

摘要

随着网络和无线传输应用的日渐频繁,对图像视频等压缩码流的可伸缩性(scalability)的要求逐渐提高。小波作为一种具有良好的时频局域性和多分辨率分析能力的变换方法,在静止图像压缩方面得到充分的认可与应用。基于小波的 JPEG2000 除了在压缩性能上大大超过基于 DCT 的传统压缩标准,还提供了多帧率、多分辨率,清晰度渐进等多种可伸缩性的解码,非常适合当今网络以及卫星遥感图像的传输和应用。

本文在小波静止图像压缩所取得成功的基础上,重点研究小波在高维图像媒体压缩编码中的若干算法。主要研究内容包括:针对场景绘制技术的图像序列的小波压缩方案,对小波压缩码流的快速解码算法,以及基于小波的视频压缩技术。论文的主要研究工作和创新点如下:

一、针对同心拼图这种新型的基于图像的场景绘制技术(image-based rendering, IBR)的特点,以及对压缩码流的随机访问解码要求,设计了基于三维小波的嵌入式压缩系统,为同心拼图的网上传输和浏览等应用提供了灵活可靠的编解码方案。由于同心拼图等 IBR 媒体是比较新兴的图形学媒体,直接应用普通的视频压缩标准不能很好的满足其对于压缩和随机解码的需求,我们的工作是该领域中较早开发的并且具有高效压缩性能的编码系统之一;

二、在第一项工作基础上,考虑同心拼图压缩场景的即时绘制特点,提出了一种渐进逆向小波合成快速算法 PIWS,有效的支持了小波压缩码流的实时解码。PIWS 快速小波合成方法不但支持了对同心拼图压缩码流的实时绘制,还可以普适的应用到小波视频压缩的快速解码应用中,为降低小波解码过程的运算量提供了一种有效的解决方案;

三、针对传统的三维小波视频压缩中对时轴相关性发掘的不足。我们提出了基于运动补偿的提升型小波(motion-compensated lifting, MCLIFT)方法实现时轴变换。通过在时轴小波变换的提升型单元结构中采用双向的宏块(macro-block)运动补偿(motion compensation),采用 MCLIFT 的时轴小波变换可以结合帧间运动信息,很好的压缩视频帧间的冗余。利用提升型算法的单元结构的运算特点,在 MCLIFT 方法可以在保证小波反变换完全重构(perfect reconstruction)的同时,利用许多现有的运动补偿技术,例如 1/2 像素运动精度和重叠块运动补偿技术,提高了时轴变换的效率。实验结果表明采用 MCLIFT 方法的小波视频编码器的压缩效率远远超出没有采用运动估计的三维小波

压缩方法,并且达到了优于 MPEG-4 标准的水准,同时还具有 MPEG-4 所不具备的帧率和压缩质量的伸缩性。此项工作在国际上同期的小波视频压缩研究中属于较早采用提升型结构来引入运动信息的几个工作之一。

四、研究了三维小波视频压缩中的运动抽线技术 (motion threading), 仔细分析了影响其压缩质量的断线问题, 并且设计了两轮运动估计技术来降低断线情况的出现。进一步结合提升型小波结构, 本文提出了高级运动抽线技术, 除了从根本上解决了断线问题, 还进一步将抽线的运动精度提高到了 $1/4$ 像素, 并且考虑了视频中常有的遮挡问题, 以及利用运动矢量的冗余性进行运动矢量估计和编码, 使小波视频的压缩效率提高到了与最先进的 H.26L 编码标准相接近的水准。

五、研究了三维小波视频压缩的时轴抖动问题。由于时轴的质量抖动可能会影响视频播放的视觉效果, 这是三维小波视频压缩中的一个重要问题。我们对三维小波视频压缩中的时轴周期性抖动现象进行了理论分析, 得出了抖动原因在于小波合成滤波对不同位置的系数施加了不同的权重的结论。由此提出了使时轴质量平稳的理论的码率控制方法。采用了这种码率控制方法的三维小波编码器具有良好的时轴平稳性, 同时对某些帧的压缩质量提高了 0.5-1.0dB。

关键词: 小波压缩, 视频编码, 提升型算法, 可伸缩性, 嵌入式编码, 运动抽线, 基于图像的场景绘制技术

Abstract

With the increasing applications in network and wireless transmission, the scalability of the encoded image or video bit-stream is becoming an important demand. Wavelet has been adopted in the still image compression as a promising transform scheme with great time-frequency localization and multi-resolution analysis property. Comparing with the current DCT-based JPEG standards, the wavelet-based JPEG2000 standards not only shows superior compression performance, but also offers scalabilities in both resolution and quality level which are very desirable in the consumer and the Internet applications.

Based on the success in the still image compression, this paper focuses on the wavelet-based compression techniques for the high-dimensional image data. The content includes: wavelet compression of the IBR (Image-Based Rendering) image array, fast wavelet synthesis algorithms, and wavelet-based video compression. The research topics and our key contributions are as follows:

I. Concentric mosaic is an emerging IBR data which can render the 3D scenery with a set of captured images. Considering the characteristic of the concentric mosaic data, we develop a 3-D wavelet transform and embedded coding system. The proposed wavelet coding system enables flexible partial decoding and bit-stream random access, which are demanded in the transmission and browsing of the concentric mosaic scenery. Since concentric mosaic is a new graphics media which processes some unique characteristic, the existing video coding standard can not satisfied the relative compression requirements. Our system is one of the earliest compression systems with high coding efficiency for the IBR data.

II. Based on the previous work, we proposed a progressive inverse wavelet synthesis (PIWS) scheme for the just-in-time (JIT) rendering of the concentric mosaic. PIWS effectively reduces the calculations of the wavelet synthesis, thereby effectively support the real-time rendering. Furthermore, the PIWS algorithm can be applied in the wavelet synthesis of the other wavelet compressed data such as video.

III. The primary weakness of the existing 3D wavelet video coder lies in the temporal filter. We proposed a motion-compensated lifting (MCLIFT) scheme for the temporal wavelet transform. Using bi-directional motion compensation in the elementary unit of the temporal lifting wavelet, MCLIFT can efficiently decorrelate the video. Many motion compensation techniques, such as $\frac{1}{2}$ -pixel and overlapped-block motion compensation (OBMC), can also be applied into the MCLIFT structure. Experimental results show that MCLIFT effectively improve the 3-D wavelet video codec so that it can achieve a superior performance to the MPEG-4 standard video codec. Furthermore, the MCLIFT codec can offer quality and frame-rate scalabilities, which are not processed by MPEG. Our work is one of the earliest works which use lifting scheme to enhance the temporal correlation.

IV. We analyze the motion threading (MTh) scheme for the 3-D wavelet video coding. In the original MTh codec, the artificial motion thread truncation problem severely affects the coding efficiency. We first proposed a two-pass motion estimation technique, and effectively reduce the occurrence of the truncated threads. After that, we extend the original motion-threading technique to the lifting wavelet structure and proposed an advanced motion threading (A-MTh) technique. This extension solves the artificial motion thread truncation problem, and enables the accuracy of motion alignment to be fractional-pixel with guaranteed perfect reconstruction. Furthermore, the mismatch problem in the motion-threading caused by occlusion or scene-change is considered. To reduce the motion cost, the correlation of the motion vectors are examined and used in the motion estimation and motion coding process. The proposed A-MTh scheme can significantly outperform the original MTh scheme, and can be competitive with the start-of-the-art H.26L video standard on coding efficiency.

V. The temporal fluctuation phenomenon in the 3-D wavelet decompressed video may make the video display annoying, thus how to solve it is important for the 3-D wavelet video coding. We investigate the periodic temporal fluctuation and find the problem is due to the unequal weights of the reconstruction filters during the wavelet synthesis. A basic theoretical rate-control scheme is proposed to solve the fluctuation based on the above analysis. The 3-D wavelet video codec with the proposed scheme has a smooth temporal performance, and improves about 0.5-1.0dB in several frames.

Key words: wavelet compression, video coding, lifting scheme, scalability, embedded coding, motion threading, Image-Based Rendering.

图表索引

图 1-1 多种图像媒体的相应压缩方法以及传输的网络条件和终端类型 (Compression schemes of various image data and the different transmission channels)	2
图 1-2 小波编码器结构 (The wavelet codec structure)	4
图 1-3 一维一级小波分解与合成示意图 (1D one-level wavelet analysis and synthesis)	5
图 1-4 图像的多级小波分解 (Multi-scale wavelet decomposition for the image data)	6
图 1-5 三种典型的小波包分解方式 (Three typical wavelet packets)	7
图 1-6 多相小波变换提升型实现 (引自[45]) Wavelet transform in the polyphase form and lifting (from [45]).....	8
图 1-7 系数与位平面的关系 (Relationship between the coefficients and the bit planes) ..	9
图 1-8 零树编码结构和扫描顺序 (The coding structure and the scan order of EZW) ...	10
图 1-9 运动信息对时间维上帧间相关性的影响 (Motion information and the temporal correlation)	12
图 1-10 基于 DCT 的视频编码标准压缩流程 (Flowchart of the DCT-based video coding standard)	13
图 1-11 采用 Lumigraph 技术的 IBR 场景的观察与拍摄方法(a)和获取的图像序列数据(b)	16
图 2-1 同心拼图的几何位置关系 (Concentric mosaic imaging geometry)	21
图 2-2 采用同心拼图的场景绘制 (Rendering with the concentric mosaic)	21
图 2-3 同心拼图的三维小波压缩系统 (3D wavelet compression system for the concentric mosaic)	22
图 2-4 一级提升性小波的正变换与反变换结构, 图右所示为基本提升型单元 (One-level forward and inverse lifting wavelet transforms and the elementary lifting units)	24
图 2-5 一级三维提升型小波 (Single scale 3D lifting.)	25

图 2-6 多级三维小波包分解 a) 2 级三维 mallat 分解 b) 2 级 x 分解+2 级(y,z) mallat 分解 c) 2 级 z 分解+2 级(x,y) mallat 分解 (Multiple-level 3D wavelet packet a) two-level mallat decomposition in all directions, b) two-level x decomposition + two-level (y,z) mallat decomposition, c) two-level z decomposition + two-level (x,y) mallat decomposition.)	25
图 2-7 树编码器状态转换图 (State transition of the tree coder)	28
图 2-8 率失真曲线的凸包络 (Convex hull of the R-D curve)	30
图 2-9 码流组装示意图 (Illustration of the bitstream assembling)	31
图 2-10 同心拼图压缩码流的选择性解压缩系统流程 (Flowchart of the selective decompression system for the 3D wavelet compressed concentric mosaic)	32
图 2-11 渐进逆向小波合成原理图 (Progressive inverse wavelet synthesis)	34
图 2-12 解码一根狭缝图像时存储单元的相应状态值 (Cache states when a single slit is accessed)	36
图 2-13 多级渐进小波合成算法结构 (Multiscale progressive inverse wavelet synthesis)	37
图 2-14 同心拼图场景: 大厅 (Concentric mosaic scene: Lobby)	38
图 2-15 同心拼图场景: 儿童 (Concentric mosaic scene: Kids)	38
图 2-16 同心拼图场景中的三种浏览方式 (Three kinds of movement in concentric mosaics)	40
图 2-17 三种浏览方式中需解码的狭缝图像数据在整体拼图序列中对应的位置, 每图 所示分别为两个场景所对应的数据 (Access slits associated with three motion modes in the concentric mosaics. The slits of two views are drawn)	40
图 2-18 PIWS 算法和另两种合成算法对平移模式的解码速度比较 (Timing curves for the translation motion of concentric mosaics)	41
图 3-1 运动补偿提升型小波 MCLIFT 视频编码器流程 (Framework of the motion compensated lifting wavelet (MCLIFT) video coder)	44
图 3-2 一级 5-3 提升性小波的正变换与反变换结构, 圈内所示为基本提升型单元 (One-level forward and inverse 5-3 wavelet transforms via lifting and the elementary lifting units (circled in the forward structure))	44
图 3-3 运动补偿提升型结构基本单元 (The elementary MCLIFT operation)	45

图 3-4 时轴 3 级提升型小波结构 (3-layer temporal lifting wavelet sturcture)	46
图 3-5 采用截断 5-3 提升型小波的 3 级 MCLIFT 时轴小波分解结构 (3-layer MCLIFT temporal wavelet decomposition with truncated 5-3 filter)	47
图 3-6 MPEG-4 的 IBBBPBBBP... 帧结构示意图 (The GOP structure of MPEG-4 IBBBPBBBP...)	48
图 3-7 帧内二维 Spac1 小波分解结构 (Intra-frame 2D Spac1 wavelet decomposition)	48
图 3-8 MCLIFT 5-3 与原始 MCLIFT 编码效率比较 (Performance comparison between MCLIFT and MCLIFT 5-3 filters)	50
图 3-9 MCLIFT 编码器和他三种编码器的性能比较 (Comparison of the Motionless 3D Wavelet Video Coder, MCLIFT, SPIHT and the MPEG-4 coder)	52
图 4-1 三维运动抽线小波技术的基本结构 (The original 3D motion-threading wavelet)	54
图 4-2 两轮估计运动抽线流程图示意图 (Flowchart of TEMTH)	56
图 4-3 运动估计中的重叠现象和整帧对应的占据图 (The overlapping phenomenon and the corresponding occupancy map for a whole frame)	56
图 4-4 压缩质量比较: 无运动三维小波编码器 (NM), 原始三维小波运动抽线编码器 (MTh), 本节提出的两轮估计运动抽线编码器 (TEMTH) 和 MPEG-4 编码器 (PSNR versus bit rate: a comparison among NM, MTH, TEMTH and MPEG-4)	60
图 4-5 基于 5-3 提升型小波与双向运动搜索的运动抽线结构 (5-3 lifting-based motion-threading (MTh) with bi-directional motion search)	62
图 4-6 1/4 像素单元提升型运算 (Quarter-pixel elementary lifting operation)	63
图 4-7 四级时轴小波分解结构 (A 4-layer temporal wavelet decomposition structure)	64
图 4-8 双向运动估计中的直接反向模式 (DirectInv mode in bi-directional motion estimation)	65
图 4-9 运动预测的 7 种选择模式 (Seven motion modes in motion estimation)	67
图 4-10 实验比较结果 (Comparisons among MC-EZBC, H.26L, original MTh and A-MTh.)	70
图 4-11 实验比较结果, 接上页 (Comparisons among MC-EZBC, H.26L, original MTh and A-MTh, cont.)	71

图 5-1 结合本文提出的码率控制方法的 TS-EBCOT 视频编码器与 3D SPIHT 和 3D EBCOT 对 MPEG-4 标准序列的压缩性能比较: a) Foreman 序列, 0.4 比特/像素, 平均峰值信噪比 PSNR(dB): TS-EBCOT: 36.1, 3D SPIHT:36.3, 3D EBCOT:36.7; b) Coastguard 序列, 0.2 比特/像素, 平均 PSNR(dB): TS-EBCOT:29.6, 3D SPIHT:29.7, 3D EBCOT:30.3; c) Akiyo 序列, 0.05 比特/像素, 平均 PSNR(dB): TS-EBCOT: 36.1, 3D SPIHT:33.9, 3D EBCOT:36.2。	82
图 1-1 四种类型的相邻点 (Four types of neighbors)	86
表 2-1 一维提升型小波变换状态表 (States for 1D progressive inverse wavelet synthesis)	34
表 2-2 一维逆向小波合成状态转换表 (State transition operations for 1D inverse lifting)	34
表 2-3 二维渐进逆向小波合成状态转换表 (States of the 2D progressive inverse wavelet synthesis)	35
表 2-4 针对同心拼图的不同小波包分解结构性能比较 (Concentric mosaic compression with different wavelet packet decomposition structures)	38
表 2-5 三维同心拼图场景压缩结果 (Results of the 3D wavelet concentric mosaic compression)	39
表 2-6 三种运动模式下的三种算法速度比较 (Overall rendering speed measured in frames per sec: VQ, RBC vs. PIWS)	42
表 3-1 两种运动补偿提升型滤波器对应的运动矢量的码率 (Bitrate for the motion vectors according to different motion-compensated lifting wavelet filter)	49
表 4-1 TEMTH 中的参数设置 (Parameters for the TEMTH coding)	59
表 4-2 不同时轴分解级对应的模式分布百分比 (Mode distribution percentage according to the temporal decomposition layers)	72
表 4-3 运动矢量码率的节省量 (Motion vector bits reduction)	74

表 5-1 Biothogonal Le Gall 5-3 和 Daubechies 9-7 滤波器, 三层以下小波分解的理论子带由低频到高频失真比例 (Theoretical distortion distribution ratio for the wavelet decomposition less than 3 layers, from the lowest pass to the highest one)	80
表 I-1 有效性预测中的上下文分类 (Contexts for the significance prediction)	86
表 II-1 Daubechie 9-7 卷积型滤波器系数 (Coefficients of the Daubechies 9-7 Filter) ..	87
表 II-2 Le Gall 5-3 卷积型滤波器系数 (Coefficients of the Le Gall 5-3 Filter)	87
表 II-3 CRF 13-7 卷积型滤波器系数 (Coefficients of the CRF 13-7 Filter)	87
表 II-4 SWE 13-7 卷积型滤波器系数 (Coefficients of the SWE 13-7 Filter)	88
表 II-5 Daubechie 6-10 卷积型滤波器系数 (Coefficients of the Daubechies 6-10 Filter)	88
表 II-6 Daubechie 10-18 卷积型滤波器系数 (Coefficients of the Daubechies 10-18 Filter)	88

第一章 绪论

1.1 信息时代的图像媒体压缩编码技术

当今计算技术, 压缩技术, 大容量存储设备以及高速网络的发展, 让互联网络上实时的多媒体传输和交互服务逐渐走进我们的生活。图像类数据(包括静止图像, 视频, 以及游戏、医疗等高维图像数据)在多媒体应用中信息量最为丰富, 而数据量也最庞大, 因此针对图像类数据的压缩一直吸引人们探索研究。压缩技术从 1948 年 Shannon 信息论的提出^{[1][2][3]}和 Huffman 熵编码(entropy coding)的设计, 中间经过了从预测编码到变换编码的几十年间的不断研究积累。20 世纪 90 年代以来, 关于静止图像以及视频的压缩技术逐渐成熟并相应的转化成产品, 最与我们生活密切相关的例如国际标准化组织(ISO)和国际电工委员会(IEC)制定的面向彩色和灰度静止图像的 JPEG 标准^[4], 面向运动图像和音频压缩存储的 MPEG-1^[5]以及针对高清晰度电视(HDTV)和 DVD 存储等高品质视频音频的标准 MPEG-2^{[6][7][8]}, 还有基于对象的视频编码标准 MPEG-4^[9]。同时, 国际电信联盟(ITU)还制定了面向低码率视频会议等应用的 H.261^[10], H.263^[11]以及正在制定的兼顾低码率和高码率视频编码的 H.26L^[12]标准。作为基于离散余弦变换 DCT 的变换类编码方法的代表, 以上标准推动了视频多媒体传输和存储的应用发展。

随着多媒体应用的日渐频繁, 对图像视频等数据在网上流化(streaming)传输播放的要求也在逐渐提高。由于网络条件以及用户终端应用的许多限制, 对图像视频等媒体的压缩已经不再是传统意义上的将给出的图像压缩成一个文件, 在解码时由整个压缩文件解成原图像/视频这样的简单, 而是需要压缩码流具有多种可伸缩性(scalability), 能够根据不同的网络带宽、稳定性和速度限制, 以及客户端等客观情况进行相应的压缩码流传输, 并解出不同分辨率、质量的图像/视频, 甚至可以只传输感兴趣的部分区域信息。同时还需要必要的错误隐藏等技术的支持。这些应用要求都促使着当今的图像类媒体压缩方法不断开拓新的研究方向。

在图 1-1 中, 我们描绘了图像的信源端压缩以及网络传输应用的示意图。图左端表示的是一些图像媒体种类以及相应的压缩方法。图右端表示不同的网络环境以及终端类型。对于局域以太网, ADSL, 窄带 ISDN 等互联网接入, 以及 GPRS 或 Wireless-LAN 无线网络接入方式, 传输速率将从几十 Kbps 到 10M/100Mbps 不等, 同时网络的错误发

生类型也有区别，例如有线网络上主要存在网络拥塞以及延时抖动引起的丢包错误（packet loss），而无线网络除了丢包错误还存在多径衰落和信道干扰造成的比特位错误（bit error）。从客户端的情况来看，处理器的速度以及显示设备的分辨率也各不相同。例如接入以太网的 PC 机，由于网络带宽环境较好，加上较强的处理器运算能力，可以选择传输较大码流得到较好的压缩质量；而接入 GPRS 网络的个人终端 PDA 除了网络条件不稳定，本身的处理器速度和显示器分辨率也有限制，因而可以选择接收较小分辨率和较低压缩质量的图像数据，满足基本的观看要求。这些不同的应用除了需要服务器具有一些相关的处理能力之外，对信源端的编码方式也提出了更高的要求。

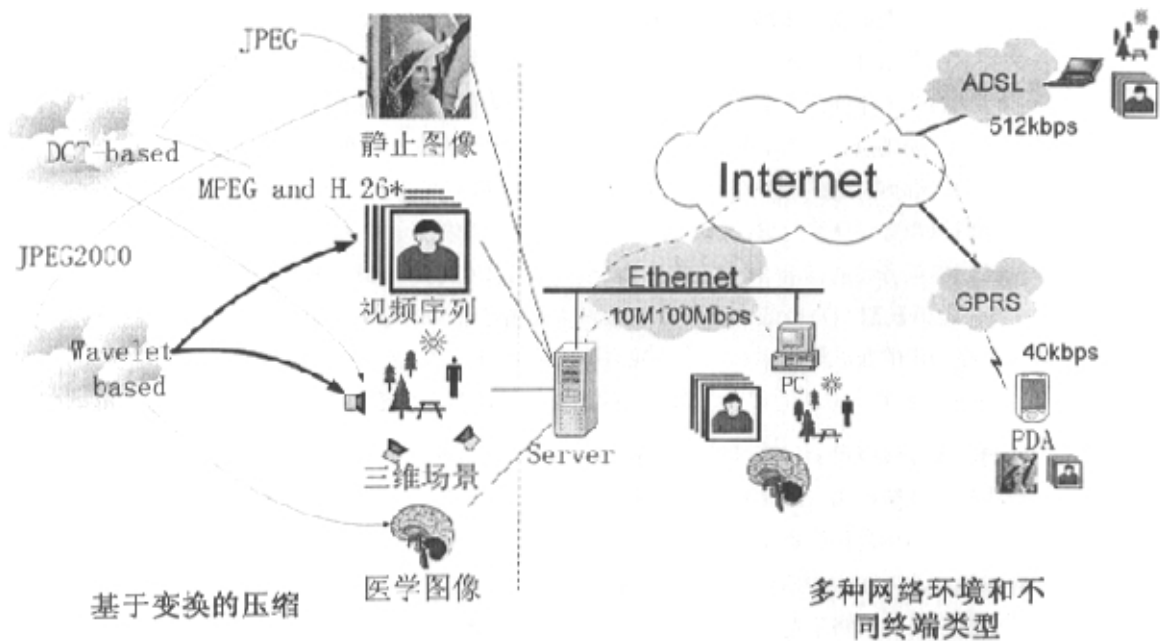


图 1-1 多种图像媒体的相应压缩方法以及传输的网络条件和终端类型（Compression schemes of various image data and the different transmission channels）

图 1-1 左端所示的图像数据包括静止图像，视频（运动图像），三维场景图像序列以及医用图像等类型。图像压缩一般采用基于变换的压缩方法，可以简单分为基于 DCT 的方法以及基于小波的方法。基于 DCT 的方法比较成熟，前面提到对静止图像有 JPEG，对视频图像也有 MPEG*和 H.26*一系列标准。传统基于 DCT 的压缩方法在对应图中所

示的多种网络应用方面具有一定的局限性,尤其是对于视频等高维图像媒体。因此许多研究者也针对网络的需求提出了码流压缩比转换(transcoding)^{[13][14][15]}技术,不同压缩比码流切换的联播(simulcast)^{[16][17][18]}以及目前致力研究的精细可伸缩性(fine granular scalability, FGS)^{[19][20]}与渐进精细可伸缩性(progressive fine granular scalability, PFGS)^{[21][22]}视频编码。然而这些技术的采用对压缩效率都会有较大程度的影响。

相比基于 DCT 的方法,基于小波的图像压缩有以下的几点优越性:首先,与傅里叶分析得到的离散余弦变换 DCT 相比,小波变换具备了时频局域性(time-frequency localization),即在得到信号的频率信息的同时,还能够知道该频率分量在时域的空间位置,例如信号的边缘、纹理信息,因而能够更精确的刻画图像内部的细节信息;其次,小波变换是对整个图像进行变换,避免了 DCT 编码中分块造成的量化重构后的块效应,视觉效果更为理想;更为重要的是,小波变换具有多分辨率分析(multi-resolution analysis)的能力,除了更符合人眼视觉系统的特征之外,还使小波压缩后的码流自身就具有压缩质量,分辨率等多方面的可伸缩性(scalability),适于网上传输等应用中针对不同需求的灵活解码^[23]。小波变换的这些特点,使其成为图像编码中极有发展前景的方法。

多年来,小波在静止图像压缩方面的研究取得了很大的进展。20 世纪 80 年代末采用过矢量量化方法(VQ)编码小波变换后的系数,研究小波图像压缩^{[24][25]},但效果低于基于 DCT 的 JPEG^[1]压缩效果。由于没有找到合适的小波系数的熵编码方法,小波压缩发展一度停顿不前。1993 年,Shapiro 提出了一种新的编码方法——嵌入式零树编码 EZW^[26](embedded zerotree wavelet),利用小波变换后各子带系数之间在空间位置上的相关性组织树的结构进行预测和熵编码,除了取得了超过 JPEG 的压缩效果,还提供了嵌入式编码(embedded coding)结构,即压缩码流可以在任意位置截断并恢复出相对码率下的图像。EZW 的出现大大推动了小波图像压缩的发展。以此基础上发展的 SPIHT^[27](set partitioning in hierarchical trees)等通过多级树集合分解等方法编码小波系数,进一步提高了压缩性能。

作为基于小波压缩的代表,1999 年底提出的 JPEG2000^[28]静止图像压缩标准,采用优化截断嵌入式块编码 EBCOT(Embedded Block Coding with Optimized Truncation)方法对小波系数组织编码^[29]。与 EZW 和 SPIHT 不同,这种方法利用的是同一子带内小波系数之间的相关性进行分块的熵编码,提供了对码流进行部分解码重构的灵活性。同时根据分块内部的率失真特性,EBCOT 实现了对分块码流组合成整体码流过程的全局优化。JPEG2000 除了在压缩性能上大大超过 JPEG,还提供了多分辨率解码,清晰度渐进的解

码, 基于视觉权重的量化以及感兴趣区域 (region of interest, ROI) 编解码等功能, 更加适合于当今网络以及卫星遥感图像的传输和应用, 成为图像压缩方法进入新阶段的标志。目前 JPEG2000 的核心压缩部分 (part 1) 已基本完成, 扩展部分 (part 2) 正在进一步完善, 同时相关的图像数据交互传输 (part 9), 三维图像体数据成像 (volumetric imaging) (part 10), 无线传输 (part 11), 以及针对数码相机等市场应用的 Motion JPEG2000 (part 3) 等部分也都处于不同的进展状态, 目前还有许多研究工作正在进行, 分别探讨视觉掩模 (visual masking)^{[30][31]}, 有线和无线信道传输^{[32][33]}, 远程交互浏览^[34], 以及 VLSI 硬件实现等方向^{[35][36]}。

基于小波在静止图像压缩上取得的巨大成功, 小波在高维图像媒体 (包括视频, 图形学数据等) 压缩中的应用受到了广泛的关注和研究。本文的研究工作就是围绕着对高维图像媒体的压缩展开的。小波高维图像媒体压缩中一些基本的技术是和小波静止图像压缩相通的, 但是在具体的特点和压缩要求上又有不同的侧重。在接下来的章节中, 我们首先通过小波静止图像压缩来介绍小波图像媒体压缩中的主要技术, 之后对高维媒体压缩的特点以及对压缩算法的要求进行阐述。

1.2 小波静止图像压缩技术介绍

对自然图像的压缩一般是个有损压缩 (lossy compression) 的过程。同普遍的基于 DCT 的图像压缩流程类似, 基本的小波编码器也包括两个主要部分, 小波变换部分和量化熵编码部分, 如图 1-2 所示。通过小波变换, 信号能量被集中到少量的低频系数中; 在量化和熵编码部分, 将小波变换得到的系数进行有损的量化去掉部分相对不敏感的信息, 之后根据信息熵的理论, 用尽可能少的比特 (bit) 来表示所需的信息。另外, 对于嵌入式的码流, 还可以基于率失真理论进行码流的优化组装^[37]。

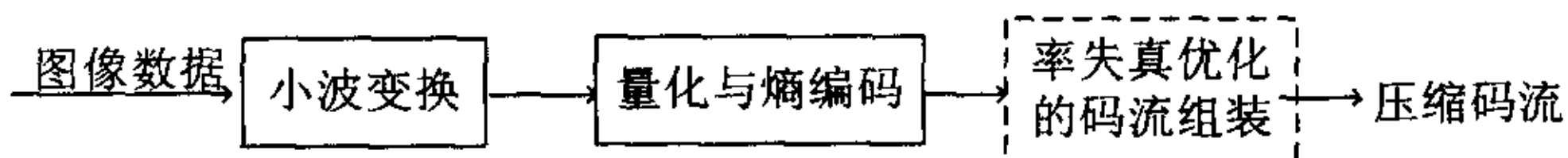


图 1-2 小波编码器结构 (The wavelet codec structure)

下面我们结合小波图像编码器结构对小波变换部分和后续的量化熵编码部分作一详细介绍。

1.2.1 小波变换

1) 小波变换原理

小波理论涉及到逼近论，调和分析和统计学等多门数学分支，然而从数字信号处理的角度，我们也可以简单的认为小波是一种具有时频局域性和多分辨特性的滤波器，通过将原始信号变换到小波域，可以有效的去除信号间的线性相关性，将信号的能量集中到少数几个幅值较大的小波系数上去，再对小波系数进行量化和熵编码，就达到了压缩的目的。为了使读者能从压缩角度更为直观的了解小波的原理和应用方法，我们略去复杂的数学推导，仅从信号处理的角度来介绍离散小波变换（discrete wavelet transform, DWT）及其在图像压缩中的应用。关于小波具体的滤波器构造和理论证明，读者可以参考^{[23][38]}。

一维的一级离散小波变换的一般形式可以用图 1-3 左来表示。这里一个信号分别通过低通和高通滤波器， H_0 和 H_1 ，再经过基于 2 的下采样，就构成了一级变换。得到的系数 $a_0(n)$ 和 $a_1(n)$ 被称为小波系数。多级小波变换是通过重复在低通分支上进行一级变换完成的。这里低通和高通滤波器是对应于一个小波。在图像压缩应用中，所用的小波需要具有对称性（symmetry）和紧支集（compact support）。同时具备这两个条件的正交（orthogonal）小波只有 Haar 小波。但 Haar 小波不连续，不能有效的表示具有一定光滑性的自然图像，导致压缩效率较差，因此在图像类媒体压缩中多采用双正交（bi-orthogonal）小波，如 Daubechies 9-7（对应的低通滤波器长度为 9，高通为 7）以及 Le Gall 5-3 小波等。Daubechies 9-7 滤波器具有较好的光滑性和近似正交性，适合于有一定光滑性的自然图像的压缩，因而被 JPEG2000 采纳作为标准滤波器；5-3 滤波器由于其整个滤波过程都可以在整数域上完成，计算速度较快，也被 JPEG2000 采纳。详细的滤波器种类请参照附录 II。在文章[39]中给出了不同小波滤波器的图像压缩性能分析。

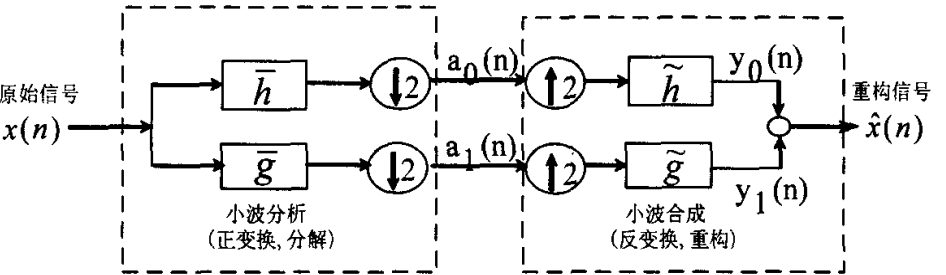


图 1-3 一维一级小波分解与合成示意图（1D one-level wavelet analysis and synthesis）

如图 1-3 所示: \bar{h} 和 \bar{g} 分别为低通和高通的滤波器, \tilde{h} 和 \tilde{g} 为对应的合成滤波器。 $a_0(n)$ 和 $a_1(n)$ 分别为下采样后得到的低频和高频子带, 采样之后的得到低频和高频子带的总长度恰好等同于信号的长度。合成的过程则与分析过程相反, 如图右所示, 低频和高频子带经过小波合成后, 可以得到与原始信号完全相同的重构信号。分析的过程又称为对信号的分解 (decompose), 合成的过程又可以称为重构 (reconstruct)。

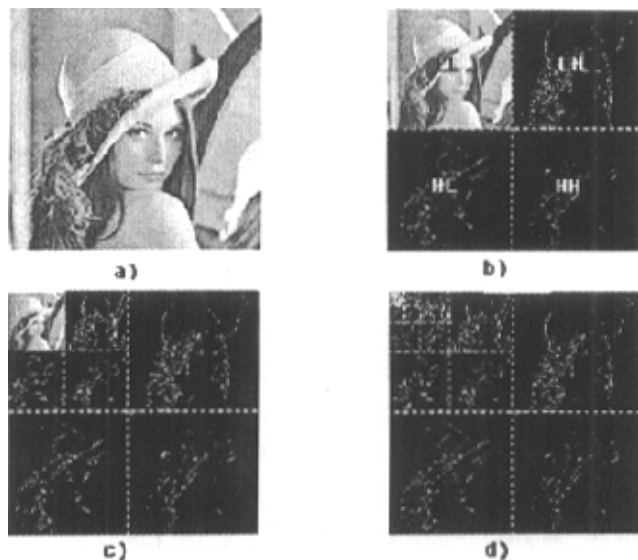


图 1-4 图像的多级小波分解 (Multi-scale wavelet decomposition for the image data)

对于图像信号, 通过对水平和垂直方向独立的进行一维小波分解, 可以将原始图像分为 4 个子带, 如图 1-4 b) 所示。对图像变换后得到的小波系数具有如下的特点: 左上角的 LL 子带代表图像低频信息, 如颜色, 变化趋势等, 右上角的 LH 子带和左下角的 HL 子带分别代表了水平方向和竖直方向的边缘和纹理等高频信息, 而对角方向的高频信息集中体现在右下角的 HH 子带中。依赖于小波具有的时频域局域性, 每个子带不但包含了能量信息, 还同时表示了能量所处的位置信息。

在小波图像压缩中, 通常会对低频子带 LL 进一步分解, 得到更低分辨率的 4 个子带, 如图 1-4 c) 所示。图 1-4 c) 和 d) 分别为 2 级和 4 级的二维小波分解系数, 可以看出多分辨下的小波系数矩阵, 高频部分由粗到细体现出了边缘信息的强度和位置, 为编码的设计提供了不同尺度和分辨率下的信息, 便于进行适当的压缩。多分辨的结构, 不但有利于能量的进一步集中, 并且还可以实现编码图像基于分辨率的可伸缩性

(scalability)，例如当网络传输带宽不足或客户端分辨率不足时，只解码对应于低分辨率的小波系数得到相应大小的图像。

上面的小波变换只是对低频分量作分解，根据实际图像的时频特性的不同，可以选择对一些高频子带作进一步分解，不同的多级分解形式对应了不同的小波包 (wavelet packet) 结构，如图 1-5。其中的 Mallat 分解就是上面所提及的，也是自然图象压缩中最为常用的分解方式，而对于包含细节较多的图像，例如医学或遥感等特殊的图像，采用 Spac1 或其他自适应的小波包分解可以提供更紧凑的表示 (compact representation) [28][40][41]。

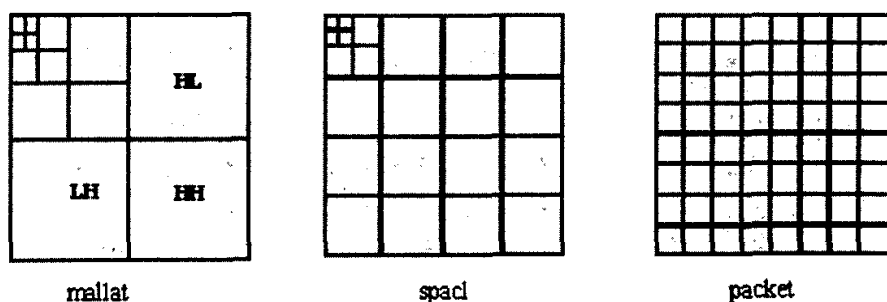


图 1-5 三种典型的小波包分解方式 (Three typical wavelet packets)

2) 第二代小波变换——提升型算法 (Lifting Scheme)

传统的基于卷积滤波的离散小波变换算法，运算量和内存的消耗都较高。1995 年 Sweldens 等提出了小波变换的另一种构造和计算方法——提升型算法 (lifting scheme) [42][43][44]，这种方法可以直接在原信号域上进行小波分解，从而摆脱了对傅里叶变换的依赖，降低了变换的运算量和内存需求。另一方面，提升型算法在区域边界，非规则采样点，以及曲面上的小波构造等方面提供了更大的灵活性，因此被称为“第二代小波” [44]。

图 1-6 显示了用提升型算法实现的一维一级小波变换的多相形式。左边的变换矩阵 $\tilde{P}(z^{-1})^t$ 表示小波分解的过程，右边的反变换矩阵 $P(z)$ 表示小波合成过程。小波变换的完全重构 (perfect reconstruction) 特性由公式 $P(z)\tilde{P}(z^{-1})^t = I$ 给出。Daubechies 和 Sweldens 指出，所有具有有限冲激响应 (finite impulse response, FIR) 的小波滤波器的变换过程 $\tilde{P}(z^{-1})^t$ 都可以分解成如图 1-6 左下角的提升型结构 [45]。提升型算法通过分层交替的步骤，在每一层只有一半系数交错的参与运算。这些系数分别对应于高频系数或

是低频系数的位置,并且每一层系数基于上一层的结果计算产生。在这种交替的结构下,每层的运算结果可以直接保存在原来的存储位置上,不必开辟新的存储空间,就可以实现直接在信号域上进行的小波变换。这种特点称之为原位(in-place)存储。同时,提升型结构将小波变换的运算量降低为卷积滤波方法的 $1/2^l$ 。本文的工作很多处利用到了这种提升型结构,更具体的内容我们将在后面的文中进一步介绍。

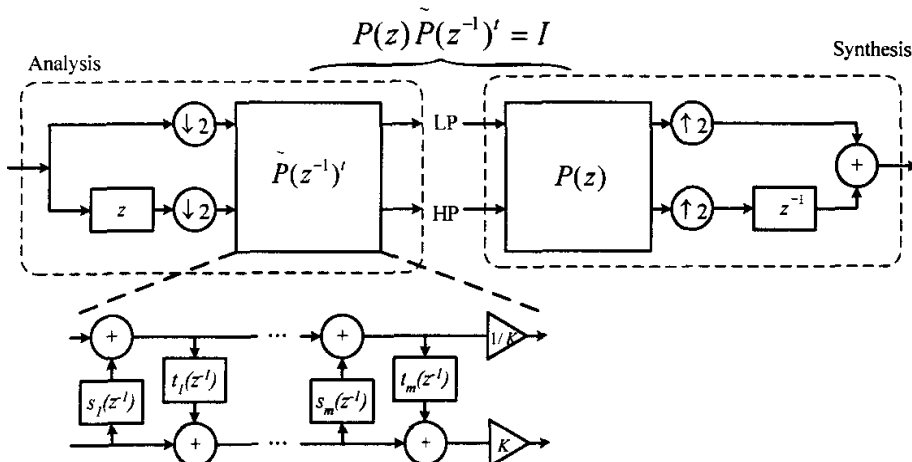


图 1-6 多相小波变换提升型实现(引自[45])Wavelet transform in the polyphase form and lifting (from [45])

1.2.2 量化和熵编码

小波变换系数的低频子带集中了较高的能量信息,高频中主要是边缘与纹理细节的位置和幅度信息。如何适当的去除人眼不敏感的信息,并根据信息的熵值,用较少的比特来表示信息,以及如何组织这些信息的优先级,使得图像的主要信息能够在压缩码流中先得到传输,以便在特定的带宽条件下丢弃相对次要的信息,就是量化和熵编码,以及码流组装中的主要任务。Xiong 等通过对基于 DCT 及小波的压缩的分析指出,对小波静止图像压缩性能的提高不在于如何设计更好的滤波器,而是合理利用和组织小波系数的相关性的量化和熵编码部分^[47]。

