

# 基于视差估计算法的多视点视频预测方案<sup>\*</sup>)

潘 峰 马力妮 杜小钰

(北京信息科技大学计算机学院 北京 100085)

**摘 要** 多视点视频编码是立体视频的关键技术,视差估计是目前多视点视频编码中常用的方法之一。探讨了多视点视频编码中的视差估计算法原理,对当前应用视差估计算法的几种典型预测结构进行了深入分析,提出一种结合分级 B 图与视点相关性的多视点编码方案,该方案在 H. 264/AVC 的编码模型 JM 下实现。实验表明,该方案在兼顾随机访问性能的同时,实现了较高的编码效率,尤其适合视点间相关性较高的运动图像序列。

**关键词** 多视点视频,视频编码,视差估计,分级 B 图,预测结构

## Multiview Video Coding Scheme Based on Disparity Estimation Algorithm

PAN Feng MA Li-ni DU Xiao-yu

(Department of Computer Science, Beijing University of Information Science and Technology, Beijing 100085, China)

**Abstract** Multiview video coding is a key technology of stereo video, and disparity estimation is the most popular algorithm of current multiview video coding schemes. The theory of disparity estimation algorithm is first investigated, and several typical coding techniques are analyzed in detail. Then a novel multiview coding scheme is proposed which is based on hierarchical B pictures and correlation property of viewpoints. Experimental results show that the proposed scheme is efficient.

**Keywords** Multiview video, Video coding, Disparity estimation, Hierarchical B pictures, Prediction structure

### 1 引言

多视点视频(multi-view video, MVV)是当前多媒体领域的研究热点,其广阔的应用前景正在受到越来越多的关注。与传统的视频不同,多视点视频系统能够提供与场景的交互能力,它允许用户在一定范围内选择任意视点或视角观看,用户可以实现对同一场景的多方位体验,这充分满足了用户的交互性和个性化需求,使用户获得更加真实的视觉感受。图 1 展示了一种典型的多视点视频系统,首先由多个摄像机同步视频采集,然后对多视点数据编码,通过存储和传输,最终在用户端根据不同应用需求解码和显示<sup>[1]</sup>。多视点视频的提出体现了下一代多媒体应用真实感、交互性和网络化的发展方向,它的立体特性和交互能力使得它有着广阔的应用前景,其编码技术也在过去的几年中得到了飞速发展。

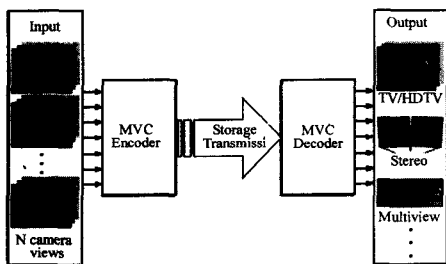


图 1 多视点视频系统

目前,ISO 组织和 ITU-T 组织的联合视频专家组(JVT, Joint Video Team of ISO/IEC MPEG & ITU-T VCEG)关于多视点视频编码(multi-view video coding, MVC)的标准化工

作已经全面展开。2002 年 5 月, JVT 小组会议正式发布了 3DAV 的需求文档<sup>[2]</sup>,并在次年 Fairfax 会议中征集 MVC 标准意见。随后的历次 MPEG 会议, MVC 作为 H. 264 的重要扩展受到密切关注,在 2005 年的 Poznan 会议上, MVC 进入 CFP 阶段(Call for Proposals)<sup>[3]</sup>,2006 年 1 月 Bangkok 会议和 2007 年 7 月 Geneva 会议中,都有多项关于 MVC 的讨论提案。此外, JVT 小组还开发了一个多视点视频应用的软件模型 JMVM (Joint Multi-view Video Model)<sup>[4]</sup>。JMVM 基于 H. 264 编码标准,采用了基于宏块的运动预测和视差估计技术,同时利用多路视频数据的视点相关性,为多视点视频编码的方案实现提供了可行验证。

由于多视点视频信号是由多个不同位置和角度的相机采集得到的,随着视点数目的增加,需要处理、存储和传输的视频数据量也随之成倍增加。而视频数据是信息量最大的一类多媒体数据,多个视点的庞大数据量给计算机处理带来了很大的困难,对数据的高效压缩成为多视点视频应用亟待解决的问题。

本文在第 2 部分首先分析了时域预测的编码方法;第 3 部分论述了多视点视频编码的时空联合预测原理,列举了几种典型的多视点编码预测结构并且分析了各方案的结构特点,第 4 部分提出一种新的基于分级 B 图的多视点视频编码方案,并且通过实验进行了性能对比分析,最后对研究的结果做了总结与展望。

