

AVS 视频编码标准技术回顾及最新进展

马思伟

(北京大学信息科学技术学院 北京 100871)

(swma@pku.edu.cn)

History and Recent Developments of AVS Video Coding Standards

Ma Siwei

(School of Electrical Engineering and Computer Science, Peking University, Beijing 100871)

Abstract AVS (audio video coding standard) is the informal name of Work group for Digital Audiovideo Coding Standard of China, which was founded by the Science and Technology Department under former Ministry of Information Industry in June 2002, approved by Standardization Administration of China. The role of the group is to establish general technical standards for the compression, decoding, processing, and the representation of digital audio-video, thereby enabling digital audio-video equipment and systems with high-efficiency and economical coding/decoding technologies. After more than ten years, AVS has established a series of video coding standards, including AVS1, AVS+ and AVS2. AVS1 and AVS2 are named from the first and second stage work of AVS, or the first and second generation standard, and AVS+ was established specially for China high definition TV broadcasting specially. AVS1 and AVS+ have been finished and widely used in various applications so far, and AVS2 is still under developing and will be released soon. This paper provides an overview of the history and recent developments of AVS video coding standards, including the key tools used in AVS and the comparison with the state-of-the-art technology, e. g. HEVC/H.265. Moreover, a brief discussion and conclusion on the future video coding are provided.

Key words video coding; audio video coding standard (AVS); transform; prediction; entropy coding; loop filter

摘要 AVS(audio video coding standard)工作组是“数字音视频编解码技术标准”工作组的简称,由国家原信息产业部科学技术司于2002年6月批准成立.工作组的任务是制(修)订数字音视频的压缩、解压缩、处理和表示等共性技术标准,为数字音视频设备与系统提供高效经济的编解码技术.经过10多年的努力,AVS已经成功制定一系列标准,如AVS1,AVS+和AVS2.其中AVS1和AVS2的说法源于AVS第1阶段和第2阶段的工作或第1代、第2代标准.AVS+是工作组专门为高清数字电视广播制定的一个标准.目前AVS1和AVS+标准已经制定完成并成功应用于许多行业中,AVS2也即将完成.将对AVS的AVS视频标准12年来的技术发展过程给以概述,并简单介绍目前最新一代视频编码标准AVS2关键技术,最后对视频编码技术的未来发展方向予以展望.

关键词 视频编码;AVS;变换;预测;熵编码;环路滤波

中图分类号 TP391

计算机开启了信息处理的数字化时代,使得各种类型的数据都可以通过数字化利用计算机进行处

理,但数字化过程中的过采样也引入了大量的数据冗余.特别是视频数据,未经压缩的原始视频数据量

之大至今仍仍是存储和传输难以承受的,因此视频压缩一直是多媒体领域的热门研究问题之一。

早在 20 世纪 50 年代就开始了针对对图像视频压缩的研究,比如针对图像统计冗余特性的哈夫曼码^[1]、基于像素的差分预测编码^[2]等。之后,20 世纪 70 年代是视频编码技术发展的一个重要阶段,出现基于块的运动补偿预测^[3]、离散余弦变换(discrete cosine transform, DCT)^[4]等,形成了经典的预测变换混合编码框架^[5],为不久 ISO/IEC MPEG 工作组和 ITU-T VCEG 工作组制定 MPEG-1, H. 261 标准奠定了基础。自 20 世纪 90 年代至今 MPEG 和 VCEG 已或独立或联合制定了一系列编码标准,包括 MPEG-2, MPEG-4, H. 263, H. 264/AVC 以及目前最新的 H. 265/HEVC (high efficiency video coding) 编码标准。但这些标准所用的技术专利多属于国外的公司企业或研究机构,尤其是 H. 265/HEVC 之前的编码标准几乎没有中国专利的影子,而国内的企业和用户如果使用这些专利技术需要支付数额巨大的专利费用。2002 年 DVD 专利收费事件充分暴露了不掌握核心技术标准就会受制于国外专利机构的状况,引起了我国政府的高度重视。同时经过多年的积累,我国已具备自主制定数字音视频技术标准的技术和人才基础,在此背景下 2002 年 12 月,信息产业部科学技术司正式发文成立“数字音视频编码技术标准工作组”(简称 AVS 工作组)^[6],以制定具有自主知识产权的视频编码标准,至今已有 12 年的历程。

AVS 工作组又细分为需求组、系统组、视频组、实现组、测试组、安全与版权组等专题组,成员覆盖了我国本领域的主要产中开发生产厂家与研究单位。工作组成员首先商议决定了制定标准的原则:先进性、现实性、兼容性和独立性。而当时音视频编码技术的进步和标准的更新换代为我国后来居上提供

了历史性的发展机遇,使制定 AVS 时坚持独立自主和先进开放的原则得以实现,于 2003 年 12 月成功完成基准档次视频编码标准的制定,并与 2006 年公布为国家标准后,此后 AVS 工作组又陆续制定了面向移动、监控等应用的编码标准,并于 2012 年通过成立 AVS 技术应用联合推进工作组制定了面向清晰数字电视广播的 AVS+ 标准,如今 AVS+ 已在数字电视广播等应用中取得重要进展。此外,2011 年底 AVS 还启动了面向超高清应用的新一代视频编码标准 AVS2 的制定工作,目前已接近完成。

本文将对 AVS 视频标准 12 年来的技术发展过程给以概述,并简单介绍目前最新一代视频编码标准 AVS2 关键技术,最后对 AVS 视频编码技术的未来发展方向予以展望。

1 AVS 视频编码技术发展历程

1.1 AVS 视频编码系列标准

自 2002~2012 年,AVS 工作组制定了面向高清数字电视广播、移动、监控等应用的一系列视频编码标准,在 AVS 工作组中习惯将这一些编码标准称为 AVS1 即第 1 代编码标准,在国家标准制定计划中正式名称为《信息技术 先进音视频编码 第 2 部分:视频》,其中通过划分成多个类和子类来满足不同类型的应用需求。特别是 2012 年,为了加快推进 AVS 在广播电视行业中的应用,国家广播电影电视总局科技司与工业和信息化部电子信息司联合发文共同成立“AVS 技术应用联合推进工作组”,并于 2012 年 7 月成功制定 AVS+ 行业标准(GT/T 257.1-2012《广播电视先进音视频编解码 第 1 部分:高清晰视频》)(注:在 AVS1 系列标准中为第 16 部分《信息技术 先进音视频编码 第 16 部分:广播电视视频》),表 1 给出了 AVS1 各类标准的制定时间表。为