传统的量化概念是将原本连续分布的变换系数按照大小归入一定的量化级(一般是均匀的),用所在量化级的代表值来代替所有在该级内的系数的值,达到信息压缩的目的。由于这种量化级是固定的,无法满足多码率的分析,后来发展起来的嵌入式编码(embedded coding),按照一组渐变的阈值对系数进行逐级编码,得到压缩精度按码流

¹在某些情况下,这个因子是 $1/4$ ^[46]。

顺序不断细化最终达到无损的压缩码流。嵌入式码流可以在任意位置截断，按照给定的码率恢复出相应精度的图像，代替传统量化达到有损压缩的目的。一般采用阈值为 2，即位平面（bit plane）^[48]的结构对小波系数扫描并进行熵编码。

	b_7	b_6	b_5	b_4	b_3	b_2	b_1	b_0	Sign
w_0	0	1	0	1	1	0	1	...	+
w_1	1	0	0	1	0	1	0	...	-
w_2	0	0	1	0	1	0	1	...	+
w_3	0	0	0	1	1	1	0	...	+
w_4	0	0	0	0	1	0	0	...	-
w_5	0	0	1	0	0	1	0	...	-
w_6	0	0	0	0	1	0	0	...	+
w_7	0	0	0	0	0	0	1	...	-

图 1-7 系数与位平面的关系（Relationship between the coefficients and the bit planes）

位平面编码的概念如下：对于某个由 2 进制表示的整型系数 $a_n, a_{n-1} \dots a_0$ ，其中 a_i 表示二进制的位，左边第一个为 1 的位 a_k 代表了该系数数值的大小量级，而后面的位则是对该数值的精度细化。如图 1-7： w_0, \dots, w_7 代表的行是系数， b_1, \dots, b_7 代表的列即为位平面。对于每个位平面 l ，若系数 w_i 在所有先前的位平面 $k > l$ 中的位都是 0，则称 w_i 在当前位平面为无效（insignificant），反之则为有效（significant）。

由于小波变换后的系数能量集中，大部分的高频系数绝对值较小，因而在先前的位平面中为零。对于大量的连续的零，可以用比较少的比特表示，因此节约了码流的开销。在每一个位平面中，根据系数间的相关性通过预测等方法可以使更多的 '0' 组合在一起进行熵编码来进一步压缩码流。位平面编码按照位平面的顺序而不是系数的顺序编码，这样就可以得到精度逐渐增加的嵌入式码流。当没有足够的码率来提供进一步细化的信息时，也能保证重要信息的先传。

对于位平面内的熵编码方法可以有很多选择，例如游程编码（run-length coding）^[49]以及算术编码（arithmetic coding）^[50]等。由于嵌入式码流的可任意截取的特点，可以根据给定码率的需要在相应位置截断得到待传输的压缩码流。

对于小波系数的编码是从子带间和子带内的系数在空间上具有相关性这个特点来考虑的。EZW^[26]和 SPIHT^[27]都属于根据子带间相关性组织树状结构的嵌入式编码方法。

在 JPEG2000 标准^[28]中, 则利用的是子带内的相关性。将各子带分成均匀大小的块, 分块独立进行编码。下面我们对这几种主要的编码技术作一简介:

1) 嵌入式零树编码 EZW (Embedded Zerotree of Wavelet)

1992 年, Lewis 首先提出了树状结构的编码方法^[51], 通过子带由低频到高频的系数相关性组织树状结构, 当某个树统计的能量小于给定的视觉阈值时, 将该树对应的所有子带系数都作为零处理, 达到压缩的目的。在这种方法下判决为零的树, 即使其中有较大的系数, 也将被置零, 对这些系数造成较大的失真。1993 年 Shapiro 在树结构的基础上提出了嵌入式零树编码^[26], 采用确定的递减的阈值来编码树结构, 保证零树中高频子带的系数都小于相应的低频子带系数, 避免了原来方法使较大高频系数失真的现象, 同时采用熵编码对所有树的状态进行编码, 达到了进一步的压缩。

在嵌入式零树编码中, 每个低频子带的系数在下一级高频的子带的相应位置上都会对应 4 个系数, 如图 1-8(a)所示。以此类推可以得到树状的结构。零树的概念基于这样的假设: 相对给定的阈值 T , 如果低频子带中的系数小于 T , 那么所有更高频子带中相对位置的系数都是小于 T 的。满足这个假设的树结构被称为“零树”。在多数情况该假设是成立的, 因此可以用很少的码字来表示零树结构中的大量系数。大于当前 T 并且没有被编码的的系数称为有效系数。

对于一组逐渐降低 (一般以 2 的倍数) 的阈值, 按图 1-8(b)中的顺序对整幅图像的子带进行扫描, 分别对检测出的有效系数、零树、以及单个的零系数赋予相应的码字并进一步进行算术编码。小波子带之间的相关性通过零树结构和熵编码得到了压缩, 并且结合逐渐降低的阈值对系数逐渐的细化, 实现了嵌入式的编码。基于嵌入式零树编码, 小波编码方法达到了超过 JPEG 的压缩性能, 是小波图像压缩的一个突破性进展。

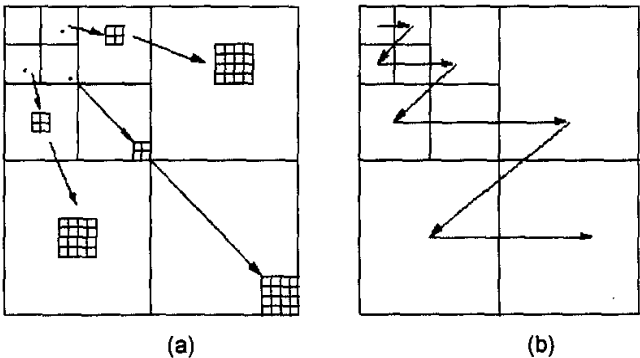


图 1-8 零树编码结构和扫描顺序 (The coding structure and the scan order of EZW)

2) 多级树集合分解算法SPIHT (Set Partitioning in Hierarchical Trees)

EZW 的出现使小波方法得到了重视, 之后的许多研究都对这种方法进行了改进, 其中最为瞩目的是 SPIHT 方法^[27]。1996 年, Said 和 Pearlman 分析了 EZW 出色性能的原因, 认为 EZW 方法的核心在于对于小波系数按不同类型集合分离的局部排序编码, 按位平面的结构细化, 以及对子带间小波系数之间关系的发掘。因此, 在 EZW 的基础上, Said 和 Pearlman 提出了将集合类型进一步分解的多级树结构 SPIHT 算法, 对 EZW 中有效系数、零树、以及单个零系数这三种类型集合继续细分, 实现了更为有序的熵编码。SPIHT 方法的性能比 EZW 方法有了进一步的提高。

3) JPEG2000中的优化截断嵌入式块编码EBCOT (Embedded Block Coding with Optimized Truncation)

1999 年, Taubman 提出了优化截断嵌入式块编码方法 EBCOT^[29]。EBCOT 与 EZW 和 SPIHT 方法的最大区别在于利用的是同一子带内系数的局部相关性, 而不是子带间系数的空间关联。EBCOT 方法首先将子带内系数分成固定大小的块, 对每个分块进行独立的位平面熵编码, 同时记录下每块相应的分段率失真曲线。在将块组合成整个码流时, 根据分块内部的率失真特性, 实现全局优化的码流组装。这种分块的编码方式优点在于: 利用了小波变换系数相关的局部性, 即较强的相关性只存在于邻近的系数中, 分块后独立的编码可以更好的发掘这种局部相关性; 并且分块结构可以按照分块的率失真特性进行全局优化的码流组装。同时, 独立的分块处理节省了内存开销, 可以灵活的支持图像的部分解码。基于 EBCOT 的以上优点, JPEG2000^[28]静止图像压缩标准采用了 EBCOT 编码方法。本文的压缩工作中的熵编码部分采用对 EBCOT 进行的三维扩展, 得到了很好的应用结果。

1.3 小波高维图像类媒体压缩的特点和研究侧重

基于小波在静止图像压缩上取得的巨大成功, 许多研究者开始关注于小波在高维图像媒体 (包括视频, 图形学数据等) 压缩中的研究工作。和现有的帧间预测结合 DCT 的方法比较, 小波高维图像压缩具有压缩质量和帧率、分辨率等多种可伸缩性, 对适应网络条件的数据传输具有新颖的意义。高维图像媒体压缩在某些方面上可以借用图像压缩中一些成功的技术, 但从研究的侧重点来看, 两者有很多不同。本文对高维图像媒体压缩的研究一方面侧重于与现今已有的视频编码技术的比较, 另一方面探讨了小波编码在新兴高维图像媒体中的应用。下面我们将结合视频和基于图像的场景绘制两类媒体的特点对小波在其中的压缩应用进行分析。

首先介绍现今高维媒体图像压缩中的主流标准。

1.3.1 高维图像类媒体压缩的主流标准简介

现有的高维图像媒体压缩标准主要针对视频媒体,包括 MPEG-1^[5]、MPEG-2^[6]、MPEG-4^[9]、H.261^[10]、H.263^[11]和 H.26L^[12]等,这些压缩标准普遍采用了帧间运动补偿预测结合 DCT 变换的方法来按照时间顺序进行编码。由于高维图像媒体具有的许多共性,视频压缩中的很多技术也可以应用于其他高维图像媒体压缩,不过对应具体的应用特点需要进行相应的编码设计。

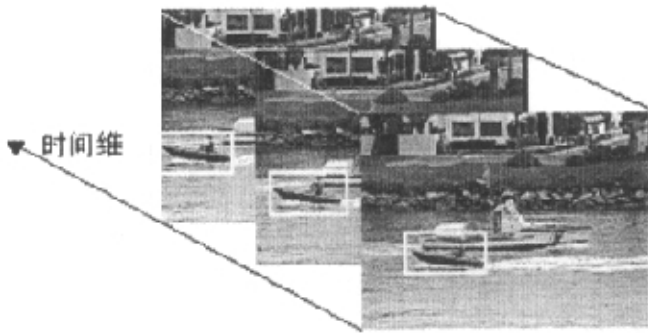


图 1-9 运动信息对时间轴上帧间相关性的影响 (Motion information and the temporal correlation)

目前的视频压缩标准都是基于帧间运动补偿预测结合 DCT 变换和熵编码实现的。基本的压缩框图如图 1-10。待编码的视频依照时间次序一帧帧的输入编码器。简单的视频帧结构包括两种类型的帧: I 帧 (intra frame) 和 P 帧。对于 I 帧是以类似图像编码的方法进行帧内编码, 对于 P 帧, 则是采用帧缓存 (frame buffer) 中前一帧图像 (I 帧或者 P 帧) 作参考帧进行基于宏块 (macro-block, MB) 的运动估计, 并将当前处理的帧与运动补偿后的参考帧差分得到预测残差 (predicted residua) 图像, 进行分块 DCT 变换和量化, 最后通过熵编码 (entropy coding) 得到压缩码流。由于在解码端只能得到压缩重构后的帧, 因此为了保证编解码端的一致性, 每一帧的参考帧都是采用前一帧的重构图像而不是原始图像。将当前量化后的 DCT 系数通过反量化和反 DCT 变换得到重构残差图像, 再与运动补偿后的参考帧相加得到重构的当前帧, 并存储在帧缓存中作为后面编码的参考帧。从图中灰色的部分看出, 编码过程包含着“预测当前帧-) 预测残差量化压缩-) 重构-) 用于下一帧参考”的循环过程。

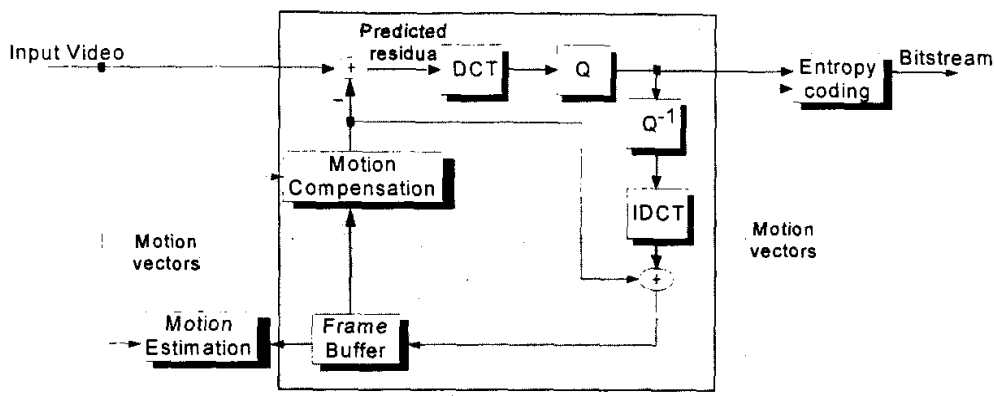


图 1-10 基于 DCT 的视频编码标准压缩流程 (Flowchart of the DCT-based video coding standard)

MPEG-4^[9]和 H.26L^[12]是视频编码标准中发展较新的两套编码器, 其中 MPEG-4 提出了面向对象的编码理念, 即对视频帧的描述是基于其中的对象 (object), 而不只是整个矩形帧, 但由于对象分割 (segmentation) 技术的限制, 目前标准中并没有应用到这一部分。在编码的细节上, H.26L 中采用了帧内预测 (intra prediction), 可变尺寸块运动搜索 (variable-size motion search), 多参考帧预测 (multi-reference prediction) 等技术, 因而在编码效率上, 大大高出了同等条件下的 MPEG-4 编码。由于技术上的许多相似性, 目前 H.26L 和 MPEG-4 成立了联合视频组 (joint video team, JVT) 共同研究视频编码的相关技术。

基于 DCT 的编码标准存在很多局限性: 首先, 编码系统中核心的部分是循环结构, 因此视频帧在用作后面的参考帧之前, 已经压缩为固定码率的码流, 因此基本的编码系统不具备压缩质量, 分辨率等可伸缩性 (scalability), 不能很好的适应网络传输中变动的网络带宽等条件。为了解决可伸缩性的问题, 目前针对 MPEG 和 H.26L 中有许多关于精细可伸缩 (fine granularity scalable, FGS)^{[19][52]}以及渐进精细可伸缩 (progressive FGS, PFGS)^{[21][55][56][57][58]}编码方法的技术提案和文章, 采用分层结构编码来得到可伸缩的码流, 编码的复杂度亦有所提高。其次, 对图像的分块 DCT 变换在压缩后会造成分块间的块效应, 影响了压缩质量和视觉效果。

人们一直关心, 是否会有更灵活有效的技术来解决视频编码中的现存问题。小波在静止图像压缩中的成功为人们带来了启示。2001 年 12 月在 Pattaya 召开的第 58 届 MPEG 大会上, 就成立了专门的探索帧间小波编码技术的特别小组 “Ad-hoc group on Exploration of Interframe Wavelet Technology in Video”^[59], 致力于研究基于小波的视频编码技术。

1.3.2 小波视频压缩技术的研究及进展

由于小波在静止图像压缩中体现的高效去相关能力和优秀的可伸缩性,对小波视频压缩的研究也开展了很多工作。小波的多分辨率可伸缩性以及嵌入式编码等许多优点是传统的视频压缩标准所不具备的,因此得到了许多研究者的重视。图 1-9 可以看出视频并不只是图像简单意义上的三维扩展。图像表现的是对静态的连续场景在空间上的采样,而视频是对动态场景在空间和时间维上采样得到的图像序列。视频中时间维上具有的帧间运动等信息的特征与静止图像内的空间信息有很大不同,例如物体运动造成的时间维上的跳变,与图像中的边缘信息跳变本质上差别很大。另外,人眼对动态信息的视觉感受与静态图像也有很大差异。利用时轴的运动信息与小波变换进行有效的结合是目前小波视频压缩的重点研究方向。

小波视频压缩的研究大体可以分为两个方向:首先是二维小波变换+时轴运动补偿的方法(2D+T),根据先进行时轴运动补偿还是先进行帧内分解的顺序对应了不同的类别^[60]。一类 2D+T 方法采用类似 MPEG 的结构对重叠块运动补偿(overlapped-block motion compensation, OBMC)^[61]预测后得到的残差图像,采用小波变换结合图像压缩中类似 EZW 和 SPIHT 的方法进行编码^{[62][63][64][65]}。但由于小波变换不擅长表示残差图像的特征,对于 MPEG 结构中对应的预测帧 P 帧(P frame)并没有得到优于 DCT 编码的结果。2D+T 的工作还有另一个方向,就是先对视频帧做二维小波分解,之后在时轴对应的子带上进行帧间运动补偿预测,由于小波子带是采样过的数据,不具有时移不变性(translation invariance),直接作子带间的预测达不到满意的精度^[66],因此需要特殊的处理,通过插值,变形(warping)或者采用相位移位和过完备(over-complete)的小波系数等手段来处理^{[67][68][69][70]}。

小波视频压缩的另一个研究方向是将视频序列当作三维数据,在时轴,水平和竖直轴向分别进行小波变换,利用用时轴的小波变换代替运动补偿预测来对帧间信息进行去相关。这种方法回避了预测编码中的循环过程,因此在解码端避免了误差漂移(error drifting)的问题,适于在如无线信道等易发错误的环境下的编码传输。最简单的方法是按照像素在每帧对应的位置直接进行时轴小波变换,如 Kim 等采用类似基于图像的 SPIHT 进行的三维扩展^{[71][72]}。基于小波潜在的可伸缩性特点,三维小波视频编码可以同时达到压缩质量,时轴帧率和帧内分辨率的可伸缩性,然而由于视频帧之间存在运动,处在相同位置的像素很可能不是来源于真正对齐的同一物体,因此时轴滤波时的帧间相关性减弱,限制了编码效率。

目前许多三维小波视频压缩工作研究集中在如何将运动信息与三维小波变换结合。Taubman 和 Zakhor 采用简单的摇镜头 (pan) 运动补偿模型在三维小波变换前进行像素的运动对齐^[73], 在多种码率下得到了比无运动模型的方法高出 0.56-1.29dB 的压缩效率。Wang 等人^[74]提出了将所有帧利用全局运动 (global motion) 通过仿射变换变形到同一坐标系, 并对变形后的帧进行三维小波变换, 对于背景运动较多的视频, 相比无运动的小波变换得到了 0.56-2.42dB 的压缩性能改进, 但由于仿射变换对像素点进行了重采样, 因此在高码率端不能实现小波反变换合成的完全重构 (perfect reconstruction)。考虑到全局运动的局限性, Ohm 提出采用传统的基于块匹配的方法预测运动矢量, 采用支集长度较短的 Haar 小波进行时轴变换, 同时仔细处理了连接和未连接的区域^[75]; Choi 和 Woods 在 Ohm 工作的基础上进一步改进了连接和未连接的处理^[76], 但是从采用了这些技术的 SPIHT 编码器的效率看来, 得到的编码增益只有 0.23dB^[72]。以上介绍的方法有些不能够实现小波反变换合成的完全重构, 有些的运动精度不足, 所以虽然结合了运动对齐, 编码效率还是不能满意。Hsiang 和 Woods 结合了分块的全局运动, 对相邻两帧的匹配块合并后进行帧内变换实现时轴小波分解, 在利用了半像素精度的运动信息同时, 能够保证完全重构^[77], 但压缩效率仍然有限。在文章[78][79]中, Secker 和 Taubman 探讨了采用提升型结构进行帧间小波变换的一些技术, 并给出了不同滤波器的性能比较, 虽然文中没有给出具体的与传统编码器的比较结果, 但是算法的分析具有很好的参考价值。

最近, MPEG 成立研究帧间小波性能的特别小组^[59], 开始积极的投入对帧间小波编码技术的研究。Chen 与 Woods 提出的 MC-EZBC 达到了很出色的编码效果^{[80][81]}。MC-EZBC 方法基于 Choi 和 Woods 的工作^[76], 采用运动补偿后的提升型 Haar 小波将视频序列每两帧分解成时轴上高频低频小波系数帧, 并解决了运动对齐中精度不足的问题, 使 MC-EZBC 达到了与 H.26L 标准^[12]接近的编码效率。然而, 由于 Haar 小波滤波器的支集长度很短, 不能很好的描述时间延续较长的帧间相关性。为了利用延续较长的相关性, Xu 和 Li 等提出了运动抽线 (motion-threading, MTh) 技术^{[82][83]}, 根据宏块 (macro-block, MB) 运动矢量将时轴的像素串连成长度不同的运动线, 并在运动线上作适形小波变换 (shape-adaptive DWT)^[84]得到时轴分解, 将分解后的时轴小波系数存储在原来像素的位置, 再对每帧图像做帧内的小波分解。运动抽线技术提供了比较灵活的利用帧间运动信息进行小波变换的结构, 在某些运动简单的视频序列中达到了与 H.26L 类似的编码效率, 但是由于其本身存在断线以及运动精度不足的问题, 因此对许多视频序列的编码效率还存在局限性。

1.3.3 基于图象的场景绘制技术IBR数据压缩

用计算机进行三维物体/场景的再现一直是人们非常关心的问题。好的三维物体/场景的再现方法能给人带来非常逼真的身临其境的感觉,因此具有很多的应用方向,如游戏,医疗成像,电影,广告,模拟军事训练,远程教育,网上购物等。近年发展的基于图像的场景绘制技术 IBR (image-based rendering) ^{[85][86]} 是解决三维场景再现问题的一种有效方法。与以往几何建模的方法不同,基于图像的方法避免了复杂的几何信息和再现物体的不真实感,只需一组按一定位置和角度采样的图片即可再现出逼真的虚拟场景。并且,基于图像的绘制方法的速度与场景的几何复杂度无关,使得这种方法更适用于表达复杂的场景。

目前的 IBR 的数据类型简单有以下几个分支。根据七维的全光函数 (plenoptic function) ^[87], 三维动态场景的全部信息可以用 IBR 方法通过在所有任意时间记录所有任意位置,任意方向,任意波长的光线获得。在不考虑时间和波长的任意性下, McMillan 将 IBR 方法简化成五维的全光函数^[88], 用来反映整个空间的静态场景。1996 年提出的 Lumigraph^[89] 和 Lightfields^[90] 方法都是将物体或观察者限制在三维的有界空间内, 将全光函数简化为四维, 其场景数据的获取方法是通过在固定的网格节点位置向外或向内拍摄图像序列获取静止场景的数据。图 1-11 给出了采用 Lumigraph/Lightfields 技术的 IBR 场景的观察与拍摄方法以及拍摄得到的图像序列数据示意图。可以看出拍摄的数据是一组非常相似的静止图片序列。Shum 和 He 提出的同心拼图 (concentric mosaics) ^[91], 将问题进一步简化到三维问题, 使观察者的运动限制在平面的圆内, 可以向圆外浏览场景。同心拼图场景的获取是通过一台架在旋臂上的相机, 在旋臂匀速旋转的过程中, 向外拍摄出沿所经过的圆周轨迹均匀采样的图片序列。同心拼图的描述详见第二章。

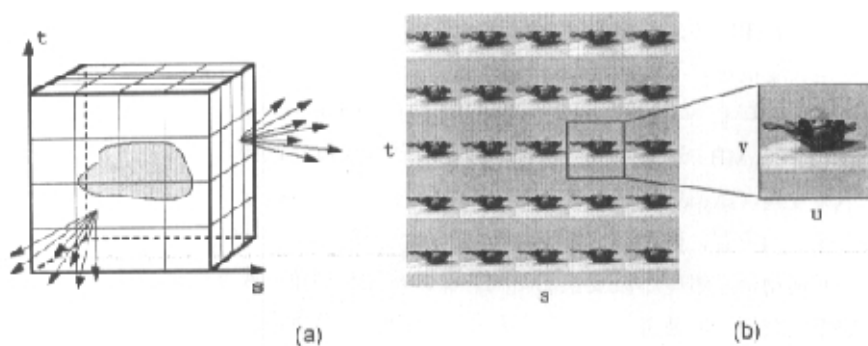


图 1-11 采用 Lumigraph 技术的 IBR 场景的观察与拍摄方法(a)和获取的图像序列数据(b)

简单说来, IBR 的数据可以看作是一组图像的序列, 或者是特殊运动的视频。由于是对场景的密集图片采集, IBR 媒体拥有庞大的数据量。因此, 要达到实时的场景传输和漫游, 对 IBR 的压缩非常必要, 其重点在于挖掘拍摄的图片序列帧内和帧间大量的信息冗余。虽然 IBR 可以看作一种特殊运动的视频, 然而由于其具有的独特特点, 传统的用于视频压缩的方法都具有相应的局限性。首先, IBR 数据是一组有序的, 相机运动参数已知的图像序列, 帧间的冗余比一般的视频帧强。其次, 每次浏览的场景都是静止图像, 而人的视觉系统 HVS (human visual system) 对静止图像的失真的容忍程度要低于运动的视频, 对压缩的要求更高。另外, 对每次观察画面的绘制只需要用到全部数据中一部分信息, 由于数据量庞大, 对 IBR 数据压缩码流的解码无法象图像那样整个解出, 同时也不能象视频一样按照图片的顺序解码, 根据在场景中运动和观察方向的随意性, 压缩编码后的 IBR 数据应该具备随机访问 (random access) 解码当前所需场景的特点。我们定义这种只解码当前场景的概念为即时绘制 (just-in-time rendering, JIT)。传统的视频编码技术的逐帧编码和循环结构不能满足这种随机解码的需求。

由于 IBR 是新兴类媒体, 因此相应的压缩方法还不多^{[92][93]}; 在^[89]和^[91]中, 都采用矢量量化 (vector quantization, VQ) 的方法分别针对 Lumigraph 和同心拼图两种 IBR 媒体进行压缩。矢量量化方法将原始数据分割成小的单元, 通过查表索引来表示每一个单元的内容, 这种方法在解码端可以达到快速的随机解码, 但压缩性能非常有限。文章^[94]中采用基于 MPEG-2 的一些技术提出的固定参考帧方法 MRF 来压缩 Lumigraph 数据, 在达到较好压缩效果的同时, 还支持了压缩码流的随机访问, 类似的技术还被进一步用于同心拼图的压缩中^[95]。

1.4 本文的研究内容及章节安排

本文对基于小波的高维图像视频媒体压缩的研究可以分为两个主要的考虑方向, 一个方向是探讨小波在新兴的同心拼图等 IBR 媒体中的压缩应用, 针对其不同的特点加以分析并提出适合的小波压缩算法; 另一个方向是结合发展已臻成熟的视频技术, 探讨小波在视频压缩中的优势和待解决的问题。本文具体的研究内容包括:

- 探讨了同心拼图 IBR 媒体的特点以及对压缩码流的随机访问解码要求, 设计了相应的基于三维小波的压缩系统。同时, 考虑了同心拼图压缩场景的即时绘制特点, 我们基于提升型小波算法结构提出了一种渐进逆向小波合成快速算法 PIWS, 实验结果表明 PIWS 方法有效的支持了压缩码流的实时解压绘制。基于三维小波的压缩系统与快速解

压缩算法 PIWS 的设计为同心拼图的网上传输和浏览等应用提供了灵活可靠的编解码方案。

- 分析了传统的三维小波视频压缩对时轴方向相关性发掘方面的不足，并结合提升型小波结构提出了采用双向重叠块运动补偿的时轴小波变换结构。接下来对三维小波压缩中的运动抽线技术 (motion threading)，我们分析了影响其压缩质量的断线问题，提出了两轮运动估计技术和高级运动抽线技术，加入了分像素 (fractional-pixel) 运动精度，并分析了视频中常有的遮挡问题，以及考虑了对运动矢量的冗余性去除，将压缩效率提高到与最先进的 H.26L 编码标准相接近的水准。基于本文的工作可以看出，小波视频压缩已经显示出其出色的压缩性能，加上其自身具备的码率，帧率等多种可伸缩性，在将来的网络视频传输中具有良好的应用前景。

- 由于时轴的质量抖动可能会影响视频播放的视觉效果，是三维小波视频压缩中应引起注意的问题。在提高小波压缩质量的同时，我们对于三维小波视频压缩中的时轴周期性抖动现象也进行了理论的分析，得出其原因在于小波合成滤波对不同位置的系数施加了不同的权重的结论。据此提出了使时轴质量平稳的理论的码率控制方法。

本文的章节安排如下：

- 第二章探讨了同心拼图这种 IBR 媒体的特点，以及对压缩算法的要求，设计了基于三维小波的压缩系统，同时提出了一种快速的小波合成算法 PIWS，有效的支持了对同心拼图压缩码流的实时解压缩和绘制。

- 从第三章开始，我们研究了小波技术在视频压缩中的应用。视频压缩的重点在于有效的将时轴运动信息和小波变换相结合。第三章提出了一种基于帧的运动补偿小波提升型结构，通过在时轴提升型结构中加入运动补偿来提高时轴变换时的相关性。

- 第四章介绍了三维小波视频压缩中的运动抽线技术的原理以及优缺点，提出了在运动抽线技术基础上发展的两项工作，并使运动抽线中的运动精度提高到 1/4 像素，同时考虑了小波视频压缩中的遮挡问题以及运动矢量冗余性。

- 第五章研究了三维小波视频压缩中存在的时轴解码质量抖动问题，并提出了相应的理论码率控制方法保证平稳的时轴质量。

- 第六章对全文进行总结，并指出了小波高维图像压缩方面将来的发展前景和一些可行的研究方向。

第二章 同心拼图媒体的三维小波压缩

近年发展的基于图像的场景绘制技术 IBR (image-based rendering) 是解决三维场景再现问题的一种有效方法, 只需一组按一定位置和角度采样的图片即可再现出逼真的虚拟场景。同以往的基于几何建模的方法相比, 基于图像的绘制方法的速度与场景的几何复杂度无关, 更适用于表达复杂的场景。

Shum 和 He 提出的同心拼图 (concentric mosaics)^[91] 是一种较新的 IBR 技术, 能够实现在平面的圆内向圆外任意的浏览场景。同心拼图场景的获取是通过一台架在旋臂上的相机, 在旋臂匀速旋转的过程中, 向外拍摄出沿所经过的圆周轨迹均匀采样的图片序列。在场景的绘制过程中, 待浏览的视野区域被分成竖条狭缝 (vertical slit) 图像, 每一条狭缝图像由获取的图像序列中相应位置的图像来进行重构。在由相机视角限定的范围内, 拍摄得到的同心拼图数据可以用来绘制出任意观察点和任意角度的视图。具体的同心拼图获取方法会在 2.1 节给出。

同心拼图的数据量十分庞大。以实际应用的一组同心拼图为例, 共拍摄了 1350 帧 320x240 的 RGB 图像, 有数据量 297MB (megabytes) 之多。因此, 要达到实时的场景传输和漫游, 对同心拼图的压缩非常必要, 其重点在于挖掘拼图帧内和帧间大量的信息冗余。虽然同心拼图可以看作慢摇镜头得到的视频序列, 然而由于其具有的独特特点, 传统的用于视频压缩的方法都具有相应的局限性。如 1.3.3 节的讨论所述, 同心拼图是相机运动参数已知的图像序列, 帧间的冗余比一般的视频帧强, 并且人的视觉系统 HVS (human visual system) 对静止图像的失真的容忍程度要低于运动的视频。从压缩码流的解码要求来看, 为了支持三维场景中观察者运动和观察方向的随意性, 压缩编码后的同心拼图码流还应该具备随机访问 (random access) 的特点, 支持当前所需场景的即时绘制 (JIT, just-in-time rendering)。

目前针对同心拼图主要的压缩方法有空间矢量量化 SVQ (spatial domain vector quantization)^[91], 以及 RBC (reference block coder) 方法^[95]。SVQ 将同心拼图分割成小的三维块, 并应用矢量量化的方法对每个块进行压缩, 可以很好的满足 JIT 的要求, 但缺点在于压缩效率很低。Zhang 和 Li 提出的 RBC 方法^[95]采用类似 MPEG-2^[6]的 DCT 结合运动补偿的方法, 充分考虑了同心拼图拍摄时的旋转运动信息。通过采用固定参考帧,

以及重新定义压缩码流的语法结构, RBC 不但在压缩效率上略微高出 MPEG-2, 还保证了 MPEG-2 不能实现的 JIT 绘制的要求。

基于对同心拼图的特点和压缩解码的要求, 本章提出了一个基于三维小波编码的同心拼图压缩系统。结合三维小波良好的去相关, 能量压缩和多分辨率的特点, 我们针对同心拼图设计了灵活的选择性局部小波合成和码流随机访问的压缩系统。本章还提出了一种能够实时绘制小波压缩后同心拼图的 JIT 引擎。由于小波滤波器的有限支集 (finite support) 的性质, 由部分的小波系数就能够解码出对应的局部重构图像, 这种局部性的特点从根本上支持了 JIT 绘制, 然而从运算的效率来看仍然不能满足实时绘制的要求。实时解码的瓶颈主要在于选择性局部小波合成模块, 因为如果对于每一个被访问的像素都需要用一定数量的小波系数来解码, 每次解码的运算量要远远高出直接对整个码流解码的平均运算量。利用提升型小波的逐层计算结构, 本章提出了一种渐进逆向小波合成 (PIWS, progressive inverse wavelet synthesis) 方法来降低解码过程的运算量, 使得应用 PIWS 方法的 JIT 解码引擎达到了实时解码的要求。

2.1 同心拼图原理

同心拼图的获取方法非常简单, 通过图 2-1 中一台架在中心固定的自动悬臂上的照相机, 在匀速旋转一周的过程中, 均匀密集采样出沿径向向外拍摄的图片序列。通过将拍摄的图片进行重组可以得到同心拼图的数据。令三维数组 $c(n, h, v)$ 代表拍出一组图片序列, 其中 n 表示图片的序号, h 和 v 分别表示所拍图片中的水平位置和竖直位置, 则定义 $F(h) = \{c(n, h, v) | v, n\}$ 为所有图片中水平位置为 h 的竖条狭缝图像拼成的一个拼接图。定义 N 为拍摄的图片数, H, V 分别为水平和竖直方向的像素个数。如图 2-1 所示, 设旋臂半径为 R , 位置 h 的狭缝图像所表示的光线与径向的夹角为 θ , 则 $F(h)$ 相当于一个虚拟狭缝照相机沿半径为 $R \cdot \sin\theta$ 的旋转轨迹径向拍出的所有狭缝图像的拼图。令 h 分别取 $0, 1, \dots, W-1$ (W 为图片的宽度) 得到的一组拼图分别对应于在不同半径的旋转轨迹上沿切线方向采集到的狭缝图像的拼接, 因此这种场景的获取方法被称为同心拼图方法。

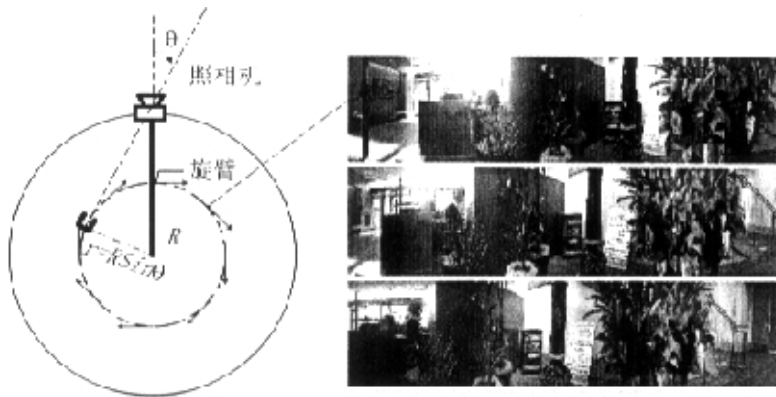


图 2-1 同心拼图的几何位置关系 (Concentric mosaic imaging geometry)

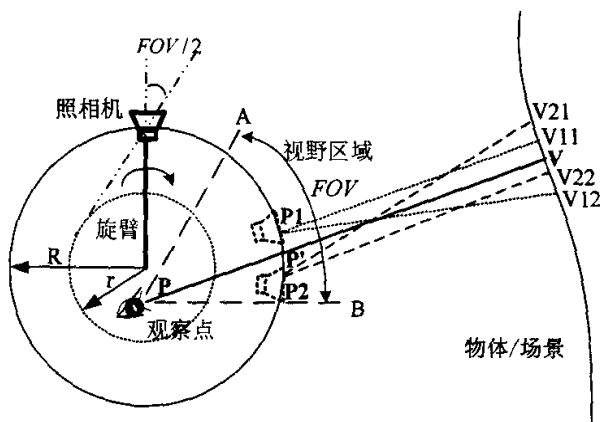


图 2-2 采用同心拼图的场景绘制 (Rendering with the concentric mosaic)

设相机的拍摄角度为 FOV ，在限定的半径为 $R \cdot \sin(FOV/2)$ 的区域中，同心拼图可以完成任意视点任意水平角度的场景绘制。如图 2-2 所示，对于视点为 P 的用户，同心拼图的绘制过程首先将视野区域 FOV 的场景图像分解成单根的竖线狭缝图像，根据光的直线原理，分别在同心拼图序列中取出相应的水平方向拍摄到的竖线图像进行拼接。例如待绘制水平光线 PV 对应的狭缝图像，在同心拼图序列中对应的图像就是 PV 被运动轨迹截取的 $P'V$ 对应的图像。由于采样的不连续性， $P'V$ 可能没有被准确的采到，在实际应用中是采用最邻近的四条狭缝图像 $P1V11$ ， $P1V12$ ， $P2V21$ 和 $P2V22$ 进行双线性插值得到，其中 $P1$ ， $P2$ 是最接近 P' 的两个采样点， $P1V11$ 和 $P1V12$ 是最接近 $P1V$ 的狭

缝图像，P2V21 和 P2V22 是最接近 P2V 的狭缝图像。另外，考虑同心拼图场景深度信息造成的透视问题，可以进一步优化狭缝图像的选择和垂直方向透视失真的解决。更为详细的关于同心拼图的描述可以参见[91]。同心拼图技术通过简单的索引和插值，可以在限定的范围内提供观察者任意位置向任意方向的浏览，因而也要求压缩系统在保证压缩效率的同时，支持绘制引擎对压缩的随机访问并解码相应的狭缝图像。

2.2 同心拼图媒体的三维小波压缩

本章应用三维小波变换方法设计了同心拼图的压缩编码系统。小波变换除了具有出色的信号去相关和能量压缩能力，其内在的多分辨率的性质也可以很好的支持传输带宽或计算速度不足时对同心拼图进行降低分辨率传输的应用。并且，由于小波编码器避免了传统视频编码器的循环过程，非常适于在无线信道等易发错误的信道中传输。

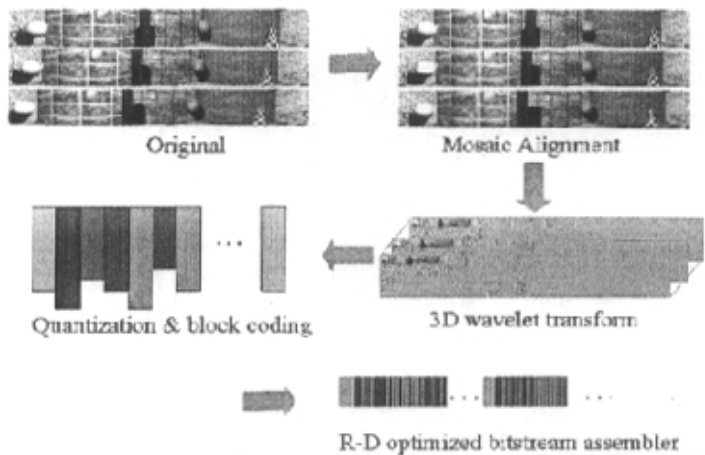


图 2-3 同心拼图的三维小波压缩系统（3D wavelet compression system for the concentric mosaic）

我们的压缩系统可以看作是 JPEG2000 图像压缩标准^[28]中 EBCOT (embedded block coding with optimized truncation) 技术的三维扩展应用。本章针对同心拼图的特点研究了三维小波编码系统的几个主要方面，包括三维提升型 (lifting) 小波，小波包 (wavelet packet) 分解方式的优化选取，以及三维分块熵编码。为了保证压缩码流在具有可伸缩性之外，还满足同心拼图绘制的随机访问要求，对于小波变换系数先进行三维分块，对于每个块进行独立的熵编码。三维小波压缩系统的流程框图如图 2-3 所示，接下来我们对具体每一步的操作进行了详细描述。

2.2.1 校准 (Alignment)

考虑相机旋转的旋转运动, 为了提高同心拼图序列帧间的相关性, 我们先对同心拼图序列作一个预处理, 对每一帧拼图进行内部水平方向上的循环移位来保证和前一帧拼图的最大匹配。匹配原则是相邻两帧经过移位后达到最小的平均绝对误差 (mean absolute error, MAE) :

$$MAE(w) = MAE[F(w), F(w-1)] = \frac{1}{N \cdot H} \sum_{n=1}^N \sum_{h=1}^H |c(n - \Delta(w), w, h) - c(n - \Delta(w-1), w-1, h)| \quad (2-1)$$

