

分类号 _____

学号 M200771397

学校代码 10487

密级 _____

华中科技大学

硕士学位论文

一种新的基于联合内插预测的 视频编码方案

学位申请人： 江文斐

学科专业： 通信与信息系统

指导教师： 刘文予 教授

答辩日期： 2009年5月25日

**A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering**

**A New Video Coding Scheme Based on Joint
Spatiotemporal Prediction**

Candidate : Wenfei Jiang

Major : Communication and Information System

Supervisor : Prof. Wenyu Liu

Huazhong University of Science & Technology

Wuhan 430074, P.R.China

May, 2009

摘 要

近年来，视频编码技术的发展趋势正在由高压缩率逐渐转移到对高质量、高分辨率、高灵活性的支持。鉴于主流的视频编码方案已经付出了惊人的复杂度但对这些最新的要求缺乏足够的适应性，有必要研究新的视频编码框架。新框架以支持可分级编码以及保证在高分辨率、高清晰度应用的性能为目标。通过对目前已有的编码机制的研究，内插预测编码（Interpolative Coding）被认为对这些需求具有较强的适应性，同时，基于点到点的内插预测编码，能够从根本上去除主流编码机制下难以回避的块效应。

内插预测编码的性能很大程度上取决于其采用的预测技术。而现有的帧内空域预测和帧间运动估计都是在各自的领域相当成功的技术，却难以对帧间/帧内相关性相对均衡的视频信号做出准确估计。本文创造性地提出了新的时空联合预测的预测方法，同时兼顾帧间和帧内相关性，该方法已被证明在绝大多数情况下优于传统的运动补偿技术，能够保证新的预测编码方案具有较高的压缩率。

基于内插预测的编码机制，本文提出了 K-L 变换/联合预测（KLT/Joint Prediction）的编码框架。除了最新的时空联合预测技术，基于图像特征的 K-L 变换也能大幅提高对视频的压缩效率。以此取代传统的运动估计/DCT 变换（ME/DCT）框架，新方案不仅能保持较高的压缩率，还能很好地满足上述的最新需求。

结合以上研究，本文提出了基于内插预测的新型视频编码方案 Joint Predictive Coding（JPC），从根本上去除块效应，完美支持分辨率可调性，同时保证较高的压缩率以及较低的复杂度。JPC 在略高于 MPEG-2 的复杂度的条件下，达到了明显高于 MPEG-2 的压缩率，并在极高的质量下达到了目前最先进的 H.264 标准的压缩水平。本文的研究不仅对当代视频技术发展趋势的有着很强的契合度，也对下一代视频编码技术的探索有积极的意义。

关键词： 视频编码，内插预测，联合预测编码，H.264，高级视频编码

Abstract

Video coding technique has been targeting high quality, high resolution and high flexibility these years. Note that popular coding schemes have spent extremely high complexity to achieve the current compression efficiency; however, they have a lack of adaptability to these new requirements. Thus, there is a need for research on new video coding frameworks. The new framework is aiming at the support of Scalable Video Coding (SVC) and good performance in the high quality/resolution applications. According to the investigation of the existing coding strategies, interpolative coding is considered to have a perfect mechanism to meet such requirements. Plus, the pixel-to-pixel mechanism is free of blocking effect, which has been a major problem of mainstream block-based coding scheme for years..

The performance of an interpolative coding scheme largely depends on its prediction technique. The existing intra-frame interpolative prediction and inter-frame motion estimation have been successful in their own applications; however, the separate use of them fails to get good results on the video that neither temporal nor spatial correlation holds the dominant position. This paper proposes a new prediction method called joint spatiotemporal prediction that considers spatial and temporal correlation simultaneously. Joint spatiotemporal prediction proves to be superior to the popular motion compensation technique in most cases.

This paper introduces a new KLT/Joint Prediction framework based on the interpolative coding mechanism. Besides joint spatiotemporal prediction, the image dependent color space transformation based on KLT also helps achieve greater compression. This guarantees that the KLT/Joint Prediction framework keeps good compression ratio and supports the scalability in resolution well.

Based on the research above, this paper proposes a new video coding scheme called Joint Predictive Coding (JPC), which bypasses the traditional Motion Estimation/DCT (ME/DCT) coding procedure. Because of its basis of pixel-to-pixel interpolative prediction, JPC is a natural scheme to meet the requirement of scalability in resolution; plus, it completely gets rid of the blocking effect. JPC achieves much higher compression efficiency than MPEG-2 by slightly higher computational complexity. Moreover, it is able to get equivalent compression ratio with H.264 at extremely high bit-rate. The research of this paper not only accords with the requirements of modern video applications, but could also be of great value to the exploration of the next-generation video coding techniques.

Keywords: video compression, interpolative prediction, Joint Predictive Coding, H.264, AVC

目 录

摘 要.....	I
Abstract.....	II
1 绪 论	
1.1 研究背景和意义	(1)
1.2 国内外研究现状	(2)
1.3 全文主要研究内容和结构安排	(4)
2 主流视频编解码框架	
2.1 主流视频编解码方案概述	(6)
2.2 视频编解码的关键技术	(9)
2.3 寻找新架构编码方案的必要性	(12)
2.4 本章小结.....	(13)
3 基于内插预测的图像编解码技术	
3.1 基于内插预测的图像编解码方案概述	(14)
3.2 Adaptive Prediction Trees 图像编解码方案.....	(15)
3.3 APT 的优势及其局限性	(22)
3.4 本章小结.....	(23)
4 时空联合预测模型和方法	
4.1 时空联合预测方法	(24)

4.2 时空联合预测的准确性分析	(26)
4.2 编码方案的预测模式配置	(32)
4.3 本章小结.....	(33)
5 基于时空联合预测的视频编码方案	
5.1 联合预测视频编码的基本框架	(34)
5.2 针对闪烁效应的滤波算法设计	(36)
5.3 与其他基于内插预测的编码方案的对比分析	(37)
5.4 率失真性能对比测试	(41)
5.5 主观质量对比测试	(43)
5.6 编码速度对比测试	(45)
5.7 本章小结.....	(47)
6 总结及展望	
致 谢.....	(46)
附 录 1 攻读学位期间发表的论文	(50)
附 录 2 文中主要中英文术语对照表	(51)
参考文献.....	(52)

1. 绪 论

1.1 研究背景和意义

视频是人类获取信息的重要载体。长久以来，巨大的视频数据量一直阻碍着视频通信的发展。然而，视频数据的高相关性，使得高度压缩成为可能，数字视频技术也成为多媒体通信的关键技术之一。经过数十年的发展，视频压缩技术日渐成熟，带来了视频会议、远程监控及数字电视、媒体存储等一系列应用。

鉴于视频编码技术的重要意义，许多的研究成果已经被转化为相应的国际标准。当前主要的视频编码标准主要有两个系列：

1. 国际电信联盟（ITU-T）制定的应用于网络通信行业的 H.26x 系列标准，如：用于 $p \times 64\text{Kbps}$ ($p=1, 2, \dots, 30$) 速率下的视频编解码标准 H.261^[1]；面向甚低码率通信的视频编码标准 H.263^[2]；以及目前最先进的 H.264^[3]。

2. 国际标准化组织（ISO）运动图像专家组（MPEG）制定的应用于媒体业务的 MPEG-X 系列标准，如：MPEG-1^[4]、MPEG-2^[5]、MPEG-4^[6]；以及 MPEG-7^[7]、MPEG-21^[8]等。

这些视频压缩标准在多媒体通信领域内得到了广泛的应用，如H.261在可视电话、H.263在视频会议、MPEG-1在VCD、MPEG-2在DVD中的应用，而H.264被选为Blu-ray Disc的强制视频编码标准之一^[9]。

近年来，随着网络带宽的不断提高，存储介质的不断扩容以及多媒体终端的升级，压缩率已经不再是考察视频编码技术的首要标准，高清晰、高分辨率以及分辨率可调性成为了主流的需求^[10]。值得注意的是，无论是 H.26x 系列和 MPEG 系列的压缩标准，都沿用同一编码框架——运动估计/离散余弦变换（ME/DCT），这一框架经过数十年的发展已经相当成熟。运动估计技术是框架内最受关注的模块之一，例如，从 H.263 到 H.264，用于运动估计的计算复杂度由 34% 猛增至 70%^[11]，考虑

到 H.264 的整体复杂度已经是 H.263 的十几倍^[12]，可以说，H.264 编码器在运动估计上付出了巨大的代价才使得输出码率减小到 H.263 的 50%，可见，通过改进运动估计技术来提高压缩效率这条路很可能走到了瓶颈。同时，传统的 ME/DCT 的编码框架与目前数字电影的高灵活性的要求契合度不足，而它所导致的块效应，至今尚无完善的解决方案。针对该状况，本文试图绕过上升空间已十分有限的传统框架，以高质量、高灵活性为目标，探索一种新的视频编码方案，开辟一条新的、更有潜力的道路。

1.2 国内外研究现状

视频编码技术经过多年的发展，已经逐渐标准化。从 1992 年发布的 MPEG-1 标准，1994 年发布的 MPEG-2 标准，1995 年 ITU-T 制定的 H.263 标准，1998 年 ISO 公布的支持基于对象编码的 MPEG-4，到由 ITU-T 的视频编码专家组（VCEG）和 ISO/IEC 的活动图像专家组（MPEG）共同成立的联合视频小组（JVT）于 2003 年公布 4 的 H.264 标准，主流视频压缩标准一直沿用着 ME/DCT 框架，陆续引入了可变速块尺寸的运动补偿、分像素精度插值、多帧参考、加权预测、整数 DCT 变换、“skip”模式与“direct”模式、多方向的帧内预测、环路滤波、基于上下文的熵编码等技术，逐步提高了编码的压缩效率，同时也提高的运算复杂度。

我国数字视音频编解码技术标准（“信息技术先进音视频编码标准”，AVS）工作组于 2002 年正式宣布成立，其最终标准草案于 2006 年宣布完成^[13]。工作组宣称其视频部分的编码效率可相当于 H.264/AVC，且复杂度有所降低。

现代视频应用要求视频数据经过一次压缩就能以多个帧率、空间分辨率或视频质量进行解码，从而可支持多种类型用户的各种不同应用要求。因此可分级视频编码（Scalable Video Coding）成为近年来国内外研究的热点，H.264 的 Scalable Extension 已于 2007 年正式公布^[14]，但 H.264 对可缩放性（scalability）的支持显然是有代价的，这源于 ME/DCT 框架与可缩放性的契合度的不足。为此，本文研究了目前主要的图像编码策略，发现内插预测编码（interpolative coding）的架构能够无

代价地兼容可分级编码，而且非常适合极高质量的编码。

内插预测编码的思路自 1978 年由 Hunt^[15]提出以来，被全球多个科研组织研究实现^[16-23]，其中文献[18][20-22]将该思路用于连续色调图像（continuous-tone image，如照片、医学图像等）的无损压缩，而文献[19]通过量化预测误差将其应用到了有损压缩。

Robinson 分别于 1997 年和 2006 年提出 Binary Tree Predictive Coding (BTPC)^[24]和 Adaptive Prediction Trees (APT)^[25]的图像编码方案，通过采用空域自适应预测技术，在各类图像的编码应用上都取得了较好的效果，APT 的压缩效率更是明显优于 JPEG^[26]而接近 JPEG 2000^[27]。视频编码方面，2000 年 Robinson、Druet 和 Gosset 提出了 Binary Tree Residue Coding^[28]，2007 年 Day 和 Robinson 提出了 Temporal Prediction Trees 视频编码方案^[29]，但由于其性能等原因未能引起广泛关注。

1.3 全文主要研究内容和结构安排

本文在国家自然科学基金项目“无线移动环境下复杂度可分级联合功率失真模型研究 (No. 60572063)”的资助下，探索性地研究了不同于传统的视频编码框架，提出一种全新的视频编码方案 Joint Predictive Coding (JPC)^[30]，它采用点到点的预测机制，从根本上去除传统框架难以逾越的块效应。本文提出的时空联合预测 (Joint Spatiotemporal Prediction)，明显优于传统的运动补偿技术；同时，JPC 所采用的 K-L 变换/联合预测(KLT/Joint Prediction)的新框架，也存在更大的上升空间。JPC 以 MPEG-2 水平的复杂度，取得了明显高于 MPEG-2 的压缩率，并在极高的码率上达到了 H.264 的水平。JPC 的基本架构如下：

- A. 利用主分量分析 (PCA) 将视频数据转到新的颜色空间；
- B. 通过运动向量搜索来获取各像素的运动信息；
- C. 以块为单位作预测模式选择；
- D. 以层序遍历图像中的像素，预测、量化、重构。
- E. 对预测误差、运动向量、模式信息等作游程哈夫曼编码。

该方案的内在结构使其对于当今视频编码的高清晰、高分辨率、可缩放性的要求具有极强的适应性。初步研究表明，该方案在高清电影、高清动画、视频游戏以及虚拟现实等领域具有值得关注的前景。

本文的结构安排如下：第二章介绍传统图像/视频编解码框架；第三章给出用于图像编码的内插预测技术，它构成了 JPC 时空联合预测的基础；第四章提出和分析时空联合预测 **Joint Spatiotemporal Prediction**；第五章详细描述本文提出的视频编码方案 **Joint Predictive Coding**；第六章总结及展望。

2. 主流视频编解码框架

2.1 主流视频编解码方案概述

数字视频序列中包含时域冗余、空域冗余、统计冗余以及心理视觉冗余等大量冗余信息。而视频编码的初衷，就是要最大限度地消除这些冗余信息，以满足早期的网络带宽要求。视频压缩根据不同冗余类型采取相应的编码技术，主要分为三类^[3]：用于消除空域冗余的变换编码，用于消除时域冗余的预测编码，以及用于消除统计冗余的熵编码等。

从视频编码标准的发展历程来看，编码技术早期的发展趋势是不断用更高的复杂度来换取更高的压缩效率，比如从 DPCM 到运动补偿，从前向预测到双向预测，从单帧参考到多帧参考等。进入 21 世纪以来，为适应无线通信、数字电影等各方面的应用，对低能耗、低失真、高分辨率等追求也逐渐成为了全球学者的研究热点。

