proposed in this paper, which can dynamically select to the inter-view prediction structure. Experimental results show that the proposed prediction structure can improve coding efficiency and random access.

Key words: Multi-view Video Coding (MVC); Random access; Correlation analysis

1 引言

多视点视频由多个摄像机从不同角度拍摄而得。它包含同一个场景内多个视点图像, 视点之间有着高度的相关性^[1,2]。因此, 除了通常在 2D 视频中那样利用时空冗余达到编码目的外, 更重要的还应消除不同视点间的空间预测冗余。多视点视频编码往往比 H.264/AVC, MPEG-2 之类的单视点视频信号编码需要具有更高的压缩率、较低编解码复杂度、时间和视点的随机访问等。

在多视点视频编码(Multi-view Video Coding, MVC)算法研究中, 预测结构是最重要的一个方面^[3], 预测结构的设计已经在 MVC 研究中得到广泛的关注。日本名古屋 Fujii 研究组最早提出了基于光线空间的多视点编码系统, 针对相机阵列的光线空

间数据提出了基于多方向(Multi-direction, M)帧预测结构。M 预测结构包括 6 种帧类型共支持 21 种帧模式, 通过率失真模型选择最佳编码模式和参考帧进行预测编码^[4-6], 这种结构对密集相机阵列多源视频有较高的编码效率, 但其计算复杂度大且不利于时间和视角的随机访问。德国 HHI 的图像通信组提出了基于分层 B 帧(hierarchical-B picture)的多视点预测结构^[7,8], 对于大部分测试序列, 都表现出较好的压缩效率, 因而被选为多视点视频编码中的参考结构^[9]。

目前 ITU-T VCEG 和 ISO/IEC MPEG 的联合视频组(JVT)所提供的多视点视频编码采用固定的视点间预测结构, 难以适应复杂情况的多视点视频编码。为了进一步提高编码效率和用户随机访问性能, 最近有学者开始对 JMVM 视点间预测结构的设计进行优化研究。Park 等人直接根据相机的摆放位置快速确定 I-视点的位置^[10], 但此方法判断 I-视点的位置误差较大, 且没有考虑随机访问性能。赵德

2008-07-28 收到, 2009-03-02 改回

国家自然科学基金(60832003, 60672052)上海市曙光计划(06SG43)和上海市教委重点项目(09ZZ90)资助课题

斌等人^[11]引入 SP/SI 切换帧、辅助表示编码，很好地缩短了随机访问性能，但压缩效率较低。并且，以上这些方法都没有考虑多视点视频序列的内容相关性随相机阵列形式、相机间距和光照等因素不同而发生变化的现象。蒋刚毅等提出了基于相关性分析的多模式多视点视频编码^[12]，从多种预测模式中选择最合适的一种，降低了编码复杂度，但压缩效率并没有提高。

本文在最近 JVT 提供的多视点视频编码参考软件(JMVC1.0)^[13]基础上，综合考虑编码效率和用户随机访问性能等因素，给出较合理的视点间预测结构，在提高编码效率的同时有较好的随机访问性能。

2 视点间预测结构

通过对 3DAV 应用场景和需求的调研及分析，JVT 对 MVC 算法性能确定了 15 项基本要求^[14]，除了编码效率，还包括后向兼容性、视点或时间随机访问性能等。而这些目标经常是相互矛盾的。如图 1 所示，我们给出了几种常见的多视点视频编码视点间预测结构^[8]。从压缩效率的角度看，JMVC 要优于 Simulcast；但从解码随机访问、并行处理支持等方面看，Simulcast 要优于 JMVC。因此，确定合适的预测结构显得尤为关键。多视点视频编码模型中，关键帧(key pictures)^[8]采用 I、P、B 帧编码的视点，分别将其称为 I-视点、P-视点、B-视点。确定视点间预测结构主要有两个方面：确定 I-视点的位置和 P-视点之间插入 B-视点的数目。

