

支持空域随机访问的多视点视频编码方法

刘延伟^{1,2)} 黄庆明^{1,2)} 季向阳¹⁾ 霍龙社³⁾ 高文³⁾

¹⁾(中国科学院计算技术研究所数字媒体研究中心 北京 100190)

²⁾(中国科学院研究生院 北京 100049)

³⁾(北京大学数字媒体研究所 北京 100871)

(ywliu@jdl.ac.cn)

摘要 基于 H.264/AVC 编码框架,提出一种支持空域随机访问功能的多视点视频编码方法,以满足快速虚拟视点绘制的需求.首先采用灵活的 GOP 时空划分来构建独立时空体,从而在每个时空体内通过限制帧间预测范围来增强解码的独立性,实现压缩域内某一帧特定区域数据的快速定位;进而通过调整预测空间来改善空域随机访问和整帧编码性能之间的关系.实验结果表明,该方法不仅在整帧压缩效率和快速空域随机访问灵活性方面达到了很好的平衡,而且能够通过传输部分码流的方式节省传输带宽.

关键词 多视点视频编码;空域随机访问;独立时空体编码

中图法分类号 TP391

Multi-view Video Coding with Spatial Random Access Functionality

Liu Yanwei^{1,2)} Huang Qingming^{1,2)} Ji Xiangyang¹⁾ Huo Longshe³⁾ Gao Wen³⁾

¹⁾(Digital Media Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

²⁾(Graduate University of Chinese Academy of Sciences, Beijing 100049)

³⁾(Institute of Digital Media, Peking University, Beijing 100871)

Abstract To obtain fast rendering effects of virtual viewpoint video, in this paper we propose a novel multi-view video coding method with spatial random access functionality based on H.264/AVC codec. An independent space-time cube coding is first proposed to limit the scope of inter-frame prediction via flexible re-organizing GOP partitions in time and space dimensions, which guarantees the independent decoding of each cube. As a result, random access of specific regions in the compressed data of a picture can be accelerated. The conflict of performance between full-frame compression and spatial random accessibility is then further ameliorated through regulating inter-prediction space. The results demonstrate that the proposed coding scheme not only makes a good trade-off between full-frame compression and spatial random access flexibility, but also supports selective streaming to save the channel bandwidth.

Key words multi-view video coding; spatial random access; independent space-time cube coding

多视点视频从不同角度表现丰富的客观世界,提供给用户从任意视点和不同视角交互浏览的特

色.基于这种独特的表现能力,多视点视频广泛应用于三维数字电视、自由视点视频通信等诸多领域^[1].

收稿日期:2007-11-12;修回日期:2008-04-14. **基金项目**:国家自然科学基金项目(60672088, 60736043);国家“八六三”高新技术研究发展计划项目(2007AA01Z315). **刘延伟**,男,1976年生,博士研究生,主要研究方向为多视点视频处理. **黄庆明**,男,1965年生,博士,教授,博士生导师,主要研究方向为多媒体技术、图像/视频分析与理解等. **季向阳**,男,1976年生,博士研究生,主要研究方向为视频压缩. **霍龙社**,男,1968年生,博士,高级工程师,主要研究方向为多媒体通信、多视点视频处理等. **高文**,男,1956年生,博士,教授,博士生导师,CCF高级会员,主要研究方向为多媒体数据压缩、计算机视觉、人工智能等.

多视点视频压缩是解决多视点视频通信中带宽限制问题的有效手段,它不仅要去掉每一路视频时间方向的统计信息冗余,而且还要压缩多路视频信号之间的相关信息冗余.针对自然采集的视频,研究人员提出了大量的多视点视频压缩算法^[2-3],其中基于 H.264/AVC 的利用分层次 B 帧的多视点视频编码机制显著提高了压缩性能,被用作多视点视频编码标准的基本框架^[2].