其中的参数定义参见 2.1 节。满足绝对误差最小的位移 $\Delta(w) - \Delta(w-1)$ 即为拼图循环移位的校准参数。

2.2.2 基于提升型小波 (Lifting Wavelet) 的三维小波变换

实际拍摄的同心拼图数据是 24 位 RGB 图像, 为了进行有效的压缩, 每一帧拼图图像 $F(w)$ 从 RGB 转换为亮度 (luminance) Y 和色度 (chrominance) 分量 U 和 V , 同时色度分量 U 和 V 的水平 and 竖直方向分辨率进行采样系数 2×2 的下采样 (down-sample), 图像尺寸削减为于原始图的 $1/4$ 。对校准后的拼图图像的每个分量分别采用三维的小波变换滤波, 将图像的能量压缩到少数较大的系数中。

由于整个同心拼图数据量过于庞大, 不能一次性的载入内存, 为了节省内存和降低解码运算量, 系统采用了带有帧/行缓存的提升型小波实现^{[42][43][44][45]}。在第一章介绍过, 提升型结构的变换过程是逐层运算的。图 2-4 所示为一维的一级双正交 (bi-orthogonal) Daubechies 9-7 小波变换 (表 II-1) 的正变换和反变换的提升型实现结构。左图的正变换结构表示, 原始信号 x_0, x_1, \dots, x_6 从提升型结构的左端输入, 右端的两列为分解输出的小波变换系数, h_i 代表高频系数, l_i 为低频系数。可以看出小波系数经过 4 层运算得到, 其中每一层只有一半的信号参与运算而被提升到高一层的状态。4 层的提升型操作可以用公式 (2-2) 描述:

$$\begin{cases} y_1(2i+1) = x(2i+1) + a[x(2i) + x(2i+2)] \\ y_2(2i) = x(2i) + b[y_1(2i-1) + y_1(2i+1)] \\ h(i) = y_3(2i+1) = y_1(2i+1) + c[y_2(2i) + y_2(2i+2)] \\ l(i) = y_4(2i) = y_2(2i) + d[y_3(2i-1) + y_3(2i+1)] \end{cases} \quad (2-2)$$

其中 $y_s(i)$ 表示提升型变换的第 s 层中间结果, 每层的提升型系数分别为: $a=-1.586$, $b=-0.052$, $c=0.883$ 和 $d=0.444$ 。

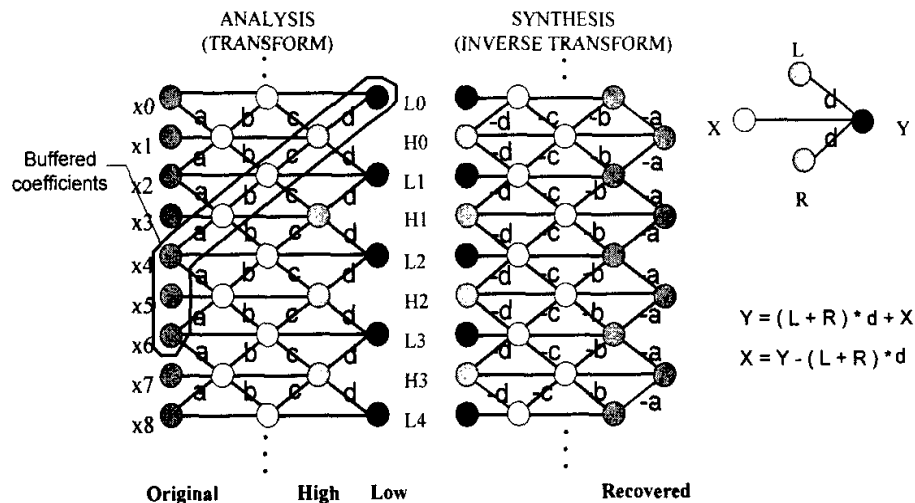


图 2-4 一级提升性小波的正变换与反变换结构, 图右所示为基本提升型单元 (One-level forward and inverse lifting wavelet transforms and the elementary lifting units)

图 2-4 右图为提升型运算的基本单元 $Y=(L+R)*d+X$, 其运算只涉及到三个节点。例如当 x_1 需要被提升, 只有相邻的 x_0 和 x_2 参与当前的单元运算。由于提升型单元可以直接反向得到提升型反变换单元 $X=Y-(L+R)*d$, 提升型小波的反变换结构也可以很容易的由正变换结构得出, 如图 2-4 中图所示。提升型基本单元的重要特点就是原位 (in-place) 运算。即, 在单位运算中, 运算过的节点的原有的值不会再被用到, 因此可以直接存储运算后得到的值, 无需另开存储空间。原位运算的特点除了用于实现帧/行缓存提升型小波之外, 还是 2.3 节中提出的渐进逆向小波合成 (progressive inverse wavelet synthesis, PIWS) 的原理基础。每个单位的提升型运算包含 2 加和 1 乘, 所以 9-7 提升型滤波器每个系数的平均运算量是 4 加和 2 乘, 相比同样 9-7 滤波器的卷积实现中平均的 8 加和 4.5 乘, 节约了一半的运算量。

对应同心拼图数据的三维方向, 有三种不同的提升型模块——帧, 行, 和水平方向。每个部件每次读入提升型变换所需的一个单位数据。对于帧, 行, 和水平提升型模块, 一个单位数据代表的分别是一帧, 一行, 和一个像素。在 9-7 双正交提升型小波中, 只需缓存 6 个单位数据就可以进行一次运算。如图 2-4 圈中所示, 缓存的 6 的单位中 3 个处在提升型运算的中间状态, 其他三个是原始数据或已经计算完毕的小波系数。之后每输入两个单位, 就进行一次 4 层的提升型操作并产生一个低通 (low pass) 和一个高通

(high pass) 系数。因此, 对于任意长度的三维数据, 帧, 行和水平方向的所有提升型运算只需略多于 6 帧的内存即可完成计算。图 2-5 给出了带有帧/行缓存的一级分解的三维提升型小波, 其中首先进行的是帧缓存, 其次是行缓存, 最后对分解得到的每一行小波系数直接进行水平方向的提升型变换得到三维分解的系数。这种帧/行缓存结构在实现上类似于图像小波变换中的行缓存结构^[28]的三维扩展。

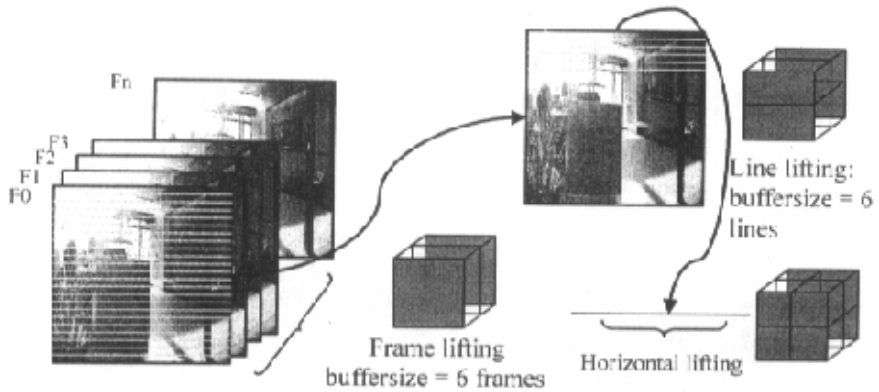


图 2-5 一级三维提升型小波 (Single scale 3D lifting.)

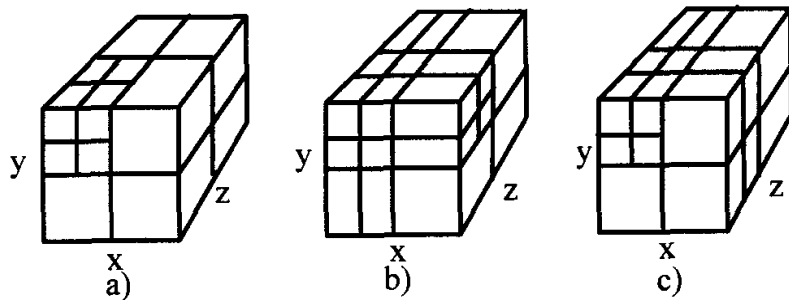


图 2-6 多级三维小波包分解 a) 2 级三维 mallat 分解 b) 2 级 x 分解+2 级(y,z) mallat 分解 c) 2 级 z 分解+2 级(x,y) mallat 分解 (Multiple-level 3D wavelet packet a) two-level mallat decomposition in all directions, b) two-level x decomposition + two-level (y,z) mallat decomposition, c) two-level z decomposition + two-level (x,y) mallat decomposition.)

同心拼图的小波压缩中需要用到多级小波包 (wavelet packet) 分解结构, 对于已产生的小波子带 (subband), 决定是否进一步的分解。图 2-6 a) 举出了 2 级三维 mallat 分解的小波包结构。或者我们也可以先对一个轴做多级小波分解后, 再对其他两个轴所

在的平面分解，如图 2-6 b)和 c)。在本章实验结果部分，我们测试了几种不同的小波包结构，并给出了针对同心拼图的经验最优结构。选定的小波包分解方式采用二叉树的遍历形式存储在发送的码流中以保证解码端的小波重构。

2.2.3 嵌入式位平面块编码 (Block-based Embedded Bit-plane Coding)

虽然前期小波变换是在整个同心拼图数据上进行的，在熵编码部分我们将已分解的小波子带分成大小相等的三维系数块，并对每个块进行独立的嵌入式熵编码，最终采用率失真优化 (rate-distortion optimization) 的方法将每个块的码流进行组装。这种分块编码和码流组装的方法具有以下优点：

1. 充分利用局部统计特征

由于统计特征在整个数据中很难达到均匀一致，通过对分块的独立编码，局部统计特性可以被发掘并用于提高编码效率。整个数据中统计特征的变化则可以通过编码之后的码流组装来描述，使之达到率失真优化的码率分配。

2. 便于解码端的码流随机访问

由于压缩码流由块的码流组装而成，每个块的信息能够单独访问，因而可以很方便的针对当前需要浏览的场景进行选择性的码流解构和解码。

3. 降低编码的内存需求

块状的结构避免了在编码时将整个同心拼图数据进行缓存。对于帧方向长度为 K 的分块方式，每个子带只需缓存 K 帧系数。每当缓存内的系数达到 K 帧，即可进行分块，量化和熵编码的操作，得到的压缩码流被缓存下来用于码流装配。比较系数所占缓存的大小，熵编码后占用的内存会大大减少。

每个块内的系数首先进行量化。我们采用一个简单的分级量化器，量化步长为 Q ，同时在零点附近有 $2Q$ 的置零区 (dead zone)。由于熵编码和码流装配的过程是嵌入式的，码率的控制是由码率组装过程来决定的，为了不在之前引入太多的失真，量化过程只是将浮点型的系数转换成整数形式以进行熵编码，即简单的取 $Q=1.0$ 。

分块系数内熵编码器的选择很灵活。本章采用嵌入式位平面的编码方法使压缩的码流易于截取。令 C 代表当前考虑的系数块， $b(i,p)$ 表示系数 i 的二进制表示中第 p 位有效位 (significant bit)。 B_p 表示第 p 个位平面 (bit plane)，由所有系数的第 p 位有效位的集合。令 P 代表位平面的总数，即对 C 中所有的系数 x ，有 $|x| < 2^P$ ，同时至少有一

个系数 x 满足 $|x| > 2^P - 1$ 。嵌入式位平面编码器按照位平面从最高有效位 $P-1$ 到 0 降序的进行编码, 因此在截取压缩码流的时候, 先保证了所有系数的最高有效位平面的优先传输, 保证了码流的解码质量。对每个系数 x_i , 如果在某个位平面 k 之前的所有位都是 0 , 即对所有的 $p > k$, $b(i, p) = 0$, 那么这个系数在第 k 平面编码时称为无效 (insignificant) 系数, 反之则称为有效 (significant) 系数。编码每个位平面时, 当前的无效系数以有效性检测的模式编码, 而有效系数则以细化 (refinement) 模式编码。在码流截取之前还是无效状态的系数值即为零。因此, 有效性检测过程的目的在于高效的表示那些在当前位平面从无效状态转为有效状态的系数, 也就是满足 $b(i, k) = 1$ 并且 $b(i, p) = 0$ ($p > k$) 的系数。嵌入式位平面编码器根据以下几个步骤对每个系数块进行编码:

步骤 1. 初始化: 记录总共的位平面数 P 并将当前状态设置在最高位。

步骤 2. 有效性检测 (significance identification) 过程: 检测并编码无效系数在位平面 p 中的对应位。

步骤 3. 细化 (refinement) 过程: 编码已经有效的系数在位平面 p 中的对应位。

步骤 4. 记录率失真 (rate distortion) 曲线: 在每个编码过程之后都同时记录当前编码的码率 R 和对应的失真 D 。

步骤 5. 对位平面 $p-1$ 重复步骤 2-4, 直至达到当前块编码的目标码率或失真。

步骤 4 中的码率 R 可以根据当前编码码流的长度得到, 失真 D 通过计算原始系数和重构系数的差得到, 具体实现中我们采用了文章[28]中给出的查找表直接查找对应的失真值, 节省了计算的时间。也可以利用文章[37]中提出的率失真优化嵌入式编码器 (rate-distortion optimized embedded coder, RDE) 来估计失真 D , 可以比直接计算达到更快的速度和更少的开销。记录下来的率失真曲线将用于后面的码流组装。

具体的位平面块编码采用了三种不同的熵编码器, 分别折衷了不同的复杂度和编码效率。不同编码器的主要区别在于对无效系数位的编码, 下面介绍具体的细节:

1) 树编码器 (Tree Coder)

在树编码中, 无效系数被组成八叉树的结构。首先, 整个块被当作一个无效状态的树根节点。在编码器从最高有效位平面向低扫描的过程中, 八叉树的节点逐渐的被分叉并最终确定有效系数的位置。基本的思路类似于三维 SPIHT 算法^[71], 但是集合的拆分是基于一个小波子带中的三维块, 而不是多级的小波子带。选择这种设计是为了数据访

问的灵活性。树编码器包含 4 个列表：无效集合列表 (LIS)，候选集合列表 (LCS)，无效系数列表 (LIP) 和有效系数列表 (LSP)。其中 LIS 中的元素是内部所有系数都为无效的集合；LCS 的元素是内部所有系数都无效，但是有至少一个系数在当前平面变为有效的集合；LIP 的元素是单个的无效点；LSP 的元素是单个的有效点。四个列表中的元素按照图 2-7 的状态转化图进行移动。状态的转换过程中：LCS 中的元素可以被拆分并根据拆分后的状态的移入 LIS, LSP, LIP 列表或者得到新的 LCS 中的元素；但是 LIS 中的元素只能移入 LCS；LIP 元素只能移入 LSP；移入 LIP 的元素将不再被移往其他的列表。虽然在概念上树编码式按照位平面顺序编码系数块，但操作都是基于这四个列表的。因此树编码器的计算复杂度在三种编码器中最低。

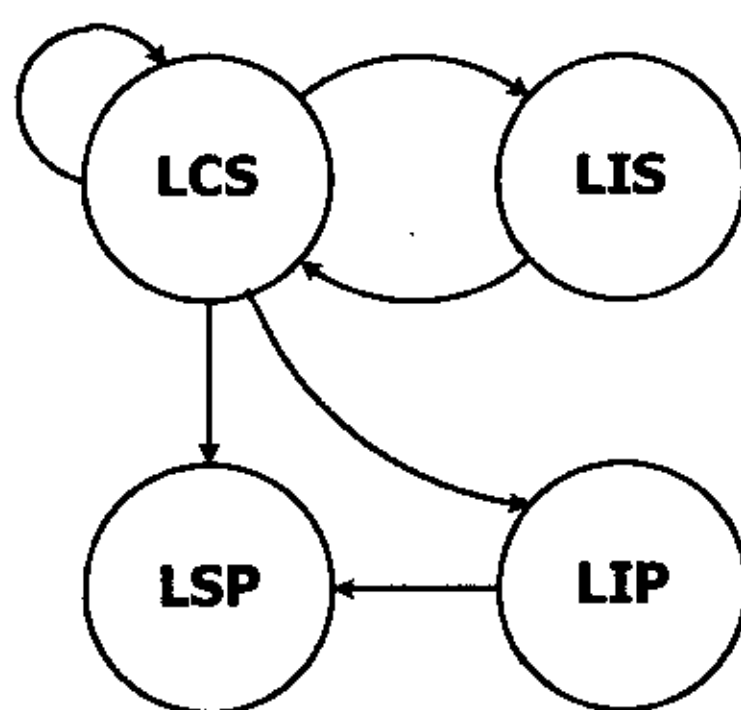


图 2-7 树编码器状态转换图 (State transition of the tree coder)

在每个位平面的编码中，首先按列表次序检查 LIS 中的集合。如果某个集合中至少有一个系数在当前平面变成有效，编码一个‘1’到码流中，并将此集合移入 LCS；否则编码一个‘0’，不移动该集合。接下来检查 LCS 中的集合。对于 LCS 中的集合 c ，由于在当前平面其内部必定至少有一个的有效系数，集合 c 被沿三维方向拆分成 8 个子集 c_1, \dots, c_8 。如果 c 的大小为 $2 \times 2 \times 2$ ，则产生 8 个子点。每个子集/点 c_i 被赋予一个状态位 s_i ，如果子集/点是有效的， $s_i = 1$ ，否则 $s_i = 0$ 。8 位的字符串 $s = \{s_i, i = 1, \dots, 8\}$ 采用 Huffman 编码。 $s_i = 1$ 的子集/点 c_i 被转入 LCS/LSP， $s_i = 0$ 的子集/点 c_i 被转入 LIS/LIP。如果子点 c_i 被转入 LSP，其符号也随即被编码。当 LCS 中所有的子集都处理完毕后，检查当前位平面编码开始时的 LIP，如果其中某些点的系数在当前位平面变为有效，编码一个‘1’到码流中，随即编码对应的符号，并将该点转入 LSP。否则对每个点编码一个‘0’。最后，细化开始时处在 LSP 中的所有系数在当前位平面对应的位值。

2) 自适应二进制 Golomb-Rice (GR)编码器

此编码器采用一种游程编码 (run-length coding) 方法, 通过自适应二进制 Golomb-Rice (adaptive binary Golomb-Rice, ABGR) 编码来标识位平面中无效系数形成的连续的零。ABGR 是[49]中方法的三维扩展。每个位平面内的系数位被分成三个类别: 至少有一个相邻点是有效系数的无效系数, 又称为是预测为有效 (predictive significant) 的系数; 26 个相邻点皆为无效的无效系数, 又称为是预测为无效 (predictive insignificant) 的系数, 以及细化位 (refine bit)。每个位平面的编码中, 我们对系数块进行三次扫描: 首先编码可预测为有效的位, 然后是可预测为无效的位, 最后是细化位。在这样分类下, 具有类似统计特性的系数位可以被聚集在一起, 更利于压缩。编码的顺序是按照经验的不同类别的率失真特性得出的。在预测为有效和预测为无效系数集合的扫描中, 从当前有效点到下一个变成有效状态的点之间的游程长度由 ABGR 编码器来编码。如果游程长度大于或等于 2^m , 则编码一个 '0' 来表示 2^m 长度的零的游程, 剩下的部分游程再重复处理。否则, 将编码一个 '1' 和 m 个比特来二进制表示游程的长度。接下来编码有效系数的符号。ABGR 原理上是对长度从 $(n-1)2^m$ 到 $n \cdot 2^m$ 之间的游程分配 $n+m$ 个比特的 Huffman 编码, 此处 m 是由[96]中的状态转换表决定的自适应的控制参数。对预测值有效和预测值无效的类, 分别采用两个具有各自控制参数 m 的独立的 ABGR 编码器进行编码。细化位不经过熵编码直接写入码流。ABGR 编码的速度也很快, 但因为每个位平面都要将所有系数扫描一遍, 所以稍慢于树编码。

3) 基于上下文的算术编码器 (Context-based Arithmetic Coder)

JPEG2000^[28]中的基于上下文的算术编码器被直接扩展作三维的应用。与 ABGR 编码器类似, 三维算术编码器也是按位平面顺序对每个平面分三个集合依次扫描: 预测为有效系数集合, 预测为无效系数集合和细化位, 但是在对预测为有效集合和预测为无效集合的扫描过程中采用的是基于上下文的算术编码器, 令具有相同上下文的系数按照分别的自适应算术编码器进行概率统计和编码。上下文的分类是根据周围的 26 个临近系数的状态决定的。我们对上下文的设计是借鉴同事的工作[82]得出的, 详细的描述在附录 I 中给出。基于上下文的算术编码器在三种编码器中具有最好的压缩性能, 但由于对上下文的计算相对来说较为费时, 因此计算复杂度也是最高的。

2.2.4 率失真优化的码流组装 (R-D Optimized Bitstream Assembler)

在所有的系数块都完成嵌入式编码后,采用码流组装根据最小化全局失真的原则对各个块的码流进行码率分配 (bit allocation)。因为色度分量 U 和 V 在小波变换前已经进行了系数为 2x2 的二维采样, 对采样分量的小波系数进行的失真度量是和亮度分量 Y 相同的。值得注意的是,如果对亮度和色度分量采用基于视觉模型的不同权重可以达到更好视觉效果。

前面的熵编码过程中已经计算并保存了每个块的率失真曲线。为了使先传输的码字对信息失真度的降低最快^[37], 需要保证率失真曲线的凸性 (即率失真斜率单调递减), 因此对每个块的率失真曲线采取凸包络 (convex hull) 的处理^[28]。过程如图 2-8, 其中实线表示原有的分段率失真曲线, 对于关系非凸的两段曲线, 去掉连接点而合并成一段 (如虚线所示)。最后得到完全凸性的率失真分段曲线。

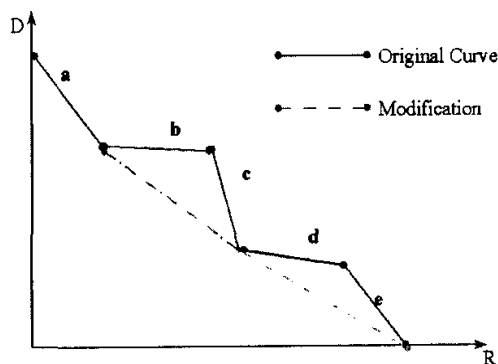


图 2-8 率失真曲线的凸包络 (Convex hull of the R-D curve)

每个块的失真还要根据所处的小波子带相应的乘上一个能量权重以利用小波域的失真估计图像空域的失真:

$$D_{Weighted} = w \cdot D_{Original} \quad (2-3)$$

此处子带权重 w 表示提升型小波合成滤波器带来的能量增益。对于双正交 9-7 滤波器, 低通滤波器的增益为 $w_L = 1.299$, 高通为 $w_H = 0.787$ 。对于在三维方向上共经过 n 个低通和 m 个高通提升型滤波的小波子带, 能量增益约为各层增益的乘积 $w = w_L^n w_H^m$, 无论提升型滤波的方向是水平, 竖直或者帧方向。

率失真理论表明,当所有的块率失真特征相同的时候,编码效率达到最优。因此码流组装的过程要求对所有块的码流截取应该位于同样的率失真斜率点上,同时使总比特数满足给定的码率。算法分为两部分:

步骤 1. 找出每个块的码流截取点

在给定的率失真斜率处截取得到的块的传输码率为斜率大于给定斜率的那部分码流的大小。实际计算中,用斜率为 λ 的斜线与率失真曲线相切的得到的切点处斜率即为 λ ,并且对应的码率就是截断得到的传输码率。

步骤 2. 找出最优的公共率失真斜率

最优的公共率失真斜率是利用区间二分(bi-section)方法查找的。我们首先设定两个率失真斜率门限值 λ_{\min} 和 λ_{\max} ,确保目标斜率处在区间 $(\lambda_{\min}, \lambda_{\max})$ 之间。

计算当前新的斜率门限值 $(\lambda_{\min} + \lambda_{\max})/2$,并根据该斜率得出对应的码率。根据得到的码率大于或者小于目标码率,相应的将门限区间的下限或者上限更新为当前的门限值。此过程不断迭代,直到所求的码率与目标码率在一定范围内足够接近,查找过程停止。由于在逼近目标码率的过程中的运算都是基于已有的率失真曲线,无需对系数块进行重新编码,整个查找的速度是非常快的。

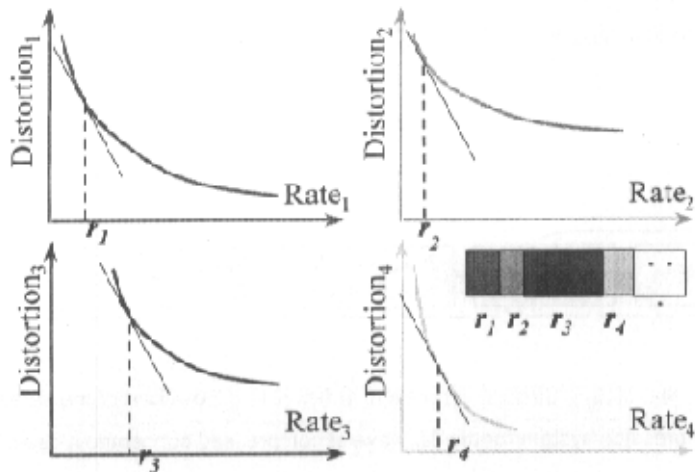


图 2-9 码流组装示意图 (Illustration of the bitstream assembling)

图 2-9 给出了码流组装过程的示意图, 图中显示了四个系数块的联合率失真斜率查找。嵌入式编码过程之后得到每个块的附加权重的率失真曲线。对所有的块的率失真曲线, 找出联合最优的率失真斜率点, 也就是同样斜率的斜线和曲线的切点, 并将码流在该点截断, 然后将所有块中在同一斜率切点截断得到的码流片段组合成最终的压缩码流。最优斜率即能够得出最接近目标码率的压缩码率的公共斜率。每个系数块的截断码率的长度也作为头信息 (header) 记录在压缩码流中, 并可以作为对压缩码流随机解码时候的索引。

2.3 用于同心拼图场景绘制的渐进逆向小波合成算法 PIWS

由于同心拼图数据量庞大, 并且对每个场景的绘制只需部分数据, 支持绘制的解压缩系统需要具有对压缩码流进行选择性的局部解码的功能。为了实现对同心拼图压缩码流的实时绘制, 本章提出了一种渐进逆向小波合成 (progressive inverse wavelet synthesis, PIWS) 方法来降低场景绘制所需的选择性解压缩的计算量。关于对整个图像数据进行快速小波分解与合成重构的算法已经有一些研究, 如^{[97][98]}, 另外还有一些工作通过分片化 (tiling) 来处理对医学图像中的三维体数据 (volume data) 的访问^{[99][100][101][102]}。为了支持灵活的数据访问, 也可以采用支集较短的滤波器例如 Haar 来降低计算量。然而, Haar 滤波器或者分片化的方法都会降低编码端的压缩质量, 甚至产生块效应, 影响视觉效果。本节提出的 PIWS 可以不受滤波器的限制, 支持对任意区域的数据的访问重构, 同时节省了小波合成的运算量, 有效支持了实时绘制的解压缩重构系统。

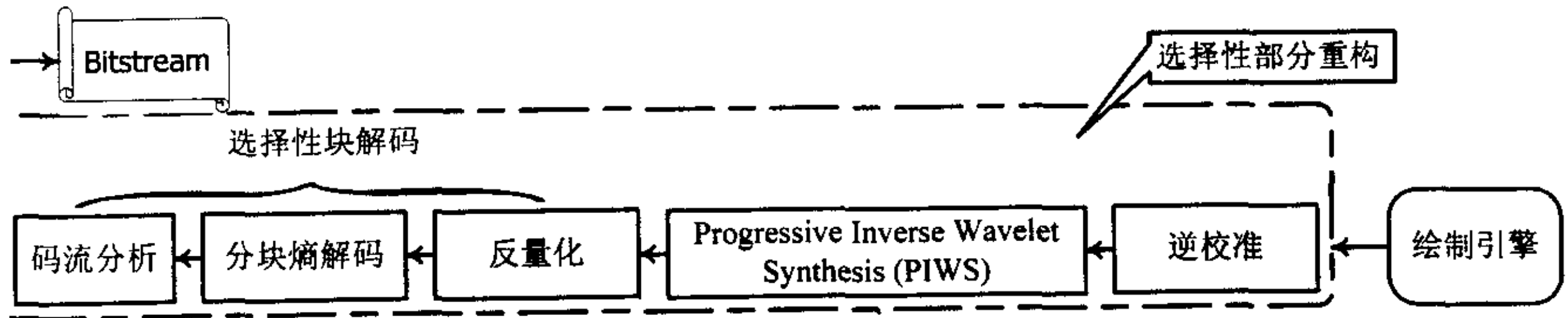


图 2-10 同心拼图压缩码流的选择性解压缩系统流程 (Flowchart of the selective decompression system for the 3D wavelet compressed concentric mosaic)

2.3.1 三维选择性解码系统

在绘制过程中，同心拼图压缩码流并不是一次性的全部解出，而是采用选择性的三维小波解压缩方法，每次只访问和解压缩与当前画面相关的部分码流。解压缩系统由绘制引擎驱动，如图 2-10 所示。绘制引擎向选择性解压缩系统申请解码与当前画面相关的狭缝图像，经过反向校准过程的重新对应，选择性解压缩系统将需求发送到选择性小波合成引擎，由其确定相关的待合成的小波系数并进行小波合成。待合成的小波系数由选择性块解码引擎提供，其中包括对码流分析进行相关码流片断的定位，以及反向的分块熵解码和反量化过程。在整个解压缩的驱动过程中，只处理与当前绘制画面相关的狭缝图像，小波系数和码流片断。

在整个系统中，反向校准的过程很容易实现，前面设计的分块编码结构为选择性块解码的实现提供了灵活性，因此选择性小波合成步骤成为解压缩过程中的瓶颈。解决方法的一种考虑是每次只解单根狭缝图像进行拼接，称为 SSD (simple slit decoder) 算法。虽然 SSD 方法很简单，但在选择性小波合成中，合成单根狭缝图像需要相邻狭缝图像的小波系数也参与运算，相邻狭缝图像的重构过程中有很多重复的计算，造成很大的解压缩负担。另一种方法是每次解码一定大小的狭缝图像块并缓存结果，这样可以一次解出多条狭缝图像，减少了平均的运算量，称为 BSD (block selective decoder) 算法。BSD 也并不是最优的，并且当一个块内只有一根需要浏览的狭缝图像时，整个块也必须全部解码，增大了运算量。本文设计的渐进逆向小波合成方法 PIWS，在提升型小波实现结构的渐进计算基础上，将提升型运算过程的中间结果按状态转换的步骤缓存起来，避免了重复计算，最大限度的减少了小波反变换的运算量。下面我们对 PIWS 方法作详细的介绍。

2.3.2 一维渐进逆向小波合成PIWS

首先从简单的一维情况来介绍 PIWS 方法。如图 2-11 所示，由于采用了提升型的小波变换结构，数据的重构是逐层计算出来的，并且每个单元的运算为原位操作，即存储单元中的数据先是小波系数，经一次提升操作到达中间状态，再经一次运算提升到完全重构状态。因此，令 m_0, m_1, \dots, m_{n-1} 表示相应位置的存储单元，如果把这些单元按表 2-1 将已达到的状态和相应数值保存下来，PIWS 算法可以用状态机的形式来表示如表 2-2。对于奇数位置和偶数位置的单元，状态转换有所区别。

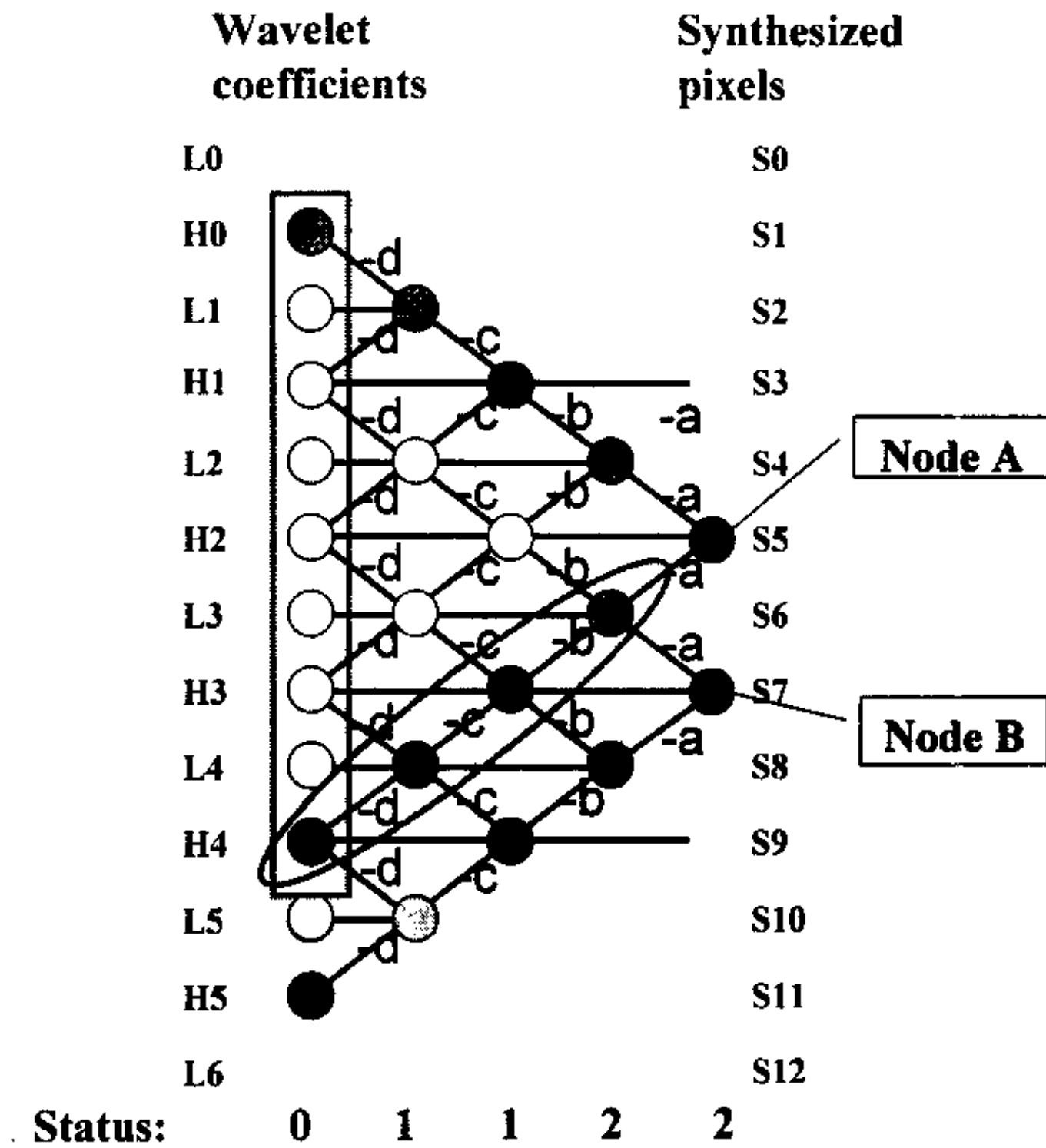


图 2-11 渐进逆向小波合成原理图 (Progressive inverse wavelet synthesis)

表 2-1 一维提升型小波变换状态表 (States for 1D progressive inverse wavelet synthesis)

状态	表示的信息
0	小波系数
1	已进行了一次提升操作
2	已重构的信号

表 2-2 一维逆向小波合成状态转换表 (State transition operations for 1D inverse lifting)

当前状态					后继状态	
i	mi	mi-1	mi+1		Mi	
偶数点	0	0	0	➡	1	
奇数点	0	1	1		1	
偶数点	1	1	1		2	
奇数点	1	2	2		2	

在将某个单元 m_i 解码重构到状态2的过程中，状态机的操作过程是要求临近两个单元 m_{i-1}, m_{i+1} 的状态提升成状态表中所需的状态1，并进一步要求下一级的临近单元达到状态0。如图 2-11，在解码A节点的同时，椭圆框对应位置的单元内的系数也参与了计算，并达到了椭圆框所标示出的相应的状态。在解码B节点的过程中（即将S7单元的状态提升到状态2），当椭圆框内的单元被要求提升某个所需状态时，由于在解码A的过程中这些单元已经达到了所要求的状态，因此可以直接使用当前的数值，从而避免了重复的计算。随着数据不断的被解码，每个存储单元都是按照0, 1, 2的状态逐渐提升。因此，对于随机访问的节点，只需要与该点状态提升所需的临近节点参与提升计算，因此大大降低了计算的负担。

从图 2-11 中可以看出，如果不讲运算结果加以存储，每次单独解码奇数位置上的一个节点，如 S5(Node A)，需相邻 9 个节点参与计算，运算量为 10 乘，20 加；对偶数位置的 S4 则需相邻 7 个节点参与计算，运算量为 6 乘，12 加；平均需要 8 乘，16 加。采用渐进逆向小波合成后，则平均解码一个节点所需的运算只有 2 乘，4 加，平均节省了 6 乘，12 加的运算。对于高维的情况，运算量的节省更有倍数的增长。

2.3.3 基于狭缝图像的同心拼图重构（二维多级渐进逆向小波合成解码）

在对三维小波压缩同心拼图数据的重构绘制中，由于每次解码的是单独的整根狭缝图像而不是单个像素点，一维的 PIWS 方法被扩展到二维的形式。令 m_{ij} 代表狭缝图像的存储单元。 m_{ij} 中的状态变化如表 2-3。

表 2-3 二维渐进逆向小波合成状态转换表 (States of the 2D progressive inverse wavelet synthesis)

No.	状态	信息
0	N	没有经过熵解码得到系数
1	H0	小波系数
2	H1	水平方向经过一级提升操作
3	H2=V0	水平方向经过二级提升操作，接竖直方向操作
4	V1	垂直方向经过一级提升操作
5	V2=S0	垂直方向经过二级提升操作，接整条狭缝图像的逆向合成
6	S _{end}	整条狭缝图像竖直方向的反变换合成

图 2-12 中表示了在同心拼图的局部场景访问中, 当某条狭缝图像被重构时, 其相邻单元到达的状态分布图。每个圆圈表示一根狭缝图像, H_i , V_i 分别表示狭缝图像达到的对应表 2-3 的状态, 空白的圆圈表示尚未进行熵解码。当 S_{end} 附近的某条狭缝图像被访问时, 相应的狭缝图像能够在解码 S_{end} 时的中间结果基础上直接进行计算, 节省了解码过程的运算量。

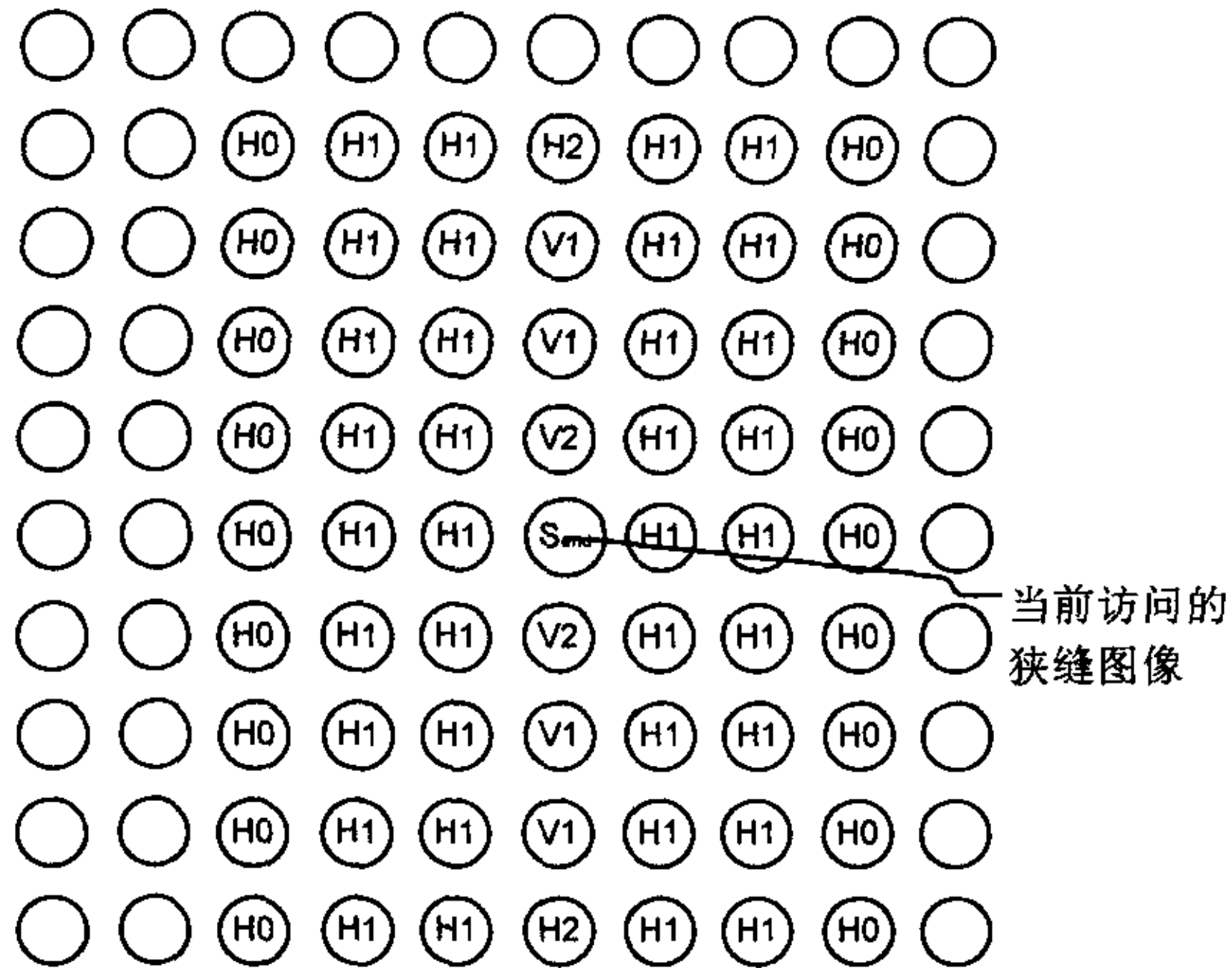


图 2-12 解码一根狭缝图像时存储单元的相应状态值 (Cache states when a single slit is accessed)

