

北京工业大学
硕士学位论文
基于H. 264的高清晰视频编码技术研究
姓名：王雁来
申请学位级别：硕士
专业：计算机应用技术
指导教师：高文;尹宝才
20040501

## 摘 要

随着数字电视、Internet 流媒体以及 DVD 视频的广泛应用, 视频压缩编码逐渐成为计算机多媒体技术的关键组成部分。数字电视以及 DVD 视频原本是在 MPEG-2 的基础之上发展完善起来的, 但是由于其技术特点已经越来越受到当前需求的局限, 所以 JVT 提出了更加高效的视频编码标准——H. 264。

MPEG-4 标准是 ISO/IEC 制定的新一代视频编码标准, 它提出了基于对象的编码方案, 允许将场景中的对象进行独立编码。其中, H. 264 作为 MPEG-4 标准中的第 10 部分, 侧重于更加高效的视频编解码, 规范视频编码技术。

文章在 H. 264 标准软件平台上, 对高清晰视频编码方法进行了研究, 包括块匹配算法、帧内预测方案以及环路滤波技术等三个方面。

首先, 对用于视频内容特征点跟踪的块匹配算法进行了优化。为了提高算法的匹配精准度, 添加了一个预处理反馈环, 利用相关系数来评估像素块的相似程度, 从而确定合适的搜索窗口大小; 并在此基础上提出了像素采样匹配, 大幅度缩减了计算复杂度。

其次, 提出了基于块内像素相关性的帧内预测方案。利用块内像素之间的相关性远大于相邻块之间的相关性的特点, 添加了 4 种帧内预测模式。此外, 利用相邻块预测模式一致的原则, 直接选择预测模式。利用块内相邻像素值进行累积插值预测, 不仅有效提高了编码效率, 而且在一定程度上降低了计算复杂度。

最后, 提出了基于内容的自适应环路滤波。不同的视频内容编码过程中会产生不同类型的块效应, 因此应该有针对性的采用不同滤波器。通过计算相邻块之间的相关性、块内差异度、边界点像素差值、运动向量差值以及 QP 等影响因子, 得到相应强度和适当范围的滤波器, 从而达到最佳的滤波效果。

实验结果表明, 针对以上三个方面的创新均在不同程度上提高了编码效率。在文章的结论部分, 对基于内容的可适应视频编码方案进行了展望。

**关键词:** 视频编码; H. 264; 块匹配; 帧内预测; 环滤波

## ABSTRACT

With the widespread adoption of technologies such as digital television, Internet streaming video and DVD-Video, video compression has become an essential component of computer multimedia technology. The success of digital TV and DVD-Video is based upon the 10-year-old MPEG-2 standard, a technology that has proved its effectiveness but is now looking distinctly old-fashioned. JVT promoted a kind of new standard named H.264 which shows more effective.

MPEG-4 is a new generation standard of video coding made by ISO/IEC. Object-oriented coding is one of the most important characteristic in MPEG-4. As the part 10 of MPEG-4, H.264 standard focuses on more effective video CODEC.

In this thesis, content based video coding scheme was investigated in the software platform of H.264, including BMA, intra prediction and loop filter.

Firstly, the block match algorithm which used for Feature Tracing was optimized. In order to improve the match result, a precondition feedback circle was added, which made use of the correlation between two blocks for comparability evaluation and by which a suitable search window was selected. Furthermore, pixels-sampled match scheme was promoted, which reduced the computation complexity largely.

Secondly, the intra prediction scheme based on adjacent pixels was put forward. For the correlation between pixels in the same block would be more than that of adjacent blocks, the former would be more accurate for pixel value prediction. Taking the correlation into consideration, four intra predict modes are promoted. Furthermore, using the principle of adjacent block mode consistency, the predict mode can be decided directly. Using the accumulative interpolation by adjacent pixels, coding efficiency and computation complexity would be improved obviously.

Finally, a kind of adaptive loop filter based on content was promoted. Different video would have different blocking effect that should be applied different filter. In order to get the suitable filter, the correlation between adjacent blocks, MSE in block, the distance of pixels, motion vector and QP value would be integrated into one factor. According to the factor, suitable filter coefficient and strength would be selected.

The experimental results show that the three creative ideas mentioned above all have good effect on video coding. In the last part, adaptive video coding based on content was prospected.

**Key words:** Video Coding, H.264, BMA, Intra Prediction, Loop Filter

## 独 创 性 声 明


本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京工业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签名:  日期: 2006-6-10

## 关于论文使用授权的说明

本人完全了解北京工业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部内容，可以采用影印、缩印或其他复制手段保存论文。

（保密的论文在解密后应遵守此规定）

签名:  导师签名: 尹宝才 日期: 2006-6-10

# 第一章 绪 论

随着数字电视、DVD 视频等高清晰影像技术的完善,以及互联网的高速发展,人们不仅对视频质量要求大为增加,而且对视频传输的需求也随之提高。视频编解码技术作为计算机多媒体技术的一个分支学科,随着 MPEG-4 标准和 H.264 标准的制定出台,近年来在国际上逐渐成为研究热点。

本章将详细阐述视频编解码技术的概况和国内外研究现状,随后将介绍论文的内容排列与组织结构等有关内容。

## 1.1 视频编码的发展历史

视频编码的历史始于 1970 年贝尔实验室中电话的出现。传统的压缩编码是建立在香农(Shannon)信息论基础上的,以经典的集合论为基础,用统计概率模型来描述信源,但未考虑信息接受者的主观特性及事件本身的具体含义、重要程度和引起的后果。因此,压缩编码的发展历程实际上是以香农信息论为出发点,不断完善的过程。