虽然多视点视频编码利用视差预测来提高压缩性能,但限制了灵活的自由视点切换能力和快速介入压缩域数据的随机访问能力.为了增强浏览的连续性,交互式多视点视频系统需要在原始采集的多个视点之间绘制新的视点.由于这种新视点的绘制通常只需要原始采集视点图像的部分数据,因此一种支持空域随机访问功能的编码方法对快速的虚拟视点绘制是十分必要的^[1,4].通过空域随机访问抽取局部图像的码流进行有选择的传输,可以避免传送一些不必要的信息,节省信道带宽.空域随机访问是指随机地介入到图像中某一局部区域,它是多视点视频编码中的一项基本需求^[1].文献[5]采用动态三维点采样的数据表示形式提供了较强的空域随机访问能力,但其表示形式仅适合于虚拟的合成视频,不适用于自然采集的多视点视频.目前,基于 H.264/AVC 的多视点编码框架只能通过整幅图像完全解码之后来支持空域随机访问功能,这种方法访问速度很慢,无法满足交互式虚拟视点绘制时快速获取数据的需求.

本文基于 H.264/AVC 的多视点编码框架提出了一种支持快速空域随机访问的编码方法.该方法进行分区域编码并限制不同区域之间的相互预测,每一区域在每一个图像组(group of pictures, GOP)内和其依赖的所有参考帧的相应区域在三维空间内构成一个独立时空体,每个独立时空体作为一个独立的编解码单元,能够支持快速的随机访问.

1 空域随机访问

现有的基于图像的虚拟视点绘制大多从相邻的视点向将要合成的虚拟视点进行数据映射,然后通过融合处理的办法来产生新的视点信号^[6].由于仅使用视点依赖的映射,且考虑到视角辐射范围的关系,在虚拟视点图像合成的过程中通常并不需要用到相邻视点的全部图像.图 1 所示为 2 种不同视角的虚拟视点合成过程,其中阴影部分表示合成中间虚拟视点图像所必需的相邻视点的部分图像,中间视点是待合成的虚拟视点.由于虚拟视点绘制对数据获取的速度要求很高,直接得到请求的压缩数据并独立解码,然后参与视点绘制是理想的数据访问方式.

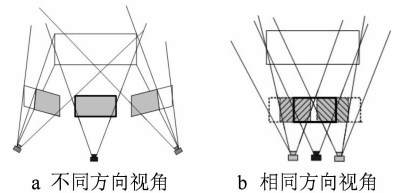


图 1 虚拟视点图像合成示意图

基于 H.264/AVC 的多视点编码框架在空域访问方面仅仅提供了有限的视点分级特性,缺少对快速帧内空域数据访问机制的支持.为了扩展此项功能,本文根据分区域编码思想,首先对编码的每一帧进行规则的矩形网格图像区划分,然后对划分的图像区进行独立时空体编码.为了平衡压缩效率和空域访问灵活性之间的关系,我们进一步提出扩展的独立时空体编码方法.图 2 a 中,输出码流描述了在 T 时刻 N 个视点的的数据记录结构, $Tile$ 表示图像区, $Suffix$ 表示填充数据, $View$ 表示视点;每一个图像区的编码形成一个数据包,访问时通过起始码来进行定位.