在同心拼图的压缩中, 由于采用多级分解的三维小波包形式, PIWS 方法也相应的扩展为多级的结构, 图 2-13 所示为三级 Mallat 分解小波变换对应的 PIWS 结构。在一级 PIWS 解码中, 若所需要达到 H_0 状态的节点所在的子带有被继续分解时, H_0 状态对应了第二级的 PIWS 引擎中的 S_{end} 状态, 因此第二级 PIWS 引擎被调用; 如果没有被继续分解, 则 H_0 状态对应了直接进行选择熵解码的结果, 即通过从码流中找到与当前所需系数相关的部分, 进行分块解码, 来得到 H_0 状态的小波系数。

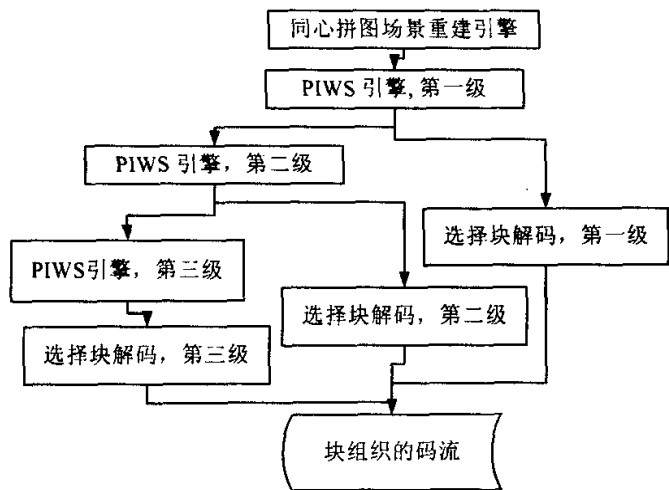


图 2-13 多级渐进小波合成算法结构 (Multiscale progressive inverse wavelet synthesis)

2.4 实验结果与分析

2.4.1 压缩性能

首先我们研究了三维小波同心拼图压缩性能。作为参考，针对同心拼图数据的三维小波包结构以及不同熵编码方法计算复杂程度对压缩性能的权衡也将在本节一一讨论。测试数据为同心拼图场景 Lobby (图 2-14) 和 Kids (图 2-15)。场景 Lobby 拍摄了 1350 帧 320x240 的图片。场景 Kids 拍摄了 1462 帧 352x288 的图片。两个场景都是在 YUV 颜色空间 (color space) 上处理，并且 U 和 V 分量进行了二维的 2x2 采样。压缩性能由重构场景与原始场景之间的峰值信噪比 (peak signal-to-noise ratio, PSNR) 来衡量：

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{N \cdot W \cdot H} \sum [f(w, n, h) - f'(w, n, h)]^2}, \quad (2-4)$$

其中 $f(w, n, h)$ 和 $f'(w, n, h)$ 分别是原始同心拼图数据和压缩重构数据。实验给出了 Y, U 和 V 分量分别的 PSNR 值，其中 Y 分量的 PSNR 值对压缩性能的衡量最为关键，并且压缩码流中约有 90% 都是 Y 分量的码流。



图 2-14 同心拼图场景：大厅（Concentric mosaic scene: Lobby）



图 2-15 同心拼图场景：儿童（Concentric mosaic scene: Kids）

首先研究不同三维小波包分解方式的压缩性能。令 x, y 和 z 轴分别代表帧，竖直，和水平方向。五种小波包分解结构评估如下：

- 结构 A: 三维的 4 级 mallat 分解，
- 结构 B: y 轴 4 级分解 + (x, z) 平面 3 级 mallat 分解，
- 结构 C: y 轴 4 级分解 + (x, z) 平面 4 级 mallat 分解，
- 结构 D: z 轴 5 级分解 + (x, y) 平面 4 级 mallat 分解，
- 结构 E: x 轴 5 级分解 + (y, z) 平面 4 级 mallat 分解。

实验中采用分块大小为 $16 \times 16 \times 16$ 的树编码器，测试数据为场景 Kids，压缩至 0.6 比特/像素 (bit per pixel, bpp)。表 2-4 显示了测试结果。可以看出全 mallat 分解（结构 A）并没有得到最好的压缩性能。通过结构 C 的进一步分解 y 轴（竖直方向），压缩效率提高了 0.5dB。4 级以上的小波包分解（结构 D 和 E）没有进一步的提高。并且先分解 z 轴（帧方向）或者 x 轴（水平方向）的结构 D 和 E 都比结构 C 的性能要差。因此，在接下来的实验中，我们统一采用结构 C，即在 4 级竖直方向分解后做帧方向和水平方向构成的平面上的 4 级 mallat 分解。

表 2-4 针对同心拼图的不同小波包分解结构性能比较（Concentric mosaic compression with different wavelet packet decomposition structures）

小波包结构	A	B	C	D	E
PSNR: Y (dB)	30.6	31.0	31.1	30.9	30.3
PSNR: U (dB)	36.4	37.1	37.3	38.4	37.6
PSNR: V (dB)	37.2	37.9	38.0	38.9	38.1

第二部分实验比较了本章提出的三维小波同心拼图编码器与 MPEG-2 压缩编码的性能。作为对比的标准编码器 MPEG-2 是采用[www.mpeg.org]处下载的 MPEG-2 编码器将整个同心拼图当作一个视频序列来进行压缩。三维小波编码器中的同心拼图首先经过校准,小波分解,之后拆分成 $16 \times 16 \times 16$ 进行独立的分块位平面熵编码,最后根据率失真优化原理组装成压缩码流。对文中给出的三种熵编码器:树编码器,自适应 Golomb-Rice 编码器,和基于上下文的自适应算术编码器都进行了比较。在 MPEG-2 中,其中第一帧作为 I 帧 (intra frame),其余的都定为 P 帧。

表 2-5 三维同心拼图场景压缩结果 (Results of the 3D wavelet concentric mosaic compression)

测试数据 PSNR	LOBBY (0.4bpp)	LOBBY (0.2bpp)	KIDS (0.4bpp)	KIDS (0.6bpp)
MPEG-2 (dB)	Y: 34.8 U: 39.9 V: 39.1	Y: 32.2 U: 38.7 V: 38.1	Y: 30.1 U: 36.6 V: 36.7	Y: 31.9 U: 38.0 V: 38.1
3D wavelet + tree coder (dB)	Y: 34.5 U: 41.6 V: 41.2	Y: 31.4 U: 40.1 V: 39.7	Y: 29.0 U: 35.8 V: 36.6	Y: 31.1 U: 37.3 V: 38.0
3D wavelet + Golomb-Rice coder (dB)	Y: 34.4 U: 41.5 V: 41.2	Y: 31.3 U: 40.0 V: 39.7	Y: 29.0 U: 35.7 V: 36.5	Y: 31.0 U: 37.0 V: 37.9
3D wavelet + arithmetic coder (dB)	Y: 35.0 U: 41.9 V: 41.5	Y: 31.9 U: 40.3 V: 39.9	Y: 29.4 U: 36.5 V: 37.2	Y: 31.5 U: 38.0 V: 38.6

场景 Lobby 压缩的目标码率分别为 0.2 bpp 和 0.4bpp。由于场景 Kids 的细节较多,压缩码率分别设在 0.4bpp 和 0.6bpp。表 2-5 给出了比较结果。可以看出树编码器的性能和 Golomb-Rice 编码器十分接近。比较起来,树编码器的计算复杂度要低于 Golomb-Rice,但是内存需求则略微高出,尤其是对高码率端。不过,由于分块的大小是 $16 \times 16 \times 16$,所以内存的消耗并不重要。因此二者相比我们倾向于树编码器。最复杂的算术编码器的压缩性能最优,最多高出了 0.5dB。采用算术编码的三维小波编码器低于 MPEG-2 平均 0.3dB,在压缩性能的角度处于基本相当的水平,而且可以通过更为精细的小波包选取等方法来进一步提高。但是 MPEG-2 的压缩码流不能够支持随机访问,因此并不适合支持同心拼图的绘制。另外,由于 MPEG-2 压缩后重构的帧的质量是随机波动的,因此虽然总的压缩性能比较满意,但是在绘制同心拼图场景的视觉效果上不能保证。比较而言,采用嵌入式块编码的三维小波编码器则具有分辨率和质量可伸缩的能力,

同时可以提供灵活的随机访问等功能，因此在网络传输和实时浏览等方面有很好的应用前景。

2.4.2 PIWS解码重构速度

下面从同心拼图浏览中的三种位置移动方式 RT（旋转）、FW（前进后退）、ST（平移）出发分别比较了 PIWS 算法和其他小波解码合成算法 SSD 与 BSD 对同心拼图数据局部解码的运算速度，以及 PIWS 三维小波同心拼图解码系统与另两种压缩方法:SVQ^[91]和基于 DCT 与运动补偿的参考块编码(RBC)^[95]解码速度。

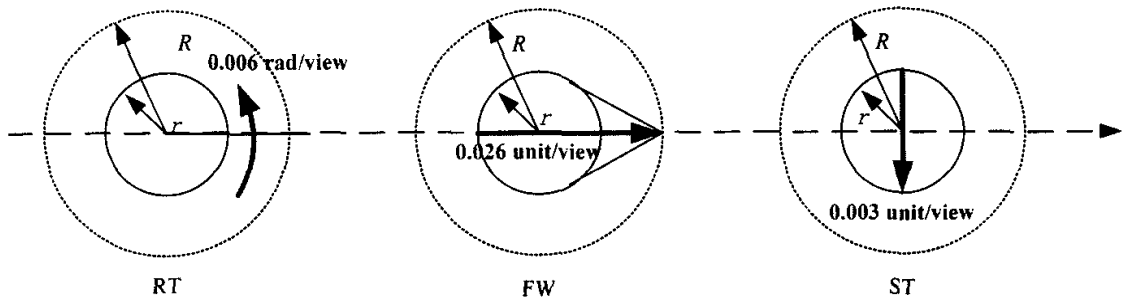


图 2-16 同心拼图场景中的三种浏览方式 (Three kinds of movement in concentric mosaics)

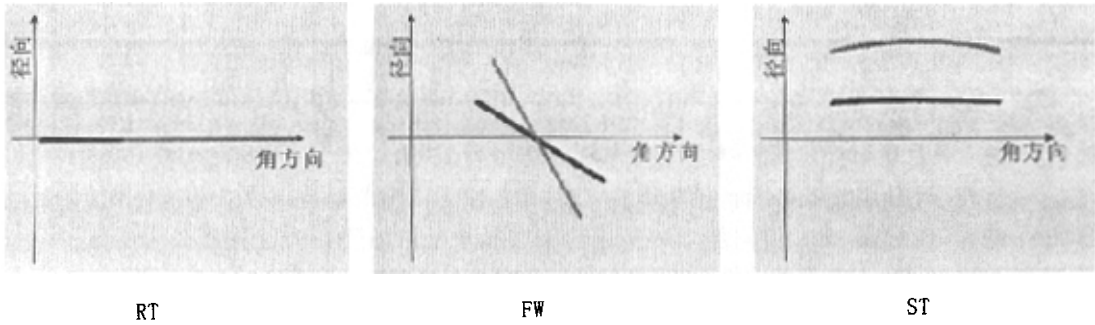


图 2-17 三种浏览方式中需解码的狭缝图像数据在整体拼图序列中对应的位置，每图所示分别为两个场景所对应的数据 (Access slits associated with three motion modes in the concentric mosaics. The slits of two views are drawn)

图 2-18 表示在 ST 模式下本文提出的 PIWS 算法和另两种小波反变换合成算法的比较: SSD 算法每次重新解码一条狭缝图像; BSD 算法每次解码一定大小的狭缝图像块。从图 2-18 中看出, SSD 算法的额外开销远远大于另两种方法, 因此解码速度也最慢, 无法支持实时绘制的要求。由于额外开销的比例相对较小, BSD 算法的速度有了显著的提高, 但性能则不够平稳, 可以看出当浏览所需数据稀疏的分布在多个块内时, 由于解码的块数增多, 运算时间会有大幅度的跳变。从实验结果来看, PIWS 解码方法(与水平轴接近重合的曲线)的计算量开销最小, 比 SSD 和 BSD 方法的重构速度快, 而且性能平稳, 很好的达到了支持用户对三维小波压缩同心拼图场景实时浏览的要求。

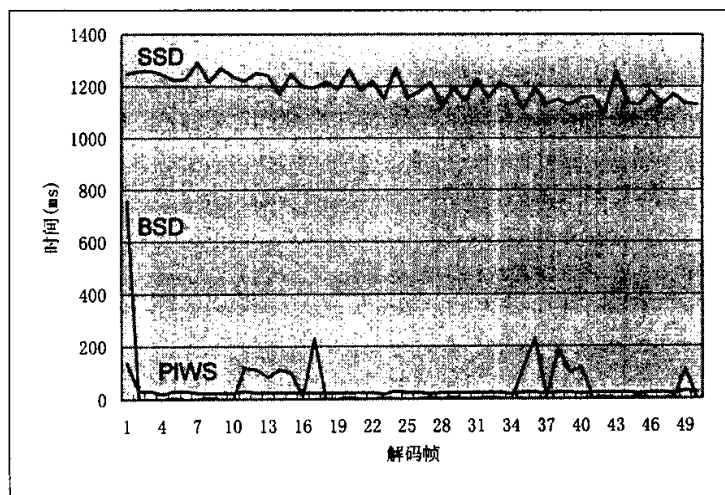


图 2-18 PIWS 算法和另两种合成算法对平移模式的解码速度比较 (Timing curves for the translation motion of concentric mosaics)

第二部分实验比较了 PIWS 三维小波同心拼图解码系统与另两种非小波压缩方法:SVQ^[91]和基于 DCT 与运动补偿的参考块编码(RBC)^[95]解码速度。以往的小波编解码运算复杂度远高出 VQ 和 DCT 变换, 在同心拼图的解码中不能很好的完成实时重构的要求。表 2-6 中的采样绘制表示绘制时采用最邻近的采样狭缝图像进行绘制; 插值绘制表示采用邻近的 4 根采样狭缝图像的插值进行绘制。结果表明, 采用 PIWS 方法的小波解码的速度仅比 VQ 慢 26.8%, 比 RBC 慢 11.5%, 在需要解码的狭缝图像最多的 ST 模式中分别慢 55.4%和 36.8%。实验结果显示, 采用 PIWS 之后, 小波局部解码达到了与基于 VQ 和 DCT 方法的解码相当量级的速度, 出色的解决了高维小波压缩解码中的运

算量庞大的问题，使三维小波压缩系统在提供码流多分辨率以及可伸缩性等功能的同时，满足了同心拼图场景实时绘制的需求。

表 2-1 三种运动模式下的三种算法速度比较 (Overall rendering speed measured in frames per sec: VQ, RBC vs. PIWS)

		采样绘制 每秒解码帧数	插值绘制 每秒解码帧数
RT	SVQ	19.7	16.3
	RBC	16.8	13.9
	PIWS	17.6	14.3
FB	SVQ	19.0	15.8
	RBC	16.4	13.9
	PIWS	15.8	13.5
ST	SVQ	17.7	14.9
	RBC	12.5	10.9
	PIWS	7.9	7.3

2.1 本章小结

本章提出了一种基于三维小波的同心拼图压缩系统。结合小波变换和分块熵编码方法，该压缩系统能够很好的支持同心拼图场景的即时解码绘制 (JIT)。并且嵌入式的压缩码流可以进行多级分辨率和多级质量的解码，对同心拼图的网络传输等应用提供了方便。实验表明在提供多种灵活解码功能的同时，本章提出的三维小波压缩系统还达到了与 MPEG-2 编码方法近似的编码效率。

为了能够对压缩码流进行快速的随机解码，本章还提出了一种渐进逆向小波合成算法 PIWS。PIWS 的原理基于提升型小波变换结构的渐进式运算特点，通过开发一种混合型缓存单元结构，将提升型运算中的各结点单元内数据根据对应的状态进行更新和存储。由于各节点的状态能够被保存和索引，避免了重复运算，大大节省了三维小波压缩数据随机选择性解码的运算量。应用 PIWS 方法保证了对小波压缩同心拼图场景进行解码绘制的实时性需求。虽然 PIWS 方法是在基于同心拼图数据，采用三维小波和双正交 9-7 提升型小波算法压缩前提下设计的，但是可以被有效的扩展到对于其它种类的提升型滤波器（如 Le Gall 双正交 5-3 提升型），其它维度的小波变换，以及不同类别的小波压缩数据（如三维纹理）的实时选择性解码重构应用中。

第三章 基于运动补偿提升型小波的视频编码

从本章开始,我们讨论小波技术在视频编码中的应用。小波视频编码侧重于研究有效的将时轴运动信息和小波变换结构相结合的方法。对于运动信息的获取一般采用与现今基于 DCT 的视频编码标准 MPEG-4 以及 H.26L 等类似的块匹配 (block-matching) 方法。我们在第一章谈到过,由于残差图像 (residual image) 的统计特征与自然图像有很大差异,将小波直接运用在对运动补偿预测得到的残差图像上进行编码的方法没有取得满意的效果。我们的工作主要研究沿三维方向进行小波变换的压缩方法,也就是说:将时间轴当作具有特定统计特征的一维信号来采用小波进行去相关。三维小波压缩不但能够提供帧率、压缩质量等多种可伸缩性。并且由于不存在预测编码的循环结构,因此回避了预测编码中的误差漂移 (error drifting) 问题,使其非常适合于无线信道等易错信道中的传输应用。三维小波变换压缩的难点在于怎样加入运动信息,由于视频中的运动具有很大的不规则性,结合时轴运动的小波变换不具有图像那样规整的矩形结构,怎样在加入运动信息的同时保证时轴小波变换的可完全重构性是主要的研究方向。

本章提出了一种基于提升型结构的运动补偿提升型小波 (motion-compensated lifting, MCLIFT) 方法,巧妙的利用提升型结构的渐进性和原位运算特点,将运动补偿结合进时轴的小波变换过程。通过在时轴小波变换的提升型单元结构中采用双向的重叠块运动补偿,视频帧可以被很好的去相关,达到压缩目的。结合相应的熵编码和比特流装配方法,采用 MCLIFT 方法的编码器压缩的视频码流可以在帧率、解码质量上具有可伸缩性。实验结果表明本文提出的 MCLIFT 小波视频编码器比传统无运动补偿的三维小波编码器压缩性能提高了平均 0.9~1.3dB,同时比 MPEG-4 的编码性能平均高出 0.2-0.6dB。

3.1 系统框架

图 3-1 为采用本章提出的运动补偿提升型小波 (motion-compensated lifting) MCLIFT 的视频编码器流程。视频帧在时轴的 MCLIFT 变换过程之后,对每一个处理后的视频帧采用二维小波分解方法得到三维方向全分解的小波系数。结合第二章中的位平面熵编码,采用 MCLIFT 的三维小波的编码码流可以达到时轴帧率 (frame rate) 以及压缩质量的充分可伸缩性 (scalability)。

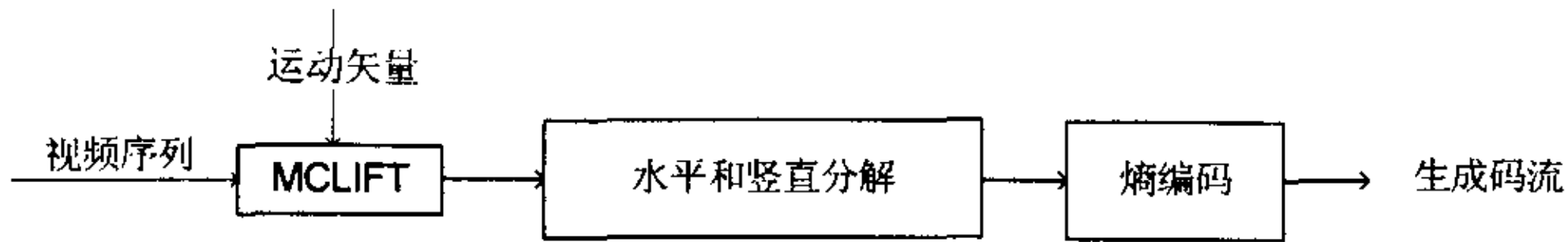


图 3-1 运动补偿提升型小波 MCLIFT 视频编码器流程 (Framework of the motion compensated lifting wavelet (MCLIFT) video coder)

3.2 运动补偿提升型小波变换

第一章中介绍过，提升型算法(lifting scheme)是一种能够有效节省存储空间和计算量的小波实现方法^[42]，并且所有的有限冲击响应(finite impulse response)小波滤波器都可以分解成相应的提升型步骤^[45]。本章我们利用提升型小波的分层结构和原位运算特点，将运动补偿信息加入到时轴滤波的提升型结构中，从而实现了时轴上的运动对齐。

考虑到视频压缩中对在线解码的时延和帧缓存空间的限制，在时轴方向采用 Le Gall 双正交 5-3 滤波器（表 II-2），在保证一定的去相关能力的同时，较短的滤波器长度对应了较少的延迟和内存消耗。图 3-1 所示为 Le Gall 双正交 5-3 小波一级正变换和反变换的提升型结构。图左的正变换结构表示，原始信号 x_0, x_1, \dots, x_6 从提升型结构的左端输入，右端的两列为分解输出的小波变换系数， H_i 代表高频系数， L_i 为低频系数。可以看出小波系数经过两层运算得到，并且每一层只有一半的信号参与运算而被提升到高一层的状态。

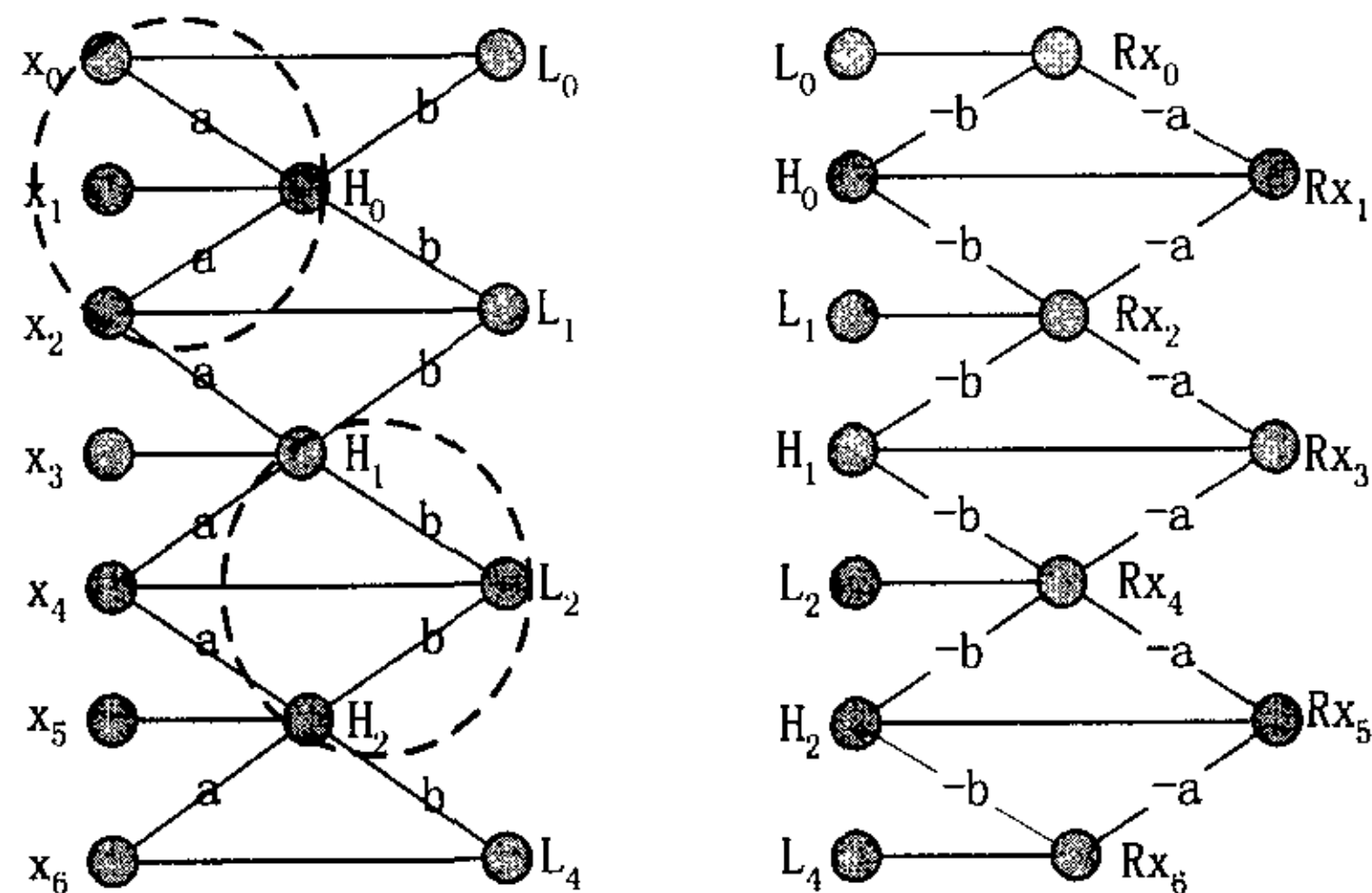


图 3-2 一级 5-3 提升性小波的正变换与反变换结构，圈内所示为基本提升型单元 (One-level forward and inverse 5-3 wavelet transforms via lifting and the elementary lifting units (circled in the forward structure))

双正交 5-3 滤波器的提升型过程可以用公式 (3-1) 描述:

$$\begin{cases} H_i = x_{2i+1} + a \times (x_{2i} + x_{2i+2}) \\ L_i = x_{2i} + b \times (H_{i-1} + H_i) \end{cases}, \text{ 其中 } a=-1/2, b=1/4; \quad (3-1)$$

归一化后的提升型实现和传统的卷积型 5-3 滤波器(3-2)是相同的:

$$\begin{cases} H_i = \frac{\sqrt{2}}{2} [x_{2i+1} - \frac{1}{2} \times (x_{2i} + x_{2i+2})] \\ L_i = \sqrt{2} \cdot [\frac{3}{4} x_{2i} + \frac{1}{4} \times (x_{2i-1} + x_{2i+1}) - \frac{1}{8} \times (x_{2i-2} + x_{2i+2})] \end{cases} \quad (3-2)$$

每一层提升型运算的基本单元如图中圈内所示, 其运算只涉及到三个节点。例如当 x_1 需要被提升, 只有相邻的 x_0 和 x_2 参与当前的单元运算, 计算后的值可以直接存储在 x_1 所在的单元中, 即原位运算。

在基于 MCLIFT 方法的视频编码器中, 时间轴输入的数据是视频帧。利用提升型小波原位运算的特点, MCLIFT 方法在时轴提升型小波的基本单元中加入运动补偿, 使得帧间相关性得到充分利用, 从而提高了时间轴上小波分析的效率。图 3-3 所示为相应于提升型基本单元设计的运动补偿提升型基本单元, 它在实现上类似于 MPEG 中的 B 帧 (B-frame)。图中 F_{2n+1} 为将通过提升型操作的当前帧, F_{2n} 和 F_{2n+2} 是参与提升型运算的与 F_{2n+1} 相邻的两帧, 可以看作是 F_{2n+1} 的参考帧。对当前帧 F_{2n+1} , 首先均匀的分成 16×16 大小的宏块 (macro-block, MB), 对于每个宏块, 采用 MPEG 中的块匹配 (block matching) 方法进行运动估计 (motion estimation) 分别在 F_{2n} 和 F_{2n+2} 中找到对应的匹配块, 并编码相关的运动矢量 $MV_{2n+1,i,j}$ 和 $MV_{2n+2,i,j}$ 。之后的 MCLIFT 单元的操作可以描述为: $MCLIFT(F_{2n+1}) = F_{2n+1} - a [MCP(F_{2n}) + MCP(F_{2n+2})]$, 其中 a 对应提升型运算中的常数, 对应于双正交 5-3 提升型滤波器的第一层 $a=1/2$; $MCP(X)$ 表示对 X 根据宏块矢量进行运动补偿得到的对 F_{2n+1} 的预测图像。

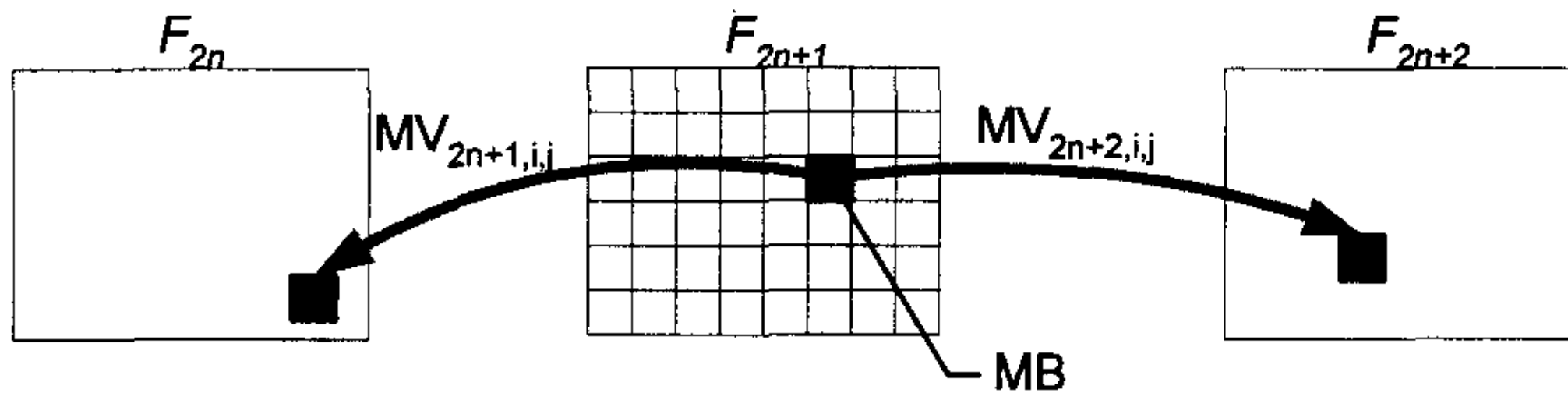


图 3-3 运动补偿提升型结构基本单元 (The elementary MCLIFT operation)

与以往的时轴提升型单元不同, MCLIFT 方法中提升型运算采用运动补偿后的图像取代了原始图像。基于提升型结构的逐层可逆性, 运动补偿后的参考帧在解码端可以得到重构, 保证了加入运动补偿后整个运动补偿提升型结构的可逆性。现有许多运动补偿技术, 例如半像素 (half-pixel) 以及 $1/4$ 像素 (fractional-pixel) 精度, 重叠块运动补偿 OBMC (overlapped-block motion compensation) 等都能够灵活的应用在 MCLIFT 方法的结构中, 从而显著提高时间轴小波变换的性能。本章的实验分析中采用了半像素精度运动矢量。由于帧内的小波变换都是在整个图像上进行, 传统的基于块的运动补偿得到的参考帧在块边界的位置上会造成明显的块效应, 不利于帧内小波变换, 采用重叠块运动补偿 OBMC 技术能够有效的去除块效应, 提高小波编码的效率。

在基于 MCLIFT 的编码器中, 时轴共进行了 3 级小波分解。图 3-4 所示为基于 5-3 滤波器的三级时轴提升型小波变换, 其中的每一个提升型单元都采用了图 3-3 中的运动补偿结构。由于在每个 MCLIFT 运动补偿单元结构中包含两组运动矢量 (motion vector) $\{MV_{2n-1,ij}\}$ and $\{MV_{2n,ij}\}$, 对于长度为 n 的信号, 图 3-2 一级 5-3 提升型变换共产生 $2n-2$ 组运动矢量。3 级变换运动矢量的个数将进一步增多, 大量的运动矢量所占的码率将影响整个编码的效率。为了减少运动矢量的数量, 在实际的编码器中, 我们采用了截断 5-3 提升型滤波器来代替原有的双正交 5-3 滤波器。在截断 5-3 滤波器中, 只保留了图 3-4 中的第一层提升型运算, 即产生高频系数的运算, 并将第二层的产生低频系数的运算截断。经过截断 5-3 提升型滤波后的高频系数与原有的 5-3 滤波器相同, 而低频系数则为原始图像信号。

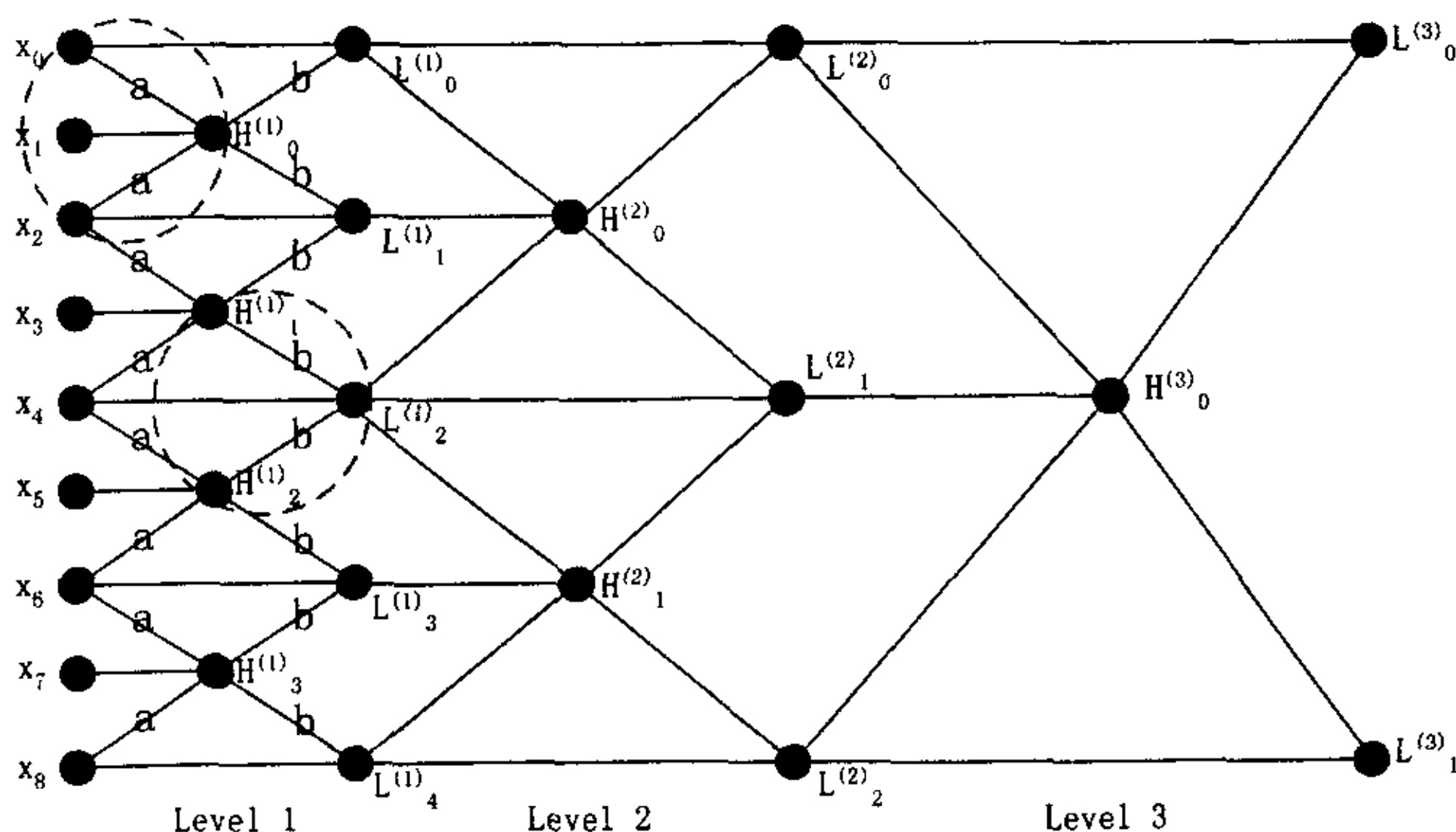


图 3-4 时轴 3 级提升型小波结构 (3-layer temporal lifting wavelet structure)

图 3-5 表示的是采用 5-3 截断提升型小波的 MCLIFT 编码器的 3 级分解帧结构。F0, F1, ..., F8 代表连续的 9 帧原始视频数据, 箭头方向表示运动估计时由被估计帧指向参考帧的方向。经过 3 级 5-3 截断提升型变换后, F0 和 F8 未被改变, 仍为原始值, 定义为 A 类; F4 以 F0 和 F8 (A 类) 为参考帧经过了一次提升型变换, 定义为 B 类; F2 和 F6 以相邻的 A 类和 B 类帧为参考进行了一次提升型变换, 定义为 C 类; 同理 F1, F3, F5 和 F7 被定义为 D 类, 以 A, B 和 C 类帧作参考帧。在变换的过程中, D, C, B 类帧分别在第一, 第二和第三级小波分解中得到。由于小波变换的嵌入式编码的特点, 量化压缩的过程是在多级变换之后进行的, 也就是不具备 MPEG 等编码标准中的循环过程, 因此 MCLIFT 中的参考帧的计算与 MPEG 中不同, 选取的是原始图像而非压缩后的重构图像。

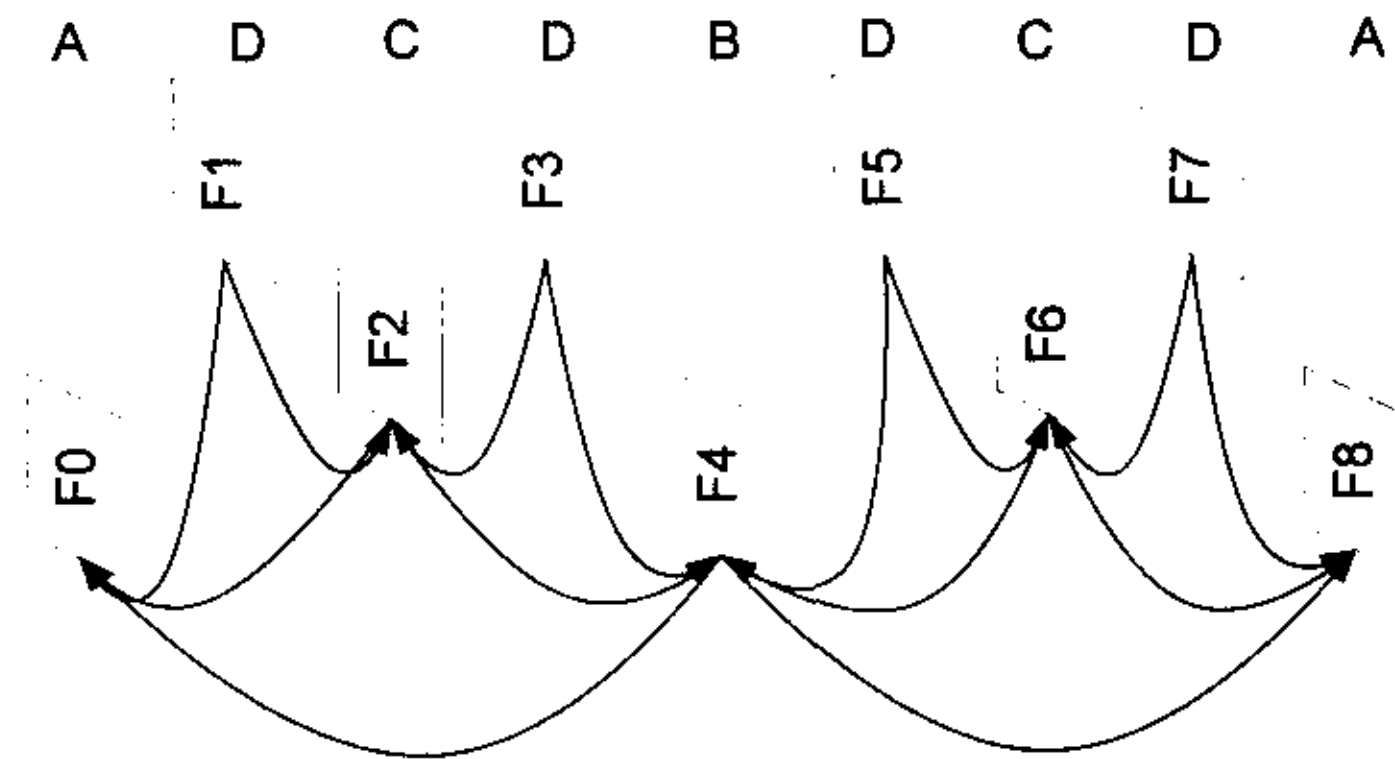


图 3-5 采用截断 5-3 提升型小波的 3 级 MCLIFT 时轴小波分解结构 (3-layer MCLIFT temporal wavelet decomposition with truncated 5-3 filter)

