

基于相关性分析的多模式多视点视频编码

蒋刚毅^{1,2)} 张云^{1,2)} 郁梅^{1,3)}

¹⁾(宁波大学信息科学与工程学院 浙江 宁波 315211)

²⁾(中国科学院计算技术研究所 北京 100080)

³⁾(北京大学视听信息处理国家重点实验室 北京 100871)

摘要 多视点视频编码方法除需具有较高编码效率外,还必须支持视点或时间的随机访问、低延时编解码、视点可分级等性能.多视点视频信号的时间、视点间相关性随相机密度、光照、对象运动等因素不同而变化.文中提出基于多视点视频信号相关性分析的多模式多视点视频编码方法,改变传统单一预测模式的多视点编码结构,将多种性能优良的预测编码模式有机结合,根据多视点视频相关性分析灵活选择合适的预测编码模式,以获得优异的编码综合性能.实验结果表明,所提出的多模式多视点视频编码方法在保证高压缩效率的前提下,可进一步降低复杂度,提高随机访问性能.

关键词 多视点视频编码;多模式;随机访问;计算复杂度;相关性分析;模式更新
中图法分类号 TP391

Multi-Modal Multi-View Video Coding Based on Correlation Analysis

JIANG Gang-Yi^{1,2)} ZHANG Yun^{1,2)} YU Mei^{1,3)}

¹⁾(Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315211)

²⁾(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)

³⁾(National Key Laboratory of Machine Perception, Peking University, Beijing 100871)

Abstract Multi-view video coding (MVC) should support view random access, temporal random access, spatial random access, low coding delay, view scalability, as well as high compression efficiency and low complexity. The correlation characteristics of the multi-view video signal is variant along the time axis, and it is influenced by video contents, illumination change, speed of moving objects and cameras, view interval, sampling frequency, and other factors. A multi-modal multi-view video coding (MMVC) scheme based on correlation analysis is proposed in this paper, which differs from the conventional MVC schemes with single prediction mode. Several MVC prediction structures with excellent performances have been properly integrated into the scheme, and are dynamically selected to encode current multi-view video according to correlation characteristics of the video. Experimental results show that the proposed MMVC scheme can reduce computational complexity and improve random access performance while maintaining high coding efficiency.

Keywords multi-view video coding (MVC); multi-mode; random access; computational complexity; correlation analysis; mode update

收稿日期: 2006-06-13; 最终修改稿收到日期: 2007-06-20. 本课题得到国家自然科学基金(60472100, 60672073)、教育部新世纪人才支持计划项目(NCET-06-0537)、教育部科学技术研究重点项目(2060059)、浙江省自然科学基金(Y105577)资助. 蒋刚毅, 男, 1964年生, 博士, 教授, 博士生导师, 主要研究领域为图像处理与视频信号编码、多媒体通信、基于视觉的智能控制. E-mail: jianggangyi@126.com. 张云, 男, 1982年生, 博士研究生, 主要研究方向为图像处理、视频信号编码与通信. 郁梅, 女, 1968年生, 博士, 教授, 主要研究领域为多媒体信号处理、编码与通信、计算机视觉及应用.

1 引 言

随着信息科学技术的不断发展,用户已不满足于只能被动地接受视频信息;而另一方面,现行的视频系统在表现自然场景时,也无法为用户提供足够的深度感、立体感.多视点视频编码(Multi-view Video Coding, MVC)是针对新一代交互式多媒体应用而提出的^[1-5],是自由视点电视、三维电视等三维音视频应用中的核心技术.近年来,人们对三维音视频标准的制定进行了探索性研究,其中具有代表性的多视点视频编码方法有顺序预测结构^[1],它通过多参考帧的顺序预测来减少相邻视点和时间域信息冗余,缓解遮挡和暴露问题,有较高的编码效率,但存在随机访问性能不佳、易导致错误传递等问题.M-picture 预测结构包括 6 类帧类型,共支持 21 种帧模式^[6,7],通过率失真优化选择最佳的帧模式和多参考帧预测编码,对小相机间距的多视点视频有较高的编码效率;但存在着计算复杂度很大、不利于随机访问等问题.此外,还有 2D Direct Mode^[1,8], View Interpolation Prediction^[1]等预测编码方法,这些方法都采用了时间、视点间联合预测的多参考帧预测方式,以提高压缩效率,但却造成复杂度成倍增加.并且,这些方法都未考虑多视点视频信号时间及视点间的内容相关性随相机密度、光照、相机及对象运动等因素不同而变化的现象,难以适应复杂情况的多视点视频编码.对于三维音视频应用,多视点视频压缩算法除了需要有较高的编码效率外,还必须支持视点或时间的随机访问、低延时编解码、视点可分级等性能^[2].NTT 公司提出的 GoGOP 方案^[9-10]使用多个帧内预测(I 帧),提高了随机访问性能,却导致编码效率较明显降低,并且所采用的 Inter GOP 多参考帧策略,增加了计算复杂度.