Table 1 History of AVS1 Standards

表 1 AVS 视频编码标准制定历程

Time	Profile	Applications and Major Coding Tools
Dec. 2003	Main	Broadcasting; 8×8 block based intra prediction, transform and deblock filter; variable block size motion compensation ($16 \times 16 \sim 8 \times 8$)
Jun. 2008	Surveillance Baseline	Video surveillance; Background-predictive picture for video coding, Adaptive Weighting Quantization (AWQ), core frame coding
Sep. 2008	Enhanced	Digital Cinema; Context Binary Arithmetic Coding (CBAC), Adaptive Weighting Quantization (AWQ)
Jul. 2009	Portable	Mobile video communication; $8 \times 8/4 \times 4$ block transform
Jul. 2011	Surveillance	Video surveillance; Background picture model based coding
May 2012	Broadcasting	High definition TV broadcasting; CBAC, AWQ, enhanced field coding

与 AVS1 相区别,AVS 工作组中将 2012 年启动的面向超清晰视频的编码标准称为第 2 代视频编码标准,即 AVS2.

1.2 AVS 视频编码关键技术

AVS 视频编码标准采用传统的基于预测变换的编码框架,可以分为预测、变换、熵编码和环路滤波 4 个主要模块,下面对每个技术模块的发展历程进行简单介绍.

1.2.1 预测

预测编码即利用信号间的相关性,用前面一个或多个信号作为当前信号的预测值,对当前信号的实际值与预测值的差进行编码.预测技术在视频编码中有着重要的应用,在现代视频编码技术中包括帧内预测和帧间预测两种,分别用于消除空域冗余及时域冗余.

帧内预测是利用图像在空间上相邻像素之间具有很强相关性的特点,目前常采用基于块的帧内预测技术,这主要是考虑到与基于块的变换量化技术的统一以及实现代价.基于块的帧内预测技术在现代视频编码标准中的应用有很多,比如 MPEG-4 标准中的 DC/AC 预测、H.264/AVC 以及 AVS 标中的多方向空间预测技术.DC/AC 预测即在变换域对 DC 和 AC 系数进行差分预测.多方向空间预测技术则是在图像空域,按照图像本身的特点选择一个最佳的预测方向,最大程度地去除空间冗余.在 H.264/AVC 中,对于亮度预测,可以是基于 4×4 , 16×16 或 8×8 块进行预测.对于 4×4 / 8×8 块有 9 种亮度预测模式,对于 16×16 块有 4 种亮度预测模式.色度是基于 8×8 块进行的,有 4 种预测模式.通过帧内预测提高了帧内编码图像的编码效率.在 AVS 基准类中,由于主要面向高清分辨率视频编码,帧内预测及变换块大小都是基于 8×8 块的,其中亮度有 5 种预测模式,色度有 4 种预测模式,这里的预测与 H.264/AVC 中的预测方法相比,其区别主要体现在对相邻像素的选取以及对相邻像素所采用的滤波器上.但在 AVS 移动类中,由于需要考虑低分辨率视频的高效编码,除了 8×8 块的帧内预测,还增加了基于 4×4 块的帧内预测及变换编码.在 AVS2 中,帧内预测通过增加更多的预测方向,编码性能得到进一步提升.

帧间预测通过块搜索匹配算法找出与当前图像中的块相对应的邻近参考帧中最相似的预测块,对当前图像块与匹配块进行差分编码,可以大大降低视频序列图像间的冗余.为了提高帧间预测的精度,

帧间预测技术在预测方向上从最简单的前向预测又发展出了后向预测和双向预测;在参考图像选择上,从一个参考帧扩展到两个以及更多参考帧,在 H.264/AVC 标准中,则可支持更加灵活的层次 B 帧帧间预测结构;在预测精度上,从整像素预测到 $1/2$ 像素预测以及 $1/4$ 像素预测,甚至 $1/8$, $1/16$ 像素预测;预测模式也更加多样来处理不同的帧间运动情况,采用从 16×16 块到 8×8 , 4×4 等划分形式的多级划分.AVS1 中的 B 帧还采用了一种新型的对称预测模式,通过前向运动矢量可以导出后向运动矢量,从而可以只编码前向运动矢量而进行双向预测,而且对 B 帧中的直接模式编码,采用了时域/空域直接预测模式相结合的方式,并在直接模式的运动矢量导出过程中进行舍入控制.AVS2 在帧间预测编码方面进行了大量研究,采用了二叉树划分、层次参考帧预测、多假设预测等编码工具,显著提高了编码性能.

1.2.2 变换

变换将空域信号变换到频域信号,有效地去除了信号的相关性,并使大部分能量集中到低频区域是图像或视频最有效的编码技术之一.根据变换的这一特点,有选择地编码部分显著的频域信号,丢弃不显著的频域信号,可以达到提高压缩效率的目的.在第 1 代视频编码标准如 MPEG-1, MPE-2 等都是采用 DCT 变换.从 H.264/AVC 开始,整数变换替换了传统的浮点 DCT 变换.与浮点 DCT 相比,整数变换具有计算复杂度低和编解码完全匹配等优点;同时,量化/反量化与变换/逆变换的归一化相结合,可以通过乘法和移位来实现量化,降低了量化/反量化的计算复杂度.AVS 也采用了整数变换,特别是在 AVS2 中采用了两层变换结构设计,进一步提升了变换编码的性能.