2.1 I-视点的位置

I-视点作为其它视点的参考视点，选好 I-视点的位置可以提高编码效率，因此不少学者对 I-视点的选择作了较深入的研究。Park 等人^[10]通过计算全局视差选择 I-视点的位置，如下：

$$MAGD(\alpha) = \frac{1}{N} \sum_{\beta=0}^{N-1} |g(\alpha, \beta)| \quad (1)$$

这里， N 表示视点数目， $g(\alpha, \beta)$ 表示视点 α 与视点 β 之间的全局视差， $MAGD(\alpha)$ 表示视点 α 与其它视点全局视差的平均值。表 1 给出了平均全局视差的计算结果，表中作灰色标记的为 I-视点的位置。当视

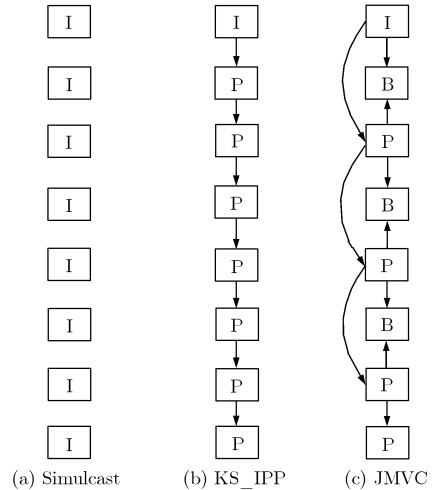


图 1 视点间预测结构

点数较多时，这种选取方法计算量太大。Sohn 等人研究得到 I-视点是接近中间位置的视点^[15]，因此，为了快速准确的确定 I-视点的位置，我们只需要对最有可能是 I-视点的中间几个视点进行相关性分析即可。

用全局视差进行相关性分析，需要进行视差估计。为了快速的确定 I-视点的位置，我们利用相关性分析函数测试两幅图像的相关性。设 A, B 分别为两幅图像， $A(i, j)$ ， $B(i, j)$ 分别代表 A, B 图像在位置 (i, j) 的灰度值， I, J 分别代表图像的高、宽， \bar{A}, \bar{B} 分别代表 A, B 图像平均像素值。两幅图像的相关性由式(2)得到。

$$C(A, B) =$$

$$\frac{\sum_{i=0}^{I-1} \sum_{j=0}^{J-1} [A(i, j) - \bar{A}] \times [B(i, j) - \bar{B}]}{\sqrt{\sum_{i=0}^{I-1} \sum_{j=0}^{J-1} [A(i, j) - \bar{A}]^2} \times \sqrt{\sum_{i=0}^{I-1} \sum_{j=0}^{J-1} [B(i, j) - \bar{B}]^2}} \quad (2)$$

那么 S_v 视点与其它 $K-1$ 个视点 S_w ($v \neq w$) 的相关性 $C(v)$ 为

$$C(v) = \frac{1}{K-1} \sum_{w=0, w \neq v}^{K-1} C(w, v) \quad (3)$$

表 2 给出了中间 3 个视点相关性分析结果，表中作灰色标记的为 I-视点的位置，与 Park 用全局视

表 1 所有视点相关性分析结果

视点	0	1	2	3	4	5	6	7
Ballroom	38.9	27.4	24.0	21.5	23.8	33.9	28.9	38.8
Exit	101.5	79.4	64.9	58.0	57.4	64.4	80.9	104.8
Race1	54.9	27.6	34.4	20.8	19.9	16.9	35.5	40.4

表2 中间3个视点相关性分析结果

视点(中间3个)	3	4	5
Ballroom	0.7862	0.7535	0.7326
Exit	0.7088	0.7171	0.6986
Race1	0.6278	0.6361	0.6453

差计算所有视点相关性分析结果一致。我们的方法不需要对所有视点进行相关性，只需要对最有可能是I-视点的中间3个视点进行相关性分析即可，且不要求全局视差就能确定I-视点的位置。