针对多视点视频信号相关性随相机密度、光照、对象运动等因素不同而变化的现象,本文提出多模式多视点视频编码结构,将多种性能优良的预测结构有机结合,根据多视点视频内容灵活选择预测编码模式,以实现率失真、低延时编解码和随机访问综合性能最优.

2 多模式多视点视频编码框架

文献[11]采用块匹配的方法对多视点视频序列进行相关性统计分析,采用的预测结构如图 1 所示.

图中的方框表示图像帧,竖直方向表示同一相机不同时刻拍摄的图像帧,水平方向则表示同一时刻不同相机拍摄的图像帧.将当前统计帧(由 P 表示)分块进行块匹配,从同一视点不同时刻的 T 帧和同一时刻不同视点的 S 帧中寻找最佳匹配块.所在参考帧的最佳匹配块越多,则表示 P 帧与该帧的相关性越大.表 1 为各多视点序列最佳匹配块在时、空域的相对分布,它反映了多视点视频序列的时、空域相关性分布,百分比值愈大,说明在该域相关性(冗余)愈大.设 R_T 和 R_S 分别为块匹配过程中时间域和空间域参考帧所占参考块的百分比, $R_T + R_S = 1$.表 1 结果显示,如 Crowd, Race1, Objects1 和 Aquarium 等的多视点视频序列^[11-14]是以时间域相关性为主的序列,这类序列的相机间距较大,运动缓慢,导致时域相关性极大(R_T 为 86.1%~91.2%);反之对于 Xmas^[11,13-14]等多视点序列,由于其相机间距小,视点间相关性极大(R_S 为 80.83%).此外, Race2, Flamenco1 和 Flamenco2 等多视点序列^[12],以时间域预测为主,但空域相关性也占一定比例(R_S 为 22.6%~32.4%),尤其在视频场景移动剧烈或环境光照影响下亮度和色度发生变化巨大时,时、空域相关性此消彼长,导致时、空相关性均衡分布.

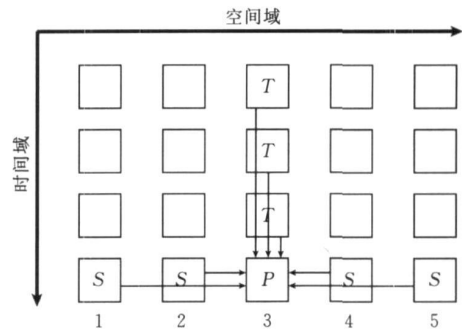


图 1 相关性分析的预测结构

表 1 各序列最佳匹配块时、空域分布

多视点序列	$R_T/\%$	$T_S/\%$
Flamenco1	77.43	22.57
Flmaneco2	74.39	25.61
Race1	91.17	8.83
Race2	67.57	32.43
Crowd	86.20	13.80
Objects1	88.63	11.37
ST_Ball	71.62	28.38
ETH_Karate	98.90	1.10
Xmas	19.17	80.83
Aquarium	86.09	13.91

基于上述分析,将多视点序列分为 3 类:(1)以时间域相关性为主;(2)以空间相关性为主;(3)时、空相关性相对均衡分布.传统采用单一模式的预测结

构的编码方法,如顺序预测、GoGOP 和 M-picture 等,对相关性迥异的多视点视频序列,均采用同一编码结构,盲目地通过多参考帧方式消除时、空冗余,虽然具有较好的编码效率,却导致了高复杂度和随机访问性能变差的问题.另外,对于复杂多变的拍摄环境,单一模式的预测结构适应性较差,如顺序预测不适合十字型等相机阵列.为了解决传统单一模式预测结构的缺陷,本文提出具有良好适应性和综合性能的多模式多视点视频编码方法.