下面以目前最先进的 H.264 标准为例，介绍主流的视频编码框架。

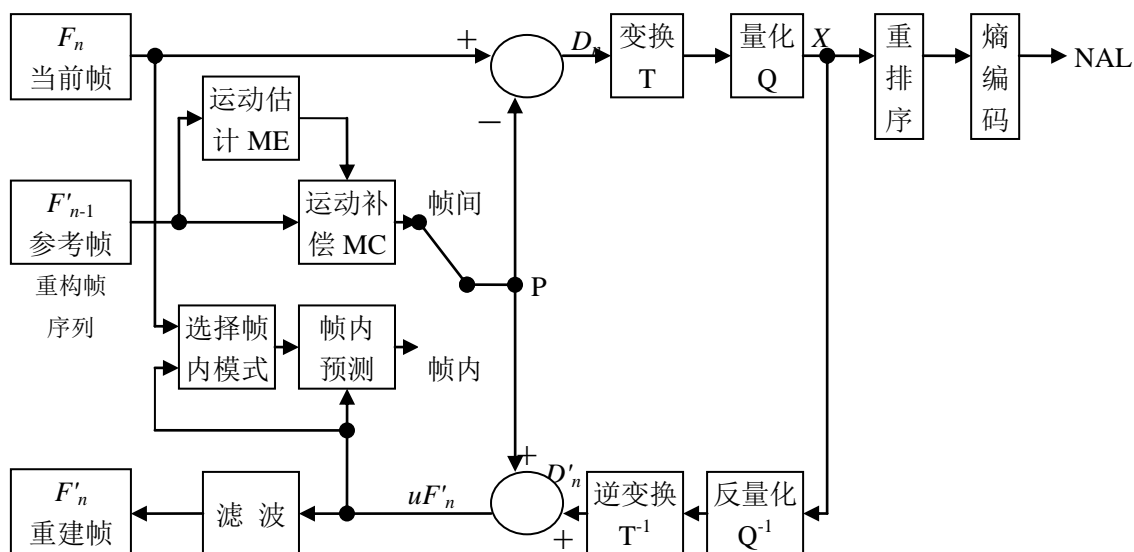


图 2.1 H.264 的编码器结构

2.1.1 编码器

H.264 结构框图如图 2.1 所示。

1) 前向路径

编码器依次对视频帧 F_n 各宏块进行编码，分别考察帧内编码和帧间编码两种模式的效果。帧内编码模式对目标宏块采用多方向的帧内预测，对误差数据块作变换、量化、熵编码。帧间编码模式通过运动估计得到参考帧中的最佳匹配块 P ，目标宏块与预测宏块 P 之间的差别即为要编码的残差 D_n 。 D_n 经过整数 DCT 变换和量化形成一系列经过量化的频域系数 X ，它们将被重排序和熵编码，残差的频域系数、运动向量以及宏块模式，量化参数等信息就够成了 H.264 码流，输出到网络抽象层 NAL。

2) 重构路径

为了构成后续帧运动估计所需的参考帧图像，量化后的宏块频域系数 X 需要被立即解码重构。经过量化的频域系数 X 被反量化 (Q^{-1})、逆变换 (T^{-1})，形成重构残差 D'_n 。由于变换和量化，所以 D'_n 和 D_n 存在误差而和解码端保持一致。预测宏块 P 加上 D'_n 即为重构宏块 uF'_n 。环路去块效应滤波器用于缓解重构图像的块效应，重建帧 F'_n 即可用作参考帧。

2.1.2 解码器

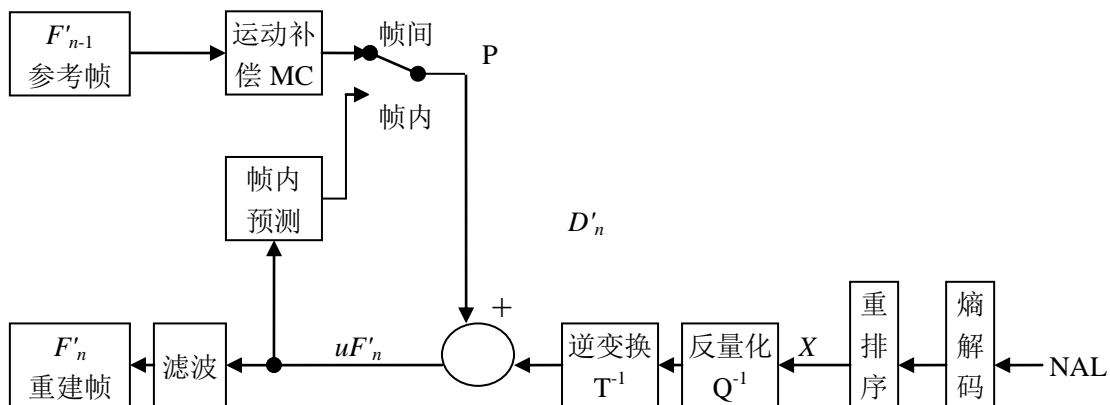


图 2.2 H.264 的解码器结构

H.264 解码器流程框图如图 2.2 所示。解码流程与编码流程中的重构路径类似。

解码器从 NAL 接收到压缩比特流, 经过熵解码和重排序, 得到量化后的频域系数 X 。 X 经过反量化和逆变换形成 D'_n (与编码器中的重构图像等同)。通过码流中的运动信息, 解码器生成一个和编码端相同的预测宏块 P 。 P 与 D'_n 相加得到 uF'_n , 再通过环路滤波器构成了 F'_n 。

2.2 视频编解码的关键技术^[32]

2.2.1 帧内预测编码

为了进一步提高帧内编码效率, H.264 标准引入了考虑图像纹理的帧内预测技术。该方法定义了不同的预测方向来匹配视频图像的局部纹理, 16×16 亮度块可选用了 4 种预测模式, 4×4 块可选用 9 种预测模式。编码过程中, 基于率失真准则从这些备选模式中选出最优模式, 以达到最佳的压缩。

2.2.2 帧间预测编码

帧间预测的基本思想是将图像序列的每一帧划分为许多互不重叠的块, 并假定宏块内所有象素的位移都相同, 对每个目标块在参考帧中根据一定的匹配准则找到最相似的块, 即匹配块, 匹配块与目标块的相对位移 (i, j) 即为运动矢量, 二者相减即为残差块。解码器根据运动矢量和残差数据即可恢复出图像帧数据。

常见的运动估计匹配准则有三种: 绝对误差和(SAD); 均方误差(MSE)准则; 归一化互相关函数(NCCF)准则, 由于 SAD 没有乘除操作, 实现简单方便, 所以使用较多。

$$SAD(i, j) = \left| \sum_{m=0}^M \sum_{n=0}^N f_k(m, n) - f_{k-1}(m+i, n+j) \right|$$

$$MSE(i, j) = \frac{1}{MN} \sum_{m=0}^M \sum_{n=0}^N [f_k(m, n) - f_{k-1}(m+i, n+j)]^2$$

$$NCCF(i, j) = \frac{\sum_{m=0}^M \sum_{n=0}^N f_k(m, n) f_{k-1}(m+i, n+j)}{\left[\sum_{m=0}^M \sum_{n=0}^N f_k^2(m, n) \right]^{\frac{1}{2}} \left[\sum_{m=0}^M \sum_{n=0}^N f_{k-1}^2(m+i, n+j) \right]^{\frac{1}{2}}}$$

考虑到运动估计的复杂度较大，往往采用快速搜索算法来获取运动向量。一般从参考帧内的相同坐标点开始，采用固定的搜索模板和搜索策略得到最佳匹配块。著名的算法有：对数搜索法(LOGS)^[33]、三步法(TSS)^[34]、钻石搜索法(DS)^[35]、^[36]、六边形搜索法(HEXBS)^[37]、降采样法^{[38][39][40]}、标准化失真搜索法(NDS)^[41]，以及UMHexagonS^[42]等。

H.264 标准支持多种块尺寸的运动估计，存在从 16×16 到 4×4 的 7 块尺寸的模式。在帧间编码过程中，编码器采用以宏块为单位遍历的方式来进行模式选择。另外，H.264 引入了半像素精度以及 1/4 像素精度的插值方法，支持更高精度的运动向量。而运动向量的空间相关性也采用了向量预测技术来去除。

2.2.3 变换与量化

视频编码往往采用离散余弦变换 (DCT)^[43]将视频信息转换到频域上。由于自然图像的能量往往主要集中在低频部分，经过 DCT 变换的视频信息更利于熵编码的压缩。此外，它所产生的高频系数，由于人眼视觉对其不敏感，为大幅量化提供了可能，易于有损压缩。

量化是视频编码中达到压缩目的的主要手段，它通常是有损的。编码器针对信号的类别（亮度/色度）、频域系数的次数，采用不同的量化步长。量化步长的选取直接影响着输出码率。早期的量化算法均采用均匀量化，而考虑人眼视觉特性的非均匀量化矩阵也被提出并用于 H.264 的 High Profile^[44]。

2.2.4 环路滤波

块效应是基于块的编码方案难以回避的难题，H.264 采用环路滤波技术来缓解 DCT 变换以及量化造成的块效应。编码器对重建帧进行滤波，并列入参考块帧序列。解码器也对解码图像进行滤波。去块效应环路滤波器能够同时提高解码图像的主观和客观质量^[45]。

2.2.5 熵编码

对于运动补偿、变换、量化之后的视频数据，编码器采用熵编码来去除其统计

冗余。熵编码首先统计各语法元素的出现频度，再根据频度构建码表，然后进行编码。常用熵编码方法有，变长编码（Variable Length Coding, VLC）^[46]和算术编码（Arithmetic Coding, AC）^[47]。

H.264 采用了基于上下文信息的自适应编码变长/算数编码（CAVLC/CABAC），具有更高的压缩率以及更高的复杂度。

2.3 寻找新架构编码方案的必要性

如前所述，随着数字视频技术的发展，高清晰、高分辨率以及分辨率可调性逐渐成为了主流的需求。1920×1080 甚至 4096×2160 这样的超大尺寸视频开始走入了人们的生活。成立于 2002 年的美国 Digital Cinema Initiatives^[48]拒绝了 H.264 而选用 JPEG 2000 作为其数字电影的标准编码器^[49]。为什么集中了目前最先进的技术、付出了极大复杂度代价的 H.264 会在与静态图像编码标准 JPEG 2000 之间的竞争中败下阵来？各项研究表明，由于胶片颗粒的影响，H.264 的运动估计无法在高分辨率视频上准确匹配，因此对高清数字电影的编码效率明显低于 JPEG2000^[50]。同时数字电影压缩要求放映设备能根据实际需要方便地改变输出分辨率，而 ME/DCT 框架对此要求(scalability)并不完全契合，因此 H.264 的可分级扩展版本（Scalable Extension）并非没有代价

为此，本文研究了目前的图像编码主要策略，发现内插预测编码（interpolative coding）是以逐级填充的方式对图像进行编码的，其码流可在任意层级末尾截断，以匹配不同的分辨率需求，因此该编码机制能够无代价地兼容可分级视频编码（Scalable Video Coding），而且该体系的编码方案在极高质量的应用上压缩效率尤为突出^[29]。鉴于内插预测编码与目前视频技术需求的契合度，这种编码框架的研究价值和发展潜力都是值得看好的。因此，设计基于内插预测的视频编码方案，不仅是必要的，也是可行的。

2.4 本章小结

本章回顾了传统视频压缩技术的发展历程，并以目前最先进的 H.264 标准为例，介绍了主流视频编码的基本框架。此外，针对视频编码的关键技术——帧内预测、运动补偿、变换、量化以及熵编码等作了详细的介绍，这些传统的技术对于构建视频压缩的新框架也具有重要的意义。最后，指出了传统视频编码框架的问题，说明了内插预测编码对高质量、可分级编码等要求的适应性，这正是本文研究该领域的动机。

3. 基于内插预测的图像编解码技术

3.1 基于内插预测的图像编解码方案概述

Interpolative Coding (内插预测编码)自从 1978 年由 Hunt 提出以来已有 30 年历史。他的方案随后被 Burt 和 Adelson^[16]所完善, 基于这种思路的图像编码器之后被多个不同的组织实现。在本系列的预测方案中, 图像中的像素点被分配到由不同采样间隔划分的层级(band)上, 而编码器以由粗到细的顺序遍历预测所有层级, 预测思路是, 交替地用方形和菱形网格的 4 个顶点来预测其中心, 而参与预测的顶点, 都属于已编码的上级层级。

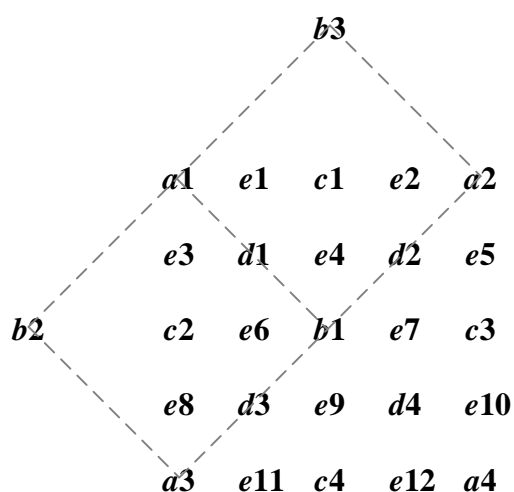


图 3.1 内插预测编码的分层结构

图 3.1 表示一幅图像中的一个 5×5 小块, 它描述了内插预测编码的层级分布结构。图像被分为从 a 到 e 五层, 每一层级的像素都由上一层级来预测。例如, 4 个标记为 a 的像素构成一个边长为 4 的正方形, 它们用以预测其中心 $b1$; 而示意图以外的所有其它 b 点也由其它 a 点预测, 这时所有的 b 点都被量化重构值 b' 取代, 可参与下一级的预测; 标 c 的点则由其所在的边长为 $2\sqrt{2}$ 且偏转 45° 的正方形的顶

点来预测，如 c_1 的预测由正方形(a_1, b_1, a_2, b_3)完成；而标 d 的像素点则由边长为 2 且未偏转的正方形顶点来预测，依此类推。不同的层级采用不同的量化参数。较高的层级（特别是 a 级），鉴于其重要性，采用较小的量化参数来量化其预测误差；而较低的层级，则采用较大的量化参数来量化。

3.2 Adaptive Prediction Trees 图像编解码方案

Adaptive Prediction Trees (APT) 是由 Robinson 2006 年提出一种基于内插预测的图像压缩方案^[25]。它所采用的空域自适应预测技术，被证明在大量不同类别的图像上都相当有效。同时，它所采用的基于主分量分析的颜色空间变换、六叉树存储结构，都对提高压缩效率起到了重要作用。

3.2.1 APT 空域自适应预测

APT空域自适应预测是APT的核心组成部分，如图3.2所示，该方法通过已编码的邻近像素集 $P_x = \{A, B, C, D, U, V\}$ 来预测目标像素 X 的灰度值。简单的双线性预测 $X_{pred} = (A+B+C+D)/4$ 通常不是最佳的预测方式。APT空域自适应预测器根据像素 $\{A, B, C, D\}$ 所构成的形状结构（定义为 $structure$ ）来决定 X 的预测公式。