2.2 B-视点的数目

一般来说，相对I帧和P帧，B帧有较高的编码效率，为了提高编码效率，我们应该尽可能地增加B-视点的数目，然而，当B-视点的数目增多之后，I-视点与P-视点以及P-视点之间全局视差增大，这样编码P-视点时视差估计很不精确。而且，随着B-视点的数目的增多，编码复杂度会明显增大，随机访问也会降低。Park等人^[10]提出基于全局视差确定P-视点之间插入B-视点的数目。假设同一时刻相邻两个视点之间的平均视差为 d ，视差估计搜索范围为 R ，插入B-视点的数目 n_B 为

$$n_B = \max \{0, \lfloor \log_2(0.94 \times R/d) \rfloor\} \quad (4)$$

但是求出的 n_B 是插入B-视点的最大数目，并非是插入B-视点数目的最佳值，插入B-视点数目应该考虑编码效率和用户随机访问等因素，在 $0, \dots, n_B$ 中找到一个最佳值。

我们综合考虑编码效率和用户随机访问等因素，确定B-视点的数目。假设用于编码的多视点视频具有 N 个视点、每个视点 M 帧。编码帧数 $S = N \times M$ 是有限的，令 V_n ($n = 0, 1, \dots, N-1$)表示第 n 个视点， $P(V_n)$ 为用户随机访问第 n 个视点的概率； $V_{n,m}$ ($n = 0, 1, \dots, N-1, m = 1, 2, \dots, M$)表示第 n 个视点第 m 帧， $X_{n,m}$ 表示在对第 $V_{n,m}$ 帧进行解码前需要的帧数， $P(V_{n,m})$ 为用户随机访问第 $V_{n,m}$ 帧的概率， $P(V_{n,m} | V_n)$ 为用户随机访问第 $V_{n,m}$ 帧的条件概率，则随机访问代价的数学期望为

$$\begin{aligned} E(X) &= \sum_{n=1}^N \sum_{m=1}^M P(V_{n,m}) X_{n,m} \\ &= \sum_{n=1}^N \sum_{m=1}^M P(V_{n,m} | V_n) P(V_n) X_{n,m} \end{aligned} \quad (5)$$

为了获得较优异的编码综合性能，插入B-视点的数目 N_B 应该满足应该满足式(6)和式(7)

$$\min E(X) = f(N_B) \quad (6)$$

$$\max \text{PSNR} = \eta(N_B) \quad (7)$$

其中 $E(X)$ 是随机访问代价函数，PSNR是率失真函

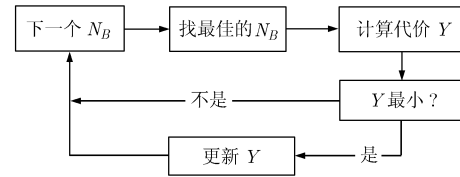
数，它们都与 N_B 有关。则优化问题模型为

$$\min Y = \lambda f(N_B) - \eta(N_B) \quad (8)$$

其中 λ 为调节参数。 N_B 满足条件式(9)

$$0 \leq N_B \leq n_B \quad (9)$$

图2给出了寻找最佳 N_B 的流程图。 N_B 从0到 n_B ，每给定一个 N_B ，编码视点间 N 帧图像就可以得到 Y ，通过比较 Y 的大小确定插入B-视点的数目 N_B 。

图2 寻找最佳的 N_B

通过以上分析，我们根据多视点视频相关性分析选择I-视点(S_I)的位置，综合考虑编码效率和用户随机访问等因素，自适应地调整视点间预测结构，表3给出了6种测试序列视点间预测结构参数。

表3 各序列视点间预测结构参数

序列	S_I	d	n_B	N_B
Ballroom	3	13	2	1
Exit	4	26	1	1
Race1	5	17	2	1
Flamenco1	4	11	2	1
Rena	42	3	4	3
Flamenco2	2	15	2	1

3 实验结果

为了验证我们提出的多视点视频编码方法具有较好的编码效率和随机访问性能，我们用MERL, KDDI, Nagoya University/Tanimoto Lab提供的测试序列Ballroom, Exit, Race1, Flamenco1, Flamenco2, Rena进行实验分析。表4给出了6个测试序列数据参数。其中，Rena序列相机间距比较小，运动缓慢；Exit, Race1等序列相机间距比较大，运动比较剧烈。表5给出了测试条件。