20 世纪 80 年代中后期,相关学科的迅速发展和新兴学科的不断出现为图像编码的发展注入了新的活力。许多学者结合模式识别、计算机图形学、计算机视觉、神经网络、小波分析和分形几何等理论探索图像编码的新途径。M.Kunt 于 1985 年提出利用人眼视觉特征的第二代图像编码技术,1988 年 M.Barnsley 提出基于迭代函数系统(IFS, Iterated Function System)的分形图像编码技术<sup>[1]</sup>,1989 年 S.Mallat、I.Daubeche 将小波分析理论应用于图像编码。90 年代初发展起来的基于模型的图像编码方法,即为现在的 MPEG-4 编码技术,已经逐渐成为视频编码的主要方法。

高效率的编码不仅可以大幅提高视频信息含量,而且还可以减小视频传输量。当前流行的几种视频编码技术有:预测编码、子带编码、小波编码以及分形编码等<sup>[2]</sup>。

一般来说,各项单独的编码技术有其优点,也有其内在的缺点。因此,任何一种单一的编码技术都不能避免所有的编码缺陷。当前视频编码发展的主要技术

方向就是“混合编码”，即综合所有的编码技术。针对不同的视频内容，分别采用不同的编码技术，甚至不同编码技术的混合。

例如，采用小波包代替单一小波基做视频编码，从而大幅提高编码效率。此外，针对特定视频作基于内容的分析，并按照信息的重要性作相应的分层处理。

从图像编码的发展趋势，我们不难看出，以后的编码技术将主要侧重于消除图像数据的视觉冗余、结构冗余、知识冗余以及统计冗余信息。如何高效去除这些冗余信息恰恰是我们所要克服的主要障碍。

## 1.2 国内外的研究发展现状

尽管上述各种编码方案层出不穷，而且也各有优势，但真正具有划时代意义的研究主流当数MPEG-4和H.264视频编解码标准。近年来，针对视频编码的研究课题主要是围绕着MPEG-4和H.264标准进行的。

国际电信联盟组织（ITU-T）提出了H.26X系列（如：H.261、H.262、H.263等）视频编码标准，主要用于网络实时传输的视频会议和可视电话等方面。该组织面向全球所有对此感兴趣的单位，在每三个月一次的会议上，讨论各个科研单位的技术提案，以此完善编码技术标准。

国际标准化组织（ISO/IEC）起草了MPEG-X（如：MPEG-1、MPEG-2、MPEG-4），主要用于视频存储（DVD）、视频广播、视频流（如互联网视频、DSL视频以及无线视频）。其中MPEG-1和MPEG-2标准在视频存储方面取得了极大成功，而所提出的MPEG-4标准则是突破了原有的编码概念，提出了基于对象的编码方法。此外，该组织还针对多媒体内容标识和普通的多媒体架构分别提出了MPEG-7<sup>[3]</sup>和MPEG-21<sup>[4]</sup>两类标准。

由于两大标准化组织所提出的视频编码标准分别针对不同的应用领域，因此必然存在着应用局限性，因此从1997年起，ISO/IEC MPEG联合ITU-T VCEG组成Joint Video Team（JVT），负责H.26L标准的制定<sup>[5]</sup>，主要目的是建立一套简单高效的视频编码标准。2001年，该系列标准已经发展到H.264版本，正式命名为AVC（Advanced Video Coding），并分别作为ITU-T的H.264标准和ISO/IEC的MPEG-4的第10部分<sup>[6]</sup>。