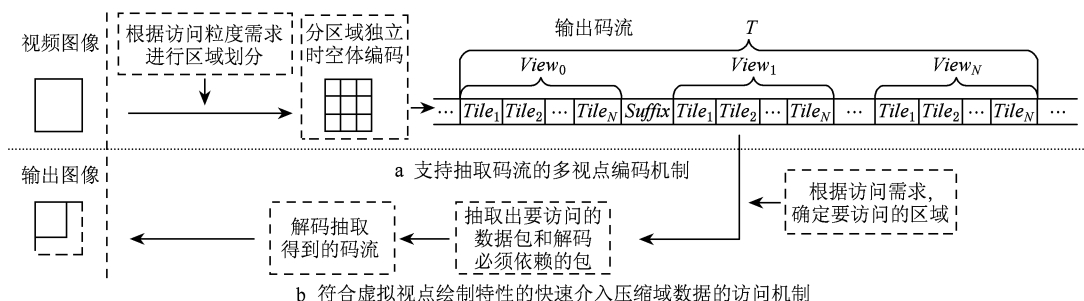


图 2 支持空域随机访问功能的多视点编码和访问机制

2 支持空域随机访问的独立时空体编码方法

2.1 规则图像区的划分

分区域编码采用规则的图像区划分. 图像区在编码时由条带组来描述, 每一图像区可以由一个或多个条带组组成, 条带组再根据灵活的宏块排序技术(flexible macroblock ordering, FMO)进行编码. 图像区的划分尺寸由空域随机访问粒度的具体要求来决定. 当采用 FMO 时, 编码器需要传送宏块与条带组之间的映射关系数据, 这增加了传输的码率. 由于划分后图像区的数量对编码性能会有很大影响, 因此规则图像区的划分首先要能够覆盖需要访问的最小区域面积, 然后尽可能地减少由划分引起的传输代价. 由文献[7]的分析可以看出, 在空域随机访问尺度小于一帧大小的前提下, 整帧编码性能随着空域随机访问尺度增大满足某种近似的单调递增关系, 如图 3 所示. 图 3 中, $f(S, P)$ 表示整帧编码性能 P 和空域访问尺度 S 之间的函数; P_F 和 S_F 分别为不采用划分区域的整帧编码性能和整帧大小; S_g 为访问需求的粒度, 它决定虚拟视点绘制时所需要的相邻视点部分图像面积的准确程度. S_g 越小, 取得的部分图像的面积越精确; 反之, 则越不精确. 根据 S_g , 可通过如下方法确定最优的随机访问尺度 S_T , 从而找到满足 S_g 的图像区划分和整帧编码性能的平衡点. 假设访问需求的粒度 S_g 的宽为 W_g 、高为 H_g , 图像宽为 W 、高为 H , 划分的图像区宽为 W_T 、高为 H_T , 则对于给定的 W_g 和 H_g , W_T 和 H_T 必须满足以下原则.

原则 1. W_T 和 H_T 要能够被 16 整除, 并且满足 $W_T = \min \{W_i | W \geq W_i \geq W_g\}$, $H_T = \min \{H_i | H \geq H_i \geq H_g\}$.

原则 2. 整幅图像不一定划分成大小完全相同的图像区, 但每个图像区都要满足原则 1.

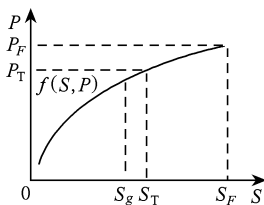


图 3 整帧编码性能和图像区划分的关系

2.2 独立时空体编码

分图像区编码把每一图像区作为一个独立的编解码实体, 需要为之设计相应的编码机制. 在多视点

编码框架中, 同一时刻一个视点的图像将作为其相邻视点的参考图像进行视差预测. 为了支持帧内部分图像的快速访问, 时间方向的运动向量预测和视点间的视差向量预测都必须限制在一定的范围之内. 为此, 我们将具有视间预测关系的多个视点在同一时间间隔内所对应的若干 GOP 定义为一个 GGOP(group of GOP). 时间方向的帧间预测和空间方向的视间预测都必须局限在该 GGOP 所确定的有限时空体内进行, 称为独立时空体编码. 如图 4 所示, T_N 表示一个 GOP 长度, V_0 的 T_0 帧中的灰色区域小图像区仅参考视点 V_0 的 T_0, T_N 帧的灰色图像区以及视点 V_1 的 T_1 帧的灰色图像区. 根据独立时空体编码原理, 独立时空体编码算法包含图像区划分、独立时空体编码以及修改插值和环路滤波处理 3 个部分. 独立时空体编码的关键步骤是如何设置预测范围. 由独立时空体的定义, 我们给出编码块范围预测准则如下:

$$\begin{cases} Hor_{min} = \max \{(-MbX \ll 4 - (16 + 8)) \ll 2, \\ ((L_A_X - MbX) \ll 4) \ll 2\} \\ Ver_{min} = \max \{(-MbY \ll 4 - (16 + 8)) \ll 2, \\ ((L_A_Y - MbY) \ll 4) \ll 2\} \\ Hor_{max} = \min \{(((MbInFrameX - MbX) \ll 4) + 8) \ll 2, ((R_D_X - MbX) \ll 4) \ll 2\} \\ Ver_{max} = \min \{(((MbInFrameY - MbY) \ll 4) + 8) \ll 2, ((R_D_Y - MbY) \ll 4) \ll 2\} \end{cases} \quad (1)$$

其中, L_A_X, L_A_Y, R_D_X 和 R_D_Y 分别表示当前编码图像区的左上角和右下角的水平 and 垂直坐标; MbX 和 MbY 分别表示当前编码宏块的以块索引度量的水平和垂直坐标; $MbInFrameX$ 和 $MbInFrameY$ 分别表示当前帧在水平和垂直方向上包含的宏块数; $Hor_{min}, Ver_{min}, Hor_{max}$ 和 Ver_{max} 分别表示水平方向和垂直方向的 1/4 像素精度预测范围的最小和最大值. 在块匹配搜索时, 通过在编码配置时设置的搜索范围 $[-p, p-1]$ 与式(1)计算得到的最大最小值之间进行适当比较, 最终确定每个编码块运动向量的范围.

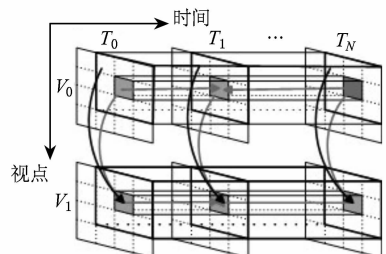


图 4 独立时空体编码原理

2.3 图像区边界处的亚像素插值和环路滤波处理

在基于 H.264/AVC 的多视点编码器中,亚像素插值提高了运动和视差估计的精度.为了配合 1/4 亚像素的上采样插值操作,在每一帧的边界处采用扩展填充处理.独立时空体编码算法在分割的图像区边界处采用了类似的扩展填充处理.编码的图像区在边界处放大 2 个整像素的宽度,然后利用重复的边界像素来进行填充,这种操作可以进一步提高图像区边界处的预测效率.图 5 所示描述了边界处详细的扩展填充操作,其中,左边是整帧 1/4 亚像素图像,右边是整帧 1/4 亚像素图像经过划分后的图像表示,中间的箭头代表了划分过程,左上部分是边界处像素扩展填充操作的放大示意图.为了使每一个图像区都能够被独立访问,我们不在图像区的边界处进行环路滤波,从而使不同的图像区相互之间割断联系.

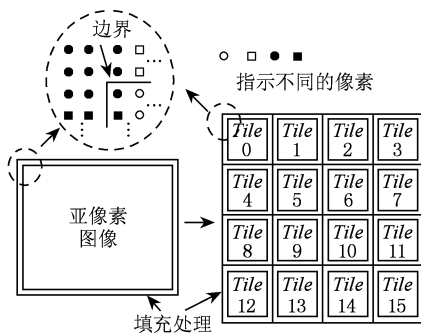


图 5 在图像区边界处附加的扩展填充处理

2.4 独立时空体时间方向长度的确定

独立时空体编码中时间方向 GOP 长度 T_N 的选择会直接影响到空域随机访问中抽取立方体的大小.根据实际空域访问速度的需求约束,通过运动预测原理对时间方向随机访问的灵活性进行定量估计,以确定 T_N 的大小.假设从已经解码的一个或几个参考帧进行预测,然后加上残差并重建图像的计算时间为 T_d ,解码一个 I 帧的计算时间为 T_I ;根据采用开环 GOP 的二重对称分层次 B 帧多视点预测编码结构,时间方向 GOP 的长度决定了最大的层次 L_{\max} .在一个 GGOP 中,具有最大参考帧数量的帧的参考帧数量^[2]为