从一级小波变换结构看来, MCLIFT 方法的帧结构和带有 B 帧的 MPEG 帧结构很类似, 然而多级 MCLIFT 和 MPEG 的结构就有所不同。我们在图 3-6 中设定了一种作为比较的基于 MPEG 的帧群结构 (group of picture, GOP) “IBBB PBBB P...”, 这种结构在整个视频序列中只有第一帧作为独立编码的 I 帧 (I-frame, intra-frame), 并且在剩下的 P 帧 (P-frame) 中, 每两个 P 帧之间插入 3 个 B 帧 (B-frame)。如图 3-6 所示, 在相邻的 9 帧 F0, F1, ..., F8 中, F0 是独立编码的 I 帧, F8 为基于 F0 为参考帧预测的 P 帧, F1, F2, F3, F5, F6, F7 为 B 帧, 分别以相邻的 I 帧或 P 帧作参考帧进行预测。

比较 MCLIFT 和 MPEG 的结构, 我们发现 MCLIFT 中 F4 是按照 B 帧编码, 同时 F1, F3, F5 和 F7 以较临近的 F2 和 F6 作为参考帧, 而 MPEG 中的 F1, F3, F5 和 F7 都是以相对较远的 F0, F4 和 F8 作参考帧。通过增加了 B 帧的数目, 同时选取较邻近的参考

帧，MCLIFT 在时轴上的相关性得到了提高。从时轴帧率的可伸缩性（frame-rate scalability）来看，这两种结构都能够实现 1/2，1/4 和 1/8 的采样帧率的应用。然而，由于 MCLIFT 采用嵌入式三维小波视频编码，编码后的码流在解码质量上也具备可伸缩性。这种综合的可伸缩性对 Internet 传输应用很有意义，并且是 MPEG 编码的单层码流（single-layer bitstream）所不具备的。MCLIFT 的缺点在于解码时的帧延迟是 MPEG 的两倍，同时解码所需的缓存为 5 帧，而 MPEG 需要 3 帧。在图 3-5 的 MCLIFT 结构中，共需要 14 次运动估计和运动补偿，这与相应的 MPEG 结构相等。

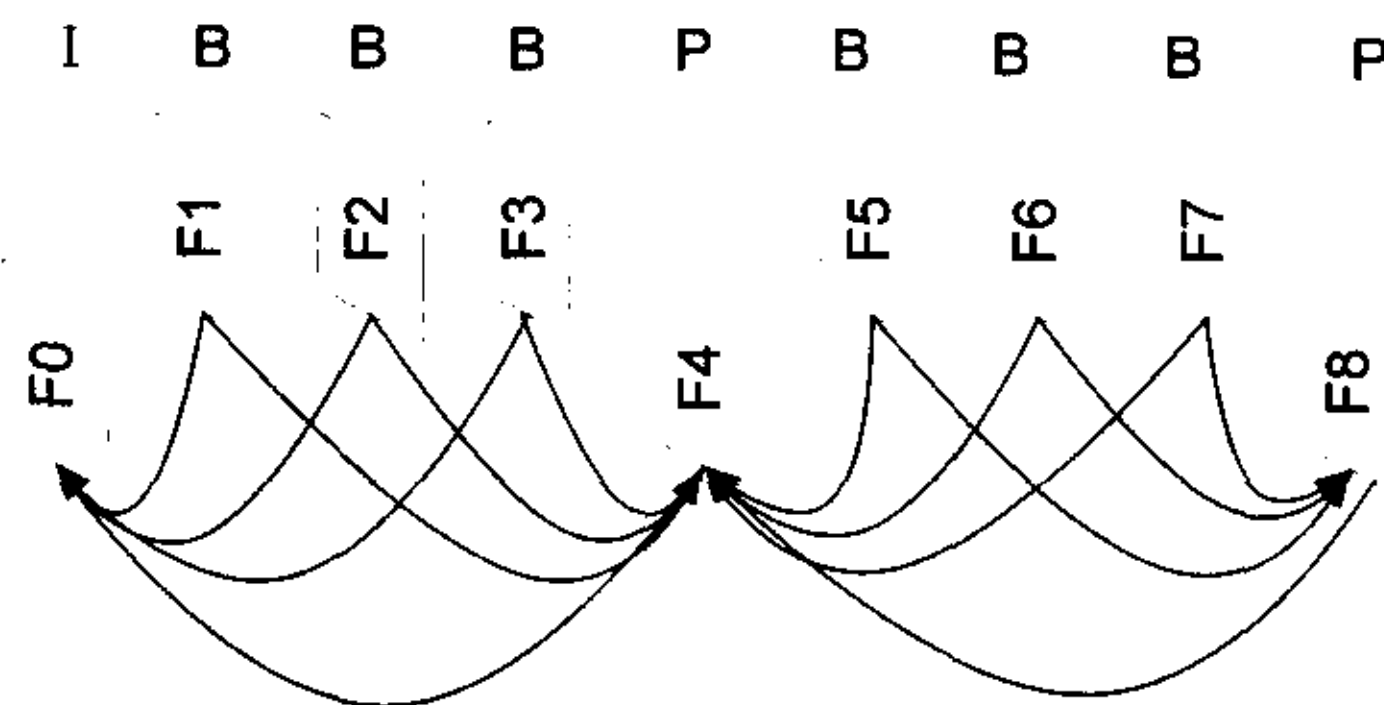


图 3-6 MPEG-4 的 IBBBPBBBP... 帧结构示意图 (The GOP structure of MPEG-4 IBBBPBBBP...)

3.3 帧间变换和熵编码

本章采用的编码系统对时轴 MCLIFT 变换过的小波系数帧再进行如图 3-7 的 Spac1 小波包结构进行帧内小波分解得到三维的小波系数，之后采用第二章中的分块自适应算术编码和率失真优化得到嵌入式的视频压缩码流。

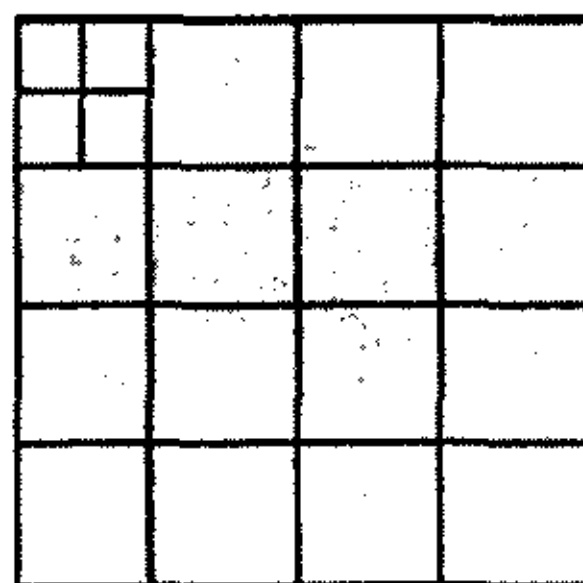


图 3-7 帧内二维 Spac1 小波分解结构 (Intra-frame 2D Spac1 wavelet decomposition)

3.4 实验结果与分析

实验结果是在 MPEG 标准测试序列 Foreman 和 Coastguard 上进行的, 两个序列都是采用 QCIF (每帧 176x144 像素) 的分辨率, 全长 300 帧, 帧率为每秒 30 帧。

第一部分的实验比较了采用截断提升型滤波器的 MCLIFT 方法和采用原有的 5-3 提升型滤波器的 MCLIFT 5-3 方法的性能。第二部分实验中, 我们将本章提出的采用 MCLIFT 的三维小波视频编码器与三种基准编码器的编码性能进行了比较。

3.4.1 不同提升型滤波器的性能

图 3-8 比较了采用截断提升型滤波器的编码器 MCLIFT 和采用原有的 5-3 提升型滤波器的编码器 MCLIFT 5-3 的性能。每一帧的运动矢量经过差分得到的预测残差都采用 MPEG-4 中 P 帧运动矢量的变长编码(variable length coding, VLC)码表进行独立编码^[9]。编码运动矢量的码率如表 3-1 所列, 可以看到由于 MCLIFT 5-3 对应的滤波器的运动矢量大约是 MCLIFT 的两倍, 因此编码运动矢量的码率也近似 MCLIFT 的两倍, 运动矢量码率的增大将降低低码端的编码效率。从图 3-8 看出, 对于 Forman 和 Coastguard, 当码率分别低于 180Kbps 和 150Kbps 时, MCLIFT 5-3 的性能低于 MCLIFT。对于 Foreman 序列, 在低码率端, MCLIFT 5-3 比 MCLIFT 低 0.8dB。为了保证低码端的性能, 目前我们决定在下面的小波编码器中采用截断的提升型滤波器 MCLIFT。尽管如此, 在 MCLIFT 5-3 中, 提升型步骤层间的运动矢量还具有很强的相关性, 如果能够利用已有的第一层提升型步骤中得到的运动矢量来预测第二层, 编码的效率有望进一步提高并超过截断型 MCLIFT。

表 3-1 两种运动补偿提升型滤波器对应的运动矢量的码率 (Bitrate for the motion vectors according to different motion-compensated lifting wavelet filter)

滤波器	MCLIFT (Kbps)	MCLIFT 5-3 (Kbps)
视频序列		
Foreman	27. 6	53. 3
Coastguard	20. 0	35. 7

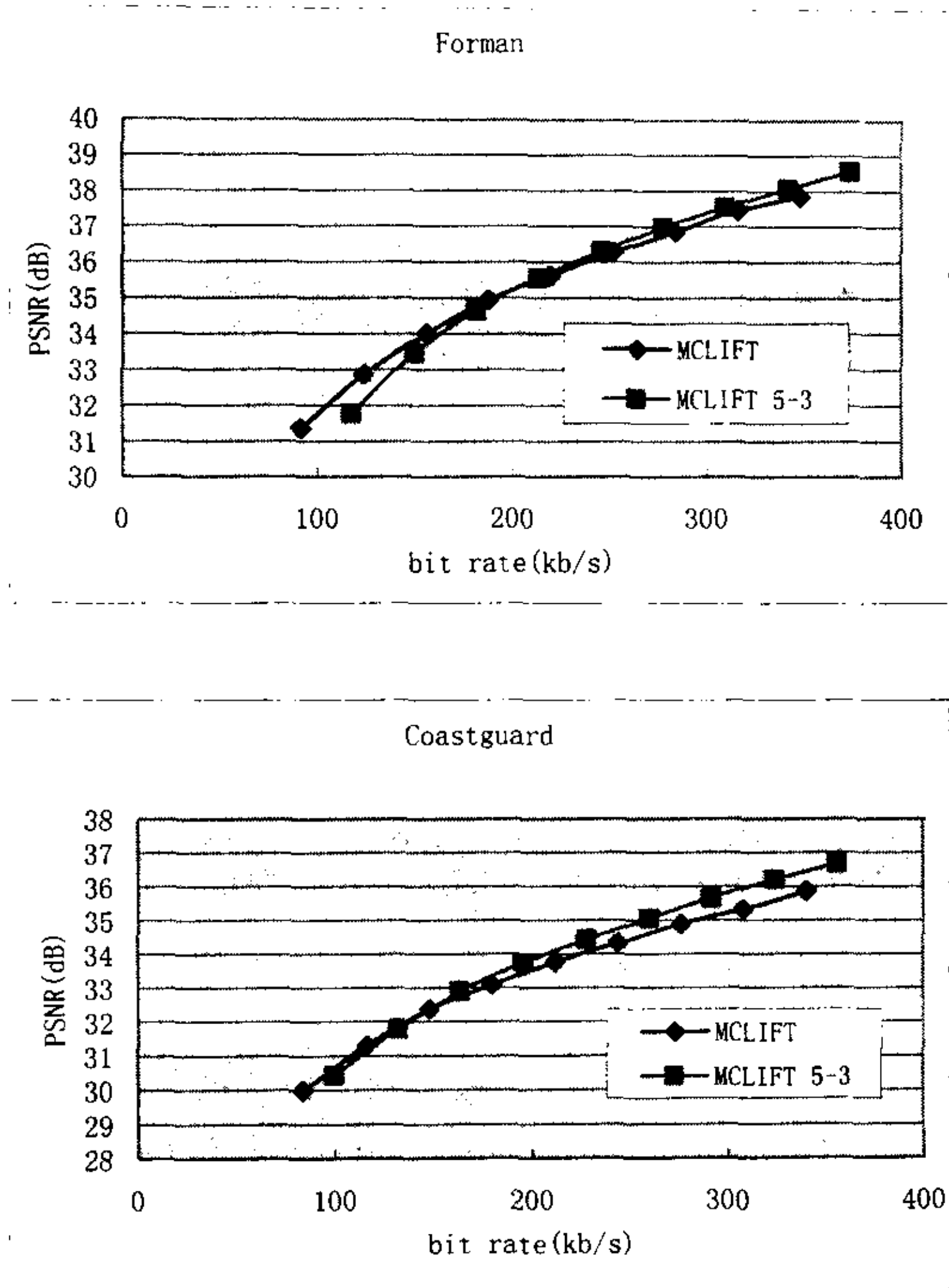


图 3-8 MCLIFT 5-3 与原始 MCLIFT 编码效率比较 (Performance comparison between MCLIFT and MCLIFT 5-3 filters)

3.4.2 采用MCLIFT的小波视频编码器与其他编码器性能比较

下面我们通过与三种基准的编码器的性能比较来衡量本章提出的 MCLIFT 方法。第一种编码器是类似第二章中采用的无运动三维小波编码器 (Motionless 3D Wavelet)，其中时间轴的小波变换按照像素相对的位置直接滤波，不考虑帧间的运动。时轴小波分解采用 3 级双正交 9-7 滤波器，帧内采用 9-7 滤波器进行 3 级二维分解。分解后的小波系数采用与 MCLIFT 相同的熵编码方法和率失真优化 (rate-distortion optimization, RDO)。

第二种编码器是著名的 SPIHT (Set Partitioning in Hierarchical Trees) 小波视频编码器^[71], 其中时轴的小波分解也没有采用运动信息。与 Motionless 3D wavelet 不同, SPIHT 的熵编码器利用的不是小波系数子带内的相关性, 而是子带间的相关性。第三种编码器是 MPEG-4 标准的 VM 16.0 版本。两种 GOP 的结构被采用。其中一种是“IPP..”, 即整个序列只有第一帧采用 I 帧, 其他的都是 P 帧, 另一种则是 3.2 节中所示的“IBBB PBBB P...”结构。为了比较的公平, 在 MCLIFT 和 MPEG-4 中都采用了半像素的和 OBMC。图 3-9 给出了比较的峰值信噪比 PSNR (peak signal to noise ratio) 对应不同码率的曲线。

我们首先比较了采用运动补偿提升型结构的编码器 MCLIFT 和其他两种无运动的三维小波编码器。由于 MCLIFT 和 Motionless 3D Wavelet 的熵编码方法完全相同, 压缩效率的区别来源于时轴的运动补偿对齐。从图中看出对于 Foreman, MCLIFT 比 Motionless 3D wavelet 在 64Kbps 到 320Kbps 的码率范围内提高了平均 1.3dB; 对于 Coastguard, MCLIFT 比 Motionless 3D wavelet 高出了 0.9dB。和 SPIHT 相比, MCLIFT 对 Foreman 和 Coastguard 分别提高了 1.9dB 和 1.3dB。由比较结果可以看出时轴上的加入了运动补偿的提升型小波对三维小波编码有很大改善。

下面我们进一步比较了 MCLIFT 与 MPEG-4 的 VM 16.0 版本。对于 Foreman, MCLIFT 优于采用“IPP..” GOP 结构的 MPEG-4 平均 0.7dB; 对 Coastguard, MCLIFT 也有 1.2dB 的增益。与采用“IBBB PBBB P..”GOP 结构的 MPEG-4 相比, 在某些比特率, MCLIFT 对于 Foreman 达到 0.6dB 的增益, 对 Coastguard 有 1.0dB 的增益。从实验结果看出, 采用加入了半像素精度 OBMC 运动补偿的提升型小波, 有效的增强了时轴方向被滤波像素的相关性, 结合嵌入式的熵编码和率失真优化, 三维小波编码器达到了优于 MPEG-4 编码标准的出色性能。除此之外, 采用 MCLIFT 方法压缩后的码流还具有在帧率和压缩质量方面的灵活的可伸缩性, 更为适应网络视频传输等应用中的需求。

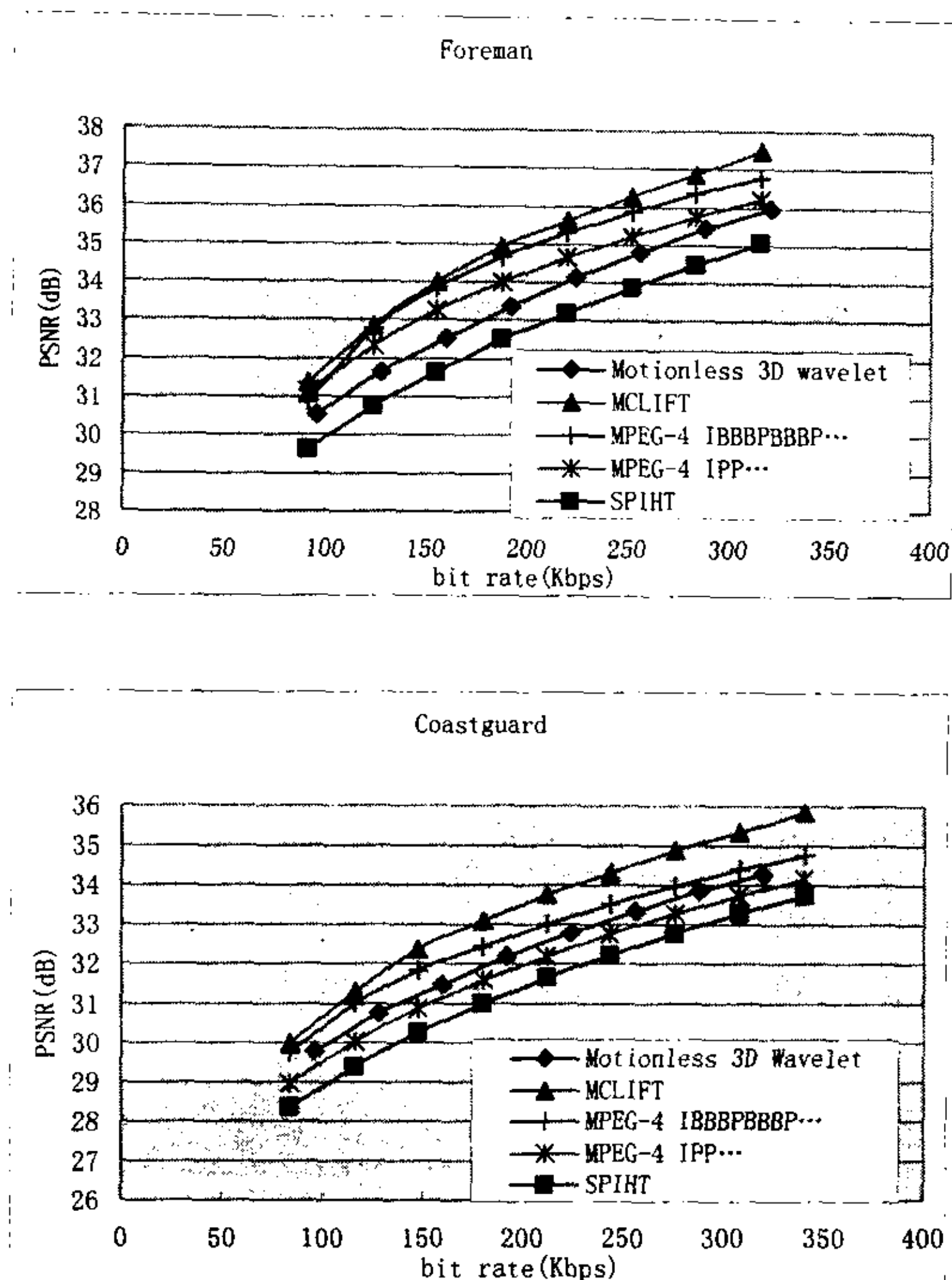


图 3-9 MCLIFT 编码器和他三种编码器的性能比较 (Comparison of the Motionless 3D Wavelet Video Coder, MCLIFT, SPIHT and the MPEG-4 coder)

3.5 本章小结

本章我们提出了一种基于运动补偿提升型小波 MCLIFT 的三维视频编码框架。通过在时轴提升型小波结构采用双向半像素精度的重叠块运动补偿, 编码性能得到了有效提高。实验结果表明基于 MCLIFT 的小波编码器达到了优于 MPEG-4 标准的编码性能。在进一步的研究中, 可以采用更高精度的 1/4 像素运动矢量 来进一步提高编码效率。由于运动补偿是双向的, 产生的运动矢量在整个码流中占据了很大的比重, 如何更好的对运动矢量进行编码来去除之间的相关性也是有待进一步研究的工作。

第四章 基于运动抽线技术的小波视频编码

运动抽线 (motion threading, MTh)^{[82][83]}是微软亚洲研究院的Jizheng Xu和Shipeng Li等提出的一种三维小波视频压缩技术。通过将时轴的像素根据宏块运动估计得到的运动矢量按一定规则连成长度不同的运动线,然后在运动线上做多级小波分解,运动抽线技术巧妙的设计了将运动信息与小波变换相结合的结构,提高了时轴小波的去相关能力,对于简单运动的视频序列达到了优于MPEG-4标准的性能。然而由于运动抽线中存在的运动线断线等问题,对于复杂运动序列的压缩效果还不尽人意。本章在运动抽线提出的基本结构上对断线问题进行了分析,并进行了两项主要的研究工作,有效的提高了小波视频编码的效率,达到了与当今压缩标准具有竞争力的压缩性能。

本章的第一项工作是通过改进运动估计时的匹配规则来降低运动线的断线情况,同时兼顾了运动匹配的准确性,得到的压缩系统对不同运动类型的视频序列都能够达到类似或高出MPEG-4的压缩性能。第二项工作针对运动抽线中断线的基本原因,提出了高级运动抽线技术A-MTh,采用提升型小波实现结构,不但从根本上解决了断线问题,同时还可以很容易的将运动精度由原来局限的整像素提高到半像素甚至1/4像素精度,进一步提高了编码效率。并且,在高级运动抽线技术中还考虑了视频运动中的遮挡问题,以及提出了运动矢量相关编码等相关技术,实验结果表明,高级运动抽线技术具有与最先进的H.26L标准相近的编码性能,同时还提供了解码质量和帧率的可伸缩性。

4.1 运动抽线技术简介与分析

Xu等提出的运动抽线 (motion threading, MTh) 技术可以很方便的将运动信息引入时轴的小波变换当中,并且不受滤波器长度的限制。图 4-1给出了基本的三维运动抽线 (MTh) 结构。其中每一竖条代表一个视频帧。从第一帧Frame₀到最后一帧Frame_{n-1}进行基于宏块 (macro-block, MB) 的运动估计,然后对每一帧所有的像素沿从左到右,再从上到下的顺序扫描,根据其所在宏块的运动矢量将沿同一运动方向的像素连成互不重合的线,因此该方法被称之为运动抽线技术。MTh技术对沿运动方向穿好的不同长度和起始点的每条运动线来进行适形小波变换 (shape-adaptive discrete wavelet transform),并将得到的小波系数直接存储在运动线所在的相同位置上。之后采用9-7滤波器对每一帧时轴滤波后的系数进行帧内的二维小波分解,并将三维分解后的系数进行熵编码得到

嵌入式的码流。由于对运动线时轴滤波的结果是直接存储在运动线对应的位置上的，因此要求运动线互不重合来保证解码端的完全重构。基于此要求，可以从图中看出共有三种类型的像素：常规像素（normal pixel）表示该像素在前一帧和后一帧都有且只有一个匹配像素；当多个像素由于块运动估计而指向下一帧的同一像素时，只有其中一个可以和目标像素连接，其他的都被标为停止像素（terminating pixel）并且相应的运动线会在该处截断；对于当前帧中没有被之前帧内任何像素所指向的一类像素，则被标为未引用（non-referred pixel）像素，并代表了新的运动线的起点。

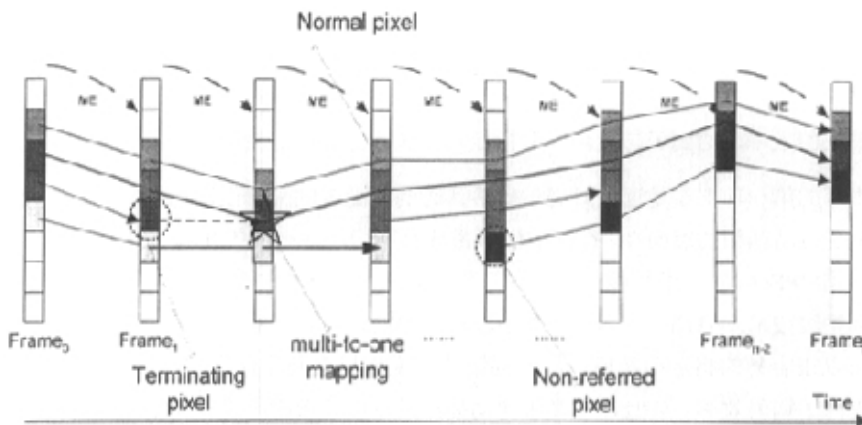


图 4-1 三维运动抽线小波技术的基本结构 (The original 3D motion-threading wavelet)

MTh 方法提供了将时轴小波变换与运动对齐（motion alignment）相结合的一种灵活的结构，在时轴方向和编码效率方面都提供了很好的可伸缩性（scalability）。运动对齐提高了时轴方向的相关性，因此和传统无运动信息的三维小波编码相比，MTh的编码效率有了显著的提高。然而在许多方面，MTh还有一些明显的局限性。首先，为了保证解码端的完全重构，运动线互不重合的限制造成了在像素多对一匹配（many-to-one mapping）的情况下，人为的对运动线进行截断。同时多对一匹配又会产生许多没有被前一帧任何像素所指向的像素，形成新的运动线。对于含有复杂的运动信息的视频序列，多对一匹配将造成大量被截断的线和新出现的线。根据文章[103]中对边界问题（boundary problem）的分析，在信号边界上的小波重构误差比信号内部要大，因此大量的断线会极大的降低编码效率。其次，为了保证完全重构，现有的MTh技术中只能采用整像素精度的运动矢量。众所周知，在传统的视频编码中，分像素（fractional-pixel）

运动矢量，包括1/2像素，1/4像素等可以极大的提高编码效率。因此，如何在保证完全重构的同时，降低断线的出现，以及采用分像素精度的运动矢量都是基于MTh研究的主要目标。

4.2 运动抽线技术中的两轮运动估计技术

根据图 4-1 已知，原始三维运动抽线小波编码结构中共有三种类型的像素：常规像素（normal pixel），停止像素（terminating pixel）和未引用（non-referred pixel）像素。每个停止像素的产生都会同时对应一个未引用像素，并增加两个运动线边界，对于运动信息复杂的视频，多对一匹配的现象增多，产生大量的停止像素和未引用像素，即运动线边界，经过小波重构之后造成的边界效应极大的降低了编码效率。事实上，不加入运动直接进行的三维小波变换是一种避免多对一匹配的极端情况，但由于没有有效的运动对齐，不利于进行小波分解。本节提出了一种折衷的方法——两轮估计运动抽线（two-pass estimated motion-threading, TEMTH），通过相关的两轮运动估计来调整运动矢量，在保证时轴上有效的运动对齐的同时，降低了多对一匹配现象的发生。

4.2.1 基于两轮估计的运动抽线（Two-pass Estimated Motion-Threading）

图 4-2 所示为提出的 TEMTH 技术的两轮运动估计流程。第一轮采用类似 H.26L 标准^[12]中的运动估计，主要目标是通过降低每个宏块对应的绝对失真总和（sum of absolute difference, SAD）来估计运动矢量。如果一帧中运动矢量为（0，0）的宏块个数超过了一定比例，该帧定为零运动帧，并将其中所有宏块的运动矢量都置零，否则该帧进入第二轮运动估计。第二轮估计是对第一轮的精细化，我们采用新的匹配原则来综合考虑预测失真大小和多对一匹配的个数。由于多对一匹配代表着不同宏块在参考帧上的对应块之间的像素重叠，采用重叠像素的个数来表征多对一匹配的情况。通过两轮估计过程，TEMTH 技术权衡了实际匹配失真的大小和运动线的截断情况，改善了压缩质量。下面对流程中的每个过程逐一进行介绍：

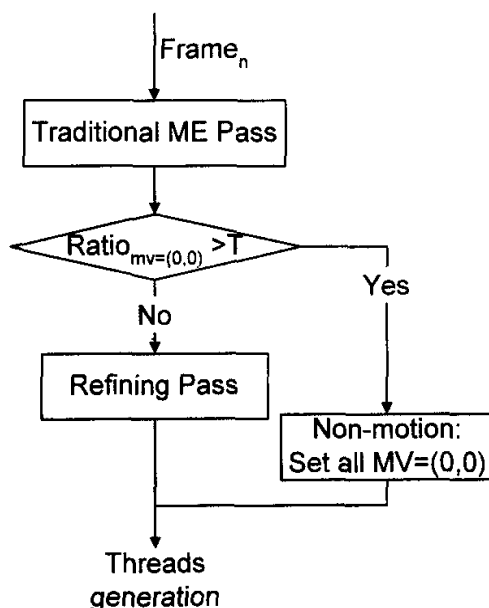


图 4-2 两轮估计运动抽线流程图示意图 (Flowchart of TEMTH)

4.2.2 标准运动估计过程

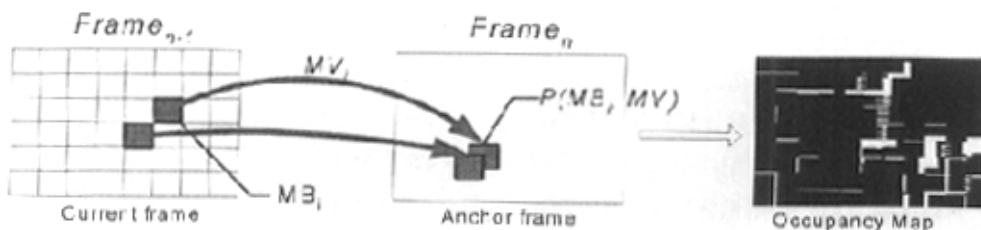


图 4-3 运动估计中的重叠现象和整帧对应的占据图 (The overlapping phenomenon and the corresponding occupancy map for a whole frame)

第一轮运动估计采用基于 16×16 的宏块全搜索运动估计得到初始化的运动矢量。对当前帧中的宏块 MB_i ，匹配过程的目标是找到代价最小的宏块，其中代价函数 $Cost = SAD + \lambda_{mv} \times Bits_{MV}$ 。此处 SAD 代表当前宏块和目标块之间的绝对误差总和，

$Bits_{MV}$ 表示编码相应的运动矢量预测误差所需要的比特数。 λ_{mv} 用来调整运动信息所需比特在代价函数中对应的率失真比例。

如图 4-3, MB_i 在目标帧中对应的匹配块位置为 $P(MB_i, MV_i)$ 。记录每一个宏块的匹配块占据的像素位置, 可以得到整帧图像对应的占据图, 如图右所示。黑色区域表示已被占据的像素, 白色表示没有和任何宏块匹配到的像素, 注意黑色区域可以是被多个宏块占据的像素。占据图可以用公式表示为:

$$Map_{\Sigma} = \bigcup_{k=1}^N P(MB_k, MV_k), \quad N \text{ 为一帧中宏块总数。} \quad (4-1)$$

占据图将被用于第二轮运动估计来计算重叠像素的个数。通过运动估计降低的 SAD 值 $\Delta SAD = SAD_{MV} - SAD_{(0,0)}$ 也将在第二轮估计中用来决定宏块扫描的顺序。

在标准运动过程后, 如果某帧内运动矢量为 $(0, 0)$ 的宏块个数超过了给定门限比例, 该帧被定义为零运动帧, 并将其中所有宏块运动矢量置零, 相应的运动线在该帧都直接按位置连接。每帧仅需分配一个比特用来标示是否是零运动帧。适度采用零运动帧不但节省了对该帧编码运动矢量的比特数, 也避免了重叠像素的产生。

4.2.3 精调运动估计过程

非零运动帧进入第二轮运动估计对第一轮的运动矢量进行精调。精调过程中, 每个宏块同样经过全搜索过程得到代价最小的匹配块。我们提出了不同于第一轮运动估计的新匹配原则来综合考虑像素的实际匹配误差和重叠像素的个数。新的代价函数定义为:

$$Cost = SAD + \lambda_{mv} \times Bits_{MV} + \lambda_{ovlp} \times Num_{ovlp}(P(MB_i, MV_i), Map_{\Sigma-i}) \quad (4-2)$$

对于宏块 MB_i , 部分占据图 $Map_{\Sigma-i}$ 表示占据图(4-1)除去第一轮中 MB_i 的贡献之后得到的部分占据图:

$$Map_{\Sigma-i} = \bigcup_{\substack{k=1 \\ k \neq i}}^N P(MB_k, MV_k), \quad i=0 \dots N \quad (4-3)$$

代价函数中的 $Num_{ovlp}(P'(MB_i, MV_i), Map_{\Sigma-i})$ 项表示第二轮估计中 MB_i 的候选匹配块 $P'(MB_i, MV_i)$ 与当前的部分占据图 $Map_{\Sigma-i}$ 中黑色区域重叠的像素个数。权重系数 λ_{ovlp} 的引入用来控制重叠像素 $Num_{ovlp}(\cdot, \cdot)$ 对代价函数的影响。由代价函数可以看出, 候选匹配块与部分占据图之间重叠像素越多, 代价函数的值越大, 最终使得具有次优 SAD 但是较少重叠像素的匹配块被选中。

根据公式(4-3)得到精调后的 MB_i 的匹配块, 并根据新的位置 $P'(MB_i, MV_i)$ 更新占据图 Map_{Σ} , 用于后面宏块的精调运动估计。由于后扫描到的宏块要基于之前宏块更新的占据图进行精调, 为了使宏块预测误差 SAD 不因新准则而产生太大的变化, 我们根据第一轮估计中 $\Delta SAD = SAD_{MV} - SAD_{(0,0)}$ 下降的速度从大到小排列第二轮中的宏块扫描次序, 保证第一轮中匹配较准的宏块的准确性。

4.2.4 实验结果与分析

我们将基于本文提出的两轮运动估计方法实现的小波编码器 TEMTH 与三种基准编码器相比较: 首先是无运动的三维小波编码器 (none-motion, NM), 第二个是文章[83]中的原始三维运动抽线编码器 (motion threading, MTh), 第三个是 MPEG-4 标准^[9]的 VM 16.0 版本。在基于小波的编码器中, 时轴方向都进行 3 级双正交 9-7 小波分解, 之后进行帧内的 3 级二维双正交 9-7 小波滤波, 得到的三维小波系数都采用第二章中的熵编码方法进行编码。MPEG-4 的设定为第一帧是 I 帧 (I frame), 其后所有帧都为 P 帧。实验采用的是 MPEG 标准 QCIF 格式 (176x144) 视频序列 Carphone, Coastguard, Foreman 和 News, 都采用总数 288 帧, 帧率 30Hz。

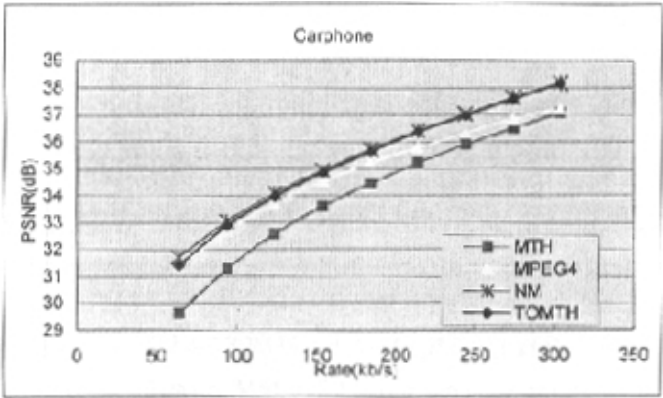
TEMTH 编码中的参数设置如表 4-1 所示。运动估计的搜索范围 S 为 16 像素, λ_{mv} 设置了比较小的值以保证运动估计的准确度, 门限 T 设为 60%, 表示包含零运动宏块比例超过 60% 的帧被定义为零运动帧。为了权衡块失真 SAD 与重叠像素的个数。对于不同的序列, 公式(4-2)中的 λ_{ovlp} 的取值应有所调整, 本实验中对不同序列我们都设置 λ_{ovlp} 为经验值 50, 普遍达到了良好的压缩效果。

表 4-1 TEMTH 中的参数设置 (Parameters for the TEMTH coding)

Parameters	λ_{mv}	T	λ_{ovlp}	Motion Range S
values	6	60%	50	16

图 4-4 给出了不同编码器对应 4 种标准序列的压缩结果，即峰值信噪比 PSNR 对比特率 bit-rate 曲线。我们首先比较 TEMTH 和其他三种编码器对运动较复杂的 Carphone 和 Foreman 序列的压缩效果。从图 4-4 看出，由于大量产生的停止像素，原始的 MTH 编码器比 MPEG-4 效果要差。TEMTH 方法有效减少了重叠像素的数目，将编码效率提高到对 Carphone 为 1.8dB，对 Foreman 为 1.1dB，达到了与 MPEG-4 相近的性能。由于 Carphone 序列的背景区域运动信息很少，TEMTH 在低码率端略低于 NM，原因在于 NM 不需要编码运动矢量信息。对于 Foreman，TEMTH 比 NM 高出达 0.8dB，这主要归于 NM 在运动对齐方面的局限性。

接下来我们比较对于简单运动的序列 News 和 Coastguard，几种编码器的压缩效率。对 News，原始的 MTH 编码器只在码率 120Kbps 以上才高出 MPEG-4，TEMTH 方法提高了 0.5dB，使得在 90Kbps 的编码效率都高出 MPEG-4。对于 Coastguard，原始的 MTH 编码器编码效率要优于 MPEG-4，改进后的 TEMTH 方法保持了该编码效率，并超出 MPEG-4 多达 2.5dB。对 Coastguard 序列，TEMTH 方法高出 NM 编码器 1.2-2.2dB，而对于运动信息很少的 News 序列，TEMTH 也达到了和 NM 相当的结果。同时，采用 TEMTH 的三维小波编码方法能够提供压缩质量和时轴帧率的可伸缩性，这一点是单码流 MPEG-4 编码器不能做到的。总结说来，本节提出的 TEMTH 编码方法对代表了不同运动特点的 4 种视频序列普遍达到了较好的压缩效果。