3.1 率失真性能比较

我们比较了4种多视点视频编码方案，各编码方案时间方向都采用分层B帧预测结构。Simulcast编码视点间没有预测，JMVC(1.0)为目前JVT所提出的多视点视频编码方法^[13]，MVC_G表示Park等人基于全局视差估计的多视点视频编码方法。MVC_

表 4 测试数据参数

测试序列	高 x 宽	相机间距	视点
Ballroom	640 × 480	20 cm	0~7
Exit	640 × 480	20 cm	0~7
Race1	640 × 480	20 cm	0~7
Flamenco1	320 × 240	20 cm	0~7
Rena	640 × 480	5 cm	38~46
Flamenco2	640 × 480	20 cm	0~4

表 5 测试条件

帧率	25
GOP 大小	8
搜索范围	64 pixel
BasisQP	28, 32, 36
ICMode	0
搜索方式	FastSearch
编码帧数	80
参考帧数	2

IBP为本文提出的新的视点间预测结构编码方法。表3给出了测试序列视点间预测结构参数。图3给出了相对Simulcast多视点视频编码方法,其它3种多视点视频编码方法的平均PSNR增益。与JMVC相比, MVC_G和本文MVC_IBP方法有较高的编码效率。其主要原因首先是I-视点的位置比较合理,其次是对运动比较缓慢的序列采用了较多的B-视点。其中, MVC_G是直接根据相机的摆放位置确定I-视点的位置,由于此方法误差比较大,对于Flamenco1序列其编码效率没有JMVC高。本文提出的MVC_IBP方法编码效率比JMVC提高0.08~0.36 dB。

3.2 随机访问评价

各方案的随机访问代价采用式(5)计算得到,我们假定各编码方案用户访问每个视点以及每帧的概

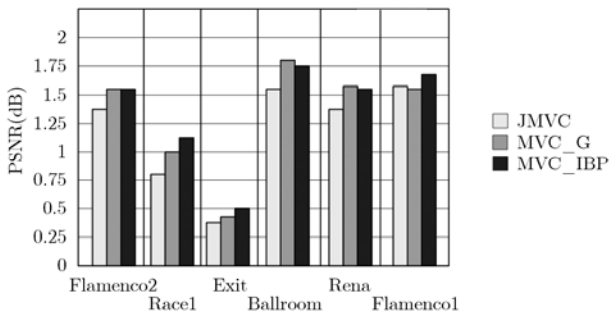


图3 平均PSNR增益

率是相同的,这样随机访问每个视点的代价由该视点解码前所需要的参考视点数进行评价。图4为各方案的平均视点随机访问代价。与JMVC相比, MVC_G和本文MVC_IBP方法有较好的视点随机访问性能。其主要原因是I-视点接近中间位置,访问其它视点所需的参考视点数较少。本文提出的MVC_IBP方法的平均视点访问代价只是JMVC (1.0)方案的65%左右。

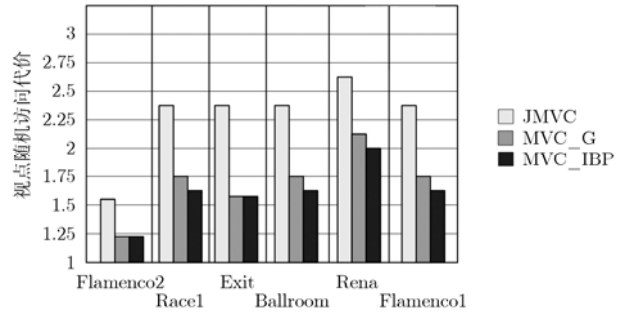


图4 随机访问代价

以上两方面的实验结果表明,与主流MVC方法相比,本文提出的MVC_IBP方法具有更高的编码效率和更好的随机访问性能。

4 结论

视点间预测结构是决定压缩效率、可伸缩性、随机访问、平行处理等多项性能的关键因素。本文综合考虑编码效率和用户随机访问等因素,提出了一种新的视点间预测结构,可以根据不同情况自适应调整视点间预测结构,在提高编码效率的同时有较好的随机访问性能。将进一步结合其它编码性能进行研究。