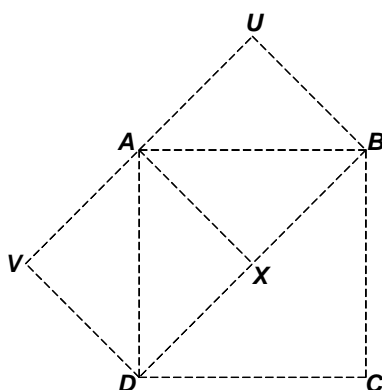


图 3.2 APT 空域自适应预测

定义 $R = (>, =, <)$ ，那么 X 周围4个点之间的关系(ARB, BRC, CRD, DRA)就有 $3^4 = 81$ 种可能，去掉某些不可能的情况（例如 $A>B>C>D>A$ ），Robinson对大量的图像进行了训练，得到不同的关系集(ARB, BRC, CRD, DRA)下与真实情况最接近的预测规

则，形成一套关系集和预测规则的对应关系。每条预测规则对应一个 $structure$ 。

APT空域自适应预测可分为两个部分。第一步，根据邻域像素的大小关系决定目标像素所在局部图像的形状特征，即 $structure$ 。第二步，根据 $structure$ 所对应的预测公式计算目标像素的预测值 X_{pred} 。该过程可由以下两个公式表示：

$$structure = Get_Structure(P_x) \quad (3-1)$$

$$X_{pred} = Adaptive_Pred(P_x, structure) \quad (3-2)$$

图 3.2 中， A, B, C, D 是距目标像素 X 最近的四个上层像素， U, V 是同级中距 X 最近的点。 (A, B, C, D) 构成一个正方形预测框，对于菱形的预测框如 (A, X, B, U) 只需将上述布局旋转 45 度，同理进行预测。

APT 空域自适应预测器的设计流程如下。

定义局部量度参数 M_i 来表征局部图像的结构特征，包括以下三种量度：

- 1) 预测像素集 $\{A, B, C, D, U, V\}$ 或者其某个子集的方差；
- 2) 预测像素集 $\{A, B, C, D, U, V\}$ 或者其某个子集的大小排列顺序；
- 3) 预测像素集 $\{A, B, C, D, U, V\}$ 的值的均等性。

对局部图像结构特征的描述：

- 1) $M_i \in \{\text{某子集或范围}\}$ 例如：方差 $<$ 某阈值
- 2) 上述不同局部量度的条件的任何逻辑组合 $\{\text{AND, OR, NOT}\}$ 。

预测等式：

$$X_{pred} = k_1A + k_2B + k_3C + k_4D + k_5U + k_6V \quad (\text{预测像素点的线性组合})$$

预测方式：

1) 选择预测器时依次考察一系列判断条件，这些判断按照给定顺序进行，第一个满足的条件所对应的计算公式被用于目标像素值的预测；

2) 可能的用于构成局部量度的子集包括： A, B, C, D, U, V 六个单独的点， $\{A, B\}, \{B, C\}, \{C, D\}, \{A, D\}, \{A, C\}, \{B, D\}$ ，四个上级邻近点 $\{A, B, C, D\}$ 以及所有六个邻近点 $\{A, B, C, D, U, V\}$ ；

3) 预测公式中的 k_i 取值为 0 或 $1/n$ ，这里 n 是非零 k_i 的个数，这意味着所有预测器的取值均为预测像素集 P_x 的某个子集的平均值。

基于以上设计，Robinson 在大量图像训练集上考察了各种形状分布，相应地找出了最佳预测规则，得出了以下自适应预测算法。

预测器依次考察以下条件，将满足条件所对应的预测公式的结果作为预测值

X_{pred} :

If $A = B = C = D$

预测值为 $X_{pred} = A = B = C = D$;

If $A = B$ AND $C = D$

If V 在图像中

$X_{pred} = V$;

Else

$X_{pred} = (A + B + C + D) / 4$

If $A = C$ AND $B = D$

If U 在图像中

$X_{pred} = V$;

Else

$X_{pred} = (A + B + C + D) / 4$

定义一个根据质量参数计算出的阈值 T ,

If $\{A, B, C, D\}$ 中的最大值和最小值之间的差值小于 T ,

$X_{pred} = (A + B + C + D) / 4$

If $\{A, B, C, D\}$ 中的三个点有相等的灰度值 P , 而剩下一个为不等于 P 的值 Q

$X_{pred} = P$;

If $B > A, D$ AND $C > A, D$ OR $B < A, D$ AND $C < A, D$

相应的分布结构为线状，定义两个预测值，山脊值 $R1$ 和山谷值 $R2$:

If $B > A, D$ AND $C > A, D$

$R1 = (B + C) / 2$;

$R2 = (A + D) / 2$;

Else

$$R1 = (A + D) / 2;$$

$$R2 = (B + C) / 2;$$

If X 的实际值更接近于 $R1$

在码流中用一个比特 1 来指示预测值为 $R1$;

Else

在码流中用一个比特 0 来指示预测值为 $R2$;

Else

$X_{pred} = \text{median}(A, B, C, D)$, 即中值预测。

3.2.2 APT 的编码流程

图 3.3 描述了 APT 的主要编码过程，解码器是编码的逆过程，只在预滤波和预测误差的量化这两个阶段与编码端的处理有所不同。APT 主要通过去除输入图像像素间的颜色轴间冗余和空域冗余，达到压缩图像的目的。编码器提供一个可在码率和失真之间灵活调整的质量参数 Q ，取值为 0~100，对应的图像质量由坏到好，包括无损。用户可以自行调整 Q 值以满足不同应用的需求。一般地，图像的率失真控制主要由量化步长 S_k 决定，因此需建立 Q 与 S_k 的关系。其中下标 k 表示第 k 层级的像素的量化步长。实验表明，采用下述经验公式具有较好的系统性能，APT 的量化器即是基于下列公式设计的。

$$S_k = \max\left(0.8^{M-k} \frac{(100-Q)^2}{100}, 1\right)$$

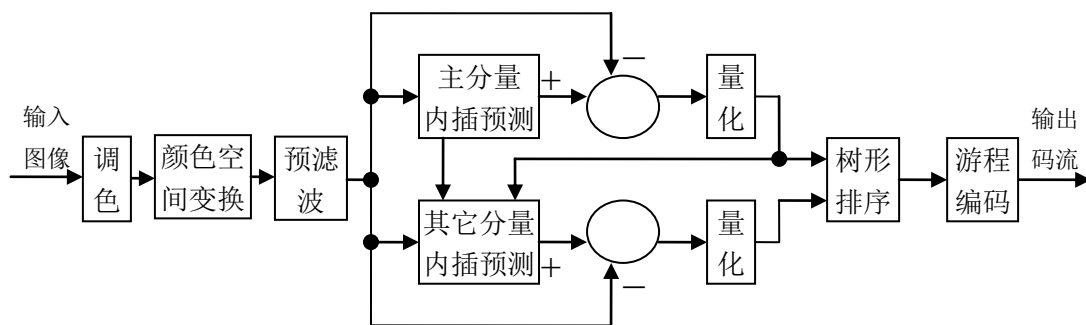


图 3.3 APT 的编码流程

下面依次对 APT 各组成模块分别作详细介绍。

1) 调色盘

某些图像所含颜色较少,即输入图像的颜色空间集所包含元素的个数远小于颜色空间所能包含的最大的元素个数,若仍采用常规方式编码,必然浪费存储空间。因此调色技术可将输入图像的颜色空间集所含元素映射到一个较小的子空间中再进行编码。

调色技术的关键在于选取合适的调色板,即映射方式。映射的基本原则是使得在调色板中相邻的元素在图像中也相邻,以保持原始图像的空域相关性尽可能不变,不影响空域冗余的去除。

编码器首先根据输入图像的色彩元素个数、出现频率以及质量参数 Q 决定是否采用调色技术。对于需要应用调色技术的图像,选取适当的算法选取调色盘进行映射,文献[51]介绍了常见的调色盘选取算法,这些方法的性能大体相当。

APT在经过大量测试后给出了启用调色的判断准则:

- a) 输入图像的颜色种类小于10种;
- b) 输入图像的颜色种类小于256种,且预先设定的 Q 值大于70;
- c) 输入图像的颜色种类小于4000种,且预先设定的 Q 值等于100,且输入图像的尺寸大小超过0.25M像素;

2) 颜色空间预处理

由于输入图像的颜色空间存在轴间冗余,即红、绿、蓝三个颜色轴之间存在着相关性,故可以通过某种数学变换将图像信号转到新的颜色空间,达到消除冗余的目的。

对于无损压缩,APT采用固定的可逆多分量变换;对于有损压缩,APT则采用主分量分析法进行颜色空间变换(即K-L变换,K-L变换已经被应用于小波变换编码中)。主分量分析通过对角化输入像素的RGB分量的互相关矩阵,获得相应的颜色变换空间的基,新的颜色空间中,主轴上具有最大的方差,实验表明,对于通用的图像集而言,采用颜色空间的主分量分析法,能够显著提高压缩效率。

3) 基于统计排序的预滤波技术

内插预测的准确度在带噪声图像上显然会大打折扣，严重影响压缩效率，更可能使得重构图像的speckling artifact（闪烁效应）更为严重。为抑制这种情况，APT通常对输入图像统计进行预滤波，该方法一定程度上降低了重建图像的客观质量，但能提升主观质量，缓解闪烁效应。

为了不使预滤波对真实的图像纹理进行平滑操作，为此，APT采用了非线性的处理办法，给出了滤波门限 $R = \min(90 - Q, 80)$ 。如果目标像素周围的象素的变化范围小于门限值 R ，则进行统计滤波。具体处理如下：对待目标像素的 3×3 邻域值进行排序，如果待编码像素值不小于排序表中第 k 大值或不大于第 k 小的值，则用对应的第 k 大或第 k 小的值来取代。这样，待编码的像素值就接近变化范围的中间值。如果 $R < 40$ ，则 $k=3$ ；其它情况， $k=4$ 。

该方法通过去除图像中可能的噪声，使之更符合APT空域预测特性，缓解了闪烁效应。

4) 主分量预测

主分量上集中了图像的大部分能量，主轴预测的准确与否，直接决定了整个系统的编码效率。

主分量的灰度值按照金字塔结构进行预测。首先对图像进行稀疏采样，作为顶层像素（如图3.1中的 $a1, a2, a3, a4$ ），它们直接利用DPCM方式保存。接着按照3.1节介绍的次序依次对其它各层（ $b \sim e$ 层）像素进行预测，直到把整个图像填满。预测误差根据所在的层级被量化，准备用于重排序和熵编码。

5) 其它分量预测技术

鉴于其它分量与主分量的相关性，APT利用主分量预测时决定的 *structure* 对其它分量进行预测。但在预测开始前会对主分量决定的 *structure* 进行修正，依次考察主轴上各 *structure* 所对应的预测值与重构值的差别，若主分量预测决定的 *structure* 不是最佳选择，其它分量将采用最佳的 *structure* 进行预测（主分量所用的 *structure* 不能修改）。实验表明，这种修正机制能将码率降低大约5%。

6) 树形结构重排序

APT对量化的预测误差按照树形结构进行重排序。其树形分布及结构如图3.4和图3.5所示。象素B1是C1和C2的父节点。对于RGB彩色图像而言，一个非叶子节点的拥有六个子节点（每个子像素的3个轴具有的3个灰度值）。编码器引入一种“终结符”表示叶子节点（可能的子树上全为零的节点），叶子节点原本应有的子节点上的则无需比特表示。APT对每一个未被终结的节点插入一个标记来指示其值本身是否为零以及是否叶子节点。这种方法能够跳过很多有规律出现的零码值，在静态图像压缩上有较好的效果。

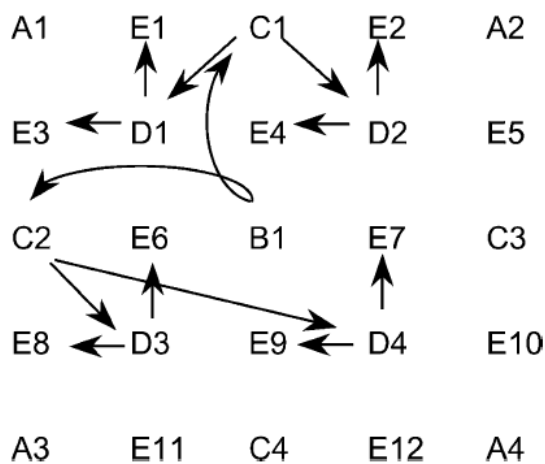


图 3.4 像素点在 APT 预测树上的分布

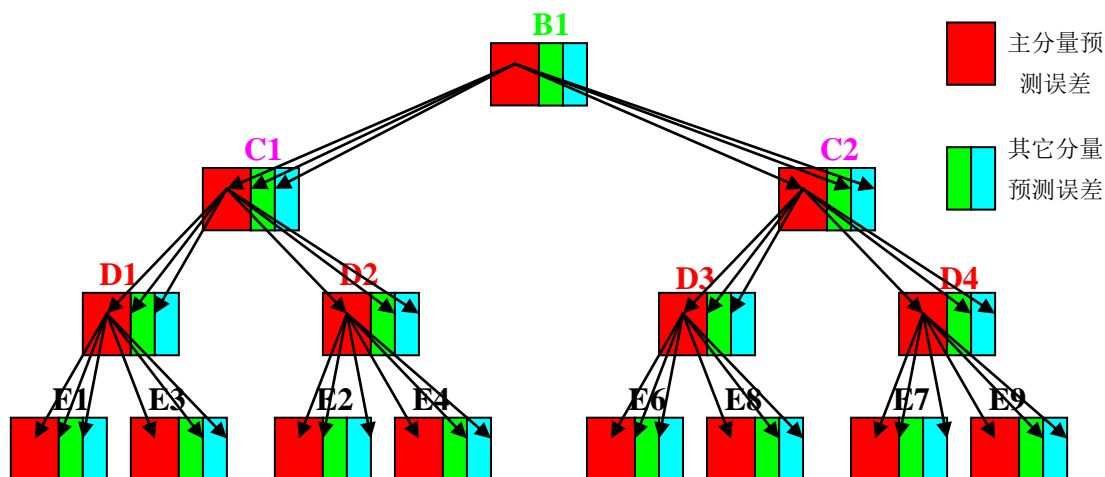


图 3.5 APT 的六叉树结构

7) 熵编码

APT的熵编码和重排序紧密结合在一起，在寻找叶子节点的同时统计预测误差的

分布特性用于构建熵编码表。编码器从顶层像素（树的根结点）开始，遍历所有六叉树上的预测误差，从最低层级开始，对所有分量为零的节点向上结合，直到遇到非零节点，插入“终结符”，即为叶子节点。

八个独立的游程哈夫曼编码器分成每四个一组同时运行：其中四个用于“方形预测”的格点，另外四个用于“菱形预测”的格点。对于每组编码器，其中之一用于编码各节点的主分量灰度值是否为零以及是否为叶子节点，另外三个用于编码三个分量的非零预测误差值。