图1-1展示了两大国际标准化组织的工作进展图。



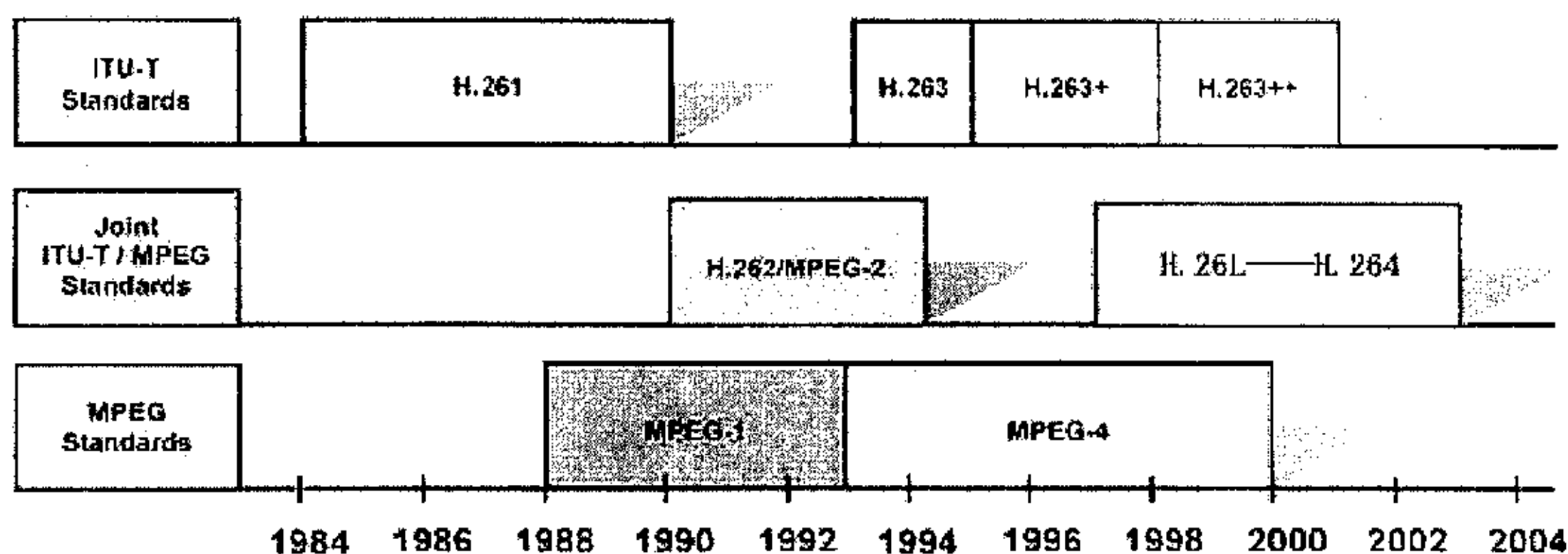


图 1-1 视频编码标准化组织进展图

Figure 1-1 Evolvement of Video Coding Standard Organization

国内方面，中国信息产业部科学技术司于2002年6月批准成立了数字音视频编解码技术标准工作组（简称AVS工作组）。该工作组的任务是：面向我国的信息产业需求，联合国内企业和科研机构，制（修）订数字音视频的压缩、解压缩、处理和表示等共性技术标准，为数字音视频设备与系统提供高效经济的编解码技术，服务于高分辨率数字广播、高密度激光数字存储媒体、无线宽带多媒体通讯、互联网宽带流媒体等重大信息产业应用。2003年系统和视频编码标准<sup>[7][8]</sup>（第一版）已经出台，该标准接纳了包括清华大学、计算所、浙江大学、北京工业大学等多家科研单位的近30多项技术提案。

AVS标准包括系统、视频、音频等三个主要标准和一致性测试等支撑标准。“第一部分：系统”，与目前广泛采用的国际标准MPEG-2兼容，并针对数字电视、光盘播放机、网络流媒体、多媒体通信等应用进行了具体的规定和定义，从框架上提供了对国内外各种主流视频、音频编码标准的支持。“第二部分：视频”是AVS标准中最复杂的部分，编码效率是MPEG-2视频的2-3倍（根据视频画面尺寸不同有所不同），超过了国际上的MPEG-4 AVC/H.264标准，而且方案简洁，实现复杂度明显比MPEG-4 AVC低，在高清晰度应用方面处于国际领先水平。“第三部分：音频”采用主流技术框架，是一套性能比国际标准或类似方案更好的音频编码方案，代表了音频编码领域的最新进展。

### 1.3 视频编码技术的本质

视频编码技术就是对要处理的数字视频用一定的规则进行变换和组合，从而

达到以尽可能少的代码来表示尽可能多的数据信息的目的。图 1-2 描绘了视频编解码的原理框图<sup>[9]</sup>。从图中可以看到，通过对视频源数据的分析，找到可以缩减的冗余信息，通过变换量化等步骤，得到尽可能简略的数据，最后进行熵编码，即可通过信道传输了。至于解码端，则是编码过程的逆过程。