### 2 时域预测

运动图像是通过在一个矩形窗口中以一定时间间隔连续拍照进行捕获的,将这一组帧序列回放就产生了运动的效果。

<sup>\*</sup>)北京市教育委员会科技发展计划面上项目(NO:KM2005112320092)。潘 峰 硕士研究生,研究方向为视频编码技术。

视频帧序列之间的变化,有可能来自物体运动,包括刚体运动(如飞驰的汽车)和可变形物体的运动(如人手的移动);也可能来自摄像机的移动,如镜头的晃动、倾斜、缩放、旋转;还有可能来自暴露的区域(如物体移开后露出的背景图像)或是灯光的色彩和明暗变化。

由于邻近帧中的景物存在着一定的相关性,因此可以通过搜索邻近帧图像中运动像素所在的位置,从而得到两帧之间的空间位置的相对偏移量。得到的相对偏移量就是运动矢量,而得到运动矢量的过程就称为运动估计。在主流的视频编码标准如 H. 264 中,都采用宏块(图像帧中  $16 \times 16$  的像素块)作为运动补偿预测的基本运算单位。宏块的运动估计过程就是在参考帧中寻找与当前宏块匹配的像素区域。当前宏块减去其最佳匹配宏块得到残差宏块,运动矢量和经过运动匹配后得到的预测误差共同发送到解码端,根据运动矢量标识的位置,从已解码的参考帧图像中寻找相应的宏块,和预测残差相加后就得到了宏块在当前帧中的位置。

运动预测有效利用了运动图像的时域相关性,是预测编码方法在运动图像处理中的典型应用。通过运动估计,可以大幅去除帧间冗余,减少视频传输的数据,达到数据压缩的目的。因此,运动估计是视频数据处理中最常采用的压缩方法之一。

### 3 多视点时空联合预测方案

多视点视频信号是在同一时刻由多个摄像机对场景或对象的多路视频采样,视点间的内容由于相机位置和角度的不同而存在一定的差异。但是,与单视点视频数据的帧间相似性类似,由于拍摄的视频场景相同,多个视点间的视频内容同样存在相似性。各视点内容的相似意味着视频数据的冗余,视点间的相似程度越高,其数据的冗余也就越大。因此,在对多视点视频信号处理时,不仅要利用时间相关性消除同一视点内时域上的冗余数据,还要考虑到各视点间冗余数据的消除。

视差估计是消除视点间的数据冗余最常采用的方法,也是多视点视频编码的关键技术<sup>[5]</sup>。所谓视差是指同一个点在两幅图像中投影点之间的位置偏差,包括各摄像机的相对运动及场景中对象的运动所产生的差异。视差估计与运动估计类似,都是利用两幅图像之间的相关性达到数据压缩的目的,运动估计是寻找两幅图像之间的时域相关性,而视差估计是寻找各视点图像之间的空域相关性。采用视差估计的压缩方法,可以提出不同的结合时间相关性和视点相关性进行时空联合预测的编码方案。

#### 3.1 联播方案

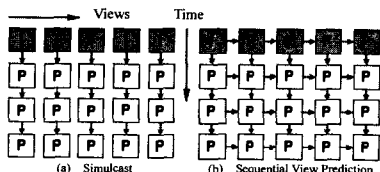


图 2

联播方案(Simulcast)不考虑视点间的关联,是一种最简单的多视点视频编码方案,这种方案是对单一视点视频编码的简单扩展<sup>[6]</sup>。图 2(a)所示的是一种典型的联播方案,与单一视点视频编码相同,各路视频均独立编码,每路视频的关键帧使用帧内编码(I-picture),其余帧参考时间轴上的前面帧作

预测进行编码(P-picture),各路视频数据互不参考。这种方案并没有利用视点间的相关性,因而未能消除视点间的数据冗余,编码效率低,通常仅被作为评价其它多视点视频编码方案的性能对比参考。

#### 3.2 Sequential View Prediction 方案

对于一系列物理位置连续的视点,各视点内容的相似程度非常高,如果利用相邻视点相关性构造编码方案,可以大大降低编码冗余。图 2(b)表示的是一种对连续视点采用“IPPP”结构进行时空预测的编码方案<sup>[7]</sup>。首个视点的视频序列正常编码,即只使用当前已经编码的图像帧通过运动估计进行预测编码;第二个视点的帧图像由首视点的帧图像通过视差估计获得,第二帧图像通过其相邻视点及前一时刻的图像共同预测编码;其余视点的编码方法类同。

这种预测方案有效地利用了视点相关性,提高了编码效率;缺点是可能造成编码误差累积,并且不能较好地支持随机访问。在码率条件允许的情况下,可以通过减少每个 GOP 的帧数,及时更新 I 帧来提高编码质量。