最后采用最高频率的无损游程哈夫曼编码来去除预测误差中的统计冗余。其基本流程如下：编码器首先找到消息队列中出现频率最高的符号 S ，首先对符号 S 和非符号 S 的游程进行编码，然后对非符号 S 序列编码。游程和符号均用标准的哈夫曼编码器进行压缩。哈夫曼编码器需两次遍历消息队列，一次提取其统计特性，构建码表，一次根据码表对消息进行编码。为提高压缩效率，APT还提供另一种不传码表的哈夫曼编码方式，即认为消息队列近似服从为拉普拉斯分布，以内置码表进行编码。

3.3 APT 的优势及其局限性

APT在图像压缩上的成功为内插预测编码赢得了更多的关注。点到点的预测机制决定了APT对图像细节纹理的保持优于其它编码方案（如JPEG 2000），却也导致了闪烁效应（speckling artifact），这是内插预测编码急需解决的难题；APT空域自适应预测的准确性保证了APT较高的压缩效率（高于JPEG而低于JPEG 2000）。值得注意的是，APT空域自适应预测器的设计是经验性的，取决于训练图像的统计特征。APT空域自适应预测器的精度只有在图像内容变化相对规则的情况下才能够保证。因此，APT在空间相关性较强的连续色调图像（continuous-tone image）上有较好效果，但在不可预测的混乱图像上，其空域自适应预测器则无法发挥作用，甚至起到适得其反的效果。

简言之，APT空域自适应预测器的使用必须以目标图像类型与训练图像集相似为前提。文献[28]在将该预测器用于视频编码时，试图用空域自适应预测对视频序列的

运动估计残差进行编码，其性能就会受很大限制。残差图像并非常规的自然图像，预测器在设计过程中并不会考虑这种类型的图像，所以预测器在残差图像上的预测精度大大降低，影响了压缩效率。关于该问题5.3.1节还将作更详尽的分析。

综上所述，在基于内插预测的视频编码器的设计过程中，必须尽可能保证APT空域自适应预测都在原始图像上进行，这正是本文下面提出的时空联合预测的设计理念。

3.4 本章小结

本章描述了用于图像压缩的内插预测编码的基本思路，介绍了该领域最成功的图像压缩方案——APT，其基本组成模块包括调色盘、颜色空间变换、预滤波、空域自适应预测器、量化器、重排序、熵编码器。重点研究了 APT 的核心技术——空域自适应预测，这是本文提出的时空联合预测的基础。

4. 时空联合预测模型和方法

4.1 时空联合预测方法

传统的视频编码器采用帧内/帧间预测来去除视频序列中的相关性，达到压缩的目的。然而，在有些情况下视频序列的时间相关性和空间相关性都不占据主导地位。为此，本文提出时空联合预测（Joint Spatiotemporal Prediction, JSP）的方法，同时考虑时间和空间相关性，理论分析和实验结果表明，时空联合预测在多数情况下比帧内/帧间预测具有更高的准确度。

如 3.3 节所述，时空联合预测的设计理念在于保证空域自适应预测必须在原始图像上进行。因此对视频数据必然要经过两次预测——当前帧和参考帧分别进行，二者之间利用运动向量相关联，即通过运动向量定位目标像素在参考帧中的相应点，对其作空域预测。而时空联合预测通过参考帧中的预测误差来估计当前帧的预测误差，这样，时空联合预测的误差就等于当前帧和参考帧的 APT 空域自适应预测误差之差。

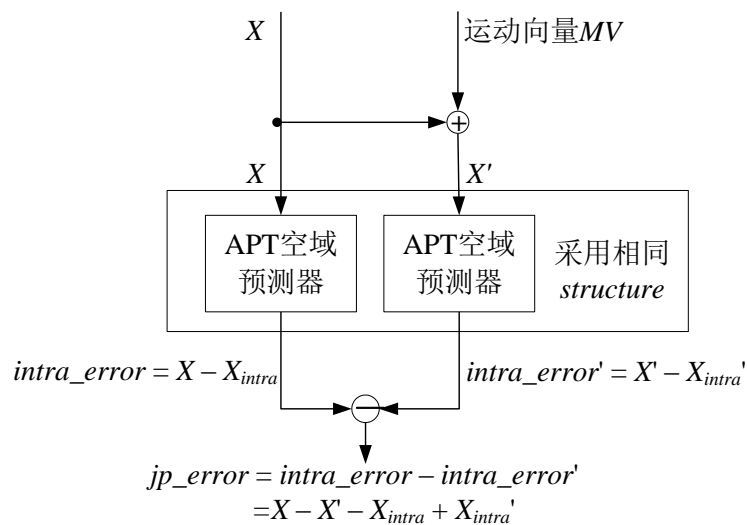


图 4.1 时空联合预测的示意图

如图 3.2 所示，像素点 $X(x, y)$ 有六个可能的临近预测点，它们构成了 X 的预测像素集 $P_x = \{A, B, C, D, U, T\}$ 。通过运动估计得到 X 的运动矢量 (mv_x, mv_y) 。同理， X 在参考帧中的匹配像素点 $X' = (x+mv_x, y+mv_y)$ 的预测像素集为 $P_{x'} = \{A', B', C', D', U', T'\}$ 。 P_x 和 $P_{x'}$ 将用于计算 X 的时空联合预测值 X_{jp} ，见图 4.1。

1) 对于当前帧中的像素点 X ，用它的预测集 P_x 进行空域自适应预测得到 *structure* 特征和预测误差。

$$\begin{aligned} structure &= Get_Structure(P_x) \\ X_{intra} &= Adaptive_Pred(P_x, structure) \\ intra_error &= X - X_{intra} \end{aligned} \quad (4-1)$$

公式(4-1)所用的函数的定义已在 3.2 节中给出。 X_{intra} 为帧内预测值；*intra_error* 为其误差。

2) 用相同的 *structure* 和 X' 的预测集 $P_{x'}$ 来估计参考像素 X' 。

例如，如果 X 的估计值是 $(A+B+C+D)/4$ ，则 X' 的估计值是 $(A'+B'+C'+D')/4$ ；如果 X 的估计值是 $(A+C)/2$ ，则 X' 的估计值是 $(A'+C')/2$ 。

$$\begin{aligned} X_{intra}' &= Adaptive_Pred(P_{x'}, structure) \\ intra_error' &= X' - X_{intra}' \end{aligned} \quad (4-2)$$

3) 时空联合预测值 X_{jp} 由参考帧得到的 X 的运动估计预测值，加上由当前帧中用 APT 空域预测得到的 X 的帧内预测值，再减去由参考帧中用 APT 空域预测得到的 X' 的预测值而得出。时空联合预测值 X_{jp} 及其误差 *jp_error* 可表示如下：

$$\begin{aligned} X_{jp} &= X' + X_{intra} - X_{intra}' \\ jp_error &= X - X_{jp} \\ &= X - (X' + X_{intra} - X_{intra}') \\ &= (X - X') - (X_{intra} - X_{intra}') \\ &= mc_error - pmc_error \end{aligned} \quad (4-3)$$

公式(4-3)中，*mc_error* 和 *pmc_error* 可以被认为分别是在真实域上运动估计的误差和预测域中的误差。

预测编码都是致力于尽可能地减小预测误差。由于 APT 在静止图像的有很好和

压缩率，而运动补偿技术在规则运动的视频序列上也有很好的效果，时空联合预测将两者的优势结合起来，其预测误差 jp_error ，在多数情况下都优于传统的运动补偿残差 mc_error 。

解码方面，解码器通过以下步骤来获得像素的重构值：

- 1) 用 P_x 预测 X , 确定 $structure$, 求出估计值 X_{intra} 。
- 2) 用 P_x' 预测 X' 来确定 $structure$ 和估计值 X_{intra}' 。
- 3) 时空联合预测值 $X_{jp} = X' + X_{intra} - X_{intra}'$ 。
- 4) 解码得到的重构值 $X_{dec} = X_{jp} + jp_error$

4.2 时空联合预测的准确性分析

目前流行的预测方法都是利用运动估计在参考帧内寻找最佳匹配块，而利用参考帧内的匹配点来估计目标像素的灰度值。但是某些视频序列中的帧间相关性是有限的，这种方法的性能就很难保证。时空联合预测的方法兼顾了时间和空间的相关性，往往能得到更准确的估计值。为了证明时空联合预测的优越性，本文对时空联合预测误差和运动估计残差的方差进行了对比研究。

定理 4.1: 只要 mc_error 和 pmc_error 的相关系数 $\rho > 0.5$ ，则不等式 $D[jp_error] < D[mc_error]$ 成立。亦即，时空联合预测误差的方差小于运动估计残差的方差。

证明：假设 $D[jp_error]$ 和 $D[mc_error]$ 分别是时空联合预测误差和运动估计残差的方差； $E[X]$ 是随机变量 X 的期望。由于运动估计残差服从零均值的拉普拉斯分布^[52]，即 $E[mc_error]=0$ 。

假设 P_i 是参与像素 X 预测的邻近点之一， k_i 是它的权重，那么由 APT 空域预测得到的估计值 X_{intra} 则可表示为：
$$X_{intra} = \sum_{i=1}^n k_i P_i,$$

由 APT 的预测规则可知 $\sum_{i=1}^n k_i = 1$ ($n=1,2,4$);

$$\begin{aligned}
pmc_error &= X_{intra} - X'_{intra} \\
&= \sum_{i=1}^n k_i P_i - \sum_{i=1}^n k_i P'_i = \sum_{i=1}^n k_i (P_i - P'_i) \\
&= \sum_{i=1}^n k_i mc_error
\end{aligned}$$

$$\therefore E[mc_error] = E[pmc_error]$$

$$\begin{aligned}
D[mc_error] &= E[mc_error^2] - E^2[mc_error] \\
&= E[mc_error^2]
\end{aligned}$$

类似的, $D[pmc_error] = E[pmc_error^2]$.

对于 $n=2$ 的情况,

$$pmc_error = k_1 \cdot mc_error_1 + k_2 \cdot mc_error_2 \quad (k_1 + k_2 = 1)$$

$$\begin{aligned}
&E[pmc_error^2] - E[mc_error^2] \\
&= E[(k_1 \cdot mc_error_1 + k_2 \cdot mc_error_2)^2] - E[mc_error^2] \\
&= (k_1^2 + k_2^2)E[mc_error^2] + 2k_1 k_2 E[mc_error_1 \cdot mc_error_2] \\
&\quad - (k_1 + k_2)^2 E[mc_error^2] \\
&= 2k_1 k_2 (E[mc_error_1 \cdot mc_error_2] - E[mc_error^2])
\end{aligned}$$

由于 mc_error_1 、 mc_error_2 、 mc_error 都服从拉普拉斯分布, 根据施瓦茨不等式可得 $E[mc_error_1 \cdot mc_error_2] \leq E[mc_error^2]$ 。

$$\therefore E[pmc_error^2] - E[mc_error^2] \leq 0$$

对于 $n=4$ 的情况, 可由数学归纳法推出

$$E[pmc_error^2] = E\left[\left(\sum_{i=1}^n k_i mc_error\right)^2\right] \leq E[mc_error^2]。$$

$$\begin{aligned}
&D[jp_error] - D[mc_error] \\
&= E[(mc_error - pmc_error)^2] \\
&\quad - (E[mc_error - pmc_error])^2 - E[mc_error^2] \\
&= E[pmc_error^2] - 2E[mc_error \cdot pmc_error]
\end{aligned}$$

定义 mc_error 和 pmc_error 的相关系数为 ρ ,

$$\begin{aligned}\rho &= \frac{E[mc_error \cdot pmc_error] - E[mc_error] \cdot E[pmc_error]}{\sqrt{D[mc_error]} \cdot \sqrt{D[pmc_error]}} \\ &= \frac{E[mc_error \cdot pmc_error]}{\sqrt{D[mc_error]} \cdot \sqrt{D[pmc_error]}} \\ \therefore E[mc_error \cdot pmc_error] &= \rho \sqrt{D[mc_error]} \sqrt{D[pmc_error]}\end{aligned}$$

带入前面的公式可得

$$\begin{aligned}& D[jp_error] - D[mc_error] \\ &= E[pmc_error^2] - 2\rho \cdot \sqrt{E[mc_error^2]} \cdot \sqrt{E[pmc_error^2]} \\ &= \sqrt{E[mc_error^2]} \cdot (\sqrt{E[pmc_error^2]} - 2\rho \cdot \sqrt{E[mc_error^2]}) \\ &\quad (\text{As proved, } E[pmc_error^2] \leq E[mc_error^2]) \\ &\leq \sqrt{E[pmc_error^2]} \cdot (\sqrt{E[mc_error^2]} - 2\rho \cdot \sqrt{E[mc_error^2]}) \\ &= (1 - 2\rho) \cdot \sqrt{E[pmc_error^2]} \cdot \sqrt{E[mc_error^2]}\end{aligned}$$

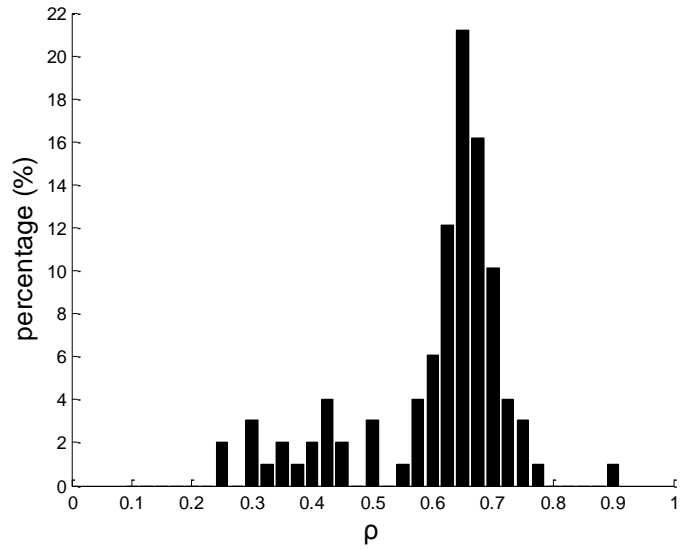
$$\rho > 0.5 \Rightarrow D[jp_error] < D[mc_error] \quad \text{证毕。}$$