$$F_{\max} = 3 \cdot L_{\max} + 2 \cdot \left\lfloor (N_{\text{view}} - 1) / 2 \right\rfloor \quad (2)$$

其中 N_{view} 表示参与编码的视点数量.在一个 GGOP 内,解码一帧所需要的最大时间 T_{\max} 近似为

$$T_{\max} \approx (F_{\max} - 2) \cdot T_d + 2T_I \quad (3)$$

其中 T_I 和 T_d 是实际编码之前通过简单实验可以确定的近似常量.假设实际空域随机访问需求约束为 $T_{\text{constraint}}$,则根据式(2),(3)便可以确定满足 $\max \{T_{\max} \leq T_{\text{constraint}}\}$ 的 T_{\max} 和 L_{\max} ,进而估计出适当的 T_N 值.

3 扩展的独立时空体编码方法

约束预测范围对整帧编码性能具有一定的负面影响.为了扼制这种影响,本文提出一种扩展的独立时空体编码技术.图 4 中,预测范围可以扩大到环绕灰色图像区较大的深色边界矩形区域.由于视差向量通常远远大于时间方向的运动向量,因此这种预测区域的放大对视间预测的性能恶化可以起到一定的抑制作用.然而,预测区域的扩大又带来了新的预测依赖问题.如图 6 所示,在解码图像区 4 之前,位于其参考帧中周围的图像区 0~3 和 5~8 必须首先解码;而当解码参考帧中这些图像区之前,它们所依赖的相应的参考帧中的图像区又必须先解码,这样在图像区之间就产生了时间和空间上的依赖关系.由于空域随机访问要求只抽取相应编码的立方体数据,使解码时有些参考帧的部分区域会因参考帧预测信号在编解码阶段不一致而导致失真,这种失真还会传递给参考它的其他部分信号.根据运动预测原理,图 6 所示参考帧中最有可能产生失真的部位是图像区 0~3 和 5~8 中靠近扩展立方体边界的部分像素.为了避免这种失真传递的影响,我们采用缩小预测空间的办法,在编码 T 时刻的图像区 4 时,在其参考帧中预测范围局限在比扩展立方体小一些的区域,如图 6 中虚线所示的较小的矩形区域.通常情况下,预测区域限制在高度和宽度上比扩展立方体小一个宏块的内部区域.

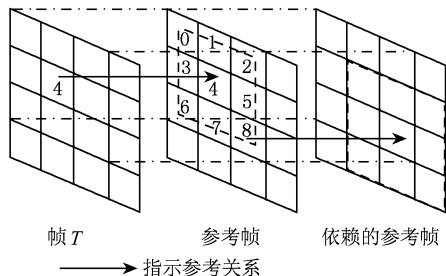


图 6 扩展的独立时空体编码技术

4 实验与分析

为了评价本文所提多视点编码算法的压缩性能

与空域随机访问之间的关系,我们修改 JMVM3.0 软件进行了若干组实验.实验序列取自于美国三菱研究院的 ballroom 和微软公司的 breakdancer^[8].这 2 个序列分辨率分别为 640×480 和 1024×768 .为了支持低延迟的数据访问,根据式(2),(3)的计算,实验中采用 GOP 长度为 4 的分层次 B 帧编码结构.

实验 1. 测试区域划分对整帧编码性能的影响.