#### 3.3 Views-Synthesis Prediction 方案

8 路视频流是常见的多视点系统模式,图 3 展示了一种对 8 路视频数据通过时空预测合成视点的编码结构<sup>[8]</sup>。将物理位置居中的一个视点作为预测主路,其第一帧 I 采用帧内编码,之后各帧  $P_i$  采用运动估计的方法进行前向预测;其它各路的数据  $P_j$  由处于同一时刻的主路数据通过视差估计预测获得。可以看到,除了主路数据的时间相关外,其它各视点的数据都没有使用时间相关性进行预测。

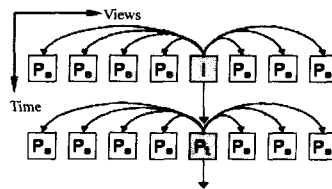


图 3 Views-Synthesis Prediction

由于各路视频数据在同一时刻只与预测主路数据相关,其它各路数据之间互不影响,因此这种方案在利用视点相关性提高编码效率的同时,也兼顾到随机访问的应用需求。

### 4 基于分级 B 图的 MVC 预测方案

JVT 提出分级 B 图(Hierarchical B pictures)的概念,将每个 GOP 中的 B 帧依据预测的先后次序视作不同的等级,较高等级的 B 帧可以继续为较低等级的 B 帧提供预测参考<sup>[9]</sup>,其优点是在保持较低码率要求的情况下可以有效提高图像质量。

本文将分级 B 图的编码方法引申到多视点编码中,提出一种基于分级 B 图的多视点视频编码方案。这种方案结合上文提到的视点合成预测方案的编码特点,同时将分级 B 图应用到多视点视频编码的 GOP 中。如图 4 所示,首先对物理位置居中的一路视频的第一帧图像采用帧内编码( $I_0$ ),然后采用单向预测获得其它各视点的帧图像( $P_0$ );对于每个视点内的 GOP,则采用分级 B 帧的编码方法,由 I/P 帧及下一个 I/P 帧双向预测得到  $B_1$  帧,再由 I/P 帧和  $B_1$  帧双向预测得到  $B_2$  帧, $B_3$  帧由相邻时刻的前后两帧双向预测得到,各视点 GOP 间的所有 B 帧互不参考。由于主路视点的物理位置居中,因此主路视点与两侧其它视点的视差相对较小,加之

各视点间 B 帧互不参考,这样就在很大程度上避免了由于视点间的预测引起的误差积累。另外,各视点的编解码只以主路视点为预测前提,其它各视点互不影响,有利于提高随机访问性能。

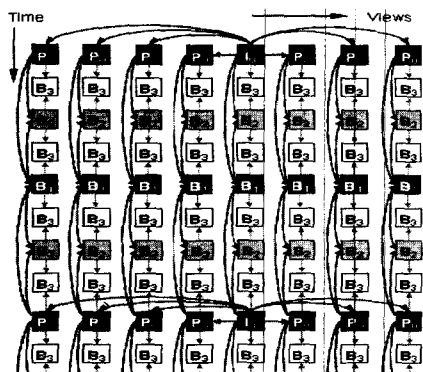


图4 本文方案

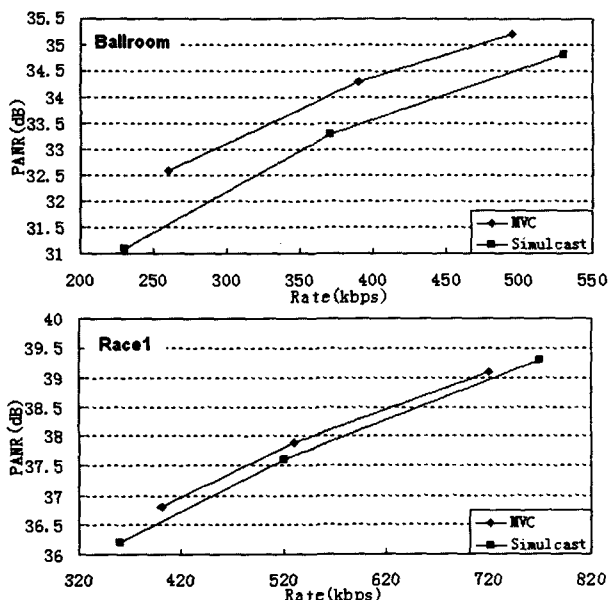


图5 率失真性能比较

本文在 H. 264/AVC 的 JM9.0 校验模型中验证方案的有效性,分别对 Ballroom 和 Race1 两个多视点测试序列进行编码实验。实验采用的测试序列 Ballroom 和 Race1 均由 8 路水平照相机阵列采集,其中 Ballroom 由 MERL 提供<sup>[10]</sup>,其相邻相机间距 19.5cm,帧率 25 帧/秒,分辨率 640×480,视点间内容相似程度较高;Race1 由 KDDI 提供<sup>[11]</sup>,其相邻相机间