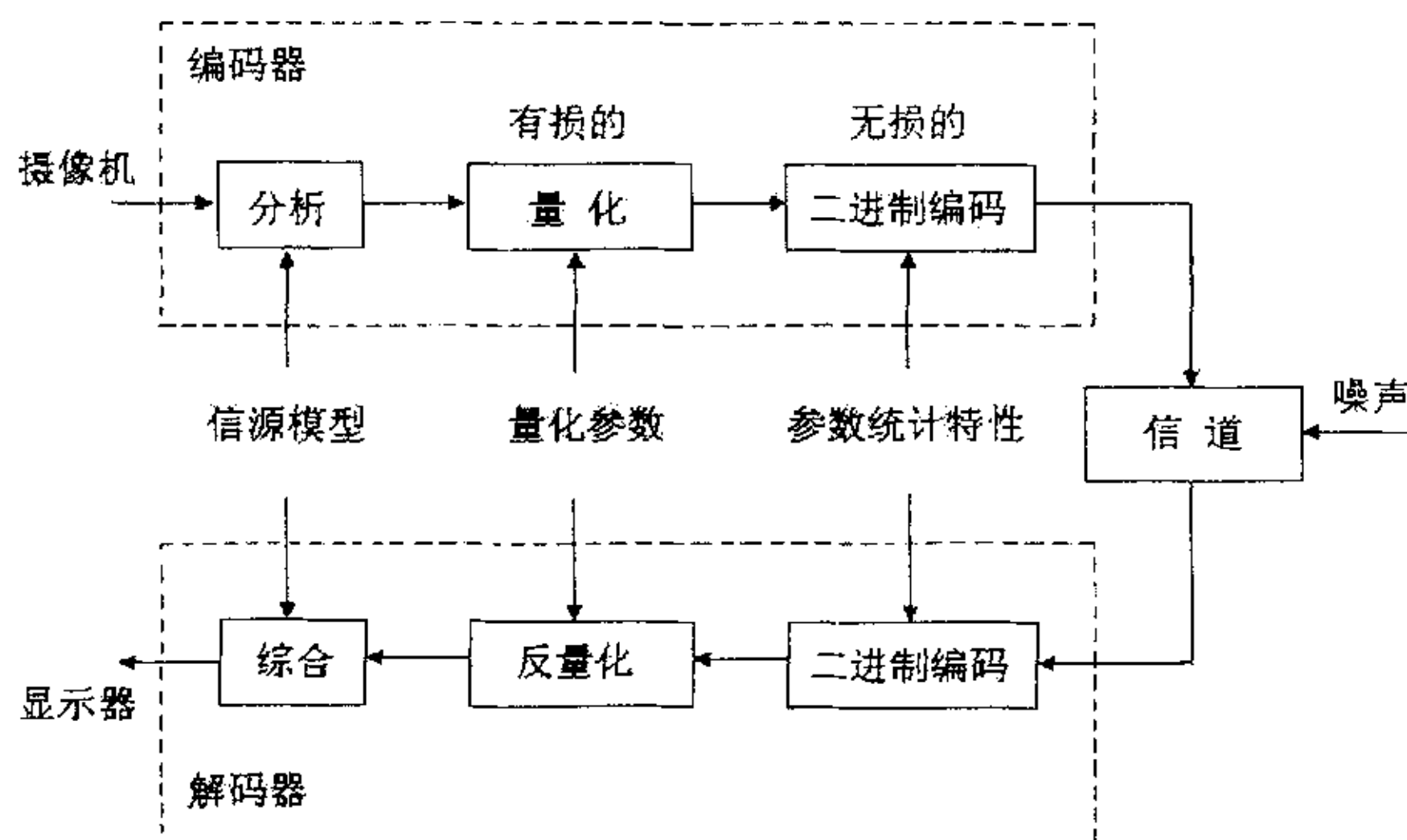


图 1-2 视频编码系统框图

Figure1-2 System Structure of Video Coding

从信息论角度来看，视频编码的本质就是去掉信息冗余，即用尽量少的数据来表达尽量多的信息内容。在数字视频编码和压缩中，有以下几种基本数据冗余：空间像素相关冗余，时间运动相关冗余，编码冗余，心理视觉冗余等多种数据冗余<sup>[10]</sup>。具体细节如下：

在视频流的每一个图像帧内部，各个像素之间必然存在着某种相关性，不同位置关系反映着不同程度大小的相互关系。空间冗余就是利用了帧内各像素之间的相互关系，通过像素之间的相关性获得其相邻像素的性质。

至于视频流的连续图像帧之间，也有很大的相关性——时间冗余。利用这一特性，可以利用已编码帧中的像素值通过插值预测等方式得到当前帧的像素值，即为帧间预测编码。

视频图像的像素值都需要用 0、1 按照一定规则表达，而不同的图像内容就需要有不同的符号序列（即码字）来表示。在实际图像中，不同信息出现的概率不同，如果所对应的码字长度都一样的话，就不能使编码产生的符号序列达到最小，就会产生编码冗余。利用编码冗余，可以减小图像编码的最后数据长度。



除了以上三种传统意义上的冗余之外，心理视觉冗余逐渐成为近年来的研究热点之一。因为人的眼睛并不是对所有视觉信息有相同的敏感度，而那些不敏感的信息就是心理视觉冗余信息。如果去掉了这种冗余信息，尽管实际上会导致定量信息的损失，但从视觉角度来看，并不会明显降低主观图像质量。

实际上，视频压缩编码的目的就是研究如何有效的去除上述冗余信息。下图为视频编码器的原理框图<sup>[11]</sup>。

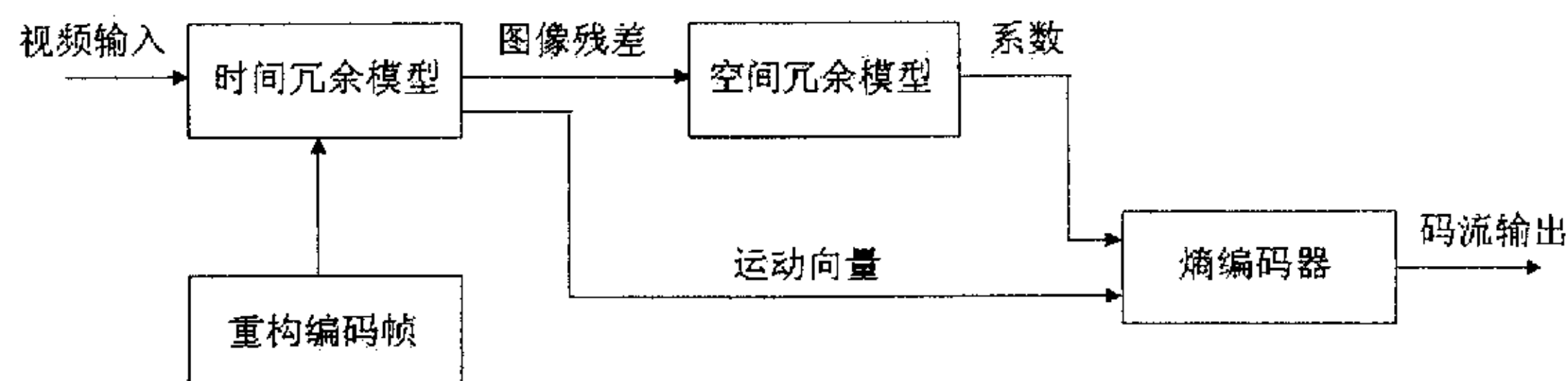


图 1-3 视频编码器原理框图

Figure1-3 Video encoder block diagram

## 1.4 课题的研究目标与研究意义

数字视频编解码研究领域当前面临的主要技术挑战包括：编码理论的上限、不同压缩算法组合的理论依据、压缩效果的有效性评价标准、码率转换的误差边界、带宽约束和带宽可变问题、面向传输安全与存储安全的压缩问题、数据表达问题等。为了解决这些问题，应当研究基于知识的模型化方法和模拟人类视听感知系统（或通俗地称为仿生方法），最终建立超过传统信息论的广义信息论。

本文所涉及的研究课题属于国家 863 项目“数字视音频编码、传输、测试与应用示范”。课题研究的主要目标是给出一套高效智能化的能够解决海量多媒体数据存储、表现、编码和传输的理论方法与算法模型，解决数字视音频编解码标准化的关键问题，以及在带宽有限或抖动的网络上有效地传输流媒体的问题等。

文章所讨论的具体研究内容包括视频编码中的块匹配算法、帧内预测编码以及环路滤波等三个方面。BMA 效率的好坏直接影响到视频运动向量估计和特征点跟踪的准确度，文章针对 BMA 的计算复杂度和匹配效果做改进与提高，缓解 BMA 的计算瓶颈；帧内预测与视频编码质量密切相关，本文对预测方案的计算复杂度进行改良，提高预测精度与预测效率；环滤波主要影响了视频的主观质量，