多模式多视点视频编码(Multi-mode Multi-view Video Coding, MMVC)框架是基于 H. 264 视频编码算法的,如图 2 所示,包含相关性分析、模式选择、

基于 H. 264 的多视点视频编码、触发更新模式等 4 个部分.在对多视点视频序列进行编码前,需要先对多视点序列的时、空相关性进行分析;然后,依据外部设定参数和序列相关性,选择多视点预测编码模式,并根据确定的编码模式将多视点数据编码成多视点视频流.当完成一个 GOP(Group Of Pictures) 的编码后,多模式算法通过一定的模式更新触发机制判断当预测模式是否已不适应于后续序列,如果“是”,则开启模式更新,进入相关性分析模块;否则仍然以当前模式编码后续多视点视频.

2.1 相关性分析

图 3 给出了 5 种多视点视频编码的预测模式.图中 I 表示帧内编码帧; D 表示视差补偿预测编码帧; P 表示运动补偿预测编码帧; P' 表示时、空双向预测编码帧,可参考 D 、 P 帧; B' 为时、空联合预测帧,可参考 D 、 P 和 P' 帧; P_s 表示时、空多参考的统计编码帧.

为了降低计算复杂度,可以先通过对一个或若干 GOP 执行图 3(e) 的统计兼编码的预测模式,并以编码过程中红色区域 P_s 帧的运动/视差矢量的比例来估算得到欲编码序列的时、空域相关性分布和时域内部相关性分布情况,从而在编码过程中实现相关性统计分析.