图 7 所示分别给出了 ballroom 和 breakdancer 序列在不同的图像区划分数量下的编码性能对比,其中 ISCC_4Tile, ISCC_12Tile 和 ISCC_16Tile 分别代

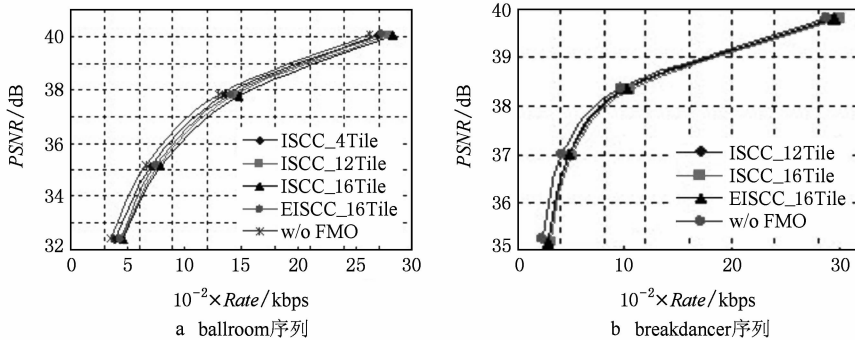


图 7 基于不同划分的码率失真性能评价

实验 2. 对比不同编码方法的空域随机访问速度和选择性传输的性能.

在不支持空域随机访问的情况下,现有的编码方法通过整帧的完全解码来实现帧内特定区域的访问,而独立时空体编码则可以通过抽取码流的方式进行访问.针对 ballroom 序列,表 1,2 给出了几种编码方式在 QP 为 22 时,部分区域抽取并解码的访问速度和率失真性能方面的对比.可以看出,本文提出的编码方法在访问速度上较原有编码方法(w/o FMO)最高可以提高约 40%;单视点数据对比我们采用了第一个视点数据,其编码并没有采用视间预测;多个视点的数据对比我们采用了多个视点平均的办法.为了避免由于引入过多的视点预测而导致严重的解码依赖,只采用 2 个视点的视间编码预测,这样可以提高整体的数据访问速度.表 1,2 中的数据同时表明,在只需要部分图像区域的情况下,传输抽取区域的率失真性能比传输整帧图像的率失真性能会好很多,能够节省大量的带宽.表 2 中,在只传输 1/4 帧的情况下,相对于整帧传输,16 个图像区的独立时空体编码最大可以节省 70%左右的码率,16 个图像区的扩展独立时空体编码最大可以节省 30%左右的码率.

表具有 4,12 和 16 个图像区划分的独立时空体编码算法;EISCC_16Tile 表示具有 16 个图像区划分的扩展独立时空体编码;w/o FMO 表示不采用图像区划分的编码.从图 7 中可以看出,与不进行图像区划分编码相比,独立时空体编码在不同的划分情况下导致整帧编码性能出现不同程度的损失.在 16 个图像区划分情况下,编码性能下降 20%左右.在整帧编码性能对比上,扩展的独立时空体编码比独立时空体编码有更好的结果.对于 16 个图像区划分的 ballroom 序列,扩展独立时空体编码比独立时空体编码性能提高约 10%.

表 1 第一个视点采用不同划分的数据访问速度与码率、失真对比

视点 0 数据	抽取 1/4 帧时间/ms	1/4 帧 PSNR/dB	1/4 帧 码率 kbps	整帧 PSNR/dB	整帧 码率/kbps
w/o FMO	14 391	40.56	3 031.85	40.14	3 031.85
4Tile_ISCC	9 117	40.01	778.44	40.13	3 081.09
16Tile_ISCC	9 765	40.82	547.94	40.12	3 180.34
16Tile_EISCC	12 062	40.55	2 098.07	40.14	3 157.49