1.2.3 熵编码

熵编码的目标是去除信源符号在信息表达上的表示冗余,也称为信息熵冗余或者编码冗余.视频编码常用的熵编码方法有两种:变长编码和算术编码.变长编码的基本思想是,为出现概率大的符号分配短码字和为出现概率小的符号分配长码字,从而达到总体平均码字最短.算术编码是另一类重要的熵编码方法,在平均意义上可为单个符号分配码长小于 1 的码字,通常算术编码具有比变长编码更高的编码效率.算术编码和变长编码不同,不是采用一个码字代表一个输入信息符号的方法,而是采用一个浮点数来代替一串输入符号.算术编码计算输入符号

序列的联合概率,将输入符号序列映射为实数轴上的一个小区间,区间的宽度等于该序列的概率值,之后在此区间内选择一个有效的二进制小数作为整个符号序列的编码码字.可以看出算术编码是对输入符号序列进行操作,而非单个符号,因此平均意义上可以为单个符号分配长度小于1的码字.由于算术编码对当前符号的编码需要依赖前一个编码符号的信息,很难并行实现,计算复杂度较高,因此在早期的视频编码标准中都采用变长编码,直到 H. 264/AVC 才引入上下文自适应的算术编码技术.在编码过程中,熵编码器利用上下文信息自主切换码表或更新符号的条件概率,较好地解决了以往熵编码技术中全局统计概率分布与编码符号局部概率分布不一致的问题,因此编码效率进一步提高.

AVS1 中有两种熵编码方案,即基于上下文的二维变长编码算法(context-based 2D variable length coding, C2DVLC)^[7]和基于上下文的二元算术编码(context-based binary arithmetic coding, CBAC)^[8]. AVS 基准类只支持 C2DVLC,而在 AVS 的加强类中使用新的上下文自适应算术编码,称为 CBAC,这两种熵编码的复杂度都比与 H. 264/AVC 对应的熵编码方法低. AVS2 中采用了算术编码,但在算术编码引擎、上下文选择等方面进行了优化,更适宜并行解码实现.

1.2.4 环路滤波

环路滤波因其处于编码环内而得名,即重建图像经过滤波后被用作参考图像以编码将来的图像,它能够在达到改进视觉质量的同时提高编码效率. H. 264/AVC 基准类中的环路滤波以宏块为单位,在每个 4×4 块边界进行水平和垂直两个方向上的滤波处理.对于亮度块,该 16×16 块又分为 16 个 4×4 的块,环路滤波需要对该块中 4 条垂直边界和 4 条水平边界进行滤波.在 AVS1 中,去块效应滤波也是以宏块为单位,但是只在 8×8 块边界进行滤波处理.相比于 AVS1, AVS2 的环路滤波技术进一步增强.除了去块效应滤波,还增添了样本偏移补偿滤波(sample adaptive offset, SAO)、样本滤波补偿(adaptive loop filter, ALF)等,能够进一步提高主观编码质量.

2 新一代 AVS2 视频编码标准

为了满足急剧增长的超高清视频压缩需求,2011 年 12 月 AVS 启动了新一代视频编码标准的

制定工作,其目标是在高清或更高分辨率下编码效率比 AVS1 的最好性能提高一倍以上,编码效率优于同期国际标准.经过近 3 年的努力工作目前 AVS2 标准已接近完成,与最新的 H. 265/HEVC 国际标准相比,二者在编码框架上相似,但每一模块都有异同,由于详细对比内容较多,限于篇幅这里仅进行简明扼要的描述,关于 H. 265/HEVC 的详细技术可以参考文献[9].

AVS2 采用了基于四叉树递归划分的大块预测/变换混合编码架构,即一帧图像首先划分成若干个最大编码单元(coding unit, CU),最大编码单元按照四叉树递归划分的方式划分成更小的编码单元,如图 1 所示(最小块为 8×8 ,整个块大小为 64×64).编码单元大小可以表示成 $L \times L$ 的格式, L 的取值可以是 8,16,32 或 64.然后编码单元又可以进一步划分成预测单元(prediction unit, PU)进行预测编码,最后对预测残差进行变换编码. AVS2 与 H. 265/HEVC 的主要区别有:1)在整个框架方面,H. 265/HEVC 没有样本补偿滤波(ALF);2)在帧内预测方面,所用的划分模式和预测方向有些区别,如 H. 265/HEVC 没有非对称预测,用了 35 种预测模式等;3)在帧间预测方面,子像素插值、运动矢量预测及编码等都有差异;3)在变换方面二者差异较多,如 AVS2 有 64×64 块的变换、二次变换,H. 265/HEVC 最大到 32×32 变换,对 4×4 帧内预测块使用离散正弦变换(DST)等;4)在熵编码方面,系数组织编码形式类似,但 AVS2 采用(run, level)的方法编码系数,H. 265/HEVC 编码非零系数重要性图(significance map).

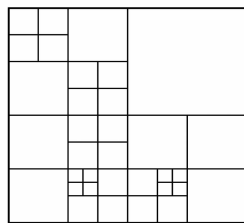


Fig. 1 Quad-tree coding unit structure in AVS2.

图1 AVS2 编码单元(CU)四叉数划分示例

下面对对 AVS2 的关键编码技术给以简单介绍.

2.1 帧内预测

AVS2 采用基于块划分的空域预测,相比于 AVS1, AVS2 增加了更多的预测方向和划分形状.首先在预测方向方面,如图 2 所示,使用了多达 33 种预测模式^[10],其中 30 种方向预测,另有 DC、平面(plane)和双线性(bilinear)插值 3 种帧内预测模式:

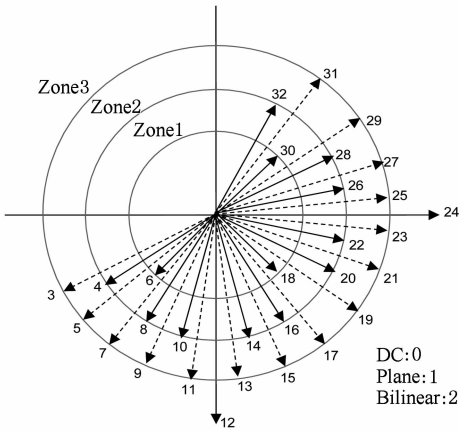


Fig. 2 Luma intra prediction modes in AVS2.