距 20cm,帧率 30 帧/秒,分辨率 320×240,场景内容较为复杂。两个测试序列分别反映方案环境良好和较差时的编码性能。实验得到的率失真曲线如图 5 所示,图中将本文方案与联播方案进行了性能对比,MVC 表示本文方案,Simulcast 表示联播方案。从图中可以看出,本文方案对于视点相关性较强的序列,可以获得比联播方案约 1dB 的增益;对于视点间相似度较低的序列,也能获得 0.2dB~0.4dB 的增益。实验结果表明,本文的方案对于视点相关性较低的图像序列,在相同码率条件下,峰值信噪比提升明显;对于视点间相关性较高的运动图像序列,编码性能则更加理想。

**结束语** 本文提出一种新的可在 H. 264/AVC 下实现的多视点编码方案,应用视差估计的压缩算法,通过时空联合预测编码和分级 B 图等途径,在兼顾较高压缩比的同时,有效提高了随机访问性能。

本文采用运动补偿和视差估计作为联合预测编码算法,虽然易于实现,但编码效率在很大程度上依赖于摄像机的空间采集密度,编码性能不稳定。此外,对于由于网络条件的限制带来的视点可分级问题,以及面向实时交互应用带来的低延时问题,都是我们下一步将要研究的内容。

### 参考文献

- [1] Matusik W, Pfister H. 3D TV: A Scalable System for Real-Time Acquisition, Transmission and Autostereoscopic Display of Dynamic Scenes. ACM Transaction on Graphics (TOG) SIGGRAPH, ISSN: 07300301, 2004, 23(3): 814-824
- [2] ISO/IEC JTC1/SC29/WG11 Requirements for 3D AV Coding [C]. DOC. N4795, USA, May 2003
- [3] ISO/IEC JTC1/SC29/WG11. Call for Proposals on Multi-view V-video Coding [C]. DOC. N7327, Poland, July 2005
- [4] ISO/IEC JTC1/SC29/WG11. Joint Multiview Video Model (JMVM) 5.0 [C]. DOC. X207, Switzerland, July 2007
- [5] Flierl M, Girod B. Multi-view Video Compression. IEEE Signal Processing Magazine [J], July 2007
- [6] ISO/IEC JTC1/SC29/WG11. Call for Evidence on Multi-View Video Coding [C]. DOC. N6720, Palma de Mallorca, Spain, 2004
- [7] ISO/IEC JTC1/SC29/WG11. Survey of Algorithms used for Multi-view Video Coding (MVC) [C]. DOC. 6909, China, 2005
- [8] Morvan Y, Farin D, de With P H. Acquisition, Compression and Rendering of Multi-view Video [R]// NXP workshop. TU/e, Netherlands, Aug. 2007
- [9] ISO/IEC JTC1/SC29/WG11. Hierarchical B pictures [C]. DOC. P014, Poland, July 2005
- [10] ftp://ftp.merl.com/pub/avetro/mvc-testseq
- [11] ISO/IEC JTC1/SC29/WG11. KDDI multi-view video sequences for MPEG 3DAV use [C]. 2004

(上接第 238 页)

- [5] Zhou Mou-yan. Deconvolution and Signal Recovery. Beijing: Defense Industry Press, 2001
- [6] Cheenama T A, Qureshi I M, Hussain A. Blind Image Deconvolution Using Space-variant Neural Network Approach. Electronics Letters, 2005, 41(6)
- [7] Yin Hujun. Blind Source Separation and Genetic Algorithm for Image Restoration // Proceedings of IEEE Conference on Advances in Space Technologies. 2006: 167-172
- [8] Cichocki A, Amari S I. Adaptive Blind Signal and Image Processing. Beijing: Electronics Industry Press, 2005
- [9] Ohm J R, Grüneberg K, Hendriks E, et al. A Realtime Hardware System for Stereoscopic Video Conferencing with Viewpoint

Adaptation. Signal Processing Image Communication, 1998, 14: 147-171

- [10] Chen Yen-wei, Nakao Z, Arakaki K, et al. Restoration of Gray Images Based on a Genetic Algorithm with Laplacian Constraint. Fuzzy Sets and Systems, 1999, 103: 285-293
- [11] Enolura T, Chen Yen-wei, Nakao Z. A Fast Image Algorithm for Image Restoration Based on a Hybrid GA and SA // Proceedings of IEEE Conference on Systems, Man, and Cybernetics, 1999, 4: 891-894
- [12] Jonson E G, Abushagur M A G. Image Deconvolution Using a Micro Genetic Algorithm. Optics Communications, 1997, 140: 6-10
- [13] Banham M R, Katsaggelos A K. Digital Image Restoration. IEEE Signal Processing Magazine, 1997, 14(2): 24-41