考虑到 $pmc_error = \sum_{i=1}^n k_i mc_error_i$ 和 APT 空域预测的准确性，通常可认为

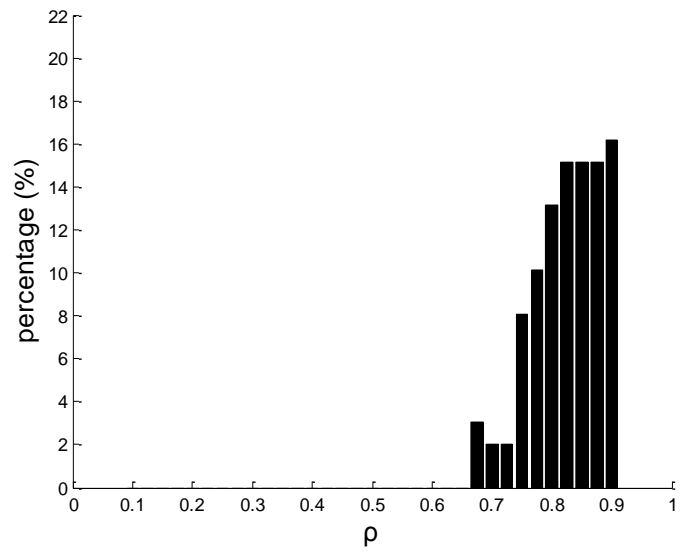
mc_error 和 pmc_error 是高度相关的，这意味着 $\rho > 0.5$ 的假设在多数情况下都成立。

本文统计了 8 个从静态到剧烈运动的 100 帧的视频序列的 mc_error 和 pmc_error 分布。此处采用经验性的方法对连续视频帧间的 ρ 值进行计算，图 4.2 列出了 ρ 值的分布。显而易见，无论是在运动非常平缓的 *News* 序列还是运动十分剧烈的 *Football* 序列， ρ 值都很少小于 0.5。因此 $D[jp_error] < D[mc_error]$ 的结论在绝大多数情况下是成立的。表 4.1 和表 4.2 给出的在 *Stefan* 和 *Bus* 序列各帧上的时空联合预测误差和运动估计残差的 SAD 之和的对比关系也证实了这一点。

因此，对于绝大多数视频序列，时空联合预测要比传统的运动估计准确。



(a) *News*



(b) *Football*

图 4.2 100 帧视频序列 *News* 和 *Football* 上的相关系数 ρ 值分布

表4.1 时空联合预测与运动补偿的绝对误差和（SAD）在100帧Stefan序列上的对比

NO.	Residual	JP_Error	NO.	Residual	JP_Error	NO.	Residual	JP_Error	NO.	Residual	JP_Error
1	293473	243259	26	186306	174982	51	234260	202024	76	278953	249647
2	288213	237354	27	172563	167248	52	228798	201066	77	271233	245099
3	262968	210602	28	126896	128524	53	237788	206547	78	259036	233543
4	267091	219184	29	132553	133189	54	262385	212613	79	256689	231291
5	281216	235941	30	202873	187767	55	269736	218697	80	253364	230499
6	319579	259898	31	239398	215288	56	262218	217765	81	282059	244474
7	310673	246337	32	261677	233600	57	262424	219528	82	258253	235878
8	319801	251220	33	260885	237600	58	260410	221641	83	272010	241146
9	334294	260174	34	259956	238512	59	267561	230989	84	323301	278236
10	332271	260340	35	277110	250935	60	280032	241665	85	317025	265322
11	345754	269490	36	284316	257558	61	284849	243654	86	332816	273625
12	338282	263896	37	281023	254058	62	279390	242471	87	364362	290747
13	327244	257794	38	277594	245731	63	295941	255643	88	375289	293724
14	318958	253490	39	273583	247044	64	289592	249718	89	396499	293874
15	329485	259619	40	275638	247835	65	282541	246924	90	444104	310622
16	322844	264922	41	281972	254782	66	278399	241877	91	387779	277436
17	310228	253727	42	278575	246466	67	291390	252241	92	478586	326679
18	312774	252567	43	266134	230351	68	284156	250823	93	412542	289830
19	295802	246668	44	214223	187912	69	270857	240425	94	458009	326021
20	273923	238728	45	171595	157489	70	270956	237421	95	444673	318269
21	284127	245258	46	220675	198458	71	249221	218400	96	365688	276277
22	266316	232766	47	279708	239664	72	251996	217567	97	384187	286247
23	254380	226946	48	267904	228844	73	258223	228082	98	351373	264580
24	164952	155634	49	221142	183602	74	272506	240354	99	353923	268238
25	169473	160846	50	201525	175224	75	277876	246827			

表 4.2 时空联合预测与运动补偿的绝对误差和 (SAD) 在 100 帧 *Bus* 序列上的对比

NO.	Residual	JP_Error	NO.	Residual	JP_Error	NO.	Residual	JP_Error	NO.	Residual	JP_Error
1	293473	243259	26	186306	174982	51	234260	202024	76	278953	249647
2	288213	237354	27	172563	167248	52	228798	201066	77	271233	245099
3	262968	210602	28	126896	128524	53	237788	206547	78	259036	233543
4	267091	219184	29	132553	133189	54	262385	212613	79	256689	231291
5	281216	235941	30	202873	187767	55	269736	218697	80	253364	230499
6	319579	259898	31	239398	215288	56	262218	217765	81	282059	244474
7	310673	246337	32	261677	233600	57	262424	219528	82	258253	235878
8	319801	251220	33	260885	237600	58	260410	221641	83	272010	241146
9	334294	260174	34	259956	238512	59	267561	230989	84	323301	278236
10	332271	260340	35	277110	250935	60	280032	241665	85	317025	265322
11	345754	269490	36	284316	257558	61	284849	243654	86	332816	273625
12	338282	263896	37	281023	254058	62	279390	242471	87	364362	290747
13	327244	257794	38	277594	245731	63	295941	255643	88	375289	293724
14	318958	253490	39	273583	247044	64	289592	249718	89	396499	293874
15	329485	259619	40	275638	247835	65	282541	246924	90	444104	310622
16	322844	264922	41	281972	254782	66	278399	241877	91	387779	277436
17	310228	253727	42	278575	246466	67	291390	252241	92	478586	326679
18	312774	252567	43	266134	230351	68	284156	250823	93	412542	289830
19	295802	246668	44	214223	187912	69	270857	240425	94	458009	326021
20	273923	238728	45	171595	157489	70	270956	237421	95	444673	318269
21	284127	245258	46	220675	198458	71	249221	218400	96	365688	276277
22	266316	232766	47	279708	239664	72	251996	217567	97	384187	286247
23	254380	226946	48	267904	228844	73	258223	228082	98	351373	264580
24	164952	155634	49	221142	183602	74	272506	240354	99	353923	268238
25	169473	160846	50	201525	175224	75	277876	246827			

4.3 编码方案的预测模式配置

视频序列的不同特点适用不同的预测方法。尽管在很多情况中，时空联合预测比运动估计更加准确，但对于某些运动非常规则或非常混乱的视频序列，直接的运动补偿和帧内预测往往会更有效。因此，多种预测模式是必需的。针对视频序列的各种特点，JPC考虑了以下4种模式：

跳过（Skip）模式：选择这种模式的块中直接复制参考帧中的相同坐标的块作为本帧的像素点。

运动补偿（MC）模式：该模式块内的像素点的被预测为参考帧中的 X' 。与传统编码中的运动估计相同。

时空联合预测（JSP）模式：该模式块内的像素点的被预测为4.1节介绍的 X_{jp} 。实验表明，常见视频序列中大多数块都选用这种模式。

帧内（Intra）模式：该模式块内的像素点被预测为 X_{intra} ，直接由APT预测得到。

为了考察除了JSP模式外其它模式存在的必要性，本文在不同的质量参数下对各类视频序列作了测试。依次去掉每一种模式，并分别计算所引起预测SAD的增量

表4.3列出了去掉每一种模式后SAD的平均增量。这表明这三种模式对预测准确性都存在很重要的影响，因此都被选用到JPC系统之中。

采用了模式选择之后，JPC的预测误差比传统的帧间/帧内预测预测残差降低了24.6%。

表4.3 去掉各模式后的效果

序列	关闭 Skip	关闭 MC	关闭 Intra
<i>Foreman</i>	2.77%	7.86%	41.38%
<i>News</i>	21.8%	131.67%	10.39%
<i>Bus</i>	0.12%	12.17%	15.56%
<i>Football</i>	2.04%	-5.77%	55.27%
<i>Stefan</i>	3.16%	13.17%	11.43%
<i>Tempete</i>	2.45%	21.11%	9.99%
Average	5.39%	30.04%	24.00%

4.4 本章小结

本章提出了一种全新的视频编码帧间预测方式——时空联合预测（JSP）。和其它的预测方式相比，JSP 最大的优势在于同时考虑了时间和空间相关性，本章从理论和实验两方面证明，时空联合预测在绝大多数情况下优于传统的运动补偿技术。

由于视频序列的活动性各不相同，单一模式的预测往往无法满足各类视频的特点，本章提出了 Skip、MC、JSP、Intra 四种编码模式并分析了它们存在的必要性。

5. 基于时空联合预测的视频编码方案

5.1 联合预测视频编码的基本框架

Joint Predictive Coding 是本文依据对传统视频编码框架和内插预测技术的分析，以时空联合预测为基础，结合 KLT/Joint Prediction 的新框架，提出的新型视频编码方案。图 5.1 列出了 Joint Predictive Coding 的主要功能模块，具体编码流程如下：

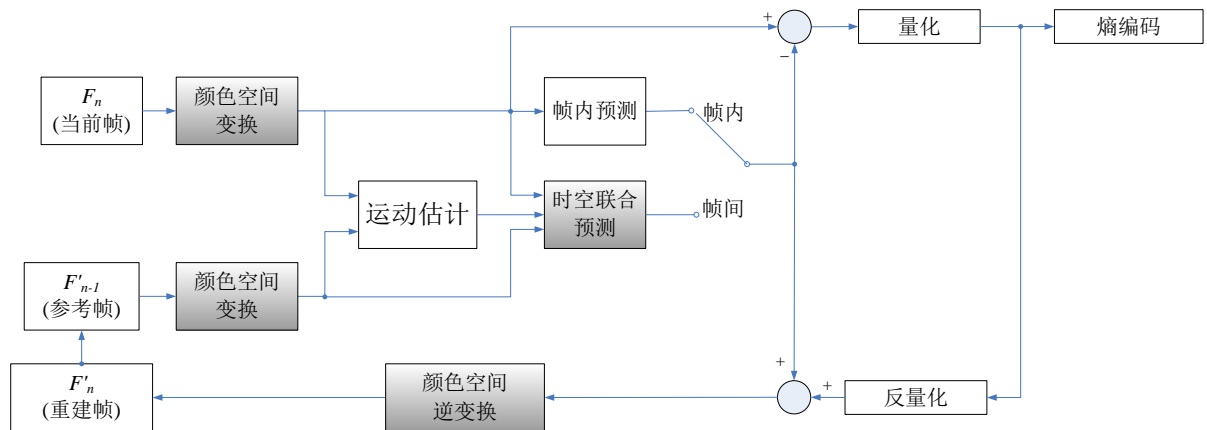


图 5.1 Joint Predictive Coding 的编码流程

1) 预处理

JPC编解码器的量化技术与APT相同，首先根据可调的质量参数计算得到各级的量化间隔。然后以4:4:4的格式从视频序列中读取一帧的YUV各分量数据（因为KLT需要所有分量的取样区间都是一样的）。为了后续的运动估计和内插预测，帧数据的边界以直接复制的方式被扩展 n 个像素（ n 由运动估计的块尺寸决定）。

2) 颜色空间变换

为获得更高的压缩率，JPC对于质量参数低于某经验值的视频帧采用主分量分析（PCA）来进行颜色空间变换。当前帧的每一个像素点的三个分量值被加到一个三维样本空间，获取其协方差矩阵，从而获取矩阵的特征值和特征向量。其特征值和特征向量用于该矩阵的对角化，得到PCA颜色空间的三个基，选取最大特征值所对应的特征向量作为主分量的基。

当前帧和参考帧都被转换为PCA颜色空间，于是当前帧在新颜色空间中拥有最大的主分量能量。因此，运动估计和Structure的判定都在主分量上完成。PCA颜色空间变换在其它的图像编码方案中也曾被用到^[53]，然而JPC将其与帧间预测相融合用于视频压缩，尚无先例。

3) 运动估计

被转换到PCA颜色空间上的参考帧首先经过六抽头滤波器插值^[54]构成半像素精度的参考帧。在当前帧的主分量上，JPC编码对每一个16×16块进行半像素精度的运动向量搜索。为每个像素点分配了一个运动向量(mv_x, mv_y)。

4) 模式选择

对于不同特征视频序列的块，JPC有4种模式，分别为：Skip、MC、JSP和Intra模式。首先，编码器根据运动估计得到的SAD的值来决定是否采用Skip模式。然后，用MC、JSP和Intra模式对当前帧进行初步预测。在每一个16×16块中计算出这三种模式的均方误差： MSE_{MC} ， MSE_{JSP} 和 MSE_{INTRA} ，然后选择MSE值最小的模式（由于帧内模式无需传输运动向量， MSE_{INTRA} 在参与比较之前要预先乘上了一个小于1的经验系数）。之后模式信息和运动向量被写入比特流中。

5) 预测和重构

JPC依次遍历视频帧的YUV三个轴，对帧内所有的像素点以层序进行地预测。估计值 X_{pred} 是由它的位置和模式确定的。预测初始化时，最高层级（ α 级）上的像素点被估计为参考像素点(X')或最近的邻点(N)。假设最高层级的取样间隔为 L ,

$$X_{pred} = \begin{cases} N(x, y - L) & (\text{mode} = \text{intra and } x = 0) \\ N(x - L, y) & (\text{mode} = \text{intra and } x \neq 0) \\ X' & (\text{mode} \neq \text{intra}) \end{cases} \quad (5-1)$$

对于其它层级的像素点，都采用步骤4)所确定的模式进行预测。

$$X_{pred} = \begin{cases} X' & (\text{Motion Compensation mode}) \\ X_{jp} & (\text{Joint Prediction mode}) \\ X_{intra} & (\text{Intra mode}) \end{cases} \quad (5-2)$$

预测误差 $error_{pred}$ 由步骤1)所得的量化参数进行量化，并重构为 $error_{rec}$ 。X的重构

值为 $X_{rec} = X_{pred} + error_{rec}$ 。在解码端， X 的输出值 $X_{dec} = X_{rec}$ 。 X_{rec} 和 X 由于误差 $error_{pred}$ 被量化而存在差异。因此， X_{rec} 而不是 X 被用于预测更低层级的像素。为充分利用轴间相关性以及节省编码时间，对第二个和第三个分量进行预测时，JPC不再作模式选择和 $structure$ 判断，而直接就采用主分量的模式和 $structure$ 了。