图 2 亮度预测块帧内预测模式

DC 模式是指当前块的所有像素都由左侧和上侧像素的平均值得到. 平面模式是指利用左侧和上侧像素值一次一个线性函数对当前像素值进行预测, 该模式适用于亮度变化平缓的区域. 双线性插值模式如图 3 所示, 先由左下和右上参考像素 A 和 B 加权预测, 得到当前预测块的右下像素 C . 然后, 利用 A, B, C 以及上侧和左侧相应行列位置的参考像素值求出两个预测值 X_v 和 X_h , 并把 X_v 和 X_h 的加权平均值作为最终预测值. 帧内预测模式的编码采用预测最可能的编码模式 (most probable mode, MPM) 进行预测编码, 根据相邻块的模式预测两个最可能的编码模式. 如果预测模式是两个最可能的模式候选之一, 则只需传输最可能模式的索引值, 否则需要编码预测模式. 色度的帧内预测模式有 5 种, 分别是 DC、水平、垂直、双线性插值和亮度导出模式 (derived mode, DM). 亮度导出模式是指色度采用和亮度一样的预测模式, 由亮度的预测模式导出而不需要单独编码.

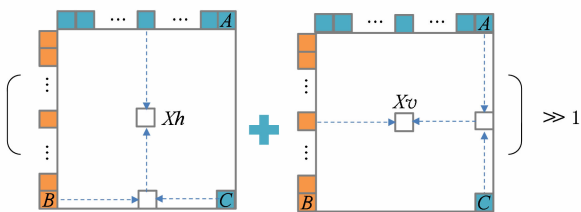


Fig. 3 Bilinear intra prediction mode in AVS2.

图 3 Bilinear 帧内预测模式

在预测单元的划分形状方面, AVS2 添加了非对称预测划分形式, 又称短距离帧内预测 (short distance intra prediction, SDIP), 是指在帧内编码中引入非方形的划分方式, 缩短平均预测距离, 从而

提高预测的精度^[11]. 具体地讲, 在 AVS2 中对 32×32 和 16×16 编码单元增加 $nN \times N$ 和 $N \times nN$ 两种划分方式, 即一个 32×32 大小的 CU 可以划分为 4 个 32×8 或 8×32 的矩形预测单元; 一个 16×16 大小的编码单元可以划分为 4 个 16×4 或 4×16 的矩形预测单元, 如图 4 所示 (最大编码单元大小为 64×64 , 非方形划分为 $8 \times 32, 32 \times 8, 4 \times 16, 16 \times 4$). 这里新增的非对称划分只适用于亮度块, 色度块不采用非对称划分预测.

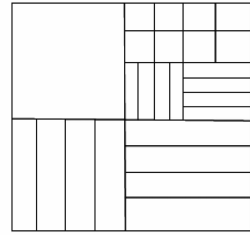


Fig. 4 Short distance intra prediction.

图 4 短距离帧内预测划分方式

2.2 帧间预测

对于帧间预测, 与 AVS1 相比 AVS2 增加了更多的预测划分方式和预测类型, 如支持前向多假设预测的 F 帧、层次 B 帧等. 特别是对监控视频, AVS2 提出的基于背景帧的预测编码技术能够显著提高一倍的编码效率. 此外在子像素插值、运动矢量预测及编码等方面也增加了新技术, 本节将进行简单介绍.

2.2.1 划分方式与预测方法

对于帧间预测, 首先在预测划分方面 AVS2 增加了更多的预测划分形式, 如图 5 所示, 其中 $2N \times nU, 2N \times nD, nL \times 2N$ 和 $nR \times 2N (n = N/2)$ 是 4 种非对称划分方式 (asymmetric motion partition, AMP), 适用于 16×16 及以上编码单元大小.

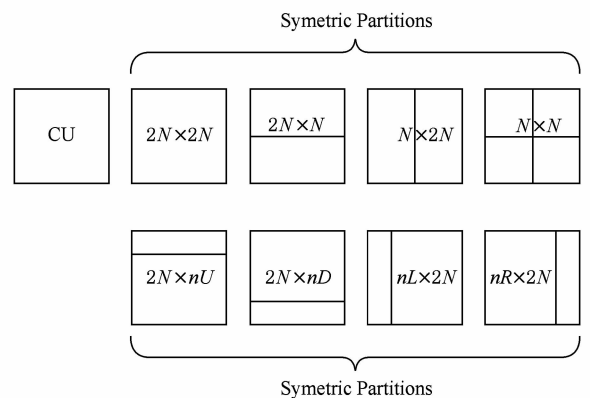


Fig. 5 Inter prediction unit partition in AVS2.

图 5 AVS2 预测单元 (PU) 划分模式

对于预测方法,除了传统的 P, B 帧所包含的前向、后向及双向预测, AVS2 还引入了前向多假设预测即 F 帧. 进一步地讲, 所谓 F 帧其实是引入了同向双假设技术的 P 帧^[12], 按照双假设预测的来源可以分为时域双假设和空域双假设两类. 时域双假设如图 6 所示, 当前块可以利用两个前向预测块的加权平均作为当前块的预测, 其中运动矢量 2 由运动矢量 1 根据时间距离导出. 空域双假设即方向性多假设预测技术 (directional multi-hypothesis prediction, DMH) 是指在进行预测时^[13], 采用预定方向上的两个预测子的复合结果作为预测值, 从而进一步减小了原始值和预测值之差. 其缺点是最优复合 MV 的决策需要经过复杂的运动估计和迭代计算, 因此复杂度较高. 如图 7 所示, 中心位置为初始预测子, 选择连线经过初始预测子的 4 种模式进行计算, 其中 1/2 像素距离和 1/4 像素距离位置的 4 种模式分别计算, 再加上初始预测值, 共有 9 种模式供选择.