文章将提出一种自适应环路滤波方案，对不同视频内容做到最优滤波。这三项技术是视频编码中的重要组成部分，将对视频编码效率产生显著影响。

随着视频编码技术研究的逐渐深入，对人们的生活逐渐产生了深远影响。视频编码技术的市场应用前景也越来越乐观，不仅在视听工业、多媒体计算机等方面有广泛应用，而且在广播电视以及未来的信息高速公路等等各个方面都有着诱人的应用前景。因此本课题的研究不仅关系到视频编码技术本身的发展，而且关系到国计民生的前途与未来，将对未来生活质量有着深远的影响与意义。

## 1.5 本文的内容安排与组织结构

本文的研究工作主要包括块匹配算法优化、基于块内像素的帧内预测方案以及自适应环路滤波技术等三个方面。在块匹配算法优化方面，主要针对传统的块匹配算法巨大的计算复杂度这一缺陷，提出了优化简便算法，利用预处理反馈以及像素采样匹配等技术，在保证了匹配精准度的前提下，大幅减少了计算复杂度。在帧内预测方面，利用块内相邻像素的相关性，提出了基于相邻像素的帧内像素预测方案，该方案改变了帧内像素预测办法以及预测顺序，通过块内像素的累积预测，提高了预测准确率。和H.264标准相比，编码效率有一定提高。至于环路滤波方面，则打破了原有的固定系数强度滤波方案，提出了一种基于内容的自适应环路滤波器。该滤波器添加了相邻块之间的相关度、块内像素的差异度、相邻块运动向量差值、相邻块QP差值等4个影响因子，针对不同的视频内容分别应用不同的环路滤波器，从而达到最佳的滤波效果。

文章分为以下几个部分：第一章简要论述视频编码的本质、研究意义、国内外的研究现状以及本文的研究主要方向和目的。第二章主要介绍面向对象的编码标准——MPEG-4，侧重于基本原理以及发展现状。第三章系统介绍最新的H.264标准，对最新的技术标准做概要的阐述。第四章提出简单化的块匹配算法，详细介绍预处理反馈环的实现以及像素采样匹配的思路。第五章提出基于相邻像素的帧内预测方案，利用块内像素相关性远远大于相邻块之间的相关性这一特点，通过累积插值，按顺序得到更加准确的像素预测值。第六章提出基于内容的自适应环路滤波器，通过添加相邻块相关系数等多个影响因子实现针对不同视频内容的自适应滤波方案。最后，总结本文所阐述的内容，并展望未来的工作。

## 第二章 面向对象的编码标准—MPEG-4

MPEG-4 标准是国际上第一类基于对象 (Object) 的交互式多媒体数据处理标准, 由国际标准化组织 (ISO) 于 1994 年开始制定, 并于 1999 年发布第一版。主要包括: 系统、视频、音频、一致性检验、参考软件和多媒体传输集成框架 (DMIF) 等六大部分组成。

采用基于对象 (object-based) 的编码、基于模型 (model-based) 的编码等第二代编码技术是 MPEG-4 编码的主要特征。其目标在于提供一种通用的编码标准, 以适应不同的传输带宽、不同的图像尺寸和分辨率、不同的图像质量, 进而为用户提供不同的服务, 满足不同处理能力的显示终端和用户个性化的需求。

### 2.1 MPEG-4 标准概述

MPEG 专家组于 1999 年 2 月正式公布了 MPEG-4 (ISO/IEC14496) V1.0 版本。同年底 MPEG-4 V2.0 版本<sup>[12]</sup>亦告出炉, 且于 2000 年年初正式成为国际标准。

MPEG-4 标准将众多的多媒体应用集成于一个完整的框架内, 旨在为多媒体通信及应用环境提供标准的算法及工具, 从而建立起一种能被多媒体传输、存储、检索等应用普遍采用的统一数据格式, 并根据不同的应用需求, 现场配置解码器, 开放的编码系统也可随时加入新的算法模块。为了支持对视频内容的访问, MPEG-4 提出了“视频对象”的概念, 其数据结构如图 2-1 所示。

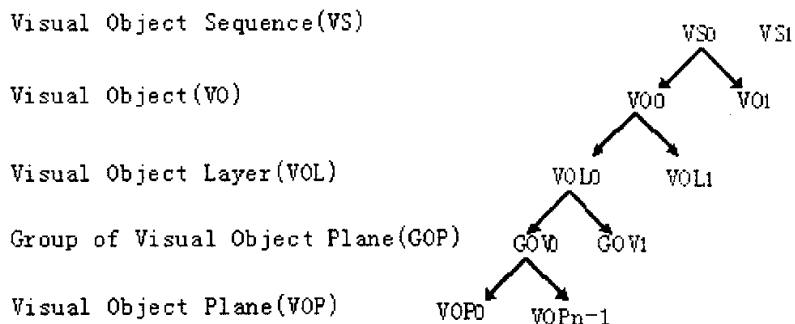


图 2-1 MPEG-4 的数据结构  
Figure2-1 Data Structure of MPEG-4

区别于传统的基于像素的视频压缩标准（例如 MPEG-1、MPEG-2、H. 261 等），MPEG-4 标准采用基于对象的视频编码方法，不仅可以实现对视频图像的高效压缩，而且还能够提供基于内容的交互功能。以前的视频编码压缩只是去掉了帧内、帧间的传统冗余信息，而 MPEG-4 则要求对视频图像做更加细致的分析，通过内容对象分割，并单独编码，从而在更大范围去掉了冗余信息。

图 2-2 演示了在 MPEG-4 下的音像场景中各个对象组合的应用实例。图中包括了 2D 的背景图像、3D 的物体（桌子、地球仪等），而且在传输到解码端以后可以选择并操作对象，如改变视角、重新安排对象顺序等交互式交流。