参考文献

- [1] Ishfaq A. Multi-view video: Get ready for next-generation television [J]. *IEEE Distributed Systems Online*, 2007, 8(3): 1-5.
- [2] Kubota A, Smolic A, Magnor M, Tanimoto M, Chen T, and Zhang C. Multi-view imaging and 3DTV [J]. *IEEE Signal Processing Magazine*, 2007, 11(4): 10-21.
- [3] ISO/IEC MPEG. Survey of algorithm used for multi-view video coding (MVC), Doc. N6909, Hongkong, China, January 2005.
- [4] Oka S, Bangchang P N, and Fujii T. Dynamic ray space coding for FTV. In *Proceeding of 3D Image Conference*, Tokyo, 2004: 139-142.
- [5] Fujii T and Tanimoto M. Free-viewpoint TV system based on ray-space representation. *SPIE Information Technologies and Communications*, 2002, 4864: 175-189.

- [6] Oka S, Fujii T, and Tanimoto M. Dynamic ray-space coding using inter-view prediction. In Proceeding of the International Workshop on Advanced Image Technology 2005 (IWAIT2005), Jeju, Korea, Jan. 2005: 19-24.
- [7] Merkle P, Muller K, Smolic A, and Wiegand T. Efficient compression of multi-view video exploiting inter-view dependencies based on H.264/MPEG4-AVC. IEEE international Conference on Multimedia and Expo (ICME'06), Toronto, July 2006: 1717-1720.
- [8] Merkle P, Smolic A, Muller K, and Wiegand T. Efficient prediction structures for multi-view video coding[J]. *IEEE Transactions on Circuits and Systems Video Technology*, 2007, 17(11): 1461-1673.
- [9] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG. Joint multiview video model (JMVM), Doc. JVT- Y207, Shenzhen, China, October 2007.
- [10] Park P K, Oh K J, and Ho Y S. Efficient view-temporal prediction structures for multi-view video coding [J]. *Electronics Letters*, 2008, 44(2): 102-103.
- [11] Liu Yanwei, Huang Qingming, Zhao Debin, and Gao W. Low-delay view random access for multi-view video coding [A]. Proceedings of IEEE International Symposium on Circuits and Systems(ISCAS 2007), Louisiana, USA, May, 2007: 997-1000.
- [12] 蒋刚毅, 张云, 郁梅. 基于相关性分析的多模式多视点视频编码[J]. *计算机学报*, 2007, 30(12): 2205-2211.
- Jiang G Y, Zhang Y, and Yu M. Multi-modal multi-view video coding based on correlation analysis [J]. *Chinese Journal of Computers*, 2007, 30(12): 2205-2211.
- [13] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG. WD 1 Reference software for MVC (JMVC) 1.0, Doc. JVT-AA212, CH, Geneva, April, 2008.
- [14] ISO/IEC MPEG. Requirements on multi-view video coding v.7, Doc. N8218, Klagenfurt, Austria, July 2006.
- [15] Sohn Kwanghoon, Kim Yongtae, and Seo Jungdong. Results on CE1 for multi-view video coding. Joint Video Team, Doc. JVT-T102, Austria, Klagenfurt, July 2006: 1-8.
- 严涛: 男, 1981年生, 博士生, 研究方向为多视点视频编码及码率控制.
- 安平: 女, 1968年生, 教授, 博士生导师, 主要研究方向为图像及3维视频信号处理.
- 沈礼权: 男, 1978年生, 博士, 研究方向为数字视频处理、编码与传输.
- 张兆杨: 男, 1938年生, 教授, 博士生导师, 主要研究方向为数字视频处理、编码与传输.