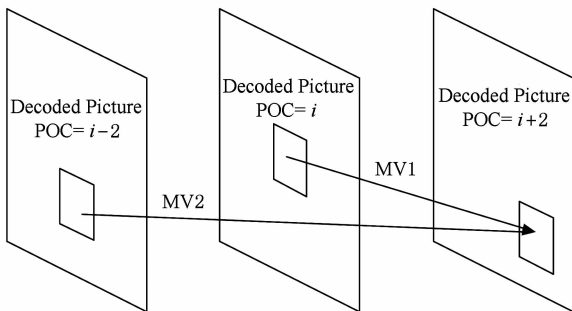


Fig. 6 Temporal forward multiple hypothesis prediction.
图 6 时域前向双假设预测模式

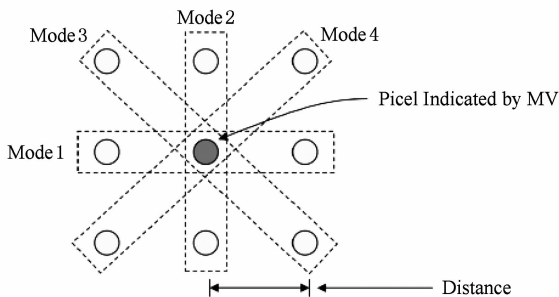


Fig. 7 Directional multiple hypothesis prediction.
图 7 DMH 模式示意图

AVS2 中帧间预测的精度, 即进行像素点运动估计匹配的精度是 1/4 像素, 所需的子像素点需要利用插值技术来得到. 如图 8 所示, $A_{i,j} (i, j = [-1..2])$ 为整像素位置, 对 a, b, c 三个位置的分像素采用横向临近整像素进行插值得到; 对 d, h, n 三个位置的

分像素采用纵向临近整像素进行插值得到; 其他位置的分像素点则是先利用横向临近整像素进行插值, 再利用纵向整像素位置的像素进行二次插值得到.

$A_{-1,-1}$					$A_{0,-1}$	$a_{0,-1}$	$b_{0,-1}$	$c_{0,-1}$	$A_{1,-1}$				$A_{2,-1}$
$A_{-1,0}$					$A_{0,0}$	$a_{0,0}$	$b_{0,0}$	$c_{0,0}$	$A_{1,0}$				$A_{2,0}$
$d_{-1,0}$					$d_{0,0}$	$e_{0,0}$	$f_{0,0}$	$g_{0,0}$	$d_{1,0}$				$d_{2,0}$
$h_{-1,0}$					$h_{0,0}$	$i_{0,0}$	$j_{0,0}$	$k_{0,0}$	$h_{1,0}$				$h_{2,0}$
$n_{-1,0}$					$n_{0,0}$	$p_{0,0}$	$q_{0,0}$	$r_{0,0}$	$n_{1,0}$				$n_{2,0}$
$A_{-1,1}$					$A_{0,1}$	$a_{0,1}$	$b_{0,1}$	$c_{0,1}$	$A_{1,1}$				$A_{2,1}$
$A_{-1,2}$					$A_{0,2}$	$a_{0,2}$	$b_{0,2}$	$c_{0,2}$	$A_{1,2}$				$A_{2,2}$

Fig. 8 Illustration of integer and fractional pixel.

图 8 整像素和分像素点示意图

2.2.2 运动矢量预测与编码

在帧间预测编码中, 为了提高压缩效率, 通常还可以对帧间预测的运动矢量 (motion vector, MV) 进行预测编码. 即对运动矢量进行预测得到运动矢量预测子 (motion vector predictor, MVP). AVS2 中普通预测模式下, 根据当前预测单元的左边、上边和右上 3 个位置的空域相邻块的 MV 来计算得到当前块的 MVP. 其中, 如果右上相邻块不可用, 则选择左上相邻块代替. 记从以上相邻块得到的 3 个 MVP 候选分别为 MVP_a, MVP_b 和 MVP_c . 对水平或垂直每个分量, 如果 3 个 MVP 候选中, 其中两个候选的符号相同, 而另一个相异, 则选择符号相同的两个候选的平均值作为最终的 MVP. 否则, 计算任意两个 MVP 候选的绝对差, 根据最小绝对差决定选取哪两个 MVP 候选的平均值作为最终的 MVP. 对于 F 或 B 帧的特殊模式 (跳过、直接模式), 对当前块 E 按照图 9 所示相邻块进行检测, 选取一个或两个运动矢量进行预测.

对差分运动矢量进行编码时, AVS2 采用了渐进的运动矢量精度编码方法 (progressive motion vector resolution, PMVR). 其依据是, 通常越靠近 MVP 的 MV 越有可能获得最优率失真代价. 因此

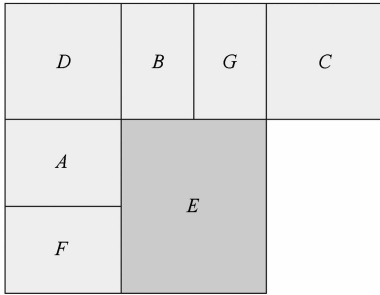


Fig. 9 Neighboring blocks used for motion vector prediction.

图 9 相邻块运动矢量预测示意图

文献[14]提出对 MVP 附近的 MV 使用 1/4 像素精度,而对远离 MVP 的 MV 使用 1/2 像素精度. 1/4 像素精度矢量只能在围绕 MVP 的一个正方形区域内使用(如图 10 所示),这个区域的大小由固定的阈值控制. 由于精度的变化, MVD 可以用更有效的方法来表示,即将水平和垂直分量的值缩小,从而减少码率.

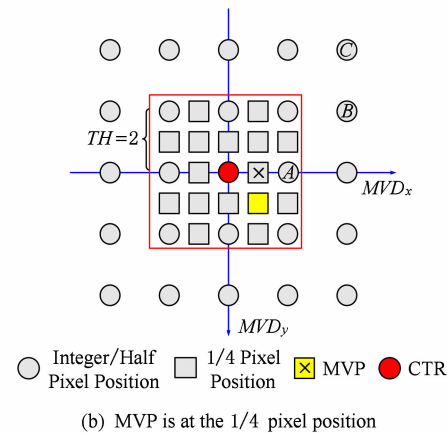
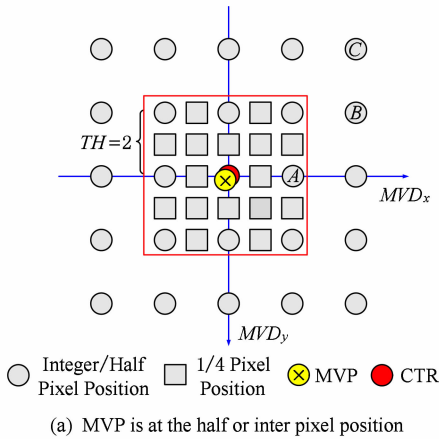


Fig. 10 Progressive motion vector resolution.