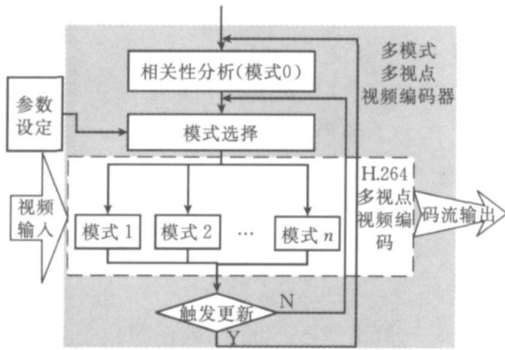


图 2 多模式多视点编码算法结构

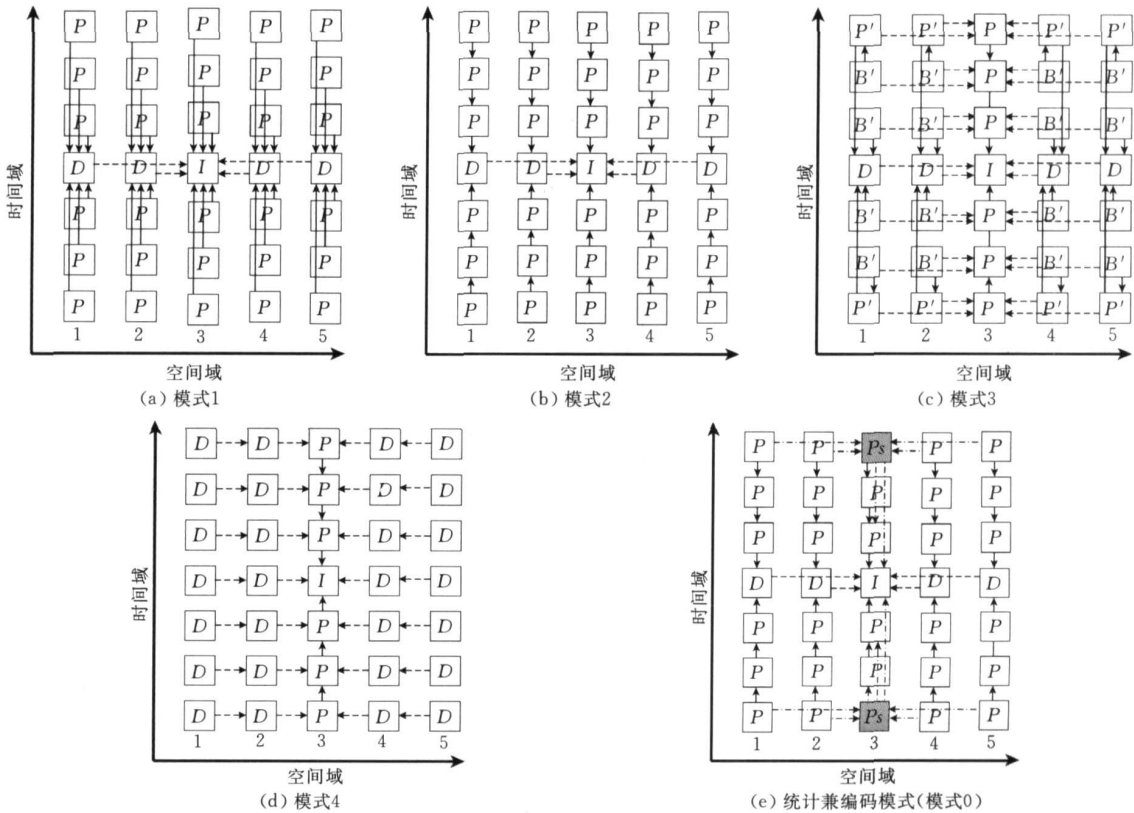


图 3 5 种多视点视频编码的预测模式

2.2 基于 H 264 的多视点预测编码模式

在 MMVC 方法中, 根据多视点视频相关性的 3 类情况, 本文设计了图 3 所示的 5 种预测编码模式. 其中, 模式 1、模式 2 是以运动补偿预测为主, 模式 2 采用时间域顺序预测模式, 而模式 1 采用时间域跳跃式预测, 具有更少的随机访问代价, 适合于运动少的多视点视频; 图 3(c) 中的模式 3 是兼顾了时、空域联合预测, 图 3(d) 的模式 4 则以视差补偿预测为主. 另外, 图 3(e) 中的模式 0 为统计兼编码的预测模式, 它是编码模式 2 和图 1 预测结构的结合, 在编码 GOP 的同时可通过图 3(e) 中的深色区域 P 帧的运动/视差矢量的比例来估算欲编码序列的相关性信息, 该模式仅在需要相关性分析时使用. 这里给出的 5 种预测模式是针对 5×7 GOP 结构, 并采用 I-center 的方案^[15].

不同于先前的顺序预测、GoGOP 和 M-picture 等方法对不同相关性的多视点视频序列均采用同一预测编码结构的情况, MMVC 方法中 5 种预测模式的选择是基于多视点视频序列的相关性来确定. 例如, 对于时域冗余为主的多视点视频序列, 主要采用运动补偿预测的方法消除冗余, 即采用模式 1 或 2. 由此, 在保证较高编码效率的前提下, 较大幅度地减小参考帧数, 能提高随机访问性能, 降低计算复杂度, 且易于部分解码与视点绘制.

2.3 预测编码模式选择

对于某些特殊的拍摄条件和应用需要, 用户可由外部直接设定多视点预测编码模式; 否则, 根据相关性分布情况选择已确定的编码模式. 图 4 为模式选择流程.

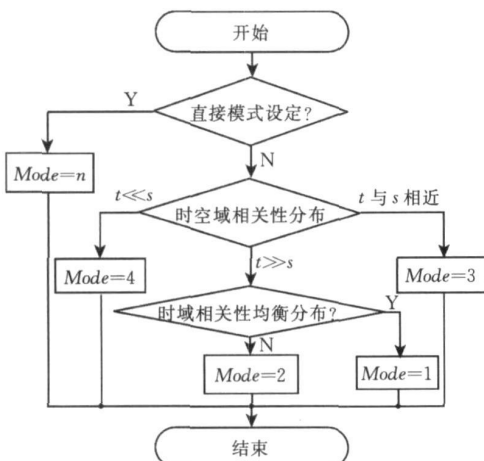


图 4 模式选择流程图

(1) 当 $R_T \ll R_S$ 时, 表示序列以视点间相关性为主, 则选择如图 3(d) 中视差补偿预测为主的模式 4

编码.

(2) 当 R_T 与 R_S 接近时, 表示序列时、空相关性分布相对均衡, 则采用如图 3(c) 中兼顾时、空域联合预测模式 3 编码.

(3) 当 $R_T \gg R_S$ 时, 表示序列以时间域相关性为主, 应进一步判断相关性在时域内部的分布情况. 当 $R_T \gg R_S$ 且在时域内部分布相对均衡, 如 Aquarium 和 ETH_Karate 等多视点序列^[12-16], 选择采用图 3(a) 中模式 1 编码; 反之, 如相关性集中分布在最邻近时刻, 如 crowd 等多视点序列^[11-12], 则选择如图 3(b) 中的模式 2 编码.