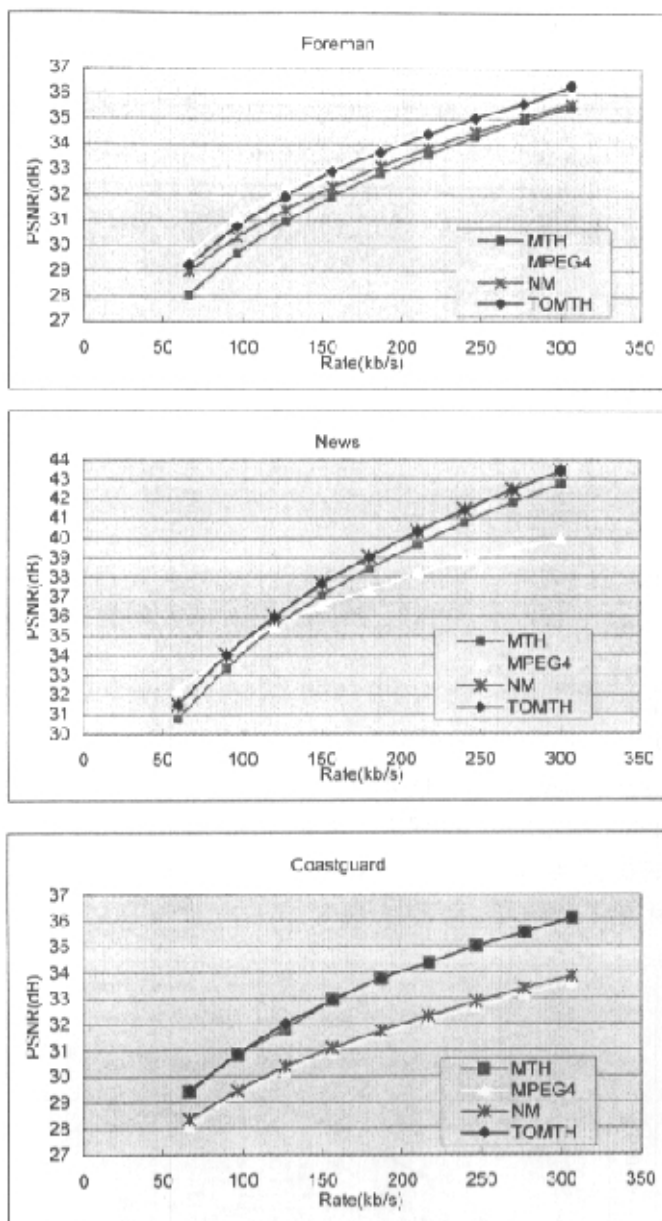


图 4-4 压缩质量比较: 无运动三维小波编码器 (NM), 原始三维小波运动抽线编码器(MTh), 本节提出的两轮估计运动抽线编码器 (TEMTH) 和 MPEG-4 编码器 (PSNR versus bit rate: a comparison among NM, MTH, TEMTH and MPEG-4)

4.2 基于提升型小波的高级运动抽线技术

在第3章基于运动补偿提升型小波的视频压缩的思想启发下,我们提出了一种基于提升型小波的高级运动抽线技术(advanced motion threading, A-MTh)。同4.2节中的解决方法不同,A-MTh不是通过减少多对一匹配情况的产生来降低断线的可能,而是从根本上避免了对运动线的人为截断。利用提升型结构,在保证完全重构的同时,彼此有重合部分的运动线可以连续的参与时轴小波变换而不会被截断。同时在提升型结构中能够很容易的引进分像素的运动矢量,将运动对齐的精度提高到了1/4像素。同时,我们还提出了一个率失真优化的模型来改进运动估计的过程。

4.2.1 基于提升型结构的1/4像素精度运动抽线

在前面几章我们提到过提升型小波变换的基本结构和特点,本节的方法是将传统的时轴运动抽线技术扩展为提升型结构,利用提升型结构的基本单元结构以及原位(in-place)运算等特点改进MTh技术。在时轴5-3提升型结构中,每一帧状态的更新都是基于提升型步骤来一层层实现的,首先提升奇数帧的状态使之成为高频小波系数帧,之后提升偶数帧的状态为低频小波系数帧。图 4-5给出了基于Le Gall 5-3提升型小波实现的MTh时轴小波变换,此处每一竖条代表一帧,每个方块代表一个像素。基于宏块的运动估计方向总是由奇数帧指向相邻的偶数帧。在每个提升型步骤中,相应的奇数帧或者偶数帧中的所有像素都沿着运动线的方向滤波并得到状态的提升。例如,要使Frame₁到达高频小波系数的状态H₀,则Frame₁中的所有像素都会与沿运动线方向上前一帧和后一帧中的一对匹配点进行单元提升型计算。

在原来MTh方法中由于多对一匹配而被截断的运动线可以通过单元提升型计算来避免截断。如图 4-5所示,在单元结构中,Frame₁中对应的原来的停止像素除了可以利用左侧(前向)的像素计算之外,还可以同时利用右侧(后向)的像素来进行计算,并不需要将运动线在右端截断,而在下一层单元计算中,多对一匹配中的被指向的像素只与前一帧中最先被扫描到的像素进行运算,不影响运算的唯一性。对于偶数帧内可能出现的未被指向的未引用像素,则利用相邻运动线的运动方向使其在前后方向都得到连接。最后,所有的像素都保证在前向和后向有像素相连。虽然这些运动线由于运动信息产生了重叠的现象,但是采用提升型结构,在保证完全重构的同时能够避免断线的发生,使传统MTh方法中影响编码效率的边界问题得到了出色的解决。

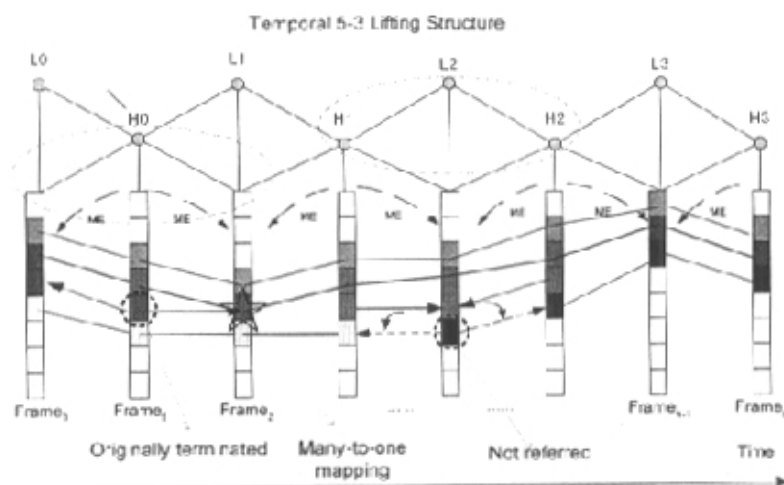


图 4-5 基于 5-3 提升型小波与双向运动搜索的运动抽线结构(5-3 lifting-based motion-threading (MTh) with bi-directional motion search)

在实际编码实现中,为了节约内存的需求,时轴方向的小波分解与合成都采用了和第二章中类似的帧缓存结构,即视频是一帧一帧的输入缓存,当帧数达到一定数量(对 5-3 滤波器为 4 帧)时,就可以进行提升型运算并输出一帧时轴小波系数,此后每输入两帧数据就输出两帧系数,并不需要等到视频帧中余下部分的到来。帧缓存结构有效节省了视频编解码过程中的内存需求。

提升型结构中也非常容易实现可完全重构的分像素精度运动抽线。从图 4-5 已知运动估计的方向都是从奇数帧指向偶数帧。结合运动对齐的单元提升型结构因此可以看作在连续三帧中进行的双向运动补偿预测。图 4-6 给出了提升 F_{2n+1} 到高频系数帧状态的单元结构。此处黑色圆圈代表整数像素,灰色圆圈代表 $1/2$ 像素,白色圆圈代表 $1/4$ 像素。实线箭头为从宏块运动矢量得到的像素运动矢量,从偶数帧向奇数帧的像素运动矢量则为由相应的实线箭头反向得到,如虚线箭头所示。在图 4-6 显示的单元结构中, F_{2n+1} 中的每个像素通过与 F_{2n} 和 F_{2n+2} 中运动补偿位置的像素的计算,提升为高频系数状态。如果在提升结构当前层中 F_{2n+1} 中像素 x_2 指向了 F_{2n} 中 x_1 和 x_2 中间的 $1/2$ 像素,那么在

提升结构的下一层，即要将 F_{2n} 提升为低频系数帧的过程中， F_{2n} 中的像素 x_1 相应的指向 F_{2n+1} 中 x_2 和 x_3 之间的 $1/2$ 像素。换句话说，相邻两帧之间对应的运动矢量是严格反向的。 $1/4$ 像素的操作和 $1/2$ 情况类似。由于单元结构中的两帧运动补偿后的参考帧在解码端可以再现，残像素运动抽线对解码的完全重构性得到了保证。

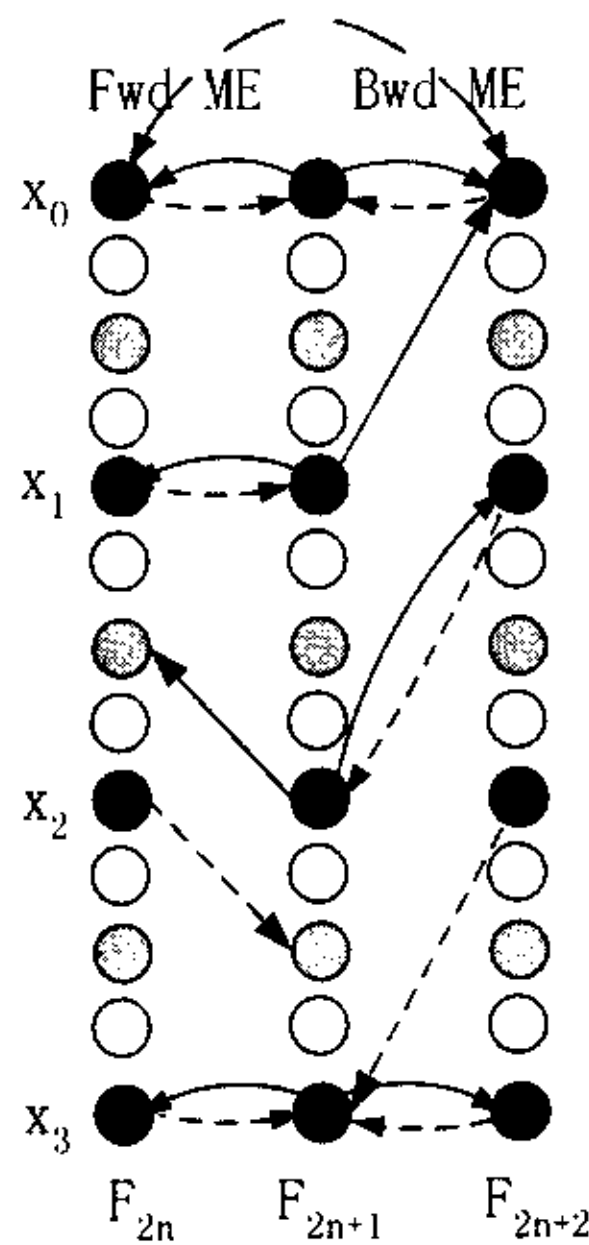


图 4-6 $1/4$ 像素单元提升型运算 (Quarter-pixel elementary lifting operation)

4.2.2 多分解级运动抽线

三维小波视频编码中，通常会在时轴方向进行多级的小波分解。图 4-7 给出了一个时轴 4 级分解的结构。每一级分解的低频子带被进一步的分解，达到时轴的可伸缩性。因此，相邻的低一级帧率降为高一级的 $1/2$ 。在原始的 MTh 中，运动矢量只在最高级进行，之后多级小波分解都是在已生成的运动线上进行。但对于低帧率级来说，直接由最高级的运动线推得的运动矢量的精度会存在误差。同时，对于帧率可伸缩的应用，对于较低的帧率仍然需要将最高帧率的运动矢量全部传输，增大了码流的负担，降低了编码效率。综合考虑帧率的可伸缩性和编码效率，本节提出的高级运动抽线技术对每一级都单独进行运动估计和运动矢量编码，注意对于低帧率级的运动估计是在相应采样帧率的

原始视频，而不是在从高级分解得到的低频系数帧上进行的，这保证了各分解级之间的相对独立。对每一层的独立的运动估计能够提高低帧率级的运动矢量精度。

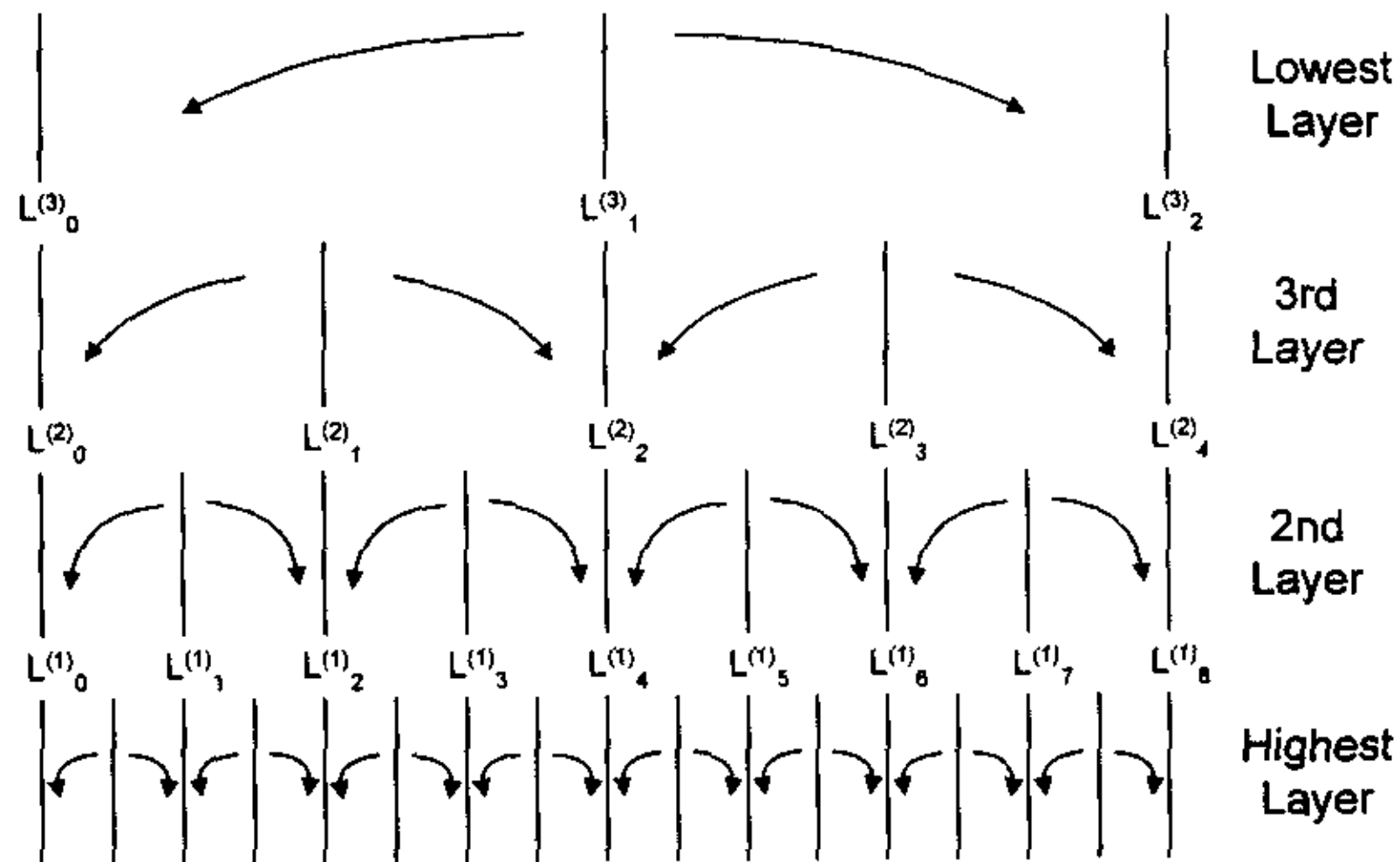


图 4-7 四级时轴小波分解结构 (A 4-layer temporal wavelet decomposition structure)

4.2.3 高级运动抽线中的率失真优化运动估计

研究表明影响运动抽线技术编码效率的主要因素是运动对齐，包括基于宏块的运动估计和运动线生成。下面我们集中考虑通过改进运动估计部分来提高高级运动抽线技术的编码效率。重点研究两个方面：其一是对运动估计中宏块匹配准确性的判断，其二是研究相邻帧和相邻层之间的运动矢量冗余来压缩运动信息的码率。我们定义了7种模式来描述不同的宏块模式，并提出一种率失真优化的运动估计方法来综合评价匹配准确性和运动冗余。

1) 错配 (mismatch) 现象分析

基于提升型结构的多级分解保证每个像素在前后帧都相连，然而事实上，某些运动现象，例如摇镜头 (panning) 和遮挡 (occlusion)，会造成在前一帧或后一帧找不到真正匹配的像素。强制连接那些不是真正匹配 (错配) 的像素将降低视频的压缩质量，并且影响视觉上的效果。因此，此类情况的运动线应该截断。

由于运动信息是基于宏块传输的，错配运动线的截断也是基于宏块的，即错配宏块中所有像素都标为停止像素。前向绝对误差总和 (sum of absolute difference, SAD) 与

后向绝对误差总和以及二者的平均值被用来决定宏块是否错配并标为停止块，运动线在错配的宏块处被截断。相应于不同的连接方式，我们定义了三种宏块模式：双向连接Bid，前向连接Fwd和后向连接Bwd。由于小波合成后在运动线边界上的重构误差要大于内部误差，因此在定义模式判定准则时，会适当倾向于双向连接Bid。由于两种截断模式Fwd和Bwd只需传输一个方向的运动矢量，因此对这两种模式的选择在避免错配的同时还可以降低运动信息的码流开销。

2) 相邻帧和相邻级之间的相关性

由于多级运动线技术中，每一级都拥有一套独立编码传输的运动矢量，因此多级运动矢量的总和造成了很大的码率压缩负担，尤其是在低码率的应用中。为了降低运动开销，我们分析了相邻帧和相邻级之间的运动矢量的冗余来进行压缩。

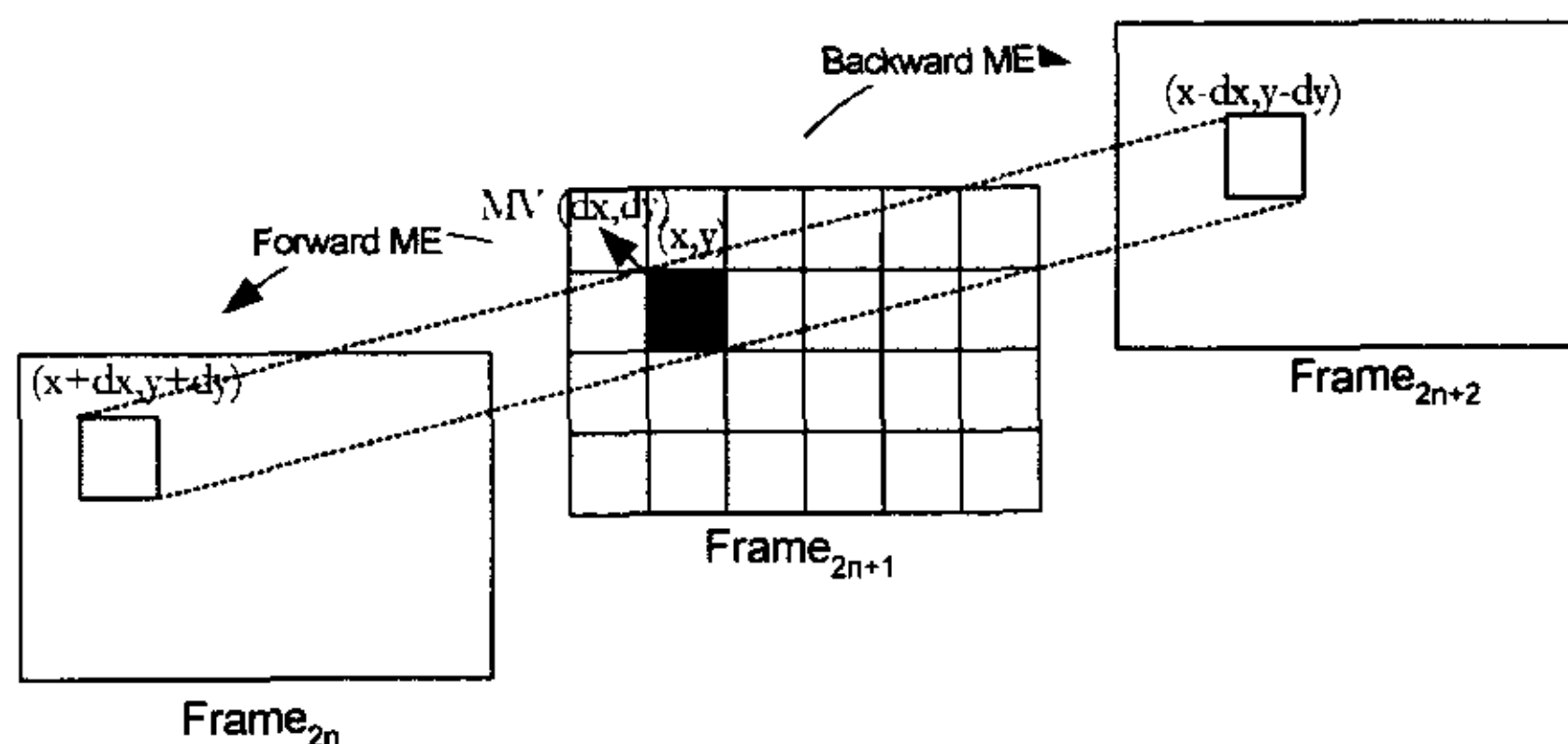


图 4-8 双向运动估计中的直接反向模式 (DirectInv mode in bi-directional motion estimation)

已知运动估计的方向都是从奇数帧指向偶数帧，定义一个前提假设：如果 F_{2n+1} 中位置 (x, y) 的宏块与 F_{2n} 中位置为 $(x + dx, y + dy)$ 的块相匹配，那么采用时轴的均匀外插可以得出该宏块在 F_{2n+2} 中将运动到位置 $(x - dx, y - dy)$ (图 4-8)。这样在运动矢量编码中，前向和后向的运动矢量拥有相同的绝对值和相反的符号，所以只需要传输一个方向的信息。采用直接反向 (DirectInv) 模式来描述此类相关性，适度选择 DirectInv 模式可以有效压缩运动矢量的码率开销。

对于级间的相关性的研究是基于时轴多级分解的结构 (图 4-7)。此处我们只研究相邻级间的相关性。在每一对相邻的级中，采用其中一级得到的宏块运动矢量来预测另

一级的宏块运动矢量。预测的方法有两种，一种是先得到较高帧率级的运动矢量，之后预测低一级的运动矢量。由于最高级对应的帧间跨度最短，因此运动估计比较低帧率级的准确度要高。但是考虑到帧率可伸缩性的应用要求，当传输帧率低于最高帧率时，最高级的运动矢量仍然必需全部传输才能产生当前帧率所需的运动信息，造成了码流的额外开销。因此由高到低的预测方式不适合于帧率可伸缩性的应用。另一种预测是由低到高的预测。由于低级的运动矢量不依赖于高级的信息，可以根据帧率级的不同选择相应的运动矢量，节省了额外的码率。此种预测方法的局限性在于对低帧率级的运动估计中被预测帧和参考帧之间的跨度较大，4级时轴分解对应的跨度为8帧，较大的跨度限制了运动估计的精度。基于此精度预测出的高帧率级运动矢量的精度也会受到影响。为了考虑帧率可伸缩性的应用要求，本章中采用了由低到高的预测方式来考虑级间的运动相关性。

3) 宏块模式与判决准则

基于以上的分析，我们设计了7种宏块模式来表征不同的运动类型和级间相关性。如图 4-9所示，每个竖条代表一帧，被预测帧被均匀分成宏块，黑色方块代表当前待估计的宏块，灰色方块代表当前块在参考帧上对应的匹配块。每种模式下面一行代表当前要进行运动估计的高帧率级，其中DirL，FT_BDL和BT_FDL三种模式用到了级间的相关性，也就是待处理宏块根据低帧率级中相应帧中相同位置的宏块的运动矢量来做运动预测；DirInv用到了帧间的相关性；Fwd, Bwd和Bid模式都是独立编码的。除了Fwd和Bwd模式，其他模式皆为运动线在前后两帧都相连的模式。每种模式定义如下：

DirL: 级间均匀相关模式。前向和后向的运动矢量的绝对值为低一帧率级相应帧中相同位置的宏块运动矢量的1/2，而符号相反，即 $MV_{fwd} = -MV_{bwd} = 1/2 MV$ 。此种模式不需要传输任何当前级的运动矢量，只需由低一级的运动矢量计算得出。该模式可以代表许多运动较匀速的类型。

FT_BDL (BT_FDL): 前向传输-后向级间相关（后向传输-前向级间相关）模式。前向（或后向）的运动矢量被传输，另一个方向的运动矢量由公式 $MV_{bwd} = MV_{fwd} \pm MV$ （ $MV_{fwd} = MV_{bwd} \mp MV$ ）计算得出，其中+号和-号的选取由前一低帧率级运动矢量MV是前向还是后向矢量决定。此种模式中 $|MV_{fwd}| \neq |MV_{bwd}|$ 。在FT_BDL和BT_FDL模式中，选择具有较小运动矢量代价的模式。

DirInv: 帧间相关模式。前向和后向的运动矢量有相同的绝对值和相反的符号，但是都不等于低一级相应的宏块运动矢量MV。由于 $MV_{fwd} = -MV_{bwd}$ ，只需传输前向的运动矢量即可得出后向值。

Bid: 双向传输模式。前向和后向的运动矢量都被编码传输。

Fwd (Bwd): 前向传输（后向传输）模式。只有前向（后向）的运动矢量被传输，相应的运动线在后向（前向）被停止。当一些特定的运动，如遮挡等现象出现时，在后向（前向）也许无法找到匹配的块，此时运动线应该在错配的方向停止。

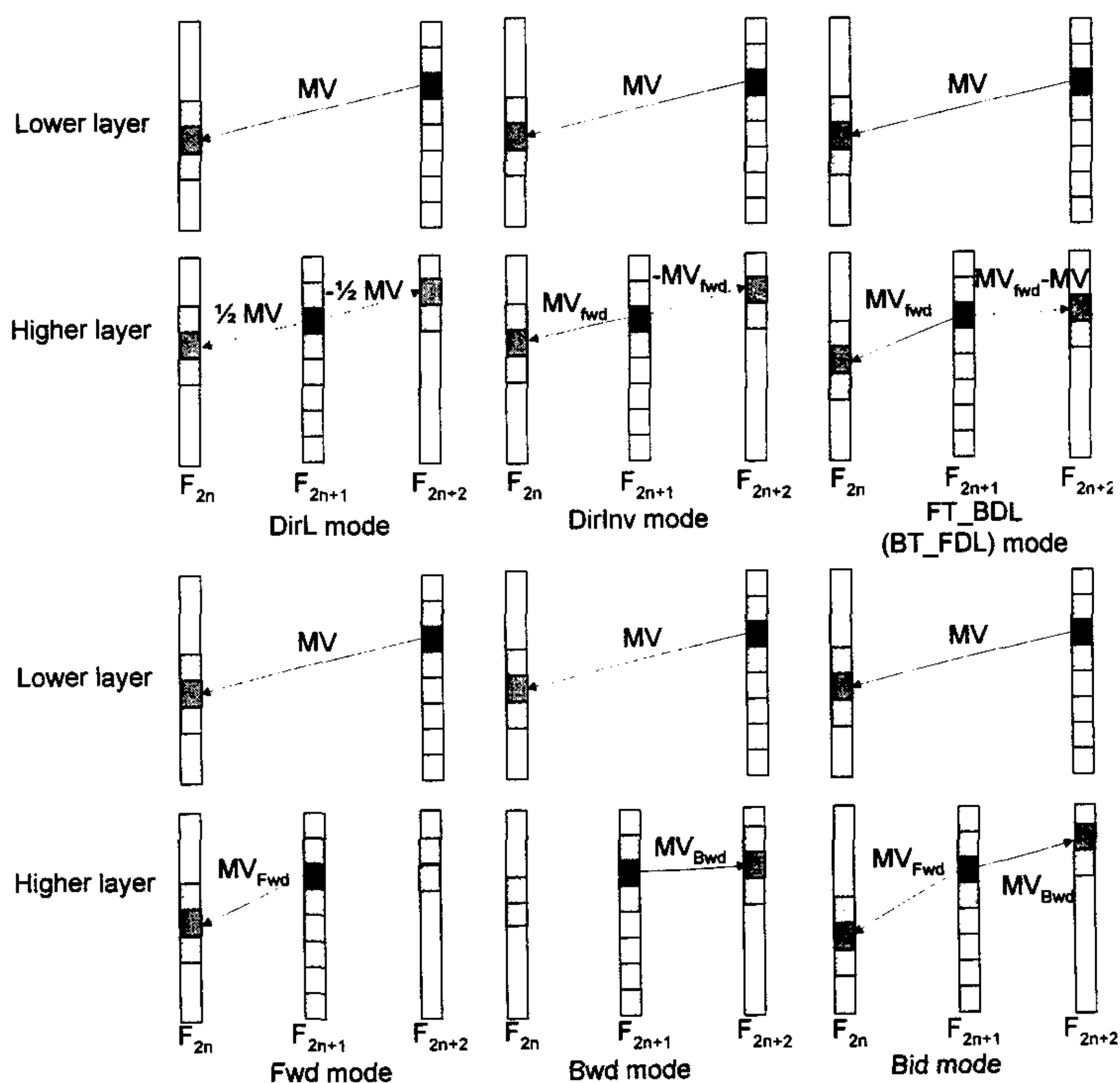


图 4-9 运动预测的 7 种选择模式 (Seven motion modes in motion estimation)

我们采用率失真优化的方法来选择运动模式。对奇数帧中的每个宏块附加一个符号来表示所选择的模式。当前处理的宏块先经过类似H.26L编码标准^[12]中P帧的方式在前向和后向做运动预测（此处只用到16x16的块搜索），产生一对前后向的运动矢量。在上述的7种模式中，基于给定的率失真判定准则选择代价最小的模式，并根据选定模式决定运动线的生成。模式判定准则为代价函数 $Cost = \eta \cdot SAD + \lambda \cdot Bits_{motion}$ 。该准则包含两部分因素：

SAD 表示当前宏块与前后向匹配块之间绝对误差和（sum of absolute difference）的平均值，用来衡量运动矢量的准确性。对于双向连接的DirL, DirInv, FT_BDL, BT_FDL和Bid模式， SAD 取前后向的平均值，对单向连接的Fwd和Bwd模式， SAD 取未被截断一端的值。Fwd和Bwd宏块中的像素代表运动线的边界，对应的小波重构误差大于运动线内部，为了避免截断造成的边界效应影响编码效率，选取参数 η 使判定准则略微倾向于双向模式。

$$\eta = \begin{cases} 1 & \text{DirL, DirInv, FT_BDL, BT_FDL, Bid} \\ 1.5 \sim 2 & \text{Fwd, Bwd} \end{cases} \quad (4-4)$$

$Bits_{motion}$ 表示编码运动矢量时用到的比特数。由于模式的分布特征非常依赖于视频序列以及时轴分解级，每个模式分配的码字是运动矢量编码过程中确定的，因此 $Bits_{motion}$ 项中没有考虑编码模式对应的比特开销。Bid模式中前后向运动矢量都需要传输，因此 $Bits_{motion}$ 统计了两个方向运动矢量的比特数，在其他模式中， $Bits_{motion}$ 只需考虑单个方向的运动矢量。 λ 参数用来控制代价准则中码率对应失真的比例。由于低码率端的率失真斜率比高端大，因此低端的 λ 应该大于高端。然而，由于运动抽线小波视频编码器的非循环式的结构，运动补偿是在码流量化压缩之前完成的，因此 λ 只能被设为固定值。综合权衡了码率高端和低端的性能，在系统实现中， λ 设为16。

对于每个宏块，选择对应代价函数取值最小的模式。运动线的组成根据对应的宏块模式和运动矢量来决定。对不同视频序列的每个时轴分解级，在运动估计过程中统计各种模式的个数，并根据模式的概率分布训练Huffman码表。之后将每个宏块选择的模式根据码表进行Huffman编码。每一时轴分解级对应的Huffman码表都写入相应压缩码流的头信息（header）中，最多占据64个bits。

4.2.4 实验结果与分析

实验结果是在6个MPEG标准的CIF格式(352x288)视频序列上进行的。所有的视频序列都包含300帧图片,帧率为30Hz。我们比较了采用本节提出的高级提升型运动抽线技术的编码器A-MTh同三种基准编码器:原始运动抽线(original MTh), MC-EZBC^[80], 和H.26L^[12]的jm17版本编码器的压缩性能。MC-EZBC的结果来源于^[81]中汇报的结果。在A-MTh和original MTh编码器中,视频序列没有被分成图片组(group of pictures, GOP),而是当作整个数据块利用帧缓存结构进行滤波。由于4层分解的最低频子带的帧间跨度为16,可以看作是大小为16的GOP结构。时轴滤波后的系数帧在帧内做3级的二维小波分解,并采用第2章中的熵编码和率失真码流组装方法进行压缩编码。在A-MTh中,运动矢量为1/4像素精度,并对应不同分解级的帧率将运动搜索的大小分别定为:最高级32像素,第二级64,其余级128。在H.26L编码器中,GOP结构设为只有第一帧为I帧(intra frame),并在每两个P帧之间插入3个B帧。运动搜索精度为1/4像素,范围为32像素。CABAC和率失真优化都是选中状态。

第一部分实验比较了几种编码器的压缩质量。从图 4-10和图 4-11可以看出,原始运动抽线技术original MTh对Coastguard序列达到了和H.26L近似的压缩性能,但是对其他序列,压缩效果很不理想,比H.26L低了约2.0~5.0dB。主要原因在于大量断线造成的边界效应以及不足的整数运动精度。本节提出的高级运动抽线技术 A-MTh对于Mobile将编码效率提高了约5.0dB,并且达到了与H.26L近似的编码效率。对Mobile序列的提高主要归功于细化的运动精度,因为该序列拥有较多的细节信息。对运动比较复杂的Foreman和Stefan序列,original MTh产生的大量断线影响了编码效率,采用提升型结构的A-MTh避免了断线造成的边界效应,编码效率提高了3.2~4.0dB。对于这两种序列,A-MTh的性能还是比H.26L低0.5~1.3dB,主要因为运动抽线对复杂运动的对齐还不够精确。但是我们应该注意到,作为嵌入式小波编码器,A-MTh可以提供帧率和压缩质量的可伸缩性,这是H.26L的单层码流所不具备的。对于Coastguard,A-MTh进一步提高了大约1.5dB,并且超出H.26L编码器 1.3dB。实验结果表明改进后的A-MTh效果十分显著。

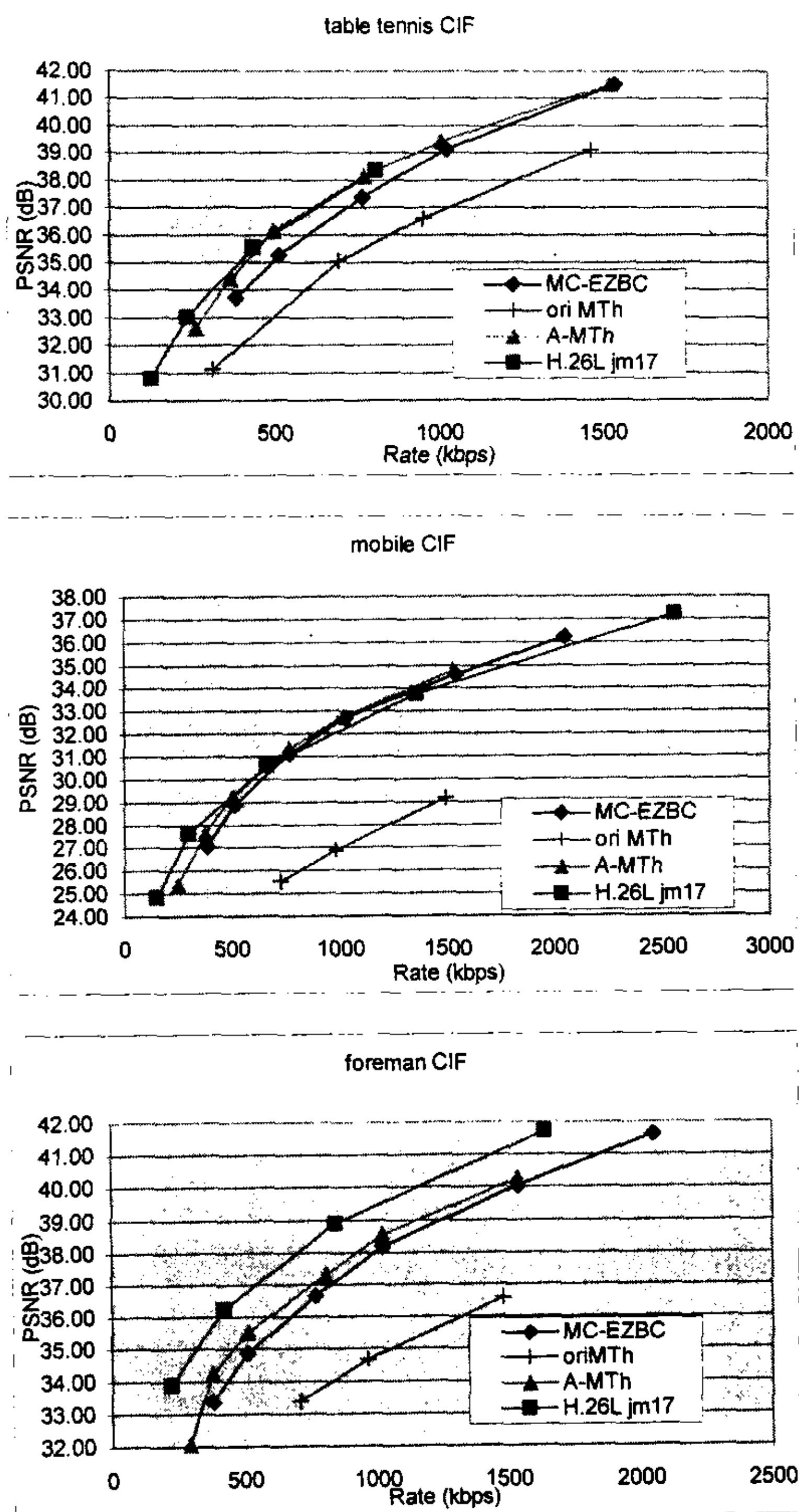


图 4-10 实验比较结果 (Comparisons among MC-EZBC, H.26L, original MTh and A-MTh.)

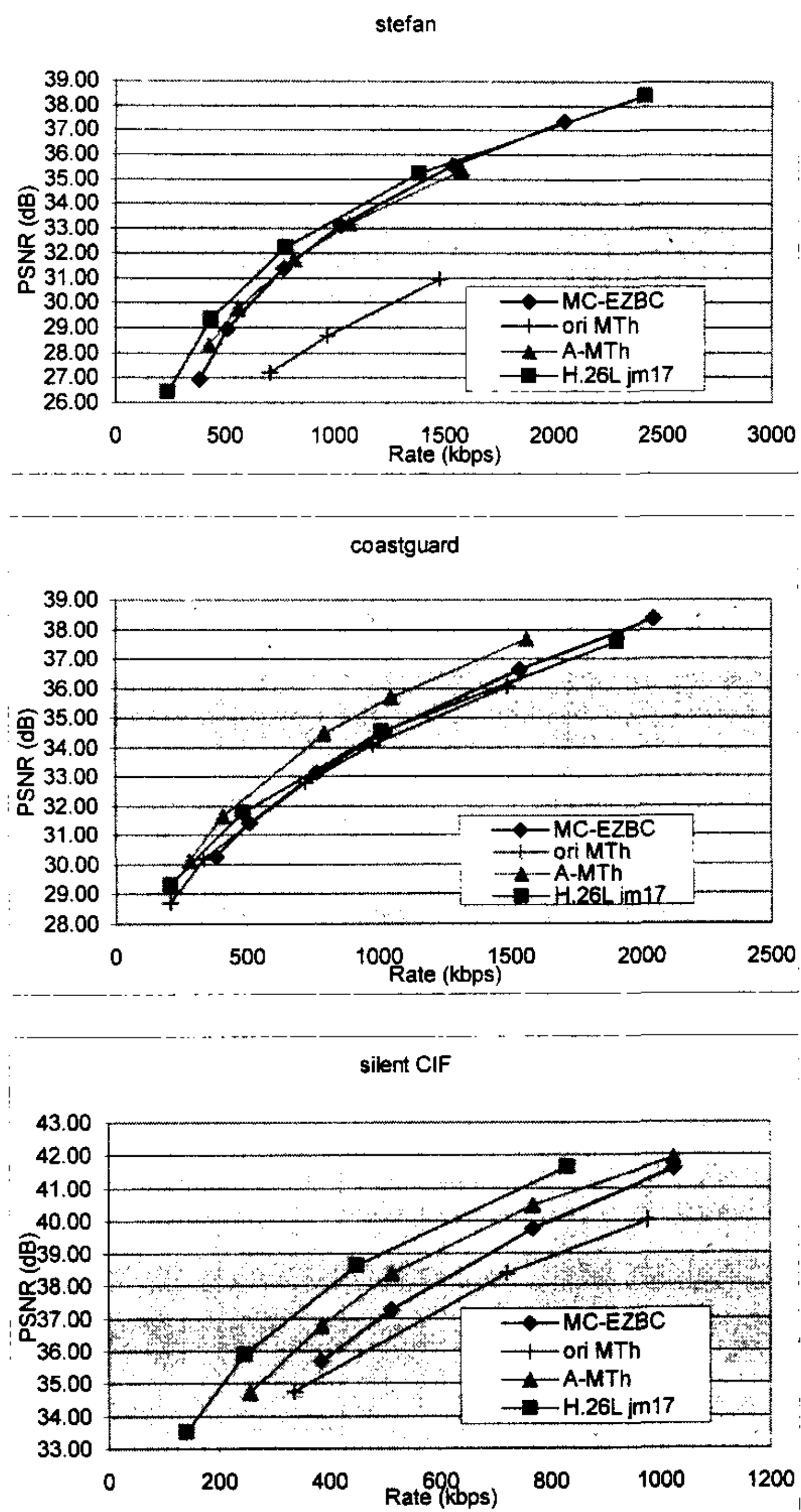


图 4-11 实验比较结果, 接上页 (Comparisons among MC-EZBC, H.26L, original MTh and A-MTh, cont.)