图 10 PMVR 中的 1/4 像素精度范围

2.3 变换

AVS2 中进行变换的基本单位称为变换单元

(transform unit, TU). AVS2 对 64×64 大小的 TU 采用两层变换结构. 首先基于 64×64 块进行一维小波变换,把变换矩阵尺度缩减到 32×32 ,然后再进行 DCT 变换. AVS2 支持 $4 \times 4 \sim 32 \times 32$ 大小的方形 DCT 变换,并且对非对称的帧内预测和非对称的帧间预测残差(第 2 层变换时)可以使用基于二叉树的非方形 DCT 变换(non-square quadtree transform, NSQT)^[15]. 引入 NSQT 技术的原因是,当变换块跨过预测块边界时,预测块的边界上有像素值跳变的现象导致变换后高频系数增多,熵编码需要耗费更多的比特,造成编码性能下降. 而非方形的变换结构可以避免跨预测块边界进行变换. AVS2 中提出对图 11(a)所示的 $2N \times N, 2N \times nU$ 和 $2N \times nD (n = N/2)$ 帧间预测模式,二级变换为采用 4 个 $2N \times 0.5N$ 的变换结构;对 $N \times 2N, nL \times 2N$ 和 $nR \times 2N$ 帧间预测模式,二级变换采用 4 个 $0.5N \times 2N$ 的变换结构,如图 11(b)所示. 此外, AVS2 对帧内预测残差的 DCT 输出进行二次变换以进一步减小冗余,这是因为帧内预测把临近像素点作为参考值,通过探测不同的方向得到预测结果,因此越靠近参考像素位置的残差就会越小,残差的分布呈现特殊的统计特性. 为了进一步减小编码冗余, AVS2 中对帧内预测残差的亮度分量的 DCT 低频变换系数进行 4×4 的二次变换^[16](对于 8×8 大小

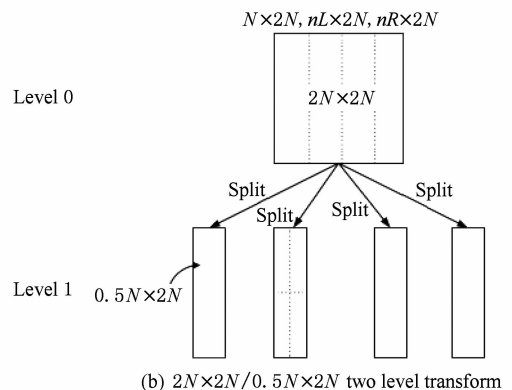
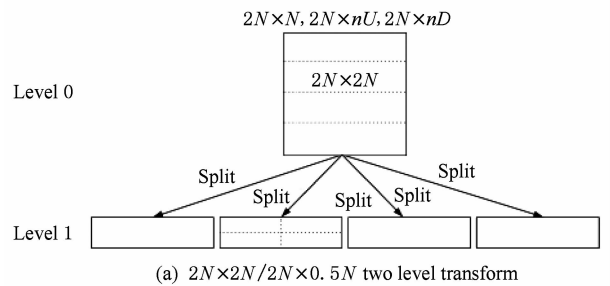


Fig. 11 Two level transform for inter prediction residual.

图 11 帧间预测残差的两层变换结构示意图

及以上块,对于 4×4 块帧内预测采用一个行列合并的变换),使得系数分布更为集中,从而提高压缩效率,如图 12 所示:

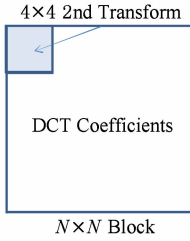


Fig. 12 Secondary transform in AVS2.

图 12 AVS2 二次变换示意图

经过变换后,AVS2 对变换系数采用以 4×4 块大小单位的两层系数组织编码方法,详见 2.4 节.

2.4 熵编码

AVS2 对高层语法如序列头、图像头、条带头采用变长编码,而对编码单元级的语法元素采用算术编码.算术编码首先需要将各语法元素的值进行二值化,得到二值码串,然后将二值码串的每一位编码到码流.对于二值化,AVS2 有 3 种二值化方法:一元码(unary code)、截断一元码(truncated unary code)

和指数哥伦布码(Exp-Golomb code).和 H. 264/AVC 类似,AVS2 中二值码串中每一位的概率模型可以是静态不变的也可以是动态更新的.与 H. 264/AVC 中的算术编码不同的是它摒弃了查表操作,通过对数域与原数域的转换来绕过乘法操作,只需加减法即可完成算术解码.因此,AVS2 不仅省去了查找表所带来的计算复杂度、硬件设计复杂度,也省去了查找表的存储空间.

AVS2 对变换系数编码采用两层组织编码机制,即每个变换单元进一步划分成若干 4×4 的系数组(coefficient group,CG)^[17],按照系数组进行系数编码.如图 13 所示, 8×8 的变换单元划分成 4 个子块, 16×16 的变换单元划分成 16 个子块, 32×32 的变换单元划分成 64 个子块.每个变换单元内的系数组按照之形扫描(Zigzag scan)方式进行编码.每个系数组内部的 16 个系数也按照 Zigzag 扫描方式进行编码.每个系数组有一个标识表示当前 CG 内是否含有非零系数.此外还需要编码两个标识,即最后一个非零系数组和最后一个非零系数在之形扫描中的位置.在系数组内部,系数按照依然按照游程/幅值(run/level)对的方式进行编码.

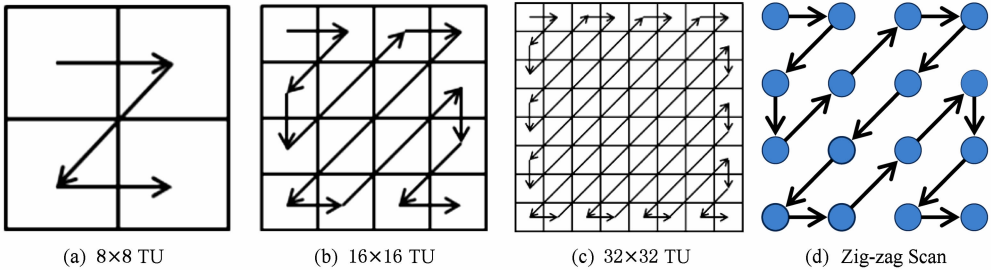


Fig. 13 Two level transform coefficients coding.