2.4 模式触发更新条件

如果频繁开启模式更新, 可能导致无谓增加计算复杂度; 而如果模式更新不及时, 则将导致压缩效率下降. 所以模式更新的策略会影响编码器性能. 可以考虑采用如下的模式更新方案:

(1) 定时更新方案: 在编码若干个 GOP 后, 触发模式更新一次. 显然, 定时更新方案存在盲目性.

(2) 基于视频内容的更新方案: 分析上述编码模式的共性, 每个 GOP 总存在一个 I 帧, 对于与该 I 帧同一时刻但不同视点的 D 帧图像仅通过视差补偿预测编码, 令该 D 帧的帧内预测块 (I 块) 数由 d_i 表示; 而对于与 I 帧同一视点但不同时刻的 P 帧图像仅通过运动补偿预测编码, 且令该 P 帧的 I 块数由 p_j 表示. 设 N_d 和 N_p 分别表示计算相关性系数 d_i 和 p_j 的 D 帧和 P 帧帧数, 则相关性表征系数 α 定义为

$$\alpha = \frac{1}{N_d} \sum_{i=0}^{N_d} d_i \setminus \frac{1}{N_p} \sum_{j=0}^{N_p} p_j \quad (1)$$

可以依据相关性表征系数 α 的取值情况, 来确定模式的更新.

2.5 随机访问评价

假设用于编码的具有 n 个视点、每个视点 m 帧的多视点视频帧总数 $s = n \times m$ 是有限的, 令 x_i 表示在对第 i 帧进行解码前需要提前解码的帧数, p_i 为用户随机访问第 i 帧的概率, 则随机访问代价的数学期望^[15] 为

$$E(X) = \sum_{i=1}^{n \times m} x_i p_i \quad (2)$$

随机访问代价是评价多视点视频编码方法随机访问性能的重要指标, 这个代价越高, 说明多视点编码方法在解码端对随机访问的支持能力越差, 为支持随机访问而消耗的资源就越多. 假设由 k_i 表示 N 个模式中第 i 个模式编码序列的百分比, 则 MMVC

的随机访问代价 $E(X)$

$$E(X)_{MMVC} = \sum_{i=1}^N (k_i \times E(X)_i) \quad (3)$$

3 实验结果

为了分析相关性系数与评估所提出的 MMVC 方法, 并确定相应的模式更新触发阈值, 对 KDDI 公司和 Tanimoto 实验室提供的 flamenco1, crowd,

race2(选取运动剧烈, 相机随车运动的视频段)、Xmas(视点间相关性大)等多视点测试序列进行统计分析. flamenco1, crowd, race2 多视点序列的相机间距为 200mm, 而对于 Xmas 我们选用其中相机间距为 9mm 的 5 个视点作为测试序列. 图 5 为其中 3 个多视点序列的 5 个视点图像. 使用 H. 264 JM 8.5 作为基本测试平台进行编程实现相关多视点视频编码方法.