在第二部分实验中，我们分析了4.2.3小节中考虑帧间和级间运动矢量相关性对节省运动信息所占码率达到的功效。首先我们统计了7种运动模式的分布。在最低的时轴分解级Lowest Layer4里，由于不存在更低帧率级作预测，因此没有用到DirL, FT_BDL和BT_FDL这三种模式。

表 4-2 给出了三个测试序列对应的每层的模式分布。可以看出，在最低帧率级，Bid模式占据主导地位，这是由于4级分解的最低级的帧间跨度有 $2^{4-1} = 8$ 帧，因此帧间运动矢量的相关性很低。随着分解级的升高，运动矢量的相关性逐渐增强，模式的分布也逐渐偏向表征帧间相关和级间相关的类型。在最高帧率级中，完全不用传输运动矢量的DirL模式占据了大约或超过50%的比例。对于包含比较复杂的面部表情运动和产生遮挡的摇镜头运动的Foreman序列，表征运动相关性的模式中的不均匀运动模式FT_BDL和BT_FDL出现的次数较其他序列频繁。对于运动比较规则的Mobile序列，模式分布则主要集中在DirL和DirInv这两种类型。Coastguard的模式分布介于Foreman和Mobile之间，由于其具有水波纹理等运动估计无法准确定位的特征信息，使得模式的分布相对分散。

表 4-2 不同时轴分解级对应的模式分布百分比 (Mode distribution percentage according to the temporal decomposition layers)

Mode Type	Lowest Layer4	Layer 3	Layer 2	Highest Layer 1
DirL	0	7.50	17.66	45.95
FT_BDL	0	34.32	34.42	20.32
BT_FDL	0	17.42	16.70	11.27
DirInv	10.44	4.89	5.57	9.25
Fwd	3.44	2.04	1.22	0.41
Bwd	5.79	2.23	1.05	0.30
Bid	80.33	31.60	23.39	12.51

a) Foreman (%)

Mode Type	Lowest Layer4	Layer 3	Layer 2	Highest Layer 1
DirL	0	37.07	64.11	71.07
FT_BDL	0	20.91	6.92	0.61
BT_FDL	0	7.90	3.05	0.43
DirInv	18.24	11.66	18.65	27.17
Fwd	1.99	0.65	0.12	0.03
Bwd	0.58	0.28	0.10	0.02
Bid	79.19	21.52	7.05	0.67

b) Mobile (%)

Mode Type	Lowest Layer4	Layer 3	Layer 2	Highest Layer 1
DirL	0	11.66	30.10	64.87
FT_BDL	0	24.09	21.20	12.18
BT_FDL	0	10.82	8.91	5.03
DirInv	19.29	23.86	19.96	11.02
Fwd	2.92	0.66	0.21	0.04
Bwd	1.23	0.97	0.26	0.03
Bid	76.58	27.95	19.37	6.82

c) Coastguard (%)

接下来我们比较了综合考虑帧间和级间相关性的运动估计和只考虑帧间相关性的运动估计这两种方法的性能差别，由此观察级间相关性对节省码率的贡献。实际的码率节省量如表 4-2所示。IntraLayer表示在运动估计中只用到了不考虑级间相关性的宏块模式，即DirInv, Fwd, Bwd和Bid。CME表示综合的运动估计，包含了所有的7种宏块模式。总的码率包括了编码运动矢量用到的码率以及表示宏块模式的码率。对每个视频序列的每个时轴分解级，宏块模式都是利用表 4-2中的分布百分比进行 Huffman编码的，并且相应的Huffman记录在每级的码流头信息中。

表 4-3 运动矢量码率的节省量 (Motion vector bits reduction)

		Foreman	Mobile	Coastguard
IntraLayer only (Kbps)	Motion rate	96.5	49.4	69.2
	Mode rate	15.0	12.5	14.7
	Total rate	111.5	61.9	83.9
CME (Kbps)	Motion rate	64.0	24.2	43.3
	Mode rate	24.1	16.9	22.0
	Total rate	88.2	41.1	65.3
Total Motion Reduction	(Kbps)	23.3	20.7	18.6
	Percentage	20.9%	33.5%	22.2%

在同等PSNR结果的情况下，表 4-3显示CME对Foreman的码率节省为20.9%，Coastguard为22.2%，Mobile为33.5%。Mobile中的运动信息比较规则，因此多数宏块都采用了不必传输运动信息的DirL模式，使得运动码率节省量较大。在384Kbps的低码率端，Mobile的运动码率节省量大约折合为0.3~0.4dB的PSNR提高。目前关于级间相关性的研究展示的是初步的结果，细致的对级间相关性的研究工作还有望取得更多的改进。

4.3 本章小结

本章研究了三维小波视频编码中运动抽线技术的优点以及影响压缩质量的断线问题，并进行了两项关键的改进工作。第一项工作通过在抽线前的运动估计过程中考虑重叠像素的数目来权衡运动矢量精度和断线情况的出现，提出了两轮运动估计方法TEMTH。实验结果表明对于传统运动抽线无能为力的运动复杂的视频序列，TEMTH将压缩质量提高了高达1.8dB，达到了与MPEG-4近似的性能。同时保持了以往对简单视频序列超出MPEG-4的优秀的压缩效率。第二项工作结合了第三章中的运动补偿提升型思想，提出了高级运动抽线技术，将运动抽线与提升型结构结合在一起，彻底解决了断线问题，并且进一步加入了1/4像素运动精度。在高级运动抽线技术中还探讨了视频帧中的遮挡问题，以及利用运动矢量的相关性，并采用率失真优化的运动估计准则，在提高运动对齐的匹配准确度和精度的同时，节省了运动矢量的编码开支。实验结果表明采用高级运动抽线技术的小波编码器达到了与最先进的H.26L标准相当的编码效率。

第五章 三维小波视频编码中的时轴质量抖动问题研究

在三维小波视频压缩的研究中,我们发现采用不同小波滤波器的不同编码方法中,都存在解压缩后视频帧质量沿时轴方向有周期性的抖动。要深入探讨这个问题,首先要从小波视频压缩中的码率控制(rate control)角度来分析。视频压缩过程中的码率控制是可以概括为:如何决定合适的编码参数,如量化步长等进行比特分配(bit allocation),使得在给定的码率下得到相对最佳的解码质量。在三维小波编码中,由于嵌入式的码流可以很容易的进行码流截断,因此码率控制的问题可以简化为如何决定小波各子带间的码率分配。在三维小波SPIHT编码方法中^[71],子带间的系数被组织成三维空间树状结构按位平面进行编码,码率控制是在熵编码过程中隐含实现的;在第二章的三维扩展嵌入式分块熵编码方法中,小波系数被分为三维的子块并进行独立的熵编码,码率控制则是根据各子块的率失真曲线进行全局的率失真优化。以上两种三维小波编码方法都是从整个视频解码全局优化的角度来考虑的,因此没有关注时间方向上帧质量平稳的问题。我们观察发现,这两种三维小波编码方法重构的视频信号在时轴上存在着峰值信噪比PSNR幅值达到0.5-2.0dB的压缩质量抖动,尤其是在运动较大的视频中。经过分析,我们发现这种时轴上的抖动是由小波合成的结构导致的。

本文中我们研究了小波系数与重构信号之间信号失真的分布关系,解释了时轴质量抖动的原因,推导出了使时轴解码性能平稳的条件。在此基础上,我们提出了一种新的三维小波视频编码率控制技术,能够对解码视频在帧内空域上进行全局优化,同时具有平稳的时轴性能。

5.1 三维小波视频编码中的周期性时轴解码质量抖动现象

5.1.1 小波系数与重构信号之间失真的分布关系

小波重构视频中的时轴帧的质量抖动实际上是时轴帧的压缩失真大小的抖动。从小波反变换的结构可以得知,重构信号(像素或视频帧)由于位置的不同会对应不同的反变换滤波器。因此,对于 l 级的小波分解,重构的信号在时轴上会有周期为 2^l 的质量抖动。从图 5-1的a)中可以看到明显的解码视频质量的周期抖动现象。事实上,这种周期的抖动不仅存在于时轴方向,同时也存在于帧内的水平和垂直方向,但由于帧内像素间

质量的抖动不会像帧间那样有明显的视觉干扰。因此我们的研究将重点分析时轴方向上一维小波变换的失真分布。

首先以一维的一级小波分解为例。令 $\hat{x} = \{\hat{x}_0, \hat{x}_1, \dots, \hat{x}_{N-1}\}$ 表示有效长度为 N 的信号, $y = \{y_0, y_1, \dots, y_{N-1}\}$ 表示信号经由双正交小波变换后得到的小波系数。令小波正变换的低通滤波器为 $h[k]$, $k \in [-N_h, P_h]$, 高通为 $g[k]$, $k \in [-N_g, P_g]$ 。则小波分解过程表示为:

$$y[2n] = \sum_{k=-N_h}^{P_h} h[k] \cdot x[2n+k] \quad \text{和} \quad y[2n+1] = \sum_{k=-N_g}^{P_g} g[k] \cdot x[2n+1+k] \quad (5-1)$$

$\hat{x} = \{\hat{x}_0, \hat{x}_1, \dots, \hat{x}_{N-1}\}$ 表示由小波系数反变换得到的重构信号。由此得出小波合成的过程可以表示为^[28]:

$$\hat{x}_n = \sum_i y_{2i} \cdot h'_{n-2i} + y_{2i+1} \cdot g'_{n-(2i+1)} \quad (5-2)$$

其中 h'_k 和 g'_k 为小波合成对应的基, 与正变换的滤波器关系为:

$$h'[k] = \alpha \cdot g[-k] \cdot (-1)^k \quad \text{和} \quad g'[k] = \alpha \cdot h[-k] \cdot (-1)^k \quad (5-3)$$

对于双正交的对称滤波器, 有 $h'_k = h'_{-k}$ 和 $g'_k = g'_{-k}$ 。

令 e_n 表示小波系数 y_n 的压缩误差, \hat{e}_n 表示重构信号 \hat{x}_n 与原信号的误差, 重构误差与小波系数误差的关系可以由公式 (5-2) 得到:

$$\hat{e}_n = \sum_i e_{2i} \cdot h'_{n-2i} + e_{2i+1} \cdot g'_{n-(2i+1)} \quad (5-4)$$

在中高码率压缩的情况下, 可以假设在每个小波子带中, 误差 e_i 具有独立同分布的特点。因此, 不同奇偶位置的重构信号和小波系数间的失真 (平方误差) 的关系可以由公式 (5-4) 计算得出:

$$\begin{cases} \hat{\sigma}_{2n} = \|e_{2n}\|^2 = (\delta^L \sum_k h'^2_{2k} + \delta^H \sum_i g'^2_{2k+1}) \\ \hat{\sigma}_{2n+1} = \|e_{2n+1}\|^2 = (\delta^L \sum_k h'^2_{2k+1} + \delta^H \sum_i g'^2_{2k}) \end{cases} \quad (5-5)$$

其中 δ^L 和 δ^H 分别是低频带和低频带小波系数误差对应的方差。

5.1.2 潜在的周期性时轴质量抖动

上一节分析了不同位置上重构信号和小波系数间失真的关系，下面我们给出一个简单的例子来说明三维小波视频解码中潜在的周期性时轴抖动。

对于均匀量化的三维小波视频编码器，不同时轴子带间的失真是相等的，即：

$$\delta^L = \delta^H = \delta \quad (5-6)$$

不考虑边界上的情况，由等式（5-5）得出：

$$\frac{\hat{\delta}_{2n}}{\hat{\delta}_{2n+1}} = \frac{\sum_k h_{2k}^2 + g_{2k+1}^2}{\sum_k h_{2k+1}^2 + g_{2k}^2} \quad (5-7)$$

对于典型的Daubechies 双正交9-7 小波滤波器一层分解，不同奇偶位置的重构信号的失真比例为 $\hat{\delta}_{2n} : \hat{\delta}_{2n+1} = 0.911073 : 1.112389$ 。相应的峰值信噪比PSNR的差为

$$10 \log \frac{\hat{\delta}_{2n+1}}{\hat{\delta}_{2n}} = 0.867 \text{ dB}$$

从以上分析可以看出解码的视频质量与其所在位置有很大的关系。因而由于不同位置上小波反变换的滤波器不同，三维小波视频编码中存在着周期的时轴质量抖动。

5.2 具有平稳时轴质量的码率控制

为了解决三维小波视频编码中潜在的时轴抖动，接下来我们提出了一种新的码率控制方法，并将其应用于三维小波视频编码。本实验基于第二章中用在同心拼图压缩的无运动三维小波编码器来进行分析。在采用的基于分块熵编码技术的编码器中，视频帧经由水平、垂直和时间三个方向的小波变化后，得到的每个子带被进一步分成固定大小的块。每个子块采用位平面技术分别进行独立的熵编码，并得到各块的率失真曲线。在熵编码之后的码流装配过程中进行码流分配得到给定的长度的码流。

原始的码率控制方法是根据率失真曲线来对所有块的码流进行相应的截断得到整个码流，当每个块的率失真曲线截断点的斜率相同时，小波系数的失真度达到理论上的全局最优。然而，如前所述，这种全局最优的码率控制导致了帧方向上质量的抖动。因此，在本章提出的码率控制方法中，结合了率失真优化和帧方向上质量平稳两方面的要

求，在沿帧方向上采用了新的码率分配方法，同时在沿水平和垂直方向上保持了率失真优化的方法。也就是对于同一时轴小波子带的分块，不论其位于任何的水平和垂直小波子带，都采用同一斜率来选取截断点；对于不同时轴小波子带的分块，按照不同的斜率来截断以调整时轴方向的平稳性。下面对时轴码率分配原则进行详细的理论推导。

5.2.1 时轴平稳条件下的理论子带失真比例

从5.1的分析中得知，对一级小波分解来说，时轴平稳性表示重构信号不同位置的失真应满足相等的条件：

$$\hat{\delta}_{2n} = \hat{\delta}_{2n+1} = \hat{\delta} \quad (5-8)$$

其中 $\hat{\delta}$ 代表重构信号的平均失真。

根据等式 (5-5) 得到：

$$\frac{\delta^L}{\delta^H} = \frac{\sum_k g'^2_{2k+1} - g'^2_{2k}}{\sum_k h'^2_{2k+1} - h'^2_{2k}} \quad (5-9)$$

该式给出了要保证时轴性能 (PSNR) 平稳，小波系数的失真需要满足的比例。

不失一般性，对于 l 级的一维小波分解，最多有 $m=2^l$ 个子带。和公式 (5-2) 类似，令 $b = \{b^0, b^1, \dots, b^{m-1}\}$ 表示相应的反变换矩阵，此处 b^i 为小波子带 i 对应的小波合成基向量。则小波反变换的过程可以表达为：

$$\hat{x}_n = \sum_k \sum_{i=0}^{m-1} y_{mk+i} \cdot b^i_{n-(mk+i)} \quad (5-10)$$

根据每个时轴子频带间独立同分布的特性，即对同一子带内系数 y_{mk+i} 和 y_i 可以假定 $\hat{e}_{mk+i} = \hat{e}_i$ 。将重构信号的位置表示为 $n = a \cdot m + j$ ， a 为任意整数。则重构信号的失真与小波系数的失真之间有如下关系：

$$\begin{aligned} \hat{\delta}_n = \hat{\delta}_{am+j} &= \sum_{i=0}^{m-1} \sum_k \hat{\delta}_{mk+i} \cdot b^i_{(a-k)m+j-i} \\ &= \sum_{i=0}^{m-1} \hat{\delta}_i \sum_{k'} b^i_{mk'+j-i} \end{aligned} \quad (5-11)$$

其中 $j = 0, \dots, m-1$ 。

从(5-11)可以看出, 重构信号的失真是和位置有关的, 对于 $n = a \cdot m + j$ 位置的系数, 重构误差只与 j 有关, 定义为 $\hat{\delta}_j$ 。令 $\hat{\Delta} = \{\hat{\delta}_0, \hat{\delta}_1, \dots, \hat{\delta}_{m-1}\}$, $\Delta = \{\delta_0, \delta_1, \dots, \delta_{m-1}\}$ 和 $S = \{\sum_k b_k^2 \delta_{mk+j-i}\}$ 分别代表重构信号的失真向量, 小波系数的失真向量以及小波反变换的矩阵, $I = \{1, 1, \dots, 1\}_m$ 代表单位向量, 则得出 I 级时轴小波分解下的时轴平稳条件:

$$\hat{\Delta} = S \cdot \Delta = I \quad (5-12)$$

如果 S 为可逆矩阵, 等式(5-12)有且有一解向量 Δ 来决定小波域的失真比例。在视频压缩中, 时轴的分解一般不超过4级。计算得出对于图像编码中常用的滤波器, 当 $1 \leq 4$ 时, S 为可逆矩阵。并且可以证明当小波包分解的频带数 $< m$ 时, 等式(5-12)可以降低维数而解得所需的子带失真比例。

5.2.2 时轴平稳的三维小波视频编码码率控制

基于第二章中的无运动三维小波编码器, 对位于同一时轴小波子带的分块, 不论其位于任何的水平和垂直小波子带, 都采用同一斜率来选取截断点; 对不同时轴小波子带的分块, 根据等式(5-12)给出的失真比例条件来调整截断点的斜率。

每个子带分块的率失真R-D性质可以近似用指数衰减的形式来描述^[104]:

$$D^n = D_{\max}^n 2^{-\beta R^n} \quad (5-13)$$

此处 D^n 为块 B^n 的编码失真, D_{\max}^n 为码率 $R^n = 0$ 时的编码失真。参数 β 为失真的衰减率。则R-D 曲线的斜率为:

$$\lambda^n = \frac{\partial D^n}{\partial R} = -\beta^n \cdot D_{\max}^n 2^{-\beta R^n} \cdot \ln 2 = -D^n \cdot \beta^n \cdot \ln 2, \quad (5-14)$$

等式(5-14)表明了R-D曲线的斜率 λ^n 与该点失真 D^n 呈线性关系。

在中高码率环境下, 各时轴子带的衰减系数 β^n 近似相等^[104]。因此可以简单的依照如下公式, 通过控制各时轴子带的率失真斜率来控制失真的比例:

$$\lambda^0 : \lambda^1 : \dots : \lambda^{m-1} = \delta_0 : \delta_1 : \dots : \delta_{m-1}, \quad (5-15)$$

此处 δ_i 为等式(5-12)给出的理论失真值, m 为时轴的小波子带个数。

采用这种斜率—失真的近似估计方法目的是可以比较简单的实现，不需要迭代的调整子带间的失真比例。在将来的工作中，可以设计更为精确的比例控制方法来精确的调整失真比例。

5.3 实验结果与分析

实验结果给出了采用本文提出的时轴平稳的码率控制方法的三维小波编码器 *TS-EBCOT* (temporally smooth EBCOT) 的编码性能。首先根据公式 (5-12) 计算给出了不同小波滤波器在不同分解方式下，小波子带失真之间的理论比例关系 (表 5-1)。从 Le Gall 5-3 和 Daubechies 9-7 滤波器的比较可以看出，对于 5-3 滤波器，为了满足失真均衡的条件，需要给低频子带更大的权重。

接下来，我们比较了 *TS-EBCOT* 与其他两种编码器 *3D SPIHT* 和 *3D EBCOT* 的性能。在 *TS-EBCOT* 和 *3D EBCOT* 中，视频数据先经过时轴的 3 级小波分解，再经过水平和垂直方向的 3 级分辨率小波分解得到小波系数；在 *3D SPIHT* 中，视频信号经过三个方向的 3 级分辨率小波分解。实验的视频数据是 MPEG-4 标准的 Foreman, Coastguard 和 Akiyo 序列，分别压缩到 0.4 比特/像素，0.2 比特/像素和 0.05 比特/像素。根据表 5-1，我们得出时轴平稳要求下三级 Daubechies 9-7 滤波器的理论时轴 R-D 斜率比例为 $\delta^{LLL} : \delta^{LLH} : \delta^{LH} : \delta^H = 2.5998 : 1.4816 : 0.8570 : 0.4947$ ，并在 *TS-EBCOT* 中应用这个比例来控制时轴率失真的截断斜率。

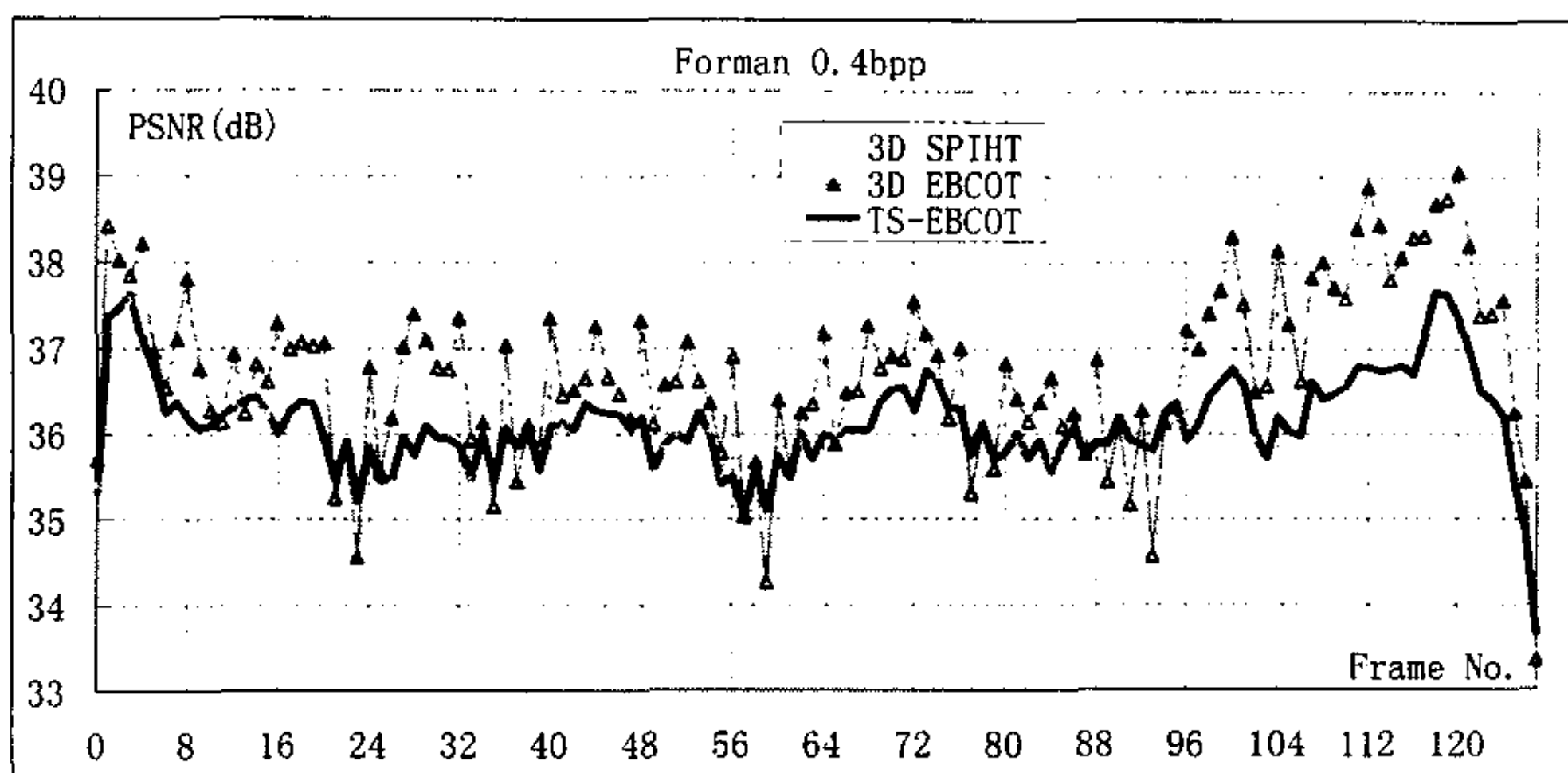
表 5-1 Biorthogonal Le Gall 5-3 和 Daubechies 9-7 滤波器，三层以下小波分解的理论子带由低频到高频失真比例 (Theoretical distortion distribution ratio for the wavelet decomposition less than 3 layers, from the lowest pass to the highest one)

	Le Gall 5-3 filter	Daubechies 9-7 filter
1 layer	1.7647:0.4706	1.2685:0.7230
2 layers	3.2258:0.8602:0.3441	1.7623:1.0045:0.5811
3 layers	6.0362:1.6097:0.6439:0.2575	2.5998:1.4816:0.8570:0.4947

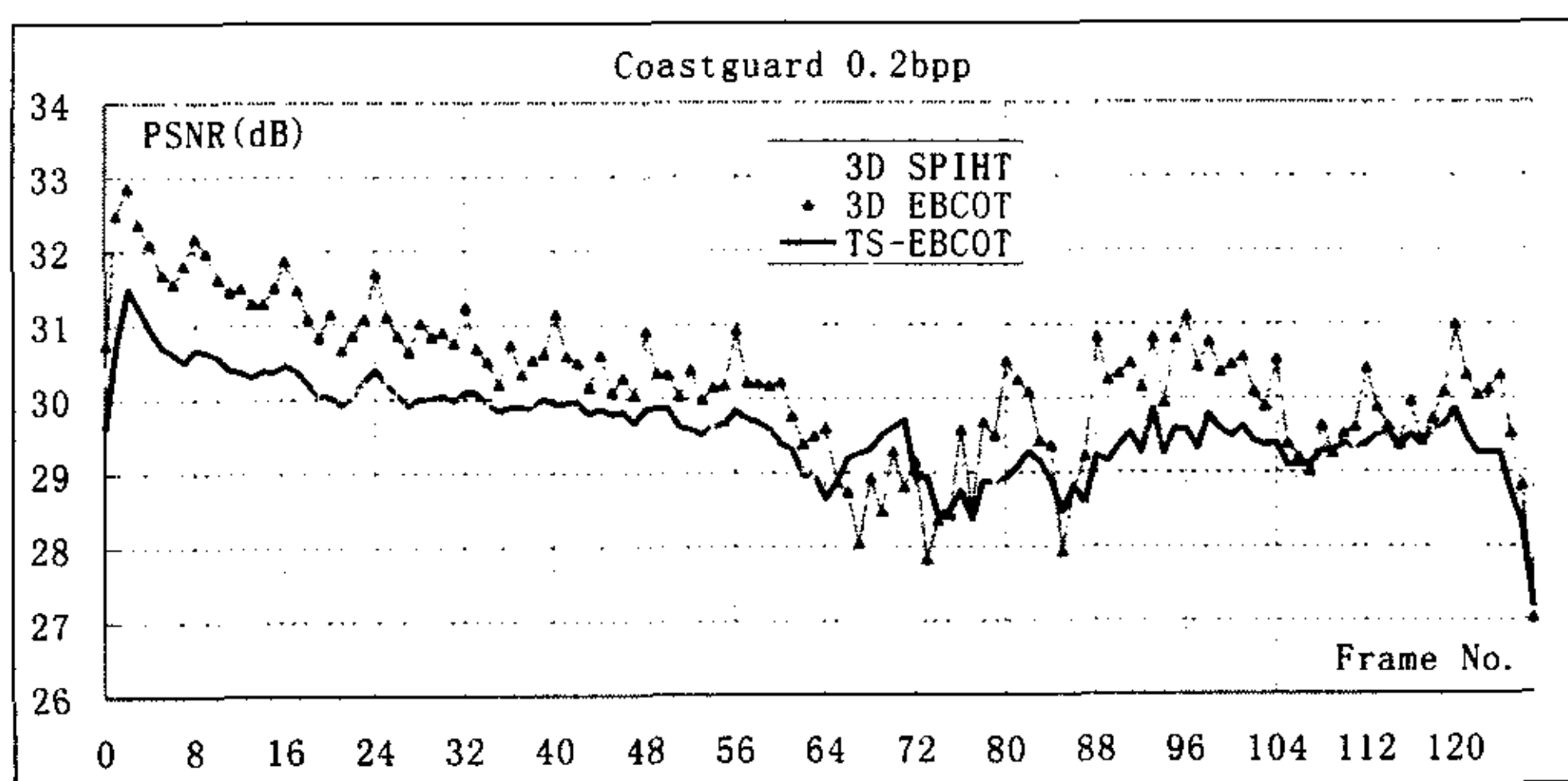
图 5-1 给出了三种编码器的峰值信噪比 PSNR 对帧序号的曲线。对 Foreman 和 Coastguard 序列，可以看出 *3D EBCOT* 和 *3D SPIHT* 编码器都有周期为 $2^3=8$ 的时轴 PSNR 抖动。采用了本文提出的码率控制方法的 *TS-EBCOT* 大大降低了周期的时轴抖动，具有较平稳的性能。从平均 PSNR 来看，*TS-EBCOT* 同样具有与 *3D SPIHT* 相近的性

能。虽然由于3D EBCOT是对整个视频全局优化的,使得TS-EBCOT 的平均PSNR比3D EBCOT 低约0.5dB,但TS-EBCOT 提供的平稳的时轴质量使得用户具有更好的视觉感受。同时,从一些3D EBCOT 和 3D SPIHT编码质量较差的一些单帧来看,TS-EBCOT 的PSNR提高了0.5-1.0dB。

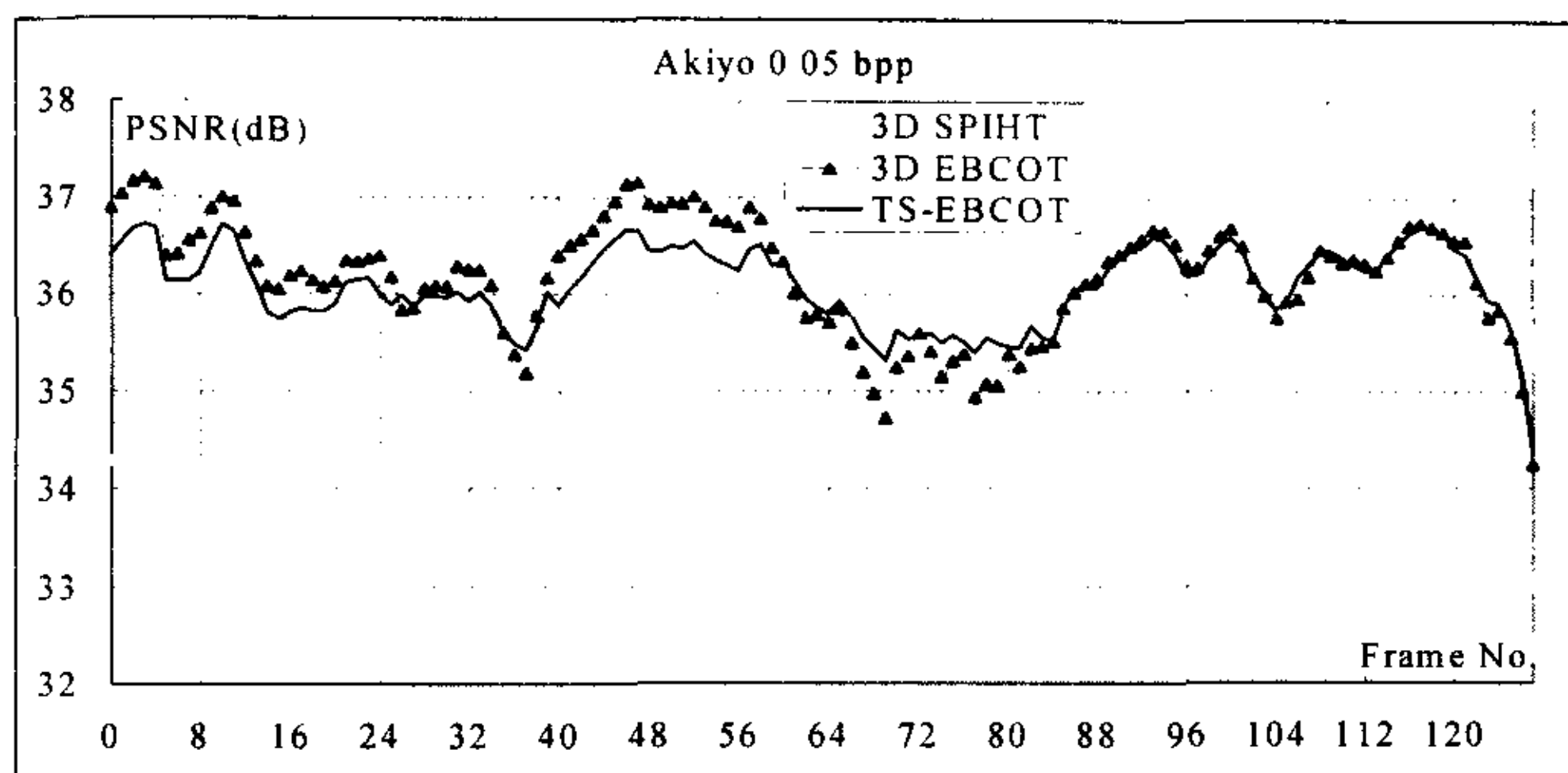
对于Akiyo 序列,由于大部分区域的运动很少,时轴上的抖动不是很明显。图 5-1c) 显示在这种低运动的情况下, TS-EBCOT 的平均压缩质量与3D EBCOT相近,并高出3D SPIHT 约2dB。



a)



b)



c)

图 5-1 结合本文提出的码率控制方法的 TS-EBCOT 视频编码器与 3D SPIHT 和 3D EBCOT 对 MPEG-4 标准序列的压缩性能比较: a) Foreman 序列, 0.4 比特/像素, 平均峰值信噪比 PSNR(dB): TS-EBCOT: 36.1, 3D SPIHT: 36.3, 3D EBCOT: 36.7; b) Coastguard 序列, 0.2 比特/像素, 平均 PSNR(dB): TS-EBCOT: 29.6, 3D SPIHT: 29.7, 3D EBCOT: 30.3; c) Akiyo 序列, 0.05 比特/像素, 平均 PSNR(dB): TS-EBCOT: 36.1, 3D SPIHT: 33.9, 3D EBCOT: 36.2。

5.4 本章小结

本章从小波反变换过程对应的滤波器结构的角度研究了小波系数压缩失真与重构信号失真之间的关系, 解释了传统三维小波编码器时轴的周期性解码质量抖动的问题。并且提出了一种新的具有时轴平稳性的码率控制方法。基于各子块率失真 R-D 的指数衰减模型, 本章提出的方法可以通过控制各小波时轴子带的失真比例来分配码率, 均衡控制解码后时轴信号的失真。实验结果表明, 结合了时轴平稳码率控制的三维小波编码器压缩的视频具有良好的时轴平稳性, 因而更加适合用户的视觉要求。同时, 对于以往小波编码中质量较差的一些帧, 压缩质量提高了约 0.5-1.0dB。

第六章 总结与展望

图像视频类媒体的压缩技术日新月异,除了获得更高的压缩效率之外,为了适合网上实时传输、播放等应用,对压缩方法的可伸缩性,易于容错纠错等方面的要求逐渐提高。基于小波变换良好的时频局域性和多分辨率分析能力,小波图像压缩可以得到传统基于 DCT 的压缩方法所不具备的许多功能,例如多分辨率解码,清晰度渐进的解码,基于视觉权重的量化以及感兴趣区域编解码等功能,更加适合于当今网络以及卫星遥感图像的传输和应用。

本文在目前小波静止图像压缩中取得成功的基础上,研究小波在高维图像视频媒体压缩中的应用。主要从针对场景图像的小波压缩以及小波视频压缩两个方面来探讨小波在高维图像媒体中的压缩技术,主要的研究工作和创新点如下:

首先,针对同心拼图这种基于图像的场景绘制技术 IBR 的特点,以及对压缩码流的随机访问解码要求,设计了基于三维小波的嵌入式压缩编码系统,为同心拼图的网上传输和浏览等应用提供了灵活可靠的编解码方案。由于同心拼图等 IBR 媒体是比较新兴的图形学媒体,直接应用普通的视频压缩标准不能很好的满足其对于压缩和随机解码的需求,我们的工作是该领域中较早开发并且具有高效压缩性能的编码系统之一;

其次,考虑同心拼图压缩场景的即时绘制特点,我们提出了一种渐进逆向小波合成快速算法 PIWS,有效的支持了压缩码流的实时解压绘制。PIWS 快速小波合成方法不但支持了对同心拼图压缩码流的实时解码绘制,还可以普适的应用到小波压缩的视频等媒体的快速解码应用中,为降低小波解码过程的运算量提供了一种有效的解决方案;

再次,针对传统的三维小波视频压缩中对时轴相关性发掘的不足。我们提出了基于运动补偿的提升型小波(motion-compensated lifting, MCLIFT)方法实现时轴变换。通过在时轴小波变换的提升型单元结构中采用双向的块运动补偿, MCLIFT 可以结合帧间运动信息,很好的对视频帧进行去相关。利用提升型算法的单元结构的运算特点, MCLIFT 方法可以在保证小波反变换完全重构(perfect reconstruction)的同时,利用许多现有的运动补偿技术,例如 1/2 像素运动精度和重叠块运动补偿技术,提高了时轴变换的效率。实验结果表明采用 MCLIFT 方法的小波视频编码器的压缩效率远远超出没有采用运动估计的三维小波压缩方法,并且达到了优于 MPEG-4 标准的水准,同时还具有 MPEG-4

所不具备的帧率和压缩质量的可伸缩性。此项工作在国际上同期的小波视频压缩研究中属于最早采用提升型结构来引入运动信息的几个工作之一。

第四,研究了三维小波视频压缩中的运动抽线技术(motion threading),仔细分析了影响其压缩质量的断线问题。并且设计了两轮运动估计技术,通过在运动估计准则中加入对重叠像素的统计,有效减少了断线的出现。进一步结合提升型小波结构,提出了高级运动抽线技术,从根本上避免了断线的发生。利用提升型结构,我们将运动抽线的精度由传统限制的整像素精度提高到 $1/4$ 像素精度,有效的提高了时轴小波变换的去相关能力。在高级运动抽线中还分析了视频中常有的遮挡问题,以及运动矢量的冗余性,利用基于率失真的准则进行运动估计,在改善估计准确度的同时节约了运动矢量的开销。本文提出的高级运动抽线技术使小波视频的压缩效率提高到了与最先进的 H.26L 编码标准相接近的水准;

最后,对三维小波视频压缩中的时轴周期性抖动现象进行了理论的分析。由于时轴的质量抖动可能会影响视频播放的视觉效果,这是三维小波压缩的视频应用中的一个重要问题。我们通过理论分析得出,抖动原因在于小波合成滤波对不同位置的系数施加了不同的权重。在此分析的基础上,我们提出了一种使时轴质量平稳的理论上的码率控制方法。采用这种方法的三维小波编码器具有良好的时轴平稳性,同时对某些帧的压缩质量提高了 0.5-1.0dB。目前这方面的分析还只是初步的结果,可以做进一步的研究。

从本文目前的工作看出,小波对高维图像视频媒体的压缩已经显示出其出色的压缩性能,同时提供了码率,帧率,分辨率等多种可伸缩性,在将来的网络传输等方面具有良好的应用前景。对于小波高维图像媒体压缩的研究工作刚刚取得了一定阶段的成果,具体还有许多方向有待进一步探索。除了本文提到的几项工作,以下列举了一些比较具有发展前景的研究方向:

从视频压缩效率看来,目前对于一些运动复杂的视频序列,小波压缩的效率比 H.26L 标准的结果还有一定差距,更进一步的对视频时轴信号分布特性的分析工作应该可以改善这方面的不足,但从过去的经验来看,有一定研究难度。对于数字电影(digital cinema)^{[105][106]}的视频类型,由于具有高清晰度,高分辨率以及高帧率的特点,对小波压缩较为有利,相关工作值得进一步研究;