本环节后，3个分量上的量化过的 jp_errors 误差矩阵已经生成，准备写入码流。当前帧被重构，转回YUV颜色空间，经过去闪烁滤波器，最后保存为参考帧。

6) 重排序和熵编码

JPC在熵编码前首先要进行重排序。预测误差按照层序被重排，以得到更多的连续零，特别是在被大幅量化的层级上，这能使变长编码获得更好的效率。

最后，三个独立的熵编码器对YUV三轴上的预测误差进行常规的哈夫曼变长编码。

7) 进入下一帧

5.2 针对闪烁效应的滤波算法设计

峰值信噪比（PSNR）是目前最常用的客观质量评价标准。然而，由于人类视觉系统的特性，PSNR的值有时并不符合人眼视觉上的质量的好坏。5.5节给出了JPC和H.264在相等PSNR下的解码图像，结果表明JPC的视频质量更好而且没有H.264的块效应。然而，正如所有的内插预测一样，JPC也有它自身的问题：如图5.2(a)所示的闪烁效应（speckling artifact）。尤其是在低码率压缩的情况下，这种闪烁噪声将十分明显，必须除去。因此，本文提出了一种去闪烁滤波器。

鉴于speckling artifact和椒盐噪声的相似性，因此统计滤波被选为内插预测编码的初步解决方案。每一个像素都经过如下处理：

对像素的 3×3 邻域各点的值进行排序。如果中心像素点的值高于/低于第 k 最大值/最小值，那么它的值就用第 k 最大值/最小值替代。这里 k 是一个经验值，在APT中它根据情况被设为3或4。

本文研究了一系列视频序列，发现闪烁噪声破坏了连续帧之间的内容的平滑度，这是视觉质量下降的关键所在。为了提高连续帧间的平滑度，JPC将每个像素点的 3×3

邻域和参考像素点的 3×3 邻域一起用于滤波处理。因此，新的滤波器对这18个像素点进行排序，如果中心像素点的值高于/低于第 $2k$ 最大值/最小值，那么它的值就用第 $2k$ 最大值/最小值替代。

视频序列的滤波通常有3种不同的使用方法。预滤波：在编码前对图像源进行处理；后滤波：在重构时对图像进行处理；环路滤波：对重构图像进行处理，并且被滤波后的帧图像被用作参考帧来代替原始帧。经过对这三种滤波方式的对比测试，环路滤波被证明比其它两种方式更有效，它不仅能提高图像的主观质量，而且在很多情况下提高了客观质量0.5-1dB。因此环路滤波的方法被加入到JPC系统。如图5.2(b)所示，虽然只是初步解决方案，但去闪烁滤波器确实在去除噪声方面有这非常明显的效果。



(a) 原始 JPC 编码的重构图像. PSNR =32.9dB

(b) 经过去闪烁滤波的图像. PSNR =32.9dB

图 5.2 去闪烁滤波器在 *Foreman* 序列第十帧上的效果

5.3 与其他基于内插预测的编码方案的对比分析

实验表明，JPC 是目前最成功的基于内插预测的视频编码方案，本节将对 JPC 和目前的另外两个基于内插预测的编码方案进行对比分析，并分析 JPC 的优势。

5.3.1 与 Binary Tree Residue Coding 的比较

Binary Tree Residue Coding (BTRC) 是由Robinson、Druet和Gosset 2000年提出

的视频编码方案^[28]，是将内插预测编码用于视频压缩的最初尝试。它将运动估计残差看作一个图像，采用BTPC（APT的前身）直接对残差图像进行编码。然而，残差已不再是自然图像，其帧内相关性被大大削弱。因此APT或BTRC对于残差图像的压缩效率远逊于在自然图像上的性能。另一方面，时空联合预测保证了仅对自然像素点进行空间自适应预测。下面将详细地解释时空联合预测为何能比BTRC获得更高的压缩率。

定义函数 $Pred(X)$ ，其返回值为第二章描述的APT空域自适应预测误差。因此，联合预测的误差是：

$$\begin{aligned} jp_error &= (X - X_{intra}) - (X' - X'_{intra}) \\ &= Pred(X) - Pred(X') \end{aligned} \quad (5-3)$$

相应的，BTRC返回的预测误差为：

$$btrc_error = Pred(X - X'). \quad (5-4)$$

尽管APT对静态图像相当有效，但在视频编码上情况就不同了：

由于参与预测的各像素的运动向量可能不同，故在连续帧中被用于预测的像素点位置可能会发生较大变化。上面定义的 $btrc_error$ 可以被分成两个部分：当前帧的预测误差 $Pred()$ 减去参考帧的预测误差 $Pred_1()$ 。注意这里把参考帧的预测函数定义为 $Pred_1()$ 而非 $Pred()$ 。

$$Pred(X - X') = Pred(X) - Pred_1(X'). \quad (5-5)$$

BTRC通过递归地使用之前编码的像素点来重构一个图像。如图5.3所示，像素 X 通过以它为中心的正方形上的4个邻近点来预测，即 A, B, C, D ，但在视频序列中，这些像素点的运动矢量很可能存在差异。令 A', B', C', D' 为 A, B, C, D 在参考帧中的对应点。因此， $Pred_1(X')$ 函数使用了4个不相关的像素点 A', B', C', D' （很可能并不构成一个正方形）来预测 X' 。换句话说，BTRC本希望用差值 $A-A', B-B', C-C', D-D'$ 来预测 $X-X'$ ，但参考帧中的 A', B', C', D' 和 X' 很可能是不相关的，显然这将导致BTRC的重构算法低效。相反JPC所采用的时空联合预测仅仅使用了 X 的运动向量，合理地应用了自然图像上的帧内和帧间相关性。实验表明，JPC在相同质量下的输出码率平均比BTRC低20.5%。

图5.4给出了JPC和BTRC在*News*和*Football*序列上的率失真曲线，证实了JPC相对于BTRC的优势。BTRC和APT并不是专门为视频编码考虑设计的，由于上述原因，图5.4所示的结果并不意外。由于时空联合预测充分利用了APT的自适应空域预测并结合了视频序列的特点，JPC取得了远胜于BTRC压缩效率。

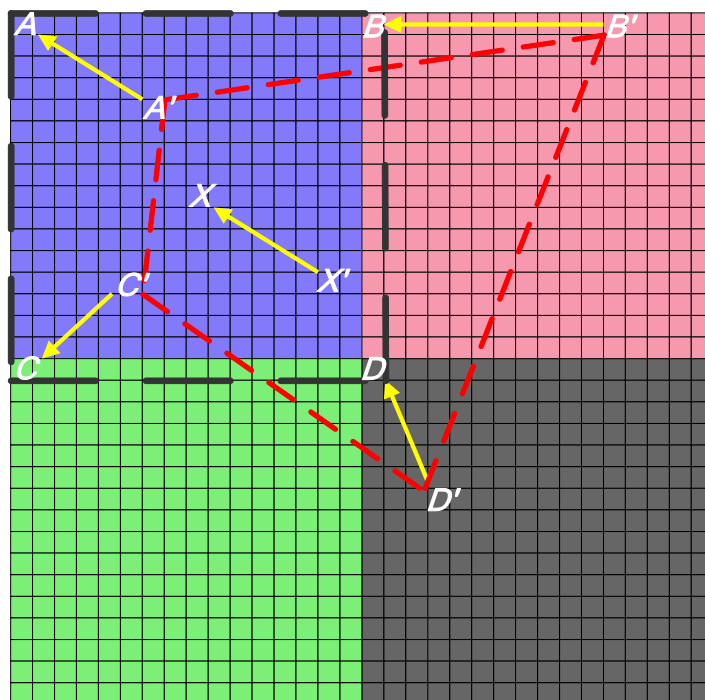
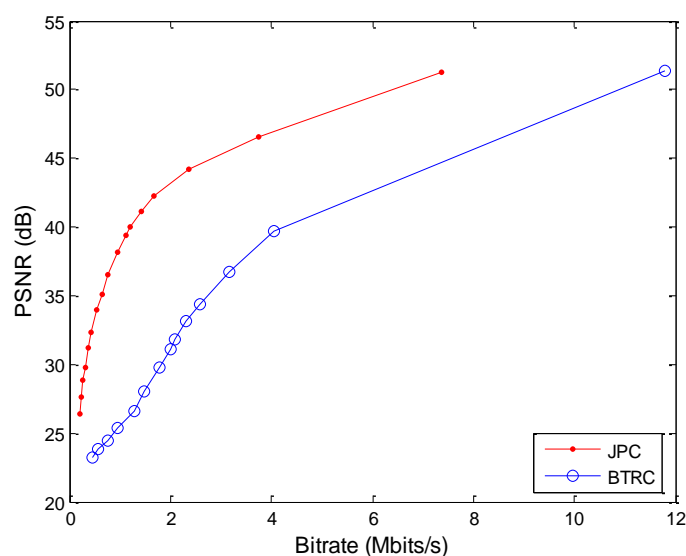
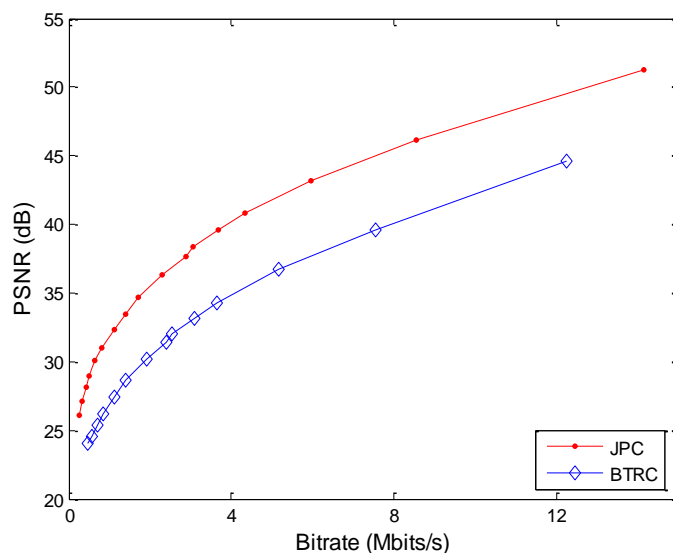


图 5.3 自适应预测在 BTRC 视频编码中效率降低的原因



(a) *News*



(b) *football*

图5.4 JPC和BTRC在压缩性能上的对比

5.3.2 与 Temporal Prediction Trees 的比较

Day 和 Robinson 于 2007 年提出了另一种基于内插预测的视频编码方案，Temporal Prediction Trees (TPT)。其主要思路是根据质量参数、运动向量等信息为视频帧计算出一个门限值，由此来决定其中各像素应当选择帧间或是帧内预测模式，实现帧间/帧内自适应的编码。其中，帧间预测与 JPC 的 MC 模式相同，采用参考帧内的对应点作为预测值；帧内预测采用 APT 空域预测器来估计目标像素。但由于这种方法仍然将帧间和帧内预测分开处理，其准确率还是低于时空联合预测。由图 5.5 可见，TPT 的压缩性能远低于 JPC。(本对比实验采用 Luma-weighted PSNR (加权 PSNR) 作为质量评价标准，其亮度权重是色度权重的 4 倍。)

TPT 性能低于 JPC 的另一个原因在于它使用的树形结构。APT 和 TPT 采用了一种所谓的六叉树结构来保存各像素点的预测误差。它们引入了一种“终结符”来指示叶子节点（可能的子树上全为零值的节点），这种方法能够跳过很多有规律出现的零码值，在静态图像压缩上有较好的效果^[55]。然而视频序列中存在大量的帧间冗余，运动估计的残差中很可能存在整块的零值。因此，大量的连续零码值使得直接

的游程编码拥有更高的效率，相反 TPT 的树形结构打乱了这种分布，同时它需要对每一个未被终结的节点插入一个标记来指示其值本身是否为零以及是否是叶子节点。基于上述分析，JPC 选用了 APT 的内插预测系统、经验性的预测规则，但抛弃了其树形结构，而直接采用标准的游程哈夫曼编码。

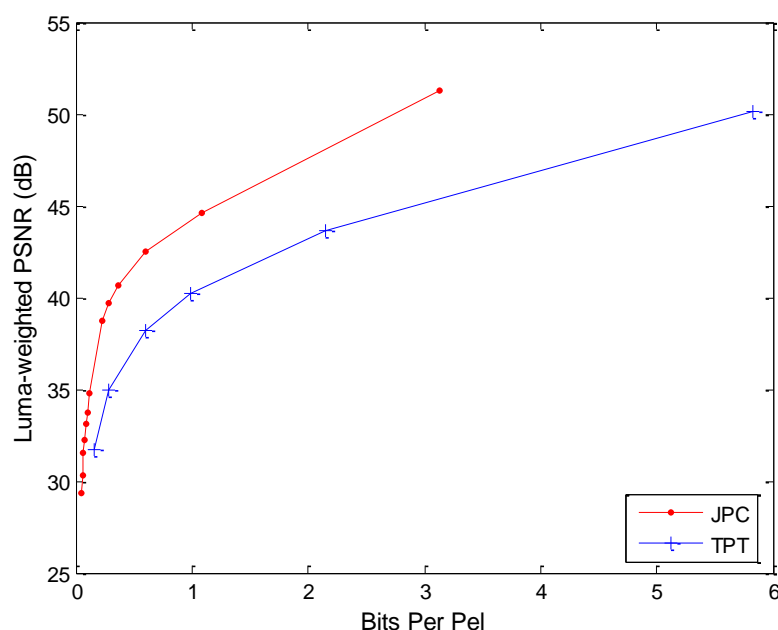


图5.5 JPC和TPT在300帧CIF (352×288) 序列*Mother and Child*上的压缩性能对比

5.4 率失真性能对比测试

本文对 JPC、MPEG-2、H.264 (high 4:4:4 profile) 在 6 个 100 帧的 CIF 视频序列，*Football*、*News*、*Foreman*、*Stefan*、*bus*、*Tempete* (352×288 30 帧/秒) 上进行了对比实验。MPEG-2 采用 MPEG-2 V1.2 参考软件^[56]，H.264 采用 JM 12.4 参考软件^[57]。编码结构选用 IPPPP...，为观察其可能的应用领域，本文选择了约从 25dB 到 50dB 的大范围质量参数。图 5.6~5.8 分别列出了 *News*、*Foreman*、*Football* 上的率失真测试结果。结果表明 H.264 性能最佳，但 JPC 也显示出较好的潜力。三种编码器在较为静态的 *News* 序列上都有较好的压缩效率；而在 *Foreman* 序列中带有快速镜头运动的部分和包含剧烈运动的 *Football* 序列上，三者压缩率都有所降低但相对

关系保持一致。另外可以发现，在极高的码率下，JPC 能够达到和 H.264 相同的压缩效率。总体来说，JPC 的率失真性能高于 MPEG-2 但尚不能与 H.264 相比。但 JPC 某些模块的改进正在进行中，包括对变块大小、多帧参考、高精度运动估计的支持。这些研究很可能进一步提高其编码性能。