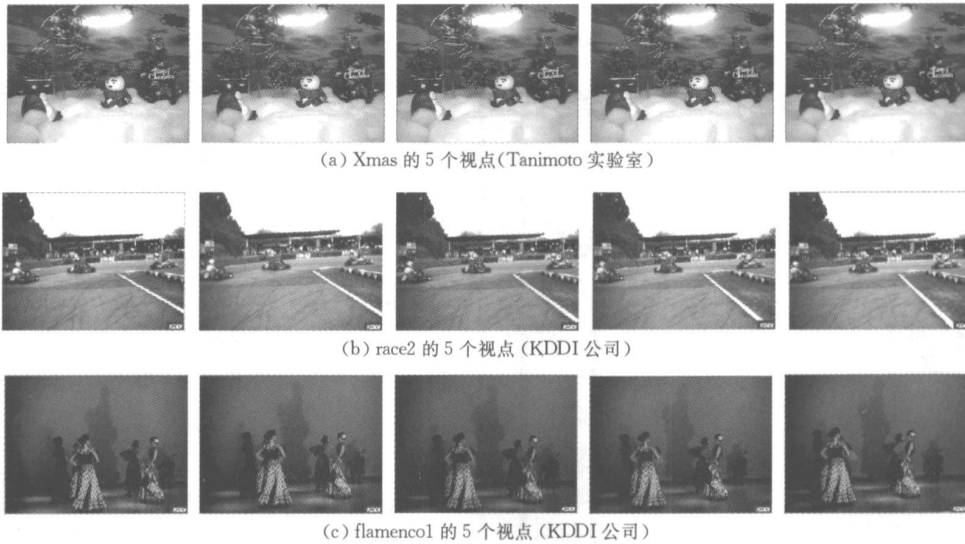


图 5 3 个多视点序列的 5 个视点

3.1 相关性表征系数 α

图 6 为 flamenco1, crowd 和 race2 序列相关性分析图, 横轴为编码帧序号, 纵轴为 R_s 与 R_t 的比值, 该值越大表示空间域相关性相对越大, 越小则表示时间相关性越大, 在 1 附近则表示时、空域相关性相近. 由图可知, crowd 和 flamenco1 序列相对平稳, 相关性为 0.2 左右, 表示空间域相关性约为时域相关性的 1/5, 时域相关性极大; race2 选取 200 ~ 270 帧, 相机随赛车转弯, 整个视频运动剧烈, 时、空域相关性相近. Xmas 相机间距 9mm 时, 实验得到实际空域相关性占 70%左右, 空间相关性很大.

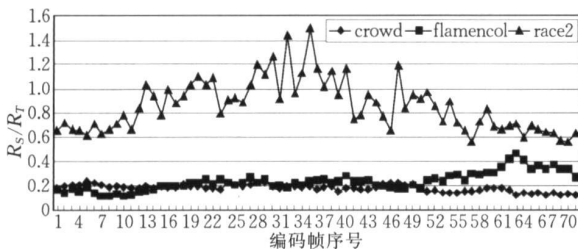


图 6 3 个多视点测试序列的相关性分析

I 帧最邻近的 2 个时间帧、2 个空间帧的 d_i 和 p_i 时的 α 信息, 即 $N_d = N_p = 2$. 其中横轴为被编码 GOP 的序号, 纵轴为相关性表征系数 α , 该值越大表示空间相关性越大. 由图 6 和图 7 可得, α 与序列相关性信息趋势一致, 接近于序列相关性的采样结果. 而 Xmas 序列的相机间距为 9mm, 空间相关性大、时间相关性弱, 其 α 均大于 3. 所以, 可以通过 α 信息表征当前 GOP 中相关性时空域分布情况. 在当前编码 GOP 的 α 值与当前编码模式不匹配时, 开启模板更新, 进行相关性统计并确定后续 GOP 相应的编码模式. 当 $\alpha < 0.35$ 时, 则选择预测编码模式 1 或模式 2; 当 $\alpha > 1.5$ 时, 则选择预测编码模式 4; 否则, 则选择预测编码模式 3.

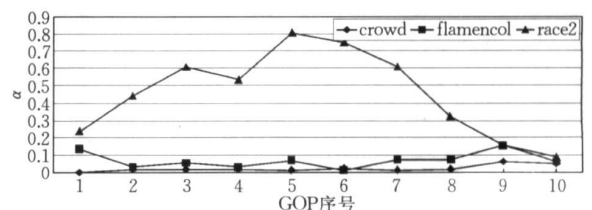


图 7 3 个多视点测试序列相关性表征系数 α

图 7 为对应 flamenco1, crowd 和 race2 序列取

3.2 率失真性能比较

图 8 给出了各方案的率失真性能比较, 其中 BSVP 和 PSVP 分别表示采用 B 帧和 P 帧的顺序预测方法, Mpicture 表示 Fujii 等人基于 M-Picture 的视频编码方法, MMVC 为本文多模式多视点视频编码方法. 对于 flamenco1 和 crowd 等不同特性的大间距测试序列, 相关性集中在时间域, BSV P,

PSVP 和本文 MMVC 方法的率失真性能一致, 而 Mpicture 方案则要差 1dB 左右; 对于运动剧烈的 race2, 空间相关性提高, 4 种方法的率失真性能基本一致. 对于 Xmas 的相机间距为 9mm 的多视点序列, 虽然以空间相关性为主, 但空域内部不集中且时间域相关性仍有较大比重, 所以 MMVC 采用模式 4, 其编码效率在高码率段略微降低.