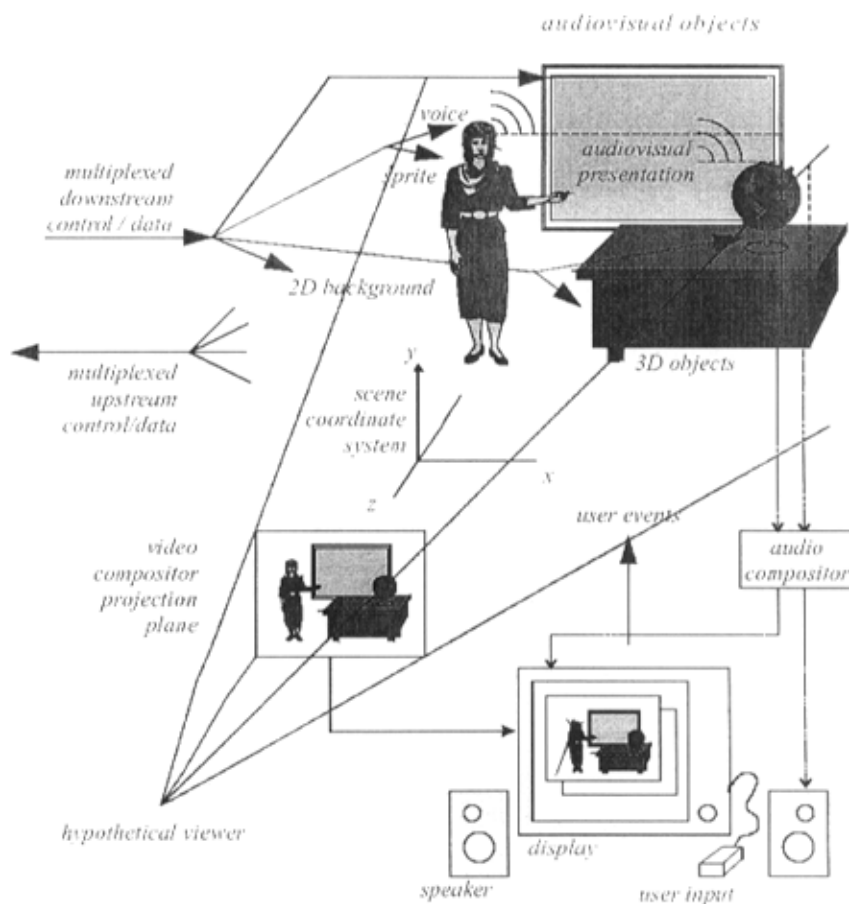


图 2-2 MPEG-4 应用示意图

Figure 2-2 Application demonstration of MPEG-4

在接收端，按照场景描述信息实现场景还原。AVO 数据流被解复用后，输入对应的解码器，恢复数据重建 AVO，然后把 AVO 从局部坐标系映射到场景全局坐

标系,按照树型结构组织起来,在接收端进行组合绘制。这里,用户可以在组合、解复用、复用、编码等四个阶段与AVO进行交互。甚至可以通过上行流发送交互信息到发送端。具体流程图2-3所示。

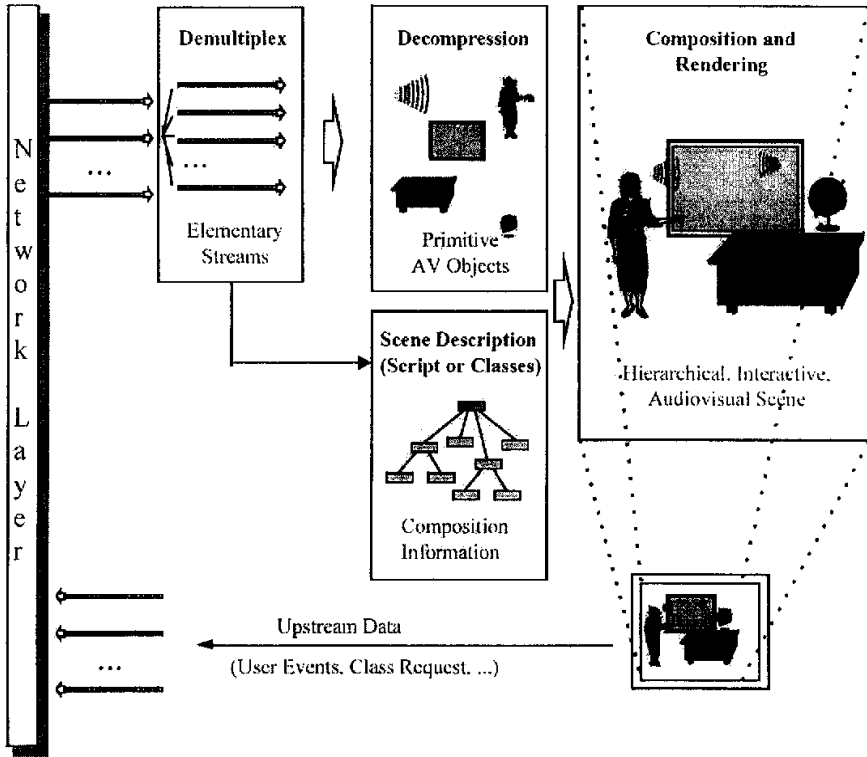


图 2-3 MPEG-4 视频接收端组件

Figure2-3 Major Components of an MPEG-4 terminal (receiver side)

## 2.2 MPEG-4 编码原理

MPEG-4 编码的基本单元是视频对象 (VO)。因为视频图像按照人眼视觉神经的特点,可以将视频内容按照纹理、运动、形状等特性,通过自动或者半自动的手段将视频图像分割为多个图像区域。每一个图像区域称为一个视频对象平面 (VOP), 同一对象连续的 VOP 称为视频对象<sup>[13]</sup>。