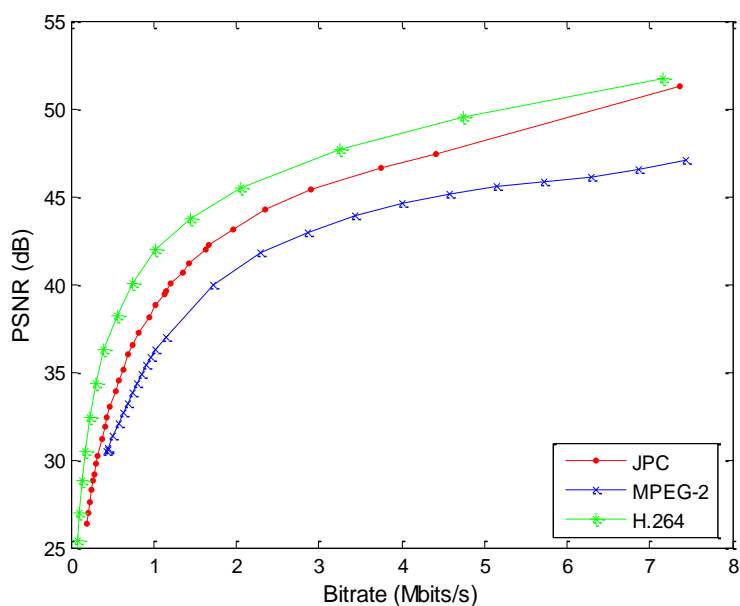


图5.6 JPC、MPEG-2、H.264在News序列前100帧上的率失真性能对比

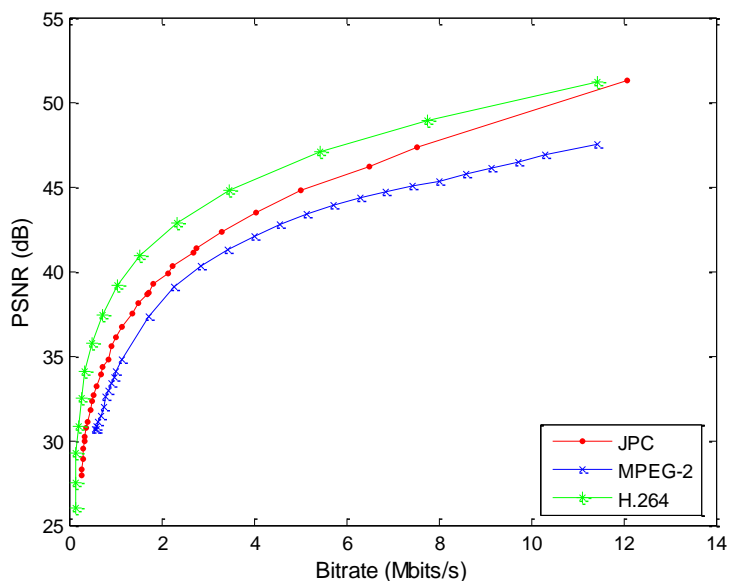


图5.7 JPC、MPEG-2、H.264在Foreman序列第130~230帧上的率失真性能对比

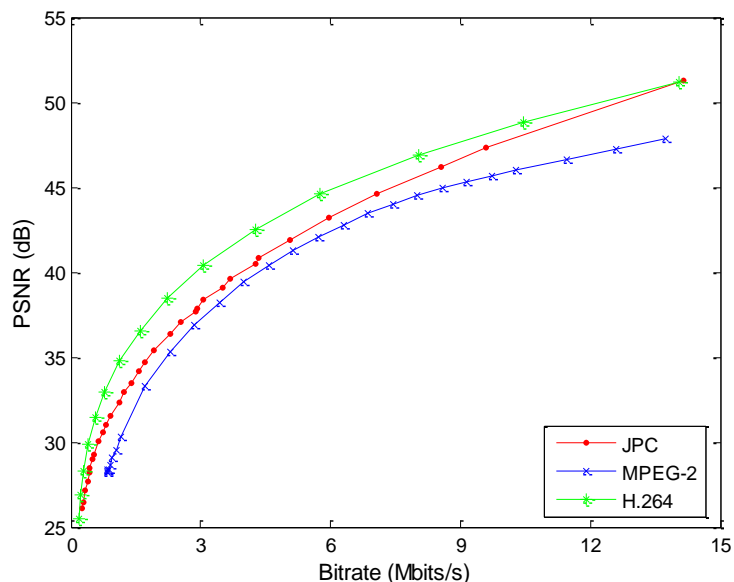
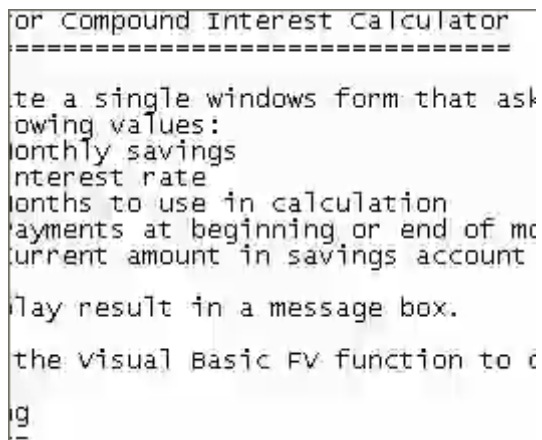


图 5.8 JPC、MPEG-2、H.264 在 *Football* 序列前 100 帧上的率失真性能对比

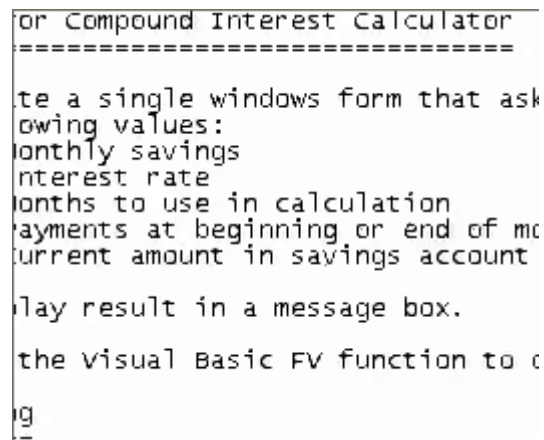
5.5 主观质量对比测试

虽然 JPC 的率失真性能尚不及 H.264，但它的主观性能在某些特定应用上往往有更好的质量。经过对大量特点各异的视频序列的测试，本文发现 JPC 和 H.264 在主观质量上各有所长。但是，JPC 在包含文字、复杂的纹理、不规则边缘的视频上具有明显的优势。

图 5.9 列出了 JPC 和 H.264 在一段纯文字视频上编码后的重构图像。图 5.9(a) 由 H.264 编码重构，加权 PSNR 为 28.0dB，字迹边界已经有些模糊；5.9(b)由 JPC 编码重构，加权 PSNR 为 27.5dB。二者对比可见 JPC 编码图像即使 PSNR 低于 H.264，其文字其清晰度也要高于 H.264。由此可推测 JPC 在幻灯片视频上也会有强于 H.264 的主观质量，而图 5.10 证实了这种推测，这说明 JPC 在压缩幻灯片视频时能够较好地保存字迹轮廓而 H.264 不能，因此录制演讲视频可能是一个有潜力的应用领域。图 5.11 列出了 H.264 和 JPC 在一段风景视频上的对比，其中 H.264 导致了明显的块效应，而 JPC 在保存边界、轮廓显示了较大的优势。

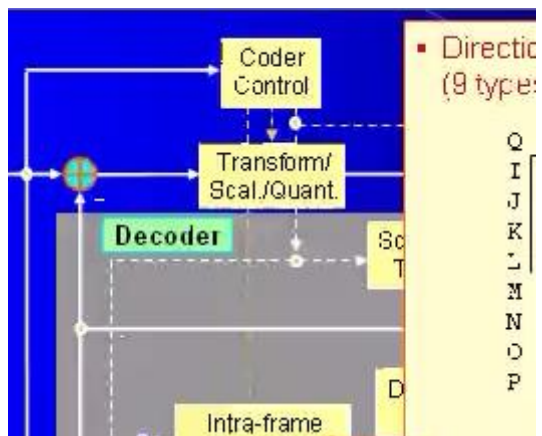


(a) H.264, Luma-weighted PSNR = 28.0dB

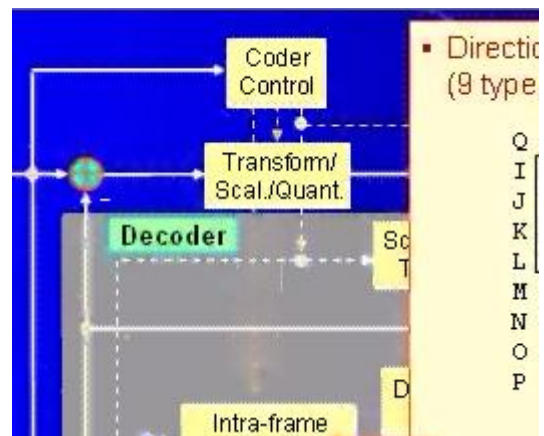


(b) JPC, Luma-weighted PSNR = 27.5dB

图 5.9 JPC 和 H.264 在一段纯文本视频序列上的主观质量对比

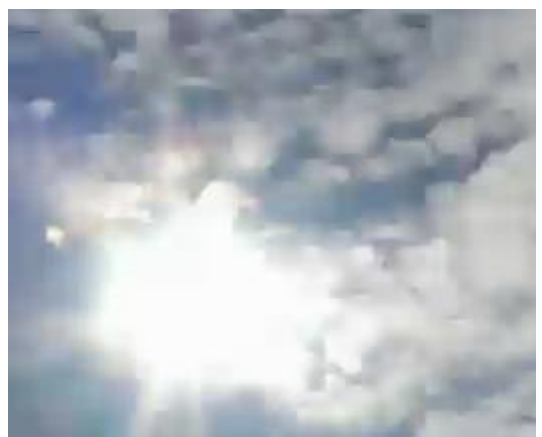


(a) H.264, Luma-weighted PSNR = 31.1dB



(b) JPC, Luma-weighted PSNR = 31.3dB

图 5.10 JPC 和 H.264 在一段幻灯片视频序列上的主观质量对比



(a) H.264, Luma-weighted PSNR = 36.0dB



(b) JPC, Luma-weighted PSNR = 36.3dB

图 5.11 JPC 和 H.264 在一段风景视频上的主观质量对比

JPC 的主要问题在于它在高度压缩时会生成时域噪声（即 speckling artifact），使之在某些时候弱于 H.264。但考虑到 H.264 应用了十分成熟的滤波技术而 JPC 的滤波只是一个初步的、未经优化的解决方案。但是，如 5.2 节所介绍的，JPC 的滤波技术在去除噪声上已经取得了不错效果，它至少说明闪烁效应不是无法解决的。鉴于内插预测编码相对于基于块的编码在基本框架上的优势，JPC 很可能在很多应用领域上取得更好的主观性能。

5.6 编码速度对比测试

本节将对 JPC、MPEG-2、H.264 的编码速度作一个简单的比较。虽然 JPC 的源码尚未经过优化，在对比测试中处于劣势，但三者的编码时间测试还是有一定参考价值。表 5.1 和表 5.2 列出了 JPC、MPEG-2、H.264 主要的存在差异的功能模块，略过了三种编码器中都基本相同的模块（如熵编码）。MPEG-2 V1.2 和 JM 12.4 参考软件分别选为 MPEG-2 和 H.264 编码器。测试在一台奔腾 4 2.6GHz CPU、512M 内存的个人计算机上进行，三种编码器都进行全搜索运动估计，而 JPC 的滤波模块被禁用。

表 5.1 参与编码复杂度比较的功能模块

JPC 编码器	H.264 编码器	MPEG-2 编码器
颜色空间变换	H.264 运动估计	MPEG-2 运动估计
JPC 运动向量搜索	H.264 模式选择	MPEG-2 模式选择
JPC 模式选择	整数变换	DCT 变换
预测与量化	量化	量化

表 5.2 参与解码复杂度比较的功能模块

JPC 解码器	H.264 解码器	MPEG-2 解码器
反量化	反量化	反量化
预测与重建	逆整数变换	DCT 逆变换
颜色空间逆变换	重建	重建

表 5.3 各编码器的平均每 I/P 帧编码时间（毫秒）

编码器	JPC		H.264		MPEG-2	
	I	P	I	P	I	P
<i>Foreman</i>	37.2	258.0	75.9	621.3	75.9	189.7
<i>News</i>	37.9	186.3	76.2	446.8	73.9	125.9
<i>Bus</i>	37.6	311.1	76.4	661.2	81.1	205.4
<i>Football</i>	37.6	280.6	77.5	745.2	75.4	207.2
<i>Stefan</i>	38.0	269.6	78.4	617.9	73.8	171.9
<i>Tempete</i>	38.8	259.2	76.1	627.2	72.0	151.7

表 5.4 各编码器的平均每 I/P 帧编码时间（毫秒）

解码器	JPC	H.264	MPEG-2
<i>Foreman</i>	24.3	32.4	63.1
<i>News</i>	29.3	22.9	61.3
<i>Bus</i>	28.0	35.9	61.8
<i>Football</i>	29.1	28.2	61.5
<i>Stefan</i>	28.7	37.6	62.4
<i>Tempete</i>	29.3	31.9	62.7

测试在各种质量参数下对 6 个视频序列进行，记录下各帧的编解码平均时间。表 5.3 列出的是三者 I 帧和 P 帧上的编码时间，而表 5.4 列出的是三者 I 帧和 P 帧上的解码时间。JPC 的编码速度平均比 H.264 快 57.6% 而比 MPEG-2 慢 39.9%，解码时间平均比 H.264 快 7.90%，比 MPEG-2 快 18.0%。结果表明，虽然未经优化，JPC 编解码器的编码速度还是证明了其竞争力。目前版本的 JPC 在熵编码上效率偏低，它必须遍历所有像素的预测误差两次——一次统计误差分布、一次进行编码。为了提高效率，可以将预测误差建模，设计熵编码表。这样 JPC 和 H.264、MPEG-2 一样，只需对码值遍历一次编码即可。

5.7 本章小结

本章详细描述了基于内插预测的视频编码方案 Joint Predictive Coding, 其功能模块包括颜色空间变换、运动估计、联合预测及量化、重排序、熵编码。讨论了内插预测编码的通病——闪烁效应, 并给出了比较有效的解决方案。分析了 JPC 相对于其它内插预测编码的优势, 从率失真性能、主观质量、编码速度三方面将 JPC、H.264、MPEG-2 作了对比测试, 结果表明, JPC 是一种相当有潜力的新型编码方案。