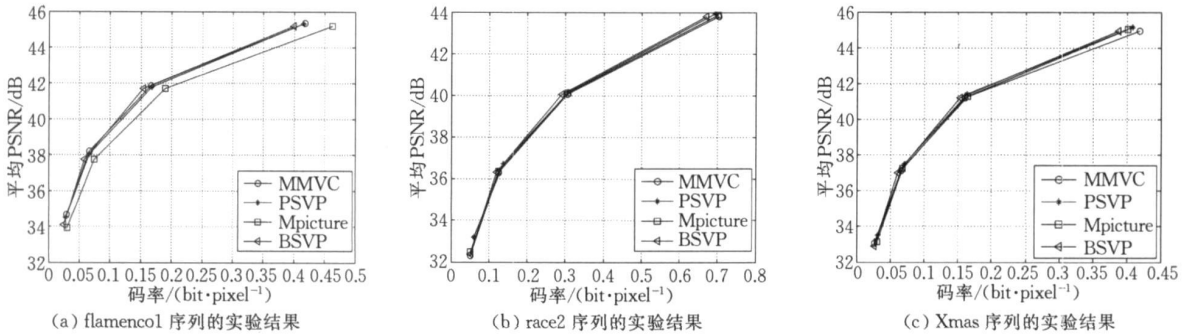


图 8 各方案的率失真性能对比结果

3.3 随机访问性能与计算复杂度分析

MMVC 中各模式编码比例 k_i 直接与实际多视点序列相关, 式(3)中的 N 为 4 (即除用于相关性分析的模式 0 外的编码模式有 4 种), 且假定各模式编码的概率相同, 即 $k_i = 0.25 (i = 1, \dots, 4)$, 各方案的随机访问代价如表 2 所示. 就随机访问性能而言, PSVP 最差, BSV P 和 Mpicture 相对好些. 本文提出的 MMVC 方法的平均随机访问代价只是其它编码方案的 24%~44%, 随机访问性能有明显提高.

表 2 随机访问代价和复杂度比较表

编码方案	$E(X)$	计算复杂度
BSVP	7.5	83
PSVP	11.0	58
Mpicture	6.0	97
MMVC	2.62	40

基于 H.264/AVC 编码框架的高精度视差补偿预测和运动补偿预测占整个多视点视频编码器的大部分计算时间, 所以通过平均编码一个 5×7 GOP 所需进行的视差补偿预测和运动补偿预测次数来表征整个编码器的计算复杂度. 各方案的计算复杂度比较如表 2 所示, 由于采用了较为复杂的预测方式, BSV P, PSVP 和 Mpicture 方案计算复杂度都很大, 尤其是 BSV P 和 Mpicture 方法. 相比其它编码方案, 本文的 MMVC 方法计算复杂度节省 31%~59%.

4 结 论

MMVC 方法根据多视点视频序列的时间、空间

域相关性自适应地选择最佳预测编码模式, 集 4 个多视点预测编码模式之所长, 在保证高率失真性能的同时, 提高了随机访问性能, 降低了计算复杂度; 由于编码时选用相对更少数量的参考帧, 因而也更易于解码端进行部分解码和绘制. MMVC 方法采用了灵活的 MVC 构架, 可通过对视点视频序列相关性分析以及阈值的调整, 选择不同预测编码模式, 以满足对率失真、随机访问或计算复杂度等的不同性能需求. 进一步的研究将解决本文中模式更新存在滞后现象等问题.

参 考 文 献

- [1] ISO/IEC JTC1/SC29/WG11 N6909. Survey of algorithms used for MVC. Hong Kong, Jan. 2005
- [2] ISO/IEC JTC1/SC29/WG11 N7282. Requirements on multi-view video coding v. 4. Poznan, July 2005
- [3] ISO/IEC JTC1/SC29/WG11 N5877. Application and requirements for 3DAV. Trondheim, July 2003
- [4] ISO/IEC JTC1/SC29/WG11 N6720. Call for evidence on multi-view video coding. Palma de Mallorca, Oct. 2004
- [5] ISO/IEC JTC1/SC29/WG11 N5878. Report on 3DAV exploration. Trondheim, Norway, July 2003
- [6] Oka S, Bangchang P N, Fujii T. Dynamic ray-space coding for FTV//Proceedings of the 3D Image Conference. Tokyo, Japan, 2004: 139-142
- [7] Oka S, Endo T, Fujii T. Dynamic ray-space coding using multi-directional picture. IEICE Technical Report, 2004, 104(493): 15-20