VOP 编码有 I-VOP、P-VOP、B-VOP 等三种模式,其中 I-VOP 为帧内 VOP, P-VOP 为预测 VOP, B-VOP 为双向预测 VOP。三类 VOP 均由 VOP 头、视频包头以及编码后的宏块组成。它们之间的关系见图 2-4 所示, I-VOP 作为帧内预测 VOP, 给 P-VOP 提供参考信息, 而 B-VOP 则需要参考前后两个方向的 I-VOP 和 P-VOP 信息。每个

VOP 编码之前, 需要生成一个定界框, 用来表示 VOP 视频帧中的位置和大小; 然后将定界框内的像素分为  $16 \times 16$  的宏块 (Macroblock), 并对各个宏块进行形状编码、纹理编码、运动信息编码 (运动估计/运动补偿) 等处理。

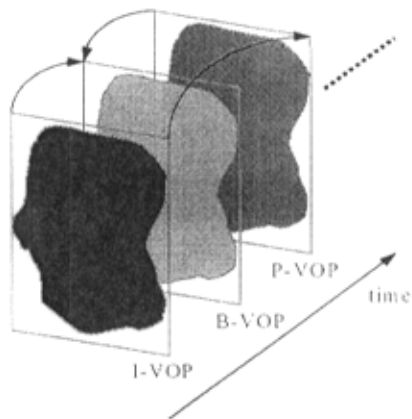


图 2-4 MPEG-4 中的 VOP 编码模式

Figure2-4 VOP Pattern of MPEG-4

图 2-5 是 MPEG-4 的编码原理框图。从图中可以看到, 视频流通过场景分析, 分割图像得到多个 VO。对每一个 VO 按照纹理、运动和形状等信息分别进行编码, 分配不同的码字, 然后将各个 VO 的码流合成为 MPEG-4 标准码流。

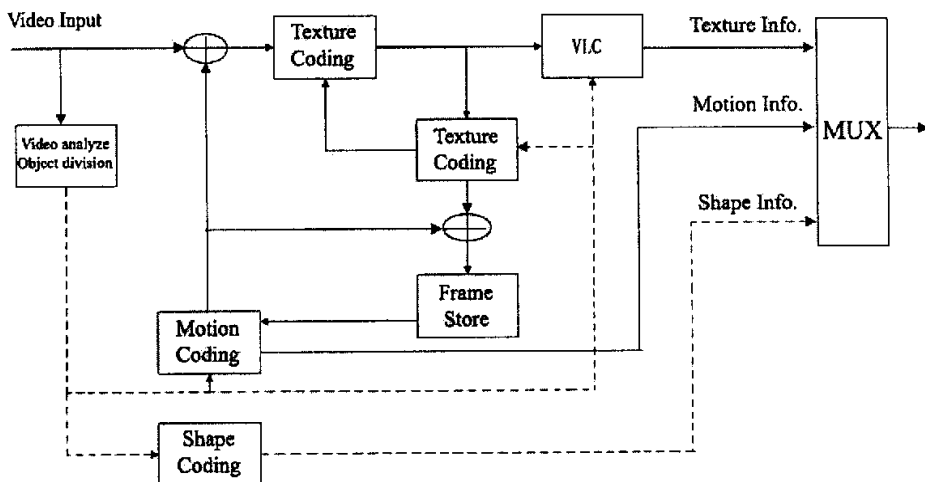


图 2-5 MPEG-4 视频编码原理框图

Figure2-5 Basic block diagram of MPEG-4 Video Coder

按照上述编码原理, 对视频序列按照图像内容进行分析, 按照对象进行分割, 然后分别编码。图 2-6 是一个在 MPEG-4 标准下的 VOP 编码样例。从图中可以看



到, 图像背景(雪山)作为一个 VO, 图像主体对象(松鼠)作为一个 VO, 两个 VO 分别按照不同的 VOP 进行编码, 然后在解码端进行合成, 即得到最终的图像。同时从这一样例可以看到, VOP 可以是任意形状的。

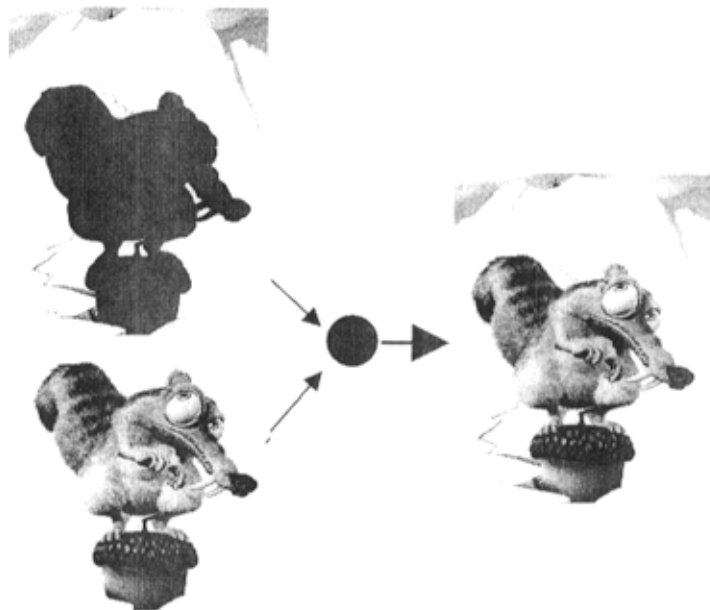


图 2-6 视频序列中的 VOP 编码样例  
Figure2-6 VOP Example of Video Sequence

### 2.2.1 形状编码

在 MPEG-4 标准中, VO 的形状信息包括二值形状信息和灰度形状信息两种, 形状信息采用位图矩阵表示, 又称为  $\alpha$  平面。采用位图表示, 不仅有较高的编码效率, 而且还有较低的计算复杂度。二值形状信息的每个元素取值有两种可能, 取决于像素和 VOP 的位置从属关系; 灰度形状信息的每个元素取值则在 0~255 之间, 其中 0 对应完全透明的元素, 255 对应完全不透明的元素。