6. 总结及展望

长久以来，视频编码的研究都注重于压缩率的提高以满足苛刻的带宽要求，但近 10 年来高质量的视觉效果逐渐成为主流需求。针对该目标，本文研究了新的基于内插预测的编码机制，提出了基于像素到像素预测的视频编码方案 **Joint Predictive Coding**，从根本上去除了传统框架下难以回避的块效应，同时，它所采用的 **KLT/Joint Prediction** 的框架，不仅存在更大的上升空间，而且全面支持可分级编码。

本文的贡献在于：

1. 提出了兼顾时间和空间相关性时空联合预测技术，该方法已经从理论和实验两方面被证明优于传统的运动补偿技术，具有更高的准确度。
2. 研究了基于像素到像素的内插预测编码机制，用以取代基于块的编码机制，从根本上去除了块效应。同时，这种点到点预测机制被证明能更好地保存图像细节纹理，也能更好地支持可分级编码。
3. 基于时空联合预测与内插预测框架，设计并实现了新架构的视频编码方案 **Joint Predictive Coding**。以 **MPEG-2** 水平的低复杂度，取得了明显高于 **MPEG-2** 的压缩率，并在极高的码率上达到了 **H.264** 的水平。

JPC 的内在结构使其对于当今视频编码的高清晰、高分辨率、可缩放性的要求具有极强的适应性。初步探索表明，该方案在高清电影、高清动画、视频游戏以及虚拟现实等领域具有值得关注的前景。**H.264** 标准在高清应用上的缺陷，对 **JPC** 来说是一个机会，这也正是 **JPC** 将来的发展方向。本文已经提出了面向高清的 **JPC** 修改草案，旨在成为高清应用领域占有一席之地的视频编码策略。

致 谢

值此研究生生涯行将结束之际，我要向多年来关心我帮助我的众多良师益友表达我最诚挚的谢意。

衷心感谢我的导师刘文予教授。是导师的教诲让我学会站得更高看世界，更让我找到了让自己出类拔萃的路。刘老师看待问题的态度、考虑问题的方式，之于治学、乃至做人，都让我受益匪浅。没有您的信任和支持，我今日所得无异于痴人说梦，是您让我的研究生生涯对得起“研究”二字。

感谢美国 Temple 大学的 Latecki 教授，Latecki 教授对于科学的不懈的追求和不灭的热情令人惊叹。通过他，我也看到了国内学术研究在严谨务实上与国外大学的差距。

我还要感谢宽带无线与多媒体系统研究中心的刘玉教授、喻莉教授，蒋洪波副教授等老师，感谢你们给我的指点和为我提供的成长机会。感谢韦耿博士、张帆博士、唐振华博士生、刘海容博士生、白翔博士生、李鹏飞硕士、王亮硕士、赵爽硕士，以及李劝男、赖重远、梁慧、叶小琼、徐小芳、王长清这些师兄师姐师弟师妹的帮助，还有赵江、陈梦、周亮、王卫明、孙智超、郭强、易友文、刘继通、沈为等朋友陪我共同度过这段时光。

深深感谢我的父母，父母的悉心照顾与深切关怀无以为报，父母对我的信任、对我的选择的尊重，更是我此生最感恩的财富。我谨将用我的一生拼搏，来回报这份信赖。

江文斐

2009年4月于喻园

附录 1 攻读学位期间发表的论文

1. Wenfei Jiang, Longin Jan Latecki, Wenyu Liu, Hui Liang, Ken Gorman. A Video Coding Scheme Based on Joint Spatiotemporal and Adaptive Prediction. IEEE Trans. on Image Processing, 2009, 18(5): 1025 - 1036.
2. Wenfei Jiang, Longin Jan Latecki, Hui Liang, Ken Gorman, Wenyu Liu, Joint Predictive Coding for HD Video Compression, International Conference on Image Processing 2009. (submitted)

附录 2 文中主要中英文术语对照表

Adaptive Prediction Trees (APT)	自适应预测树编码
Advanced Video Coding (AVC)	高级视频编码
Binary Tree Predictive Coding (BTPC)	二叉树预测编码
Binary Tree Residue Coding (BTRC)	二叉树残差编码
Context Adaptive Variable Length Code (CAVLC)	基于上下文自适应变长编码
Context Adaptive Binary Arithmetic Code (CABAC)	基于上下文自适应算数编码
Continuous-tone Image	连续色调图像
Discrete Cosine Transform (DCT)	离散余弦变换
Interpolative Coding	内插预测编码
Joint Predictive Coding (JPC)	联合预测编码
Joint Spatiotemporal Prediction (JSP)	时空联合预测
Mean Square Error (MSE)	均方误差
Motion Estimation (ME)	运动估计
Motion Vector (MV)	运动向量
Normalized Cross Correlation Function (NCCF)	归一化互相关函数
Principal Component Analysis (PCA)	主分量分析
Scalable Video Coding (SVC)	可分级视频编码
Speckling Artifact	闪烁效应
Sum of Absolute Difference (SAD)	绝对误差和
Temporal Prediction Trees (TPT)	时域预测树编码
Video Home System (VHS)	家用录像系统

参考文献

- [1] ITU-T H.261. Video codec for audiovisual services at P*64kbit/s, 1990.
- [2] ITU-T H.263. Video codec for low bit rate communication, 1996.
- [3] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 14496-10 (E) AVC). July 2004.
- [4] ISO/IEC 11172. Information technology coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s.
- [5] ISO/IEC 13818. Information technology-generic coding of moving pictures and associated audio.
- [6] ISO/IEC 14496-3. Information technology-generic coding of audio-visual object-part 3: audio. Vancouver Canada, July 1999.
- [7] ISO/IEC JTC1/SC29/WG11 N2330. MPEG-7 Draft Proposal Package Description (PPD). Dublin, July 1998.
- [8] ISO/IEC JTC1/SC29/WG11 N3162. First ideas on defining a Multimedia Framework (version 0.2), Maui, Dec. 1999.
- [9] Wikipedia, H.264/MPEG-4 AVC, <http://en.wikipedia.org/wiki/H.264>, 2009.
- [10] M. Smith, J. Villasenor, Intra-frame JPEG2000 vs. Inter-frame Compression Comparison: The benefits and trade-offs for very high quality, high resolution sequences, in: SMPTE Technical Conference and Exhibition, Pasadena, California, USA, 2004:1~9.
- [11] Wei Pu, Yan Lu, Feng Wu, Joint Power-Distortion Optimization on Devices with MPEG-4 AVC/H.264 Codec, in: Proc. IEEE International Conf. on Communications 2006, 1: 441 - 446.
- [12] 万俊青, 王兴国, 基于 H.264 标准的桌面视频会议系统设计, 电视技术, 2005(2): 73 - 76.
- [13] 数字音视频编解码技术标准工作组, AVS 简介, <http://www.avs.org.cn/>, 2009.

- [14] H. Schwarz, D. Marpe, and T. Wiegand, Overview of the scalable extension of the H.264/MPEG-4 AVC video coding standard, *IEEE Trans. on Circuits Syst. Video Technol.*, 2007,17(9):1103–1120.
- [15] B. R. Hunt, Optical Computing for Image Bandwidth Compression: Analysis and simulation, *Appl. Opt.*, 1978. 15: 2944-2951.
- [16] Peter J. Burt, and Edward H. Adelson, The Laplacian Pyramid as a Compact Image Code, *IEEE Trans. on Comm.*, 1983, 31(4): 532 - 540.
- [17] W Woods, S D O'Neil, Subband Coding of Images, *IEEE Trans on Acou., Spee. And Signal Process.*, 1986, 34(5): 1278-1288.
- [18] P. Roos, M. A. Viergever, M. C. A. van Dijke, and J. H. Peters, Reversible intraframe coding of medical images, *IEEE Trans. Med. Imag.*, 1988, 7(12): 328–336.
- [19] L. Arnold, Interpolative coding of images with temporally increasing resolution, *Signal Process.*, 1989, 17: 151–160.
- [20] P. G. Howard and J. S. Vitter, New methods for lossless image compression using arithmetic coding, *Inf. Process. Manag.*, 1992, 28(6): 765–779.
- [21] G. R. Kuduvalli and R. M. Rangayyan, Performance analysis of reversible image compression techniques for high-resolution digital teleradiology, *IEEE Trans. Med. Imag.*, 1992, 11: 430–445.
- [22] P. G. Howard and J. S. Vitter, Fast progressive lossless image compression, in *Proc. Data Compre. Conf., Snowbird, UT, USA*, 1994.
- [23] E. A. Gifford, B. R. Hunt, and M. W. Marcellin, Image coding using adaptive recursive interpolative DPCM, *IEEE Trans. on Image Process.*, 1995, 4(8): 1061–1069.
- [24] J. A. Robinson, Efficient general-purpose image compression with binary tree predictive coding, *IEEE Trans. on Image Process.*, 1997, 6(4): 601–608.
- [25] J. A. Robinson, Adaptive prediction trees for image compression, *IEEE Trans. on Image Process.*, 2006,15(8): 2131 – 2145.
- [26] ISO/IEC JTC 1/SC 29/WG 1 FCD 14495 - public draft, Lossless and nearlossless coding of continuous tone still images (JPEG-LS), 1997.

- [27] ISO/IEC CD 15444-1, Information technology—JPEG 2000 image coding system—Part 1: Core coding system.
- [28] J. A. Robinson, A. Druet, and N. Gosset, Video Compression With Binary Tree Recursive Motion Estimation And Binary Tree Residue Coding, *IEEE Trans. on Image Process.*, 2000, 9(7): 1288 - 1292.
- [29] M G Day, J A Robinson, Residue-free video coding with pixelwise adaptive spatio-temporal prediction, *IET Image Processing*, 2008,2(3):131-138.
- [30] Wenfei Jiang, Longin Jan Latecki, Wenyu Liu, Hui Liang, Ken Gorman. A Video Coding Scheme Based on Joint Spatiotemporal and Adaptive Prediction, *IEEE Trans. on Image Process.*, 2009, 18(5): 1025 - 1036.
- [31] Yao Wang, Jorn Ostermann, Ya-Qin Zhang. *Video Processing and Communication*. Person Education, 2002.
- [32] Iain E G Richardson, H.264 / MPEG-4 Part 10 White Paper, www.vcodex.com, 2003.
- [33] Jain J R, Jain A K. Displacement measurement and its application in interframe image coding. *IEEE Trans. Commun.*, 1981, 29(12): 1799- 1808.
- [34] Koga T, Linuma K, Hirano A, et al. Motion compensated inter frame coding for video conferencing in Proc. Nat. Telecommunication Conf. New orleans: G5.3.1-G5.3.5.
- [35] Tham J Y, Ranganath S, Ranganath M, et al. A novel unrestricted center- biased diamond search algorithm for block motion estimation. *IEEE Trans. on Circuits Syst. Video Technol.*, 1998, 8(4): 369- 377.
- [36] Zhu S, Ma K. A new diamond search algorithm for fast block matching motion estimation, *IEEE Trans. on Image Process.*, 2000, 9(2): 287- 290.
- [37] Zhu C, Lin X, Chau L P. Hexagon-based search pattern for fast block motion estimation. *IEEE Trans. on Circuits Syst. Video Technol.*, 2002, 12(5): 349- 355.
- [38] Liu B, Zaccarin A. New fast algorithms for the estimation of block motion vectors. *IEEE Trans. on Circuits Syst. Video Technol.*, 1993, 3(2): 148- 157.
- [39] Choi K. T, Chan S C, Ng T S. A new fast motion estimation algorithm using

- hexagonal subsampling pattern and multiple candidates search in Proc. Int. Conf. Image Process., London, 1996: 497-500.
- [40] Chan Y L, Siu W C. New adaptive pixel decimation for block motion vector estimation . IEEE Trans. on Circuits Syst. Video Technol., 1996, 6(1): 113- 118.
- [41] Cheung C K, Po L M. Normalized partial distortion search algorithm for block motion estimation. IEEE Trans. on Circuits Syst. Video Technol., 2000, 10(3): 417-422.
- [42] Chen Z, Zhou P, He Y. Fast Integer and Fractional Pel Motion Estimation for JVT, JVT- F017r.doc, Joint Video Team (JVT) of ISO/IEC MPEG & ITU- T VCEG, 6th Meeting. Awaji, Island, Japan, 2002.
- [43] K. R. Rao, P. Yip. Discrete Cosine Transform. Academic Press, 1990.
- [44] Gary J. Sullivan, Pankaj Topiwala, Ajay Luthra. The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions. SPIE Conference on Applications of Digital Image Processing XXVII, 2004.
- [45] P. List, A. Joch, J. Lainema, G. Bjntegaard and M. Karczewicz. Adaptive deblocking filter. IEEE Trans. on Circuits Syst. Video Technol., 2003, 13: 614-619.
- [46] J. Massey, Variable-length codes and the Fano metric, IEEE Transactions on Information Theory, 1972, 18(1):196~198.
- [47] P. G. Howard, J. S. Vitter. Analysis of arithmetic coding for data compression. Informat. Process. Manag., 1992, 28(6): 749-763.
- [48] Digital Cinema Initiatives, AboutDCI, <http://www.dcinovies.com>, 2009.
- [49] A. Bilgin, M. W. Marcellin, JPEG2000 for Digital Cinema, ISCAS 2006, Island of Kos, Greece, pp. 3878~3881, May 2006.
- [50] Boxin Shi; Lin Liu; Chao Xu, Comparison between JPEG2000 and H.264 for digital cinema, IEEE International Conference on Multimedia and Expo. 2008:725-728.
- [51] W. Zeng, J. Li, and S. Lei, An efficient color re-indexing scheme for palette—Based compression, in Proc. Int. Conf. Image Processing, Vancouver, BC, Canada, 2000: 476–479.
- [52] I-Ming Pao, Ming-Ting Sun, Modeling DCT coefficients for fast video encoding,

- IEEE Trans. on Circuits Syst. Video Technol., 1999, 9(4): 608-616.
- [53] C. L. Yang, L. M. Po, D. H. Cheung, and K. W. Cheung, A novel ordered-SPIHT for embedded color image coding, in Proc. IEEE Int. Conf. Neural Networks and Signal Processing, Nanjing, China, 2003: 1087-1090.
- [54] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, Overview of the H.264 / AVC Video Coding Standard, IEEE Trans. on Circuits Syst. Video Technol., 2003, 13(7): 560–576.
- [55] APT Online Reference Code, [Online].
<http://www.intuac.com/userport/john/apt/index.html>
- [56] MPEG-2 reference software, MPEG-2 ver. 1.2, MPEG Software Simulation Group, [Online]. <http://www.mpeg.org/MSSG>
- [57] JM Software, JM 12.4, H.264/AVC Software Coordination, [Online].
<http://iphome.hhi.de/suehring/tml/>