表 2 2 个视点采用不同划分的访问速度与码率、失真对比

2 个视点 平均值	抽取 1/4 帧时间/ms	1/4 帧 PSNR/dB	1/4 帧 码率 kbps	整帧 PSNR/dB	整帧 码率/kbps
w/o FMO	15 812	40.55	2 776.57	40.17	2 776.57
4Tile_ISCC	11 734	39.89	710.37	40.16	2 840.66
16Tile_ISCC	12 265	39.98	506.99	40.15	2 956.79
16Tile_EISCC	13 138	40.56	1 878.97	40.17	2 886.02

实验 3. 评价本文方法对主观质量的影响.图 8 所示为分别采用不分图像区编码、分 16 个图像区的独立时空体编码、扩展独立时空体编码的一帧重构图像以及抽取后的部分图像的主观质量情况.可以看出,主观质量几乎没有差别,尤其是抽取区域的边界处几乎没有什么失真.

由实验 1,2 可以看出,通过设计最优的图像区域划分,本文方法能够在压缩性能和访问的灵活性之间达到一个很好的权衡. 3 个实验的结果分析表明,抽取部分码流的率失真性能高于整帧的率失真性能,而且抽取出的部分图像具有较好的重构质量,独立时空体编码思想完全可以支持视点依赖的部分码流抽取,满足快速视点绘制的需求.

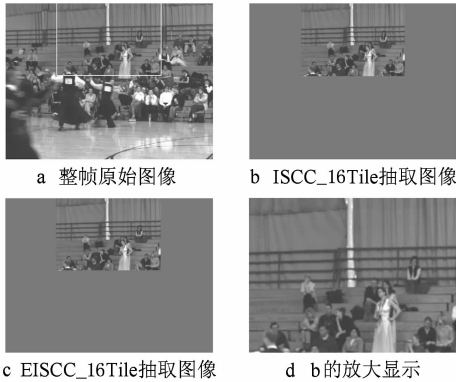


图 8 主观质量比较 (ballroom 序列)

5 结 论

本文根据区域划分编码思想,提出了一种支持快速空域随机访问的多视点视频编码方法. 与已有的多视点编码方法相比,该方法能够支持快速的部分码流抽取和解码,并可以使得压缩性能和随机访问速度之间得到一个较好的折中. 在交互式多视点视频系统中,本文方法不仅能够做到快速地从压缩域读取数据来支撑交互式虚拟视点视频的合成,还具有选择性地传输部分码流节省带宽的优势. 我们

下一步的研究重点是改进压缩码流的记录结构,以进一步提升局部码流的访问速度.

参 考 文 献

- [1] ISO/IEC JTC1/SC29/WG11 N8064—2006 Multiview video coding requirements [S]
- [2] ISO/IEC JTC1/SC29/WG11 m12945—2006 Multiview coding using AVC [S]
- [3] He Renlong, Jiang Gangyi, Fu Randi, *et al.* A multiview video coding method with adaptive selection of reference frame modes [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2007, 19(9): 1132-1137 (in Chinese) (贺人龙, 蒋刚毅, 符冉迪, 等. 基于自适应参考帧模式选择的多视点视频编码方法[J]. *计算机辅助设计与图形学学报*, 2007, 19(9): 1132-1137)
- [4] Tong X, Gray R M. Interactive rendering from compressed light fields [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, 13(11): 1080-1091
- [5] Lamboray E, Würmlin S, Waschnusch M, *et al.* Unconstrained free-viewpoint video coding [C] // *Proceedings of IEEE International Conference of Image Processing*, Singapore, 2004: 3261-3264
- [6] Shum H Y, Kang S B, Chan S C. Survey of image-based representations and compression techniques [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, 13(11): 1020-1037
- [7] Leung R, Taubman D. Transforms and embedded coding techniques for maximum efficiency and random accessibility in 3D scalable compression [J]. *IEEE Transactions on Image Processing*, 2005, 14(10): 1632-1646
- [8] ISO/IEC JTC1/SC29/WG11 N8019—2006 Description of core experiments in MVC [S]