图 13 两层变换系数扫描方式

2.5 环路滤波

环路滤波是指对重建后的图像进行滤波处理之后再作为参考帧进行预测或显示处理,可以提高编码的主观质量或客观质量. AV2 中的环路滤波包括 3 个模块,即去块效应滤波、样本偏移滤波和样本补偿滤波.

首先对于去块效应滤波,AVS2 中基于 8×8 块对亮度和色度在编码单元块、预测单元块边界和变换单元块边界进行滤波,如图 14 所示(实线为块垂直边界,粗虚线为水平边界,细虚线为下一滤波单元待滤波的边界).滤波过程中,按照先垂直后水平的顺序依次检查垂直边界和水平边界是否需要滤波.边界的滤波由边界滤波强度控制,如图 15 所示,

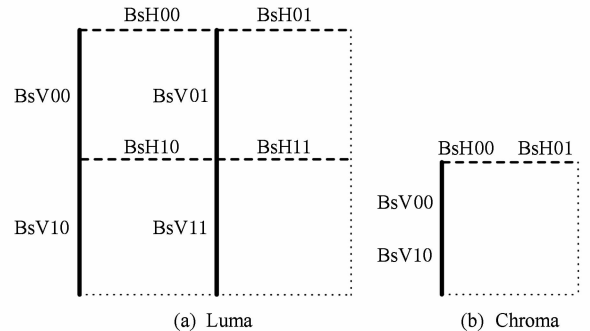


Fig. 14 Deblock filtering boundary for luma and chroma.

图 14 滤波单元中需要滤波的边界示意(4:2:0 格式)

如果强度为 0 则不滤波;强度为 1 和 2 时则只对 p_0 和 q_0 进行滤波;强度为 3 时则需对 p_0, p_1, q_0, q_1 进

行滤波;强度为 4 时需要对 $p_0, p_1, p_2, q_0, q_1, q_2$ 全部进行滤波。

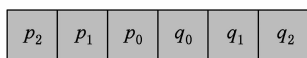


Fig. 15 Neighbouring samples for deblock filtering.

图 15 水平或垂直边界样本

自适应样本偏置滤波是指在去块效应滤波之后基于像素进行的非线性滤波操作,可以在线性滤波器之余进一步增强重建图像的编码质量^[18]. 自适应样本偏置滤波对于增强边缘锐度、抑制条带伪影(banding artifacts)和振铃效应(ringing artifacts)有显著效果. 自适应样本偏置滤波单元与由最大编码单元经过偏移和扩展得到,但不超过当前最大编码单元所在的条带或图像边界. 自适应样本偏置滤波分为区间模式和边缘模式两种. 区间模式时查表得到滤波偏置值,叠加到相应像素位置. 边缘模式分为 4 种,如图 16 所示. 根据当前滤波后样本的值与其相邻的滤波后样本的值的的关系查表得到偏置值,对当前像素进行滤波。

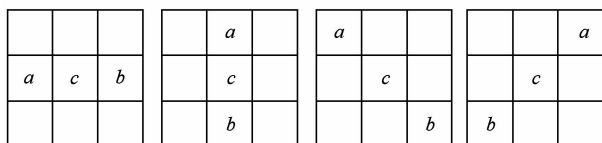


Fig. 16 Edge modes of sample adaptive offset filtering.

图 16 自适应样本补偿边缘模式

样本补偿滤波是对自适应样本偏置滤波滤波后的重建图像进行维纳滤波,以进一步减小编码失真,提高参考图像质量^[19]. AVS2 采用如图 17 所示的 7×7 十字加 3×3 方形的中心对称滤波器,每个滤

波器只包含 9 个不同的系数 $C_0 \sim C_8$. 对于亮度分量,将图像按照最大编码单元对齐的规则均匀划分为 16 个区域并训练对应的 16 套滤波器系数,而对于色度分量分别只训练一套滤波器系数. 对于亮度分量的 16 套滤波器系数,编码端根据率失真性能进行自适应合并,得到最优的滤波器数目. 亮度和色度分量通过帧级标志和最大编码单元级标志来分别控制滤波器是否被使用。

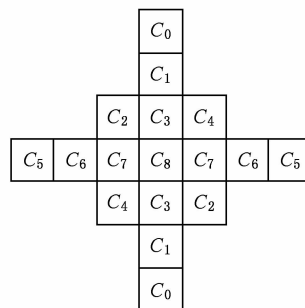


Fig. 17 Filter shape of adaptive loop filter.

图 17 样本补偿滤波器形状

2.6 性能测试

使用 AVS2 参考软件 RD7.0 的编码性能与 AVS1, H. 265/HEVC 对比结果如表 2 所示. 在随机访问(random access, RA)、低延时(low delay, LD)和全帧内编码(all intra, AI)3 种配置下,所采用的测试序列都是 AVS 的通用测试序列,其中 LD 下的测试序列为监控视频序列. 从表 2 数据可以看出, AVS2 比 AVS1 能显著提升编码效率,对于监控视频序列, AVS2 的编码性能比 HEVC 有显著提升. 这里需要指出 AVS2 尚未最终完成,最后所采用的技术及性能或与本文中报告的数据可能会略有差异。

Table 2 Performance Comparison Between AVS2 and AVS1, H. 265/HEVC

表 2 AVS2 与 AVS1, H. 265/HEVC 对比性能测试

%

Sequence	AI		RA		LD
	AVS2/AVS1	AVS2/HEVC	AVS2/AVS1	AVS2/HEVC	AVS2/HEVC
UHD	27.70	1.00	47.20	-3.60	NA
1200p	NA	NA	NA	NA	21.70
1080p	34.20	-0.70	52.60	-2.00	NA
720p	30.20	0.20	53.10	-3.40	NA
SD	NA	NA	NA	NA	12.30
Overall	30.70	0.17	50.97	-3.00	17.00

Note: NA is means available.