- [8] Lee Y, Lee Y, Kim D. Multi-view video coding using illumination change-adaptive motion estimation and 2-D direct mode//Proceedings of the 6th Pacific-Rim Conference on Multimedia. Jeju Island, Korea, 2005; 396-407
- [9] Fujii T, Tanimoto M. Free-viewpoint TV system based on ray-space representation//Proceedings of the SPIE. Boston, Massachusetts, 2002; 4864; 175-189
- [10] Kimata H, Kitahara M, Kamikura K et al. System design of free view point video communication//Proceedings of the 2004 International Conference on Computer and Information Technology. Wuhan, China, 2004; 52-59
- [11] Fecker U, Kaup A. Statistical analysis of multi-reference block matching for dynamic light field coding//Proceedings of the 10th International Fall Workshop Vision, Modeling, and Visualization. Erlangen, Germany, 2005; 445-452
- [12] ISO/IEC JTC1/SC29/WG11. M10533. KDDI multi-view video sequences for MPEG 3DAV use. Munich, 2004
- [13] Fecker U, Kaup A. H. 264/AVC-compatible coding of dynamic light fields using transposed picture ordering//Proceedings of the 13th European Signal Processing Conference (EUSIPCO 2005). Antalya, Turkey, 2005
- [14] Tanimoto M, Fujii T. Test sequence for ray-space coding experiments//Proceedings of the MPEG Document M10408. Hawaii, USA, 2003
- [15] Jiang G, Yu M, Zhou Y et al. New multi-view video coding scheme for 3DAV systems//Proceedings of the 2006 Picture Coding Symposium. Beijing, 2006
- [16] Wurmlin S, Waschbusch M, Gross M. ETH-REAL: A real-world test data set for 3DAV EE2. MPEG Document M10580. Munich, Germany, Mar. 2004



JIANG Gang-Yi, born in 1964, Ph. D., professor. His major research interests include image processing and video coding, multimedia communication, vision based intelligent control.

ZHANG Yun, born in 1982, Ph. D. candidate. His major research interests include image processing, video coding and communication.

YU Mei, born in 1964, Ph. D., professor. Her major research interests include multimedia signal processing, coding and communication, computer vision and its applications.

Background

Multi-view video is a kind of new multimedia which provides stereoscopic impression and interactive function, and multi-view video coding (MVC) is one of the key techniques of 3D Audio-Visual (3DAV). Multi-view video signals show different temporal and interview correlations due to the influence of camera interval, illumination, motion of cameras and objects, and so on. In MVC, motion estimation and motion compensation is employed to eliminate temporal correlation within a single view video, while disparity estimation and compensation is adopted to reduce inter-view correlation among neighboring views. Various view-temporal prediction structures are proposed for MVC based on the multi-reference frame prediction technique. Even though different temporal and spatial reference frames are selected by these prediction modes, the prediction structure is fixed for each prediction mode and it does not adapt to variations of the correlation characteristics of the multi-view video signal. In order to achieve better rate-distortion performance, complex prediction structure is usually needed for prediction mode with a fixed structure. In other words, many reference frames including temporal and spatial frames are used without considering the correlation variation of the multi-view video sequence. This may result in a huge increase of computational complexity, and the ability of random access and partial decoding are also decreased. On the other hand, if a simple prediction structure is used, the correlation among the multi-view video signal may not be exploited sufficiently by the encoder so that high rate-distortion performance is unable to

be achieved. Moreover, different from traditional video coding schemes, MVC should support view random access, temporal random access, spatial random access, low coding delay, view scalability, as well as high compression efficiency and low complexity. However, some of these requirements are conflicting to one another, which means that a prediction mode with fixed structure is not flexible to meet different requirements for multi-view video codec.

This work is supported by the Natural Science Foundation of China (Nos 60472100, 60672073), the Program for New Century Excellent Talents in University (NCET-06-0537), Natural Science Foundation of Zhejiang Province (grant No Y105577), and the Key Project of Chinese Ministry of Education (grant No 206059). The proposed multimodal multi-view video coding scheme is based on analysis of temporal and interview correlations. Different prediction modes are designed for multi-view video signals with different correlation characteristics. Suitable prediction mode is dynamically selected from the designed modes according to the correlation characteristic of current multi-view video signals. The goal of this project is to propose the framework of multimodal multi-view video coding scheme based on analysis of temporal and interview correlations, the corresponding analysis method of temporal and interview correlations, the design of multiple prediction modes, mode selection and updating, etc. so as to obtain high smooth coding efficiency and optimize the integrative efficiency of the multi-view video codec.