从小波视频编码的实用性看,文中指出过小波压缩的最大优点在于良好的可伸缩性,但是加入帧间运动补偿后会影响到帧内分辨率的可伸缩性,因此这方面还需要相应的工作来实现。同时,虽然小波压缩视频自身已经具有帧率可伸缩性,但如何评价低帧率

解码时的解码客观质量和视觉效果等工作也是必须的。对于小波压缩计算量较大的问题,文中提出了普适的 PIWS 快速解码算法,针对视频压缩的具体要求的快速小波运算将更快的推动小波的产业化。另外小波压缩码流对不同网络,无线信道传输的码率分配,数据打包,差错控制等工作也是未来长远发展中的可能研究方向。

从小波在高维图像压缩的应用看,在本文所述的同心拼图压缩工作之后,Wu 等又通过智能拼接极大的提高了同心拼图序列的相关性,取得了出色的压缩效率^[107]。由于小波压缩提供的嵌入式编码,选择性随机解码等灵活性,具体的应用方向除了本文所述的同心拼图 IBR 媒体之外,还可以扩展到许多其它的高维媒体数据,例如 IBR 技术的 Lumigraph 数据,以及医学图像的三维体数据(volume data)等。

近年来国际上关于小波高维图像压缩的研究工作一直被热切的关注着,除了各大国际会议和杂志上涌现的许多相关论文,MPEG 小波压缩研究组^[59]的成立进一步体现了业界的重视。虽然小波方法目前有算法复杂度等方面的局限性,但是随着研究的进步和硬件的发展,小波压缩更适于现代通信的可伸缩性和嵌入式等优点将逐步显现出来,让我们继续努力并拭目以待。

附录 I 算术编码的上下文分类

本附录对2.2.3节中的基于上下文的算术编码(context-based arithmetic)中上下文的分类方法做一简要的描述。令 $x(i, j, k)$ 表示 (i, j, k) 位置的系数点, 图 I-1给出了上下文分类时考虑的四种类型的相邻点: 水平相邻点 h , 竖直相邻点 v , 帧间相邻点 n , 以及对角相邻点 d 。对每种类型都分别统计对应的有效系数(significant coefficient)的个数。我们将 h, v, n 和 d 聚类成10种上下文类型用于算术编码中。根据不同小波子带经过的高频分解次数的总和, 采用三个不同的对应表进行上下文分类, 如图 I-1所示。

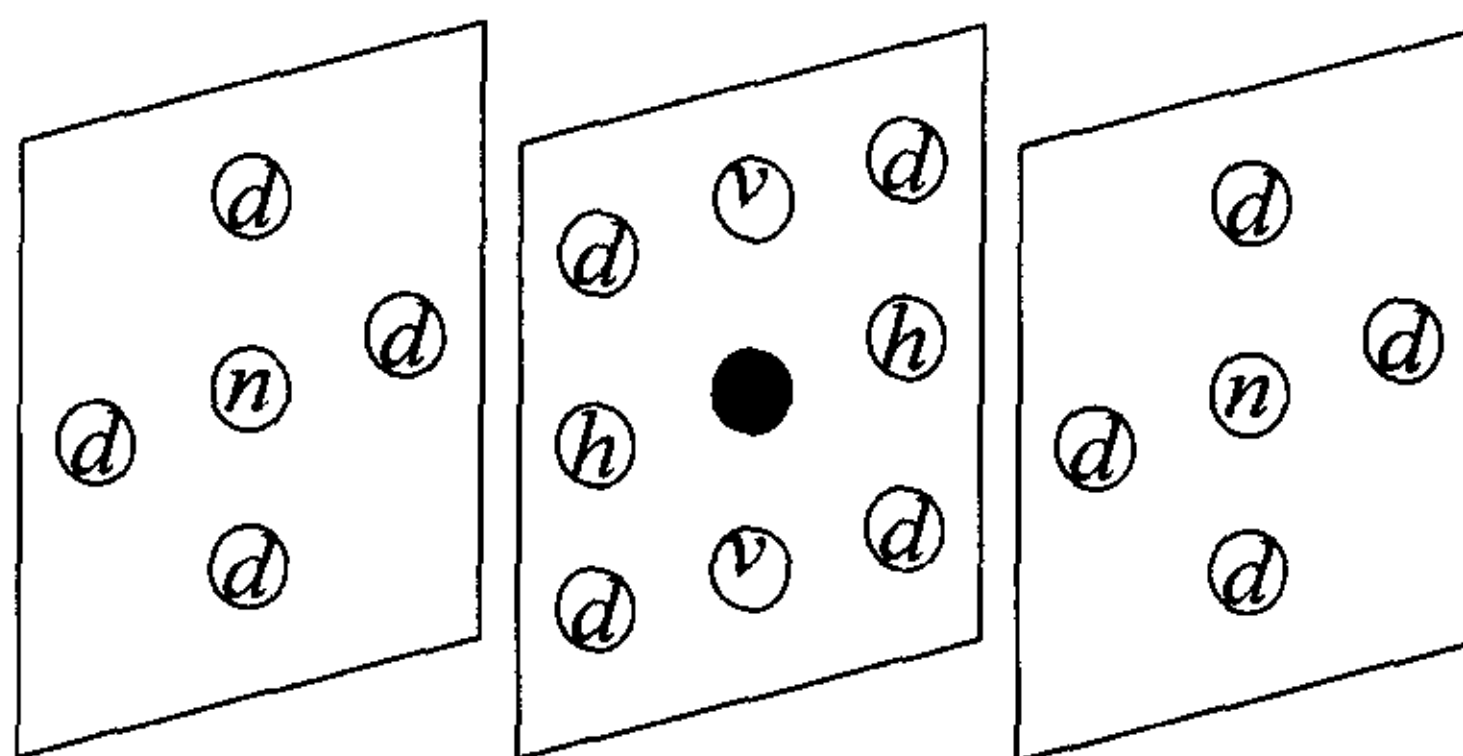


图 I-1 四种类型的相邻点 (Four types of neighbors)

表 I-1 有效性预测中的上下文分类 (Contexts for the significance prediction)

At most one high pass filter					Two high pass filters				Three or more high pass filters		
h	v	n	d	Context	h	$v+n$	d	Context	d	$H+v+n$	Context
2	x	x	x	0	2	x	x	0	6~12	x	0
1	1~2	x	x	0	1	3~4	x	0	4~5	3~6	1
1	0	1~2	x	1	1	1~2	4~12	1	4~5	0~2	2
1	0	0	x	2	1	1~2	0~3	2	2~3	4~12	3
0	2	0	x	3	1	0	4~12	3	2~3	2~3	4
0	1	0	x	4	1	0	0~3	4	2~3	0~1	5
0	0	1~2	x	5	0	3~4	x	5	0~1	4~12	6
0	0	0	3~12	6	0	1~2	4~12	6	0~1	2~3	7
0	0	0	2	7	0	1~2	1~3	7	0~1	1	8
0	0	0	1	8	0	0	4~12	8	0~1	0	9
0	0	0	0	9	0	0	0~3	9			

附录 II 图像视频压缩中常用的小波滤波器

本附录给出了图像视频压缩中常用的一些双正交小波滤波器类型的卷积系数，供读者参考。表中 \bar{h} 表示低通滤波器， \tilde{h} 为相应的重构滤波器，见1.2.1节。数据引自文献[28]。

表 II-1 Daubechie 9-7 卷积型滤波器系数 (Coefficients of the Daubechies 9-7 Filter)

i	\bar{h}	\tilde{h}
0	0.6029490182363579	1.115087052456994
1	0.2668641184428723	-0.5912717631142470
2	-0.07822326652898785	-0.05754352622849957
3	-0.01686411844287495	0.09127176311424948
4	0.02674875741080976	

表 II-2 Le Gall 5-3 卷积型滤波器系数 (Coefficients of the Le Gall 5-3 Filter)

i	\bar{h}	\tilde{h}
0	6/8	1
1	2/8	-1/2
2	-1/8	0

表 II-3 CRF 13-7 卷积型滤波器系数 (Coefficients of the CRF 13-7 Filter)

i	\bar{h}	\tilde{h}
0	164/256	1
1	80/256	-9/16
2	-31/256	0
3	-16/256	1/16
4	14/256	
5	0	
6	-1/256	

表 II-4 SWE 13-7 卷积型滤波器系数 (Coefficients of the SWE 13-7 Filter)

i	\bar{h}	\tilde{h}
0	348/512	1
1	144/512	-9/16
2	-63/512	0
3	-16/512	1/16
4	18/512	
5	0	
6	-1/512	

表 II-5 Daubechie 6-10 卷积型滤波器系数 (Coefficients of the Daubechies 6-10 Filter)

i	\bar{h}	\tilde{h}
0	0.5575435262285	0.869813136679233
1	0.03372823688575	-0.188640851913885
2	-0.09127176311425	-0.095087384971863
3		-0.009884638967935
4		0.02674875741081

表 II-6 Daubechie 10-18 卷积型滤波器系数 (Coefficients of the Daubechies 10-18 Filter)

i	\bar{h}	\tilde{h}
0	0.536628801791641	0.440781829293449
1	0.054299075394283	-0.115519002860397
2	-0.111388018824611	-0.060571607153814
3	0.000058297264620	0.009733420188029
4	0.020401844374067	0.021802742673181
5		0.001787592313644
6		-0.006683900685050
7		0.000001928418995
8		0.000674873932507

参考文献

- [1] C.E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. tech. J.*, vol.27, pp.379-423, July, 1948
- [2] C.E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. tech. J.*, vol.27, pp.623-656, Oct, 1948
- [3] C.E. Shannon, "Coding Theorys for a Discrete Source with a Fidelity Criterion," *IRE Nat. Conv. Rec.*, pt.4, pp.142-163, March. 1959
- [4] ISO/IEC. International Standard 10918-1: Information technology - digital compression and coding of continuous-tone still images: Requirements and guidelines, 1990. (JPEG).
- [5] ISO/IEC International Standard 11172: Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbps, 1993. (MPEG-1).
- [6] ISO/IEC International Standard 13818-2: Information technology - Generic coding of moving pictures and associated audio information: Video, 1995. (MPEG-2 Video).
- [7] J.L.Mitchell, W.B.Pennebaker, C.E.Fogg, and D.J.Legall. *MPEG video compression standard*, New York:Chapman & Hall, USA, 1996
- [8] V.Bhaskaran and K.Konstantinides, *Image and video compression standards*. London: Kluwer Academic Publishers, United Kindom, 1997
- [9] ISO/IEC. International Standard 14496: Information technology - coding of audio-visual objects - part 2: Visual, 1999. (MPEG-4 Video).
- [10] ITU-T. Recommendation H.261: Video codec for audiovisual services at px64 kbits, 1993
- [11] ITU-T. Recommendation H.263: Video coding for low bit rate communication, 1998.
- [12] ITU-T, "H.26L test model long tern number 8 (TML-8) draft0," *Q.6/SG16 (VCEG)*, June 28, 2001.

- [13] S. Tamer and G. Mohammed, "Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats," *IEEE Trans. Multimedia*, vol. 2, issue 2, pp.101-110, June 2000.
- [14] G. Keesman, R. Hellinghuizen, F. Hoeksema, and G. Heideman, "Transcoding of MPEG bitstreams," *Signal Processing: Image Communication*, vol. 8, pp. 481-500, Sept. 1996.
- [15] A. Vetro; C. Christopoulos; H. Sun, "Video transcoding architectures and techniques: an overview", *IEEE Signal Processing Magazine*, Vol. 20 Issue: 2 , pp: 18 -29, Mar 2003
- [16] B. Furht, R. Westwater, J. Ice, "Multimedia broadcasting over the Internet. I", *Multimedia, IEEE*, Vol. 5, Issue: 4, pp: 78-82, Oct-Dec 1998
- [17] B. Girod, N. Farber, and K. Stuhlmuller, "Internet video streaming," *Multimedia Communicatioions*, pp.547-556, 1999, Springer.
- [18] S. Y. Cheung, M. H. Ammar, and X. Li, "On the use of destination set grouping to improve fairness in multicast video distribution," *IEEE Fifteenth Annual Joint Conference of the IEEE Computer Societies, Networking the Next Generation, INFOCOM'96*, vol.2, pp. 553-560, 1996.
- [19] W. Li, "Bit-plane coding of DCT coefficients for fine granularity scalability," *ISO/IEC JTC1/SC29/WG11, MPEG98/M3989*, Atlantic City, NJ, Oct. 1998.
- [20] H. Radha, Y. Chen, K. Parthasarathy, and R. Cohen, "Scalable Internet video using MPEG-4," *Signal Processing: Image communications*, vol. 15, no. 1-2, pp. 95-126, Sept. 1999.
- [21] S.P. Li, F. Wu, and Y. Q. Zhang, "Study of a new approach to improve FGS video coding efficiency," *ISO/IEC JTC1/SC29/WG11, MPEG99/M5583*, Maui, Dec. 1999.
- [22] F. Wu, S.P. Li, and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding," *IEEE Trans. Circuits and Systems for Video Technology, Special Issue on Streaming Video*, vol. 11, no. 3, March 2001
- [23] S. Mallat, *A wavelet tour of signal processing*, ISBN012466606X, San Diego : Academic Press, 1999.
- [24] M. Barlaud, P. Mathieu, M. Antonini, "Wavelet transform image coding using vector quantization", *Multidimensional Signal Processing Workshop Sixth*, 1989 ,pp: 103 -104

- [25] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, "Image coding using vector quantization in the wavelet transform domain", *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, vol.4, pp: 2297 -2300, 3-6 Apr. 1990
- [26] J.M.Shapiro,"Embedded Image Coding Using Zerotrees of Wavelet Coefficient", *IEEE Trans. Signal Processing*, vol. 41, pp. 3445-3462, Dec.1993.
- [27] A. Said, and W. A. Pearlman, "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees", *IEEE Trans. CSVT*, vol. 6, no. 3, June 1996, pp. 243-250
- [28] ISO/IEC JTC1/SC29/WG1. FCD 15444-1: Information technology - JPEG2000 image coding system, March 2000. available at <http://www.jpeg.org/FCD15444-1.htm>.
- [29] D. Taubman, "High Performance Scalable Image Compression with EBCOT", *Image Processing, IEEE transaction on*, vol.9, No. 7, July 2000
- [30] H. R.Wu, D.M. Tan, B. Qui, Z. Yu, "Adaptation of vision model based distortion metrics to perceptual image coding", *Digital Signal Processing, DSP 2002. 14th International Conference on*, Vol. 1 , pp. 181-184, 2002
- [31] C. Charrier, T. Eude, "A new visual masking tool for JPEG2000", *Data Compression Conference, 2003. (DCC 2003)*, 2003 pp: 418 -418
- [32] A. Natsu, D. Taubman, "Unequal protection of JPEG2000 code-streams in wireless channels", *IEEE Global Telecommunications Conference 2002. GLOBECOM '02.* , Vol. 1 , pp. 534 -538, 2002
- [33] W. Yu, J. Fritts, T. Sun, "An efficient packetization algorithm for JPEG2000", *International Conference on Image Processing. 2002. Proceedings.* , Vol. 1 , pp. 208-212, 2002
- [34] D. Taubman, "Remote browsing of JPEG2000 images", *International Conference on Image Processing. 2002*, Vol: 1 , pp: 229-232, 2002
- [35] S. Movva, S. Srinivasan, "A novel architecture or lifting-based discrete wavelet transform or JPEG2000 standard suitable or VLSI implementation", *16th International Conference on VLSI Design, 2003. Proceedings*, pp. 202-207, 2003

- [36] C.-T. Huang, P.-C. Tseng, L.-G. Chen, "Flipping structure: an efficient VLSI architecture for lifting-based discrete wavelet transform", *Asia-Pacific Conference on Circuits and Systems*, 2002. APCCAS '02. , Vol. 1 , pp. 383-388, 2002
- [37] J. Li and S. Lei, "An embedded still image coder with rate-distortion optimization", *IEEE Trans. On Image Processing*, Vol. 8, No. 7, pp. 913-924, Jul. 1999.
- [38] I. Daubechies. *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Series in Applied Mathematics. SIAM, Philadelphia, 1992.
- [39] J. D. Villasenor, B. Belzer, J. Liao, "Wavelet filter evaluation for image compression", *IEEE Transactions on Image Processing*, Volume: 4 Issue: 8 , pp. 1053 -1060, Aug 1995
- [40] K. Ramchandran, M. Vetterli, "Best wavelet packet bases in a rate-distortion sense", *IEEE Transactions on Image Processing*, Volume: 2 Issue: 2 , pp. 160-175, Apr 1993
- [41] K. Ramchandran, Z. Xiong, K. Asai, M. Vetterli, "Adaptive transforms for image coding using spatially varying wavelet packets", *IEEE Transactions on Image Processing*, Volume: 5 Issue: 7 , pp. 1197 -1204, Jul 1996
- [42] W. Sweldens, "The lifting scheme: A new philosophy in biorthogonal wavelet constructions," *Wavelet Applications in Signal and Image Processing III*, pp. 68-79, Proc. SPIE 2569, 1995.
- [43] W. Sweldens, "The lifting scheme: a custom-design construction of biorthogonal wavelets", *J. of Appl. And Comput. Harmonic Analysis*, 3(2):186-200, 1996
- [44] W. Sweldens, "The lifting scheme: a construction of second generation wavelets", *SIAM J. of Math. Analysis*, 29(2):511-546, 1997
- [45] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps", *J. Fourier Anal. Appl.*, vol. 4, pp. 247-269, 1998.
- [46] J. Reichel, "Complexity related aspects of image compression", 2001, *Ph.D thesis EPFL*, No 2326.
- [47] Z. Xiong, K. Ramchandran, M. T. Orchard and Y.-Q. Zhang, "A comparative study of DCT- and wavelet-based image coding", *IEEE Trans. On Circuits and Systems for Video Technology*, Vol.9, No.5, pp.692-695, Aug.1999

- [48] J. W. Schwartz and R. C. Barker, "Bit-plane encoding: A technique for source encoding", *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-2, no. 4, pp. 385-392, 1966
- [49] E. Ordentlich, M. Weinberger, and G. Seroussi, "A low-complexity modeling approach for embedded coding of wavelet coefficients", *Data Compression Conference (DCC'98)*, Salt Lake City, Utah, Mar. 1998.
- [50] M. Rabbani and P. W. Melnychuk, "Conditioning contexts for the arithmetic coding of bit planes", *IEEE Tran. Signal Processing*, Vol. 40, no. 1, Jan., 1992
- [51] A.S. Lewis, G. Knowles, "Image compression using the 2-D wavelet transform", *IEEE Transactions on Image Processing*, Volume: 1 Issue: 2, pp. 244 -250, Apr 1992
- [52] W. Li, F. Ling, and X. Chen, "Comparison of FGS with simulcast," *ISO/IEC JTC1/SC29/WG11, MPEG99/M5083*, Sept. 1999.
- [53] H. Jiang and G. M. Thayer, "Using frequency weighting in FGS bit-plane coding for natural video," *ISO/IEC JTC1/SC29/WG11, MPEG99/M5489*, Dec. 1999.
- [54] W. Li, "Frequency weighting for FGS," *ISO/IEC JTC1/SC29/WG11, MPEG99/M5589*, Dec. 1999.
- [55] S.P. Li, F. Wu, and Y. Q. Zhang, "Experimental results with progressive fine granularity scalable (PFGS) coding," *ISO/IEC JTC1/SC29/WG11, MPEG00/M5742*, March 2000
- [56] X.Y. Sun, F. Wu, S.P. Li, W. Gao, and Y-Q. Zhang, "Macroblock-based progressive fine granularity scalable (PFGS) video coding with flexible temporal-SNR scalabilities," *Proc. ICIP2001*, pp.1025-1028, Thessaloniki, Greece, Oct. 7-10, 2001.
- [57] R. Yan, F. Wu, S.P. Li, R. Tao, and Y. Wang, "Efficient video coding with hybrid spatial and fine-grain SNR scalability," *Proceedings of SPIE, Visual Communications and Image Processing 2002(VCIP 2002)*, Volume 4671, pp. 850-859, San Jose, California, USA, Jan. 20-25, 2002.
- [58] Y. He; F. Wu; S.P. Li; Y.Z. Zhong; S.Q. Yang, "H.26L-based fine granularity scalable video coding", *IEEE International Symposium on Circuits and Systems*, 2002. ISCAS 2002., Volume: 4, pp. 548-551, 2002
- [59] N4474, "Ad hoc Group on Exploration of Interframe Wavelet Technology in Video", http://mpeg.telecomitalia.com/meetings/pattaya/pattaya_ahg.htm

- [60] Y.-Q. Zhang, S. Zafar, "Motion-compensated wavelet transform coding for color video compression", *IEEE tran. on Circuits and Systems for Video Technology*, Vol: 2 Issue: 3 , Sept. pp: 285 –296.1992
- [61] M.T.Orchard and G.J.Sullivan, "Overlapped block motion compensation: an estimation-theoretic approach", *IEEE Trans. on Image Processing*, Vol.3, No.5, pp.693-699, Sep. 1994
- [62] A. Martucci, I. Sodagar, T. Chiang, Y.-Q. Zhang, "A zerotree wavelet video coder", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume: 7 Issue: 1 , pp. 109-118, Feb 1997
- [63] D. Marpe, H.L. Cycon, "Very low bit-rate video coding using wavelet-based techniques", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume: 9 Issue: 1 , pp. 85 -94, Feb. 1999
- [64] K. Shen, E.J. Delp, "Wavelet based rate scalable video compression", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume: 9 Issue: 1 , pp. 109-122, Feb 1999
- [65] K.K. Lin, R.M. Gray, "Video residual coding using SPIHT and dependent optimization", *Data Compression Conference*, 2001. Proceedings. DCC 2001, pp. 113-122, 2001
- [66] A. Nosratinia and M. Orchard, "A multi-resolution framework for backward motion compensation", *Proc. SPIE Electronic Imaging*, San Jose, CA, pp. 190-200, February 1995
- [67] X. Yang and K. Ramchandran, "Scalable wavelet coding using aliasing-reduced hierarchical motion compensation", *IEEE Trans. on Image Processing*, Vol.9, pp.778-791, May 2000
- [68] M. Wien, "Hierarchical wavelet video coding using warping prediction", *International Conference on Image Processing*, 2000, Volume: 3 , pp. 142-145, 2000
- [69] H.W.Park and H.S. Kim, "Motion estimation using low-band-shift method for wavelet-based moving picture coding", *IEEE Trans. on Image Processing*, Vol.9, pp.577-587, April 2000
- [70] X. Li, L. Kerofsky, S. Lei, "All-phase motion compensated prediction in the wavelet domain for high performance video coding", *International Conference on Image Processing*, 2001, Volume: 3 , pp. 538-541, 2001

- [71] B.-J. Kim, W.A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)", *Data Compression Conference, DCC '97*, pp. 251-260, Snowbird, UT, Mar 1997
- [72] B.-J. Kim; Z.X. Xiong; and W.A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.: 10, pp: 1374 -1387, December 2000.
- [73] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video", *IEEE Trans. on Image Processing*, Vol. 3, No. 5, pp. 572-588, Sept. 1994.
- [74] A. Wang, Z. Xiong, P. A. Chou, and S. Mehrotra, "3D wavelet coding of video with global motion compensation," *Data Compression Conference (DCC'99)*, pp: 404 -413, Snowbird, UT, Mar. 1999.
- [75] J. R. Ohm, "Three-dimensional subband coding with motion compensation", *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 559-571, Sept. 1994.
- [76] S. Choi, and J. W. Woods, "Motion-compensated 3-D subband coding of video", *IEEE Trans. on Image Processing*, vol. 8, pp. 155-167, Feb. 1999.
- [77] S.T. Hsiang, J.W. Woods : "Invertible three-dimensional analysis/synthesis system for video coding with half-pixel accurate motion compensation", *SPIE VCIP 99*, Vol. 3653, pp. 537-546
- [78] A. Secker, D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting", *International Conference on Image Processing 2001* , Volume: 2 , pp. 1029-1032, Oct 2001
- [79] A. Secker, D. Taubman, "Highly scalable video compression using a lifting-based 3D wavelet transform with deformable mesh motion compensation", *International Conference on Image Processing. 2002*, Volume: 3 , pp. 749-752, 2002
- [80] P. Chen, and J. W. Woods, "Improved MC-EZBC with quarter-pixel motion vectors", JVT proposal, *ISO/IEC JTC1/SC29/WG11, MPEG2002/M8366*, Fairfax, VA, May 2002
- [81] P. Chen, and J. W. Woods, "Exploration Experimental Results and Software", JVT proposal, *ISO/IEC JTC1/SC29/WG11, MPEG2002/M8524*, Klagenfurt, AT, July 2002

- [82] J. Xu, S. Li and Y. Zhang, "Three-dimensional shape-adaptive discrete wavelet transforms for efficient object-based video coding", *IEEE/SPIE Visual Communications and Image Processing (VCIP) 2000*, Perth, June 2000
- [83] J. Xu, Z. Xiong, S.P. Li, and Y.-Q. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3D ESCOT)", *Applied and Computational Harmonic Analysis*, pp.290-315, 2001.
- [84] S.P. Li, W. Li, "Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume: 10 Issue: 5 , pp. 725 -743, Aug 2000
- [85] S.B. Kang, "A Survey of Image-based Rendering Techniques", *Videometrics VI (SPIE International Symposium on Electronic Imaging: Science and Technology)*, vol. 3641, San Jose, CA, 23-29 Jan., 1999,
- [86] P. Debevec, L. McMillan, "Image-based modeling, rendering, and lighting", *Computer Graphics and Applications*, IEEE , Volume: 22 Issue: 2 , pp. 24-25, Mar/Apr 2002
- [87] E. H. Adelson, and J. R. Bergen, "The plenoptic function and the elements of early vision", *Computational Models of Visual Processing*, Chapter 1, Edited by Michael Landy and J. Anthony Movshon. The MIT Press, Cambridge, Mass. 1991.
- [88] L. McMillan and G. Bishop, "Plenoptic modeling: an image-based rendering system", *Computer Graphics Proceedings, Annual Conference series (SIGGRAPH'95)*, pp. 39-46, Aug. 1995.
- [89] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph", *Computer Graphics Proceedings, Annual Conference series (SIGGRAPH'96)*, pp. 43-54, New Orleans, Aug. 1996
- [90] M. Levoy and P. Hanrahan, "Light field rendering", *Computer Graphics Proceedings, Annual Conference series (SIGGRAPH'96)*, pp.31-42, Aug. 1996.
- [91] H.-Y. Shum and L.-W. He, "Rendering with concentric mosaics", *Computer Graphics Proceedings, Annual Conference series (SIGGRAPH'99)*, pp.299-306, Aug. 1999.
- [92] J. Li, H.Shum and Y.-Q. Zhang, "On the compression of image based rendering scene", *Proceedings of ICIP 2000*, Vancouver, Canada, Sep. 2000

- [93] T.H. Chen, "From image and video compression to computer graphics", *2000 IEEE International Conference on Image Processing*, Volume: 2, pp. 9-12, 2000
- [94] C. Zhang and J. Li, "Compression of Lumigraph with multiple reference frame (MRF) prediction and just-in-time rendering", *IEEE Data Compression Conference*, pp.253-262, Snowbird, Utah, Mar.2000.
- [95] C. Zhang, and J. Li, "Compression and rendering of concentric mosaics with reference block codec(RBC)", *SPIE Visual Communications and Image Processing 2000*, SPIE 4067-05, Perth, Australia, Jun. 2000.
- [96] JPEG-LS AHG, "JPEG-LS part 2 CD version 1.0", *ISO/IEC JTC1/SC29/WG1N1557*, Maui, Hawaii, Dec. 1999.
- [97] E. Ordentlich, D. Taubman, M. Weinberger, G. Seroussi, M. W. Marcellin, "Memory efficient scalable line-based image coding", *IEEE Data Compression Conference*. Snowbird, Utah, March 1999.
- [98] P. Cosman and K. Zeger, "Memory constrained wavelet-based image coding", *Signal Processing Letters*, vol. 5, pp 221-223, Sept. 1998.
- [99] I. Ihm and S. Park, "Wavelet-based 3D compression scheme for very large volume data", *Graphics Interface*, pp. 107-116, Jun. 1998.
- [100] I. Ihm, and S. Park, "Wavelet-based 3D compression scheme for interactive visualization of very large volume data", *Computer Graphics Forum*, Vol.18, No.1, pp.3-15, Mar. 1999.
- [101] C. Bajaj, I. Ihm, and S. Park, "Making 3D texture practical", *Pacific Graphics '99*, pp.259-268, Seoul, Korea, Oct. 1999.
- [102] F. F. Rodler, "Wavelet based 3D compression with fast random access for very large volume data", *the Seventh Pacific Conference on Computer Graphics and Applications*, Seoul, Korea, Oct. 1999.
- [103] J. Xu, Z. Xiong, S.P. Li, and Y.-Q., Zhang, "On boundary effects in 3-D wavelet video coding", *Proc. of Symposium on Optical Sci. and Tech.* San Diego, CA, July 2000
- [104] P. Cheng, J. Li and J. Kuo, "Rate control for embedded wavelet video coder", *IEEE Trans. On Circuit and System for Video Technology*, Vol. 7, No. 4, pp. 696-702, Aug. 1997.

- [105] P. D. Lubell, "A coming attraction: digital cinema", *IEEE Spectrum*, Volume: 37 Issue: 3, pp. 72-78, Mar 2000
- [106] P. Chen, J.W. Woods, "Video coding for digital cinema", *2002 IEEE International Conference on Image Processing*, Volume: 1, pp. 749-752, 2002
- [107] Y. Wu, C. Zhang, J. Li, "Smart rebinning for the compression of concentric mosaic", *IEEE Transactions on Multimedia*, Volume: 4 Issue: 3, pp. 332-342, Sep 2002
- [108] Y. Wang, J. Ostermann, Y.-Q. Zhang, *Video processing and communications*, ISBN: 0130175471, Upper Saddle River, NJ : Prentice Hall, 2002.
- [109] S. SKodras, C. Christopoulos, T. Ebrahimi, "The JPEG 2000 still image compression standard", *IEEE Signal Processing Magazine*, Volume: 18 Issue: 5, pp. 36-58, Sep. 2001
- [110] B.E. Usevitch, "A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000", *IEEE Signal Processing Magazine*, Volume: 18 Issue: 5, PP. 22-35, Sep 2001
- [111] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Englewood Cliffs, NJ, Prentice Hall, 1995

作者在攻读博士学位期间的研究成果

获奖论文

- [1] Yunnan Wu, Lin Luo, Jin Li, and Ya-Qin Zhang, "Rendering of 3D wavelet compressed concentric mosaic scenery with progressive inverse wavelet synthesis (PIWS)", *SPIE/IEEE Visual Communications and Image Processing (VCIP 2000)*, pp. 31-42, Perth, Australia, Jun. 2000. 获 SPIE/IEEE VCIP 国际会议 2000 年度最佳学生论文奖(won the Best Student Paper award.) (EI indexed)

SCI, EI 检索论文

- [2] Lin Luo, Jin Li, Shipeng Li, and Zhenquan Zhuang, "A motion compensated lifting approach for 3D wavelet video coding", 2003 年 1 月, 计算机科学与技术学报 (*Journal of Computer Science and Technology*), pp. 214-222, (英文期刊)。
- [3] Lin Luo, Yunnan Wu, Jin Li and Ya-Qin Zhang, "3D wavelet compression and progressive inverse wavelet synthesis rendering of concentric mosaic", *IEEE Trans. on Image Processing*, July 2002, pp. 802-816

EI 检索论文

- [4] Lin Luo, Feng Wu, Shipeng Li, and Zhenquan Zhuang, "Advanced lifting-based Motion-Threading (MTh) techniques for 3D wavelet video coding", invited paper, to appear in *SPIE/IEEE Visual Communications and Image Processing (VCIP2003)*, Lugano, Switzerland, Jul. 2003
- [5] Lin Luo, Feng Wu, Shipeng Li, and Zhenquan Zhuang, "Layer-correlated motion estimation and motion vector coding for the 3D-wavelet video coding", to appear in *IEEE International Conference on Image Processing (ICIP2003)*, Barcelona, Spain, Sep. 2003
- [6] Lin Luo, Feng Wu, Shipeng Li, Zhenquan Zhuang, and Ya-Qin Zhang, "A two-pass optimal motion-threading technique for 3D wavelet video coding", *IEEE Intl. Symposium on Circuits and Systems (ISCAS 2002)*, pp.819-822, Scottsdale, AZ, May 2002
- [7] Lin Luo, Yunnan Wu, Jin Li, and Ya-Qin Zhang, "Compression of concentric mosaic scenery with alignment and 3D wavelet transform", *SPIE Image and Video Communications and Processing (IVCP 2000)*, vol. 3974, pp.89-100, San Jose, CA, Jan. 2000.

其他论文

- [8] 罗琳, 许继征, 庄镇泉, 汪祖媛, 李世鹏, "一种时域质量平稳的三维小波视频编码率控制技术", Vol.8, No.1, pp.27-31, 2003 年 2 月, 电路与系统学报,
- [9] 罗琳, 吴韵楠, 庄镇泉, 李劲, "基于渐进逆向小波变换方法的压缩同心拼图的实时绘制", 2003 年 3 月, 科大学报。

- [10] Lin Luo, Jin Li, Shipeng Li, Zhenquan Zhuang, and Ya-Qin Zhang, "Motion compensated lifting wavelet and its application in video coding", *IEEE Intl.Conf. on Multimedia and Expo (ICME 2001)*, pp. 481-484, Tokyo, Japan, Aug. 2001.
- [11] Lin Luo, Jizheng Xu, Shipeng Li, and Zhenquan Zhuang, "Rate control with smoothed temporal distortion for a 3D embedded wavelet video coder", *IEEE International Conference on Information, Communications & Signal Processing (ICICS 2001)*, CD release, Singapore, Oct. 2001

待投论文

- [12] Lin Luo, Feng Wu, Shipeng Li, and Zhenquan Zhuang, "Advanced lifting-based Motion-Threading (MTh) techniques for 3D wavelet video coding", to be submitted for *IEEE Tran. on Circuit and Systems on Video Technology*

攻读博士学位期间参与的科研项目与学术交流

- [1] 参加中国科学技术大学与微软亚洲研究院的合作研究项目：“视频媒体的小波压缩算法研究”
- [2] 参与 1999 年微软亚洲研究院暑期实习生计划：“同心拼图的三维小波压缩”
- [3] 由微软亚洲研究院资助，参加在日本东京召开的 IEEE International Conference on Multimedia & Expo 2001 年度国际会议并作题为“Motion compensated lifting wavelet and its application in video coding”的论文讲演
- [4] 由微软亚洲研究院资助，参加在澳大利亚珀斯的 SPIE/IEEE Visual Communications and Image Processing 2000 年度国际会议并作题为“Rendering of 3D wavelet compressed concentric mosaic scenery with progressive inverse wavelet synthesis (PIWS)”的论文讲演，该论文获得此次会议最佳学生论文奖
- [5] 由微软亚洲研究院资助，参加在美国新奥尔良召开的 ACM SIGGRAPH 2000 图形学年度国际会议
- [6] 获得 2000 年度微软亚洲研究院颁发的“微软学者奖（Microsoft Fellowship）”
- [7] 曾担任 IEEE Transactions on Circuits and Systems for Video Technology, Journal of Visual Communication and Image Representation, IEEE Visualization' 2000 conference 等国际知名期刊和会议的论文审稿人。

致谢

博士生涯，一直潜心在学海中探索遨游，值此论文草成之际，终于可以将心绪沉淀，细细品味多年的求学历程，谨向所有帮助和支持我的人，送上衷心的感谢与祝福。

饮水思源，师从庄镇泉教授真是我的幸运。导师治学严谨，在学术上目光如炬的敏锐洞察力和对研究方向的宏观驾驭能力，令人十分钦佩。从课题的选题立意、观点的凝聚提炼、内容的研究分析，直至文章的组织推敲，导师对每一环节倾心的指点历历在目。更敬佩导师教书育人的风范，时常聆听导师教诲，谓研究者在开拓创新、活跃思维的同时，更应严以律己、刻苦钻研。对此甚有感悟，日后必将加倍努力，踏实进取，以不负恩师殷切的期望。还要感谢导师为我创造了良好、宽松的研究实习条件和环境，以及生活上的诸多关心，使我得以顺利完成学业。感激之情，言语不尽。

难忘微软，在微软亚洲研究院参加合作研究项目的经历丰富充实。感谢充满提携后生情怀的张亚勤院长，他所给予的热情鼓励和大大力支持，令我信心倍增，勇于迎接挑战。感谢李开复院长、张宏江院长和沈向洋院长为我提供学术交流的机会，让我开阔视野，增长见识。有幸来到微软先后得到李劲博士、李世鹏博士和吴枫博士的悉心指导和热忱帮助，在此深表谢意。作为刚刚步入科研殿堂的年轻学子，一心渴望能够探求科学的真谛，是李劲博士以他渊博的学识，周密细致的科研精神和创新灵动的思维启迪我，并毫无保留地向我传授他分析解决问题的心得和提高论文质量的经验，以及有关科研方面的种种体会，使我能够加足马力，驶向在科研征途上成长进步的快车道。尽管他现已在大洋彼岸，仍然诲人不倦地赐教于我，让我感动不已。此后，吴枫博士和李世鹏博士在视频压缩方面给予我许多有关先进技术、前沿理念的详细指点，帮助我在科研道路上明确方向、夯实基础、迈稳脚步，使整个研究得以顺畅、圆满地完成。他们在研究中充满睿智、积极进取的拼搏精神，而在日常相处中又热情开朗、轻松幽默，让我领略到完美的人格魅力。在微软亚洲研究院访问期间得到许多师长的指点与关怀，令我终生难忘，并将激励我在科研领域永不停步。

感念科大，给予我起飞动力的前进基地。从少年班到攻读博士学位，九年的科大生活记录着我的成长经历。真诚感谢母校的栽培。感谢系里的冯焕清老师、郭立老师、朱领娣老师、戴蓓倩老师、周佩玲老师和施泽生老师，谢谢他们多年来对我的热情支持与关心爱护。感谢少年班的胡薇薇老师，叶国华老师等全体老师的辛勤培育。

感谢实验室的老师 and 同学：李斌、汪祖媛、陈贤富、李海鹰、高清维、解光军、陶亮等诸位博士，李煊、李军、简兵、杨俊安、叶学义、田巍、张云超、邹谊、顾理、李动恒、庄连生、李霄寒、傅庆东、陈超、孙卫强等各位同学，谢谢大家的真诚相助，让我的博士生涯中充满着友谊，并留下了难忘的记忆。

感谢在微软研究院相识的朋友们。对曾经合作过的吴韵楠致以特别的谢意。顾烈、Alexis M. Tourapis 博士、李江博士、朱斌博士、阎蓉、罗翀、孙晓艳、李娜、徐清清、吴皓、陈瑾、许茜、许继征、赵俊、杨志杰、焦锋、蔡华、余可曼、沈国斌、孙宏辉、郭子华、周剑平、李岩、董昕、王立峰、王天树、陈宏、张波、向哲、潘锦辉、张察、邢桂伟、苏中、朱兴泉、刘天明、季竺、王贵锦、贺玉文、王琪、周游、马立然、杨凡、熊瑞勤、李江红...，大家无私的协作与尽情的欢乐情景，在我的心中将永远珍藏。

岁月流逝，难以抹去我对母校的深情眷恋，难以忘怀所拥有的美好回忆。感谢在我的人生成长经历中，所有关怀和曾经帮助过我的人们，祝愿大家健康快乐。

本文献给我最敬爱的父母和亲人，是他们教导我诚实做人，培育我勤勉刻苦的品质，感谢他们一直以来对我的默默奉献和无尽的关爱，祝愿他们永远幸福平安。

缩略词索引

bps	bits per second
CABAC	Context-based Adaptive Binary Arithmetic Coding
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
EZW	Embedded Zerotree Wavelet
EBCOT	Embedded Block Coding with Optimized Truncation
GOP	Group of Picture
IBR	Image-Based-Rendering
IEC	International Electrotechnical Commission
ISO	International Standardization Organization
ITU	International Telecommunications Union
JIT	Just-In-Time
JPEG	Joint Picture Experts Group
MB	Macro-block
MC	Motion Compensation
ME	Motion Estimation
MPEG	Moving Picture Experts Group
OBMC	Overlapped-Block Motion Compensation
PSNR	Peak Signal to Noise Ratio
R-D	Rate-Distortion
SAD	Sum of Absolute Difference
SPIHT	Set Partitioning in Hierarchical Trees
SVQ	Spatial-domain Vector Quantization
VQ	Vector Quantization
WLAN	Wireless Local Access Network