3 总结与未来展望

迄今为止 AVS 是我国最具影响力的数字视频编解码标准制定组织,已经成功制定了一系列具有自主知识产权的编码标准.本文首先对 AVS 视频标准的编码技术的发展历程进行了简要回顾,并对新一代视频编码标准 AVS2 进行了详细介绍. AVS2 相比上一代编码标准节省了约一半的码率,有效解决了高分辨率视频的编码效率问题,具有广阔的应用前景.

而随着 HEVC/H. 265, AVS2 新一代标准的制定完成,新一轮技术储备已经悄然开始,对于未来视频发展方向我们认为基于模型的编码具有更大的潜力.目前,传统的基于预测/变换框架的编码技术面临着效率瓶颈问题,而基于模型的编码通过挖掘数据内部的知识冗余信息可以进一步提高编码效率,并且还有很大的提升空间.特别是针对某一特定应用领域的视频编码,如面向智能监控的视频编码^[20],结合这些应用领域特性建立相应的模型不仅能提高编码效率,还能进一步丰富编码理论,AVS 中提出的基于背景建模的监控视频编码即是很好的实例之一.

同时,传统的编码方法主要是从数字信号处理的角度进行压缩处理,虽然一定程度上消除了部分视觉冗余,但由于人眼的稀疏特性,还有大量的视觉冗余信息可以去除,而随着对人眼视觉系统的在生理和心理上研究进展的进一步深入,在结合视觉特性的编码方法方面取得革命性的突破不是没有可能的.另外,云计算时代的到来改变了传统数据的存储和处理的方式,其强大的分布式并行计算能力为海量数据的即时搜索及分析提供了可能.在这一背景下,图像/视频数据的编码与表示也可以另辟蹊径,文献^[21]提出了一种基于 SIFT(scale invariant feature transform)的图像压缩方法,即只需传输很少数据量的 SIFT 特征和高密度下采样的图像到云端,由云平台对云存储数据进行特征匹配检索,然后根据检索到的存储资源进行图像重建即可得到高分辨率高质量的图像.这一新型图像编码方法在云存储时代或许是值得深入研究的图像/视频编码方向之一,可以应用于移动视觉搜索、移动视频服务等场景中.

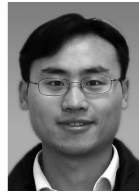
总之,高效的视频编码技术依然有着巨大的应用需要,特别是针对监控视频、云图像/视频研究新型高效的压缩方法,不再仅仅是基于数字信号处理

的方法进行编码,而是与视觉模型、内容理解等相结合,对于智能监控、视频搜索等具有十分重要的意义.

参 考 文 献

- [1] Huffman D. A method for the construction of minimum redundancy codes [J]. Proceedings of the IRE, 1952, 40(9): 1098-1101
- [2] Harrison C. Experiments with linear prediction in television [J]. The Bell System Technical Journal, 1952, 31(4): 764-783
- [3] Taki Y, Hatori M, Tanaka S. Inter frame coding that follows the motion [C] //Proc of Institute of Electronics and Communication Engineers of Japan (IECE) Annual Convention. 1974: 1263
- [4] Ahmed N, Natarajan T, Rao K. Discrete cosine transform [J]. IEEE Trans on Computers, 1974, C-32(1): 90-93
- [5] Netravali A, Stuller J A. Motion compensated transform coding [J]. The Bell System Technical Journal, 1979, 58(7): 1703-1718
- [6] AVS working group website [OL]. [2014-10-13]. <http://www.avs.org.cn>
- [7] Wang Q, Zhao D, Gao W. Context-based 2D-VLC entropy coder in AVS video coding standard [J]. Journal of Computer Science & Technology, 2006, 21(3): 315-322
- [8] Zhang L, Wang Q, Zhang N, et al. Context-based entropy coding in AVS video coding standard [J]. Signal Processing: Image Communication, 2009, 24(4): 263-276
- [9] Sullivan G J, Ohm J R, Han W, et al. Overview of the high efficiency video coding (HEVC) standard [J]. IEEE Trans on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668
- [10] Piao Y, Lee S, Kim I, et al. Unified design of directional intra prediction for AVS2 [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1303_Luoyang/contrib/AVS_M3054.doc
- [11] 余全合, 曹潇然, 李蔚然, 等. 短距离帧内预测技术 [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1309_Shenyang/contrib/M3171.docx
- [12] 凌勇, 虞露. F 帧 CE: 一种前向双假设预测图像块编码技术 [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1404_Beijing/contrib/M3326.doc
- [13] Kim I, Lee S, Piao Y, et al. Directional multi-hypothesis prediction (DMH) for AVS2 [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1306_Taicang/contrib/M3094.zip
- [14] 马俊铨, 马思伟. 渐进的运动矢量精度 [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1303_Luoyang/contrib/AVS_M3049.doc

- [15] 李蔚然, 袁媛, 曹潇然, 等. Non-square quad-tree transform [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1309_Shenyang/contrib/M3153.docx
- [16] 林和源, 童怡新, 王颂文. M3296 帧内编码二次转换改进提案 [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1404_Beijing/contrib/M3296.doc
- [17] Wang J, Wang X, Ji T, et al. Two-level transform coefficient coding [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1212_Beijing/contrib/M3035.doc
- [18] Piao Y, Lee S, Kim C, et al. Sample adaptive Offset for AVS2 [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1309_Shenyang/contrib/M3197.doc
- [19] Zhang X, Si J, Wang S, et al. Adaptive loop filter for AVS2 [OL]. [2014-10-13]. ftp://124.207.250.92/public/avsdoc/1404_Beijing/contrib/M3292.doc
- [20] Zhang X, Liang L, Huang Q, et al. An efficient coding scheme for surveillance videos captured by stationary cameras [C] //Proc of SPIE Conf on Visual Communications and Image Processing. Huangshan: SPIE, 2010: 77442A-1-10
- [21] Yue H, Sun X, Wu F, et al. SIFT-based image compression [C] //Proc of IEEE Int Conf on Multimedia and Expo. Los Alamitos, CA: IEEE, 2012: 473-478



Ma Siwei, born in 1979. PhD. Professor at the Electrical Engineering and Computer Science School, Peking University. His research interests include video coding, processing and transmission.