图 2-7 为二值形状信息的 VOP 示意图。在 VOP 上覆盖界定框(Binary alpha Blocks), 每个 BAB 独立编码; 然后按照界定框和 VOP 的交集定义具体数值, 具体数值代表了界定框的位置是否从属于 VOP。如果 BAB 上的每一个数值为 255, 则称为不透明块(Opaque Block); 反之, 若 BAB 上的每个数值为 0, 则称为透明块(Transparent Block)。二值 alpha 平面可用临界信息进行基于上下文的算术编码(CAE)。

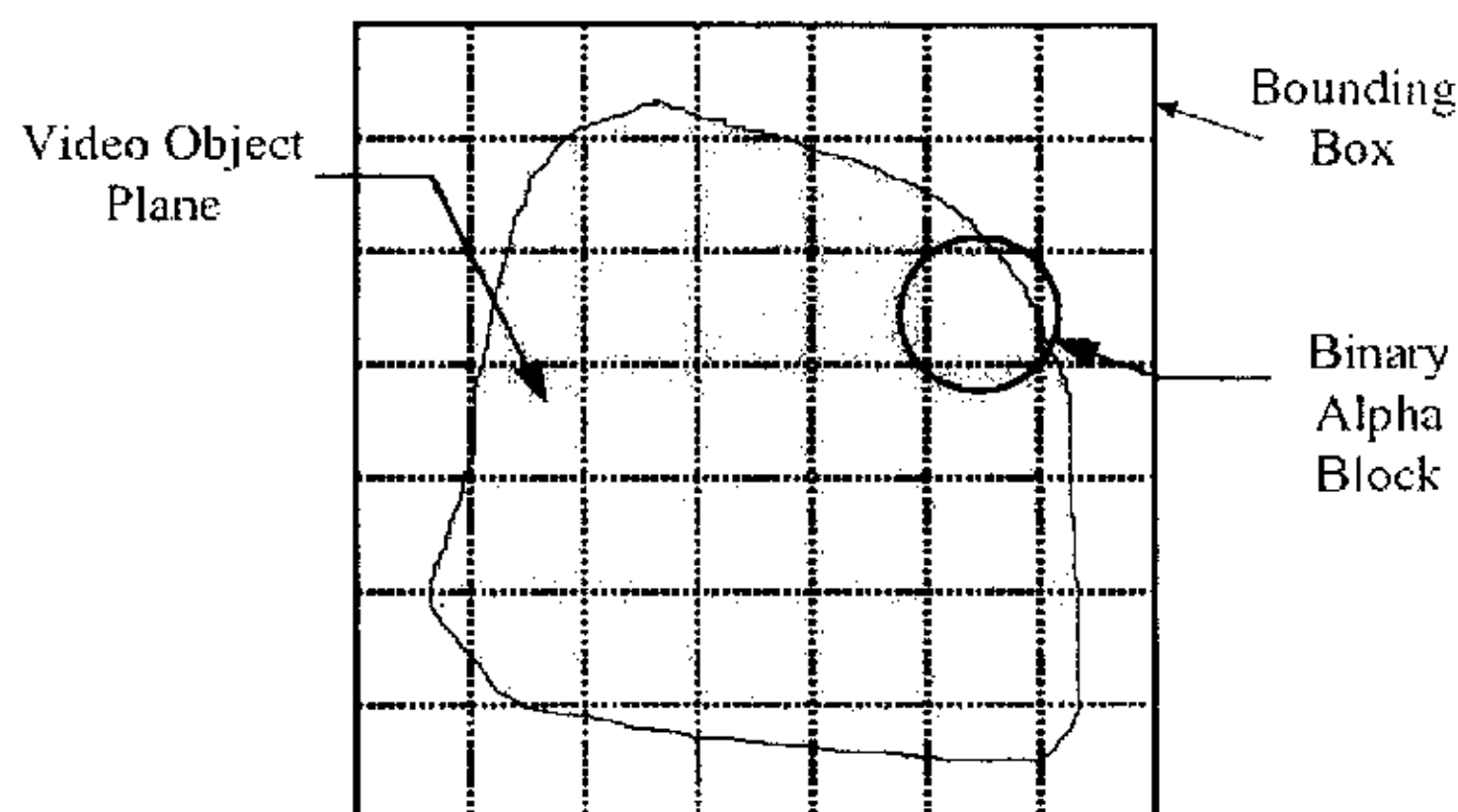


图 2-7 二值形状信息的 VOP 示意图

Figure 2-7 VOP with Binary Shape Information

二值形状编码的缺点在于无法标定每个块的透明程度。因此,在主层次(Main Profile)上, MPEG-4 提出了灰度形状信息编码方案,取值范围从 0 到 255,所以灰度形状编码可以提供足够多的对象透明程度表示信息。灰度 alpha 平面可用运动补偿加 DCT 变换类似纹理编码一样进行编码。

### 2.2.2 纹理编码

视频信号的 YUV 分量均包含了 VOP 的纹理信息。对于 I-VOP 来说,纹理信息直接包含在亮度和色度分量中,而对于其他两类 VOP,则包含在运动补偿后的残差中。纹理编码的工具具有可伸缩性静态纹理编码、可伸缩性形状编码、纹理容错控制以及小波变换等,它们共同有机组成了高级可伸缩性纹理编码。(具体组成参见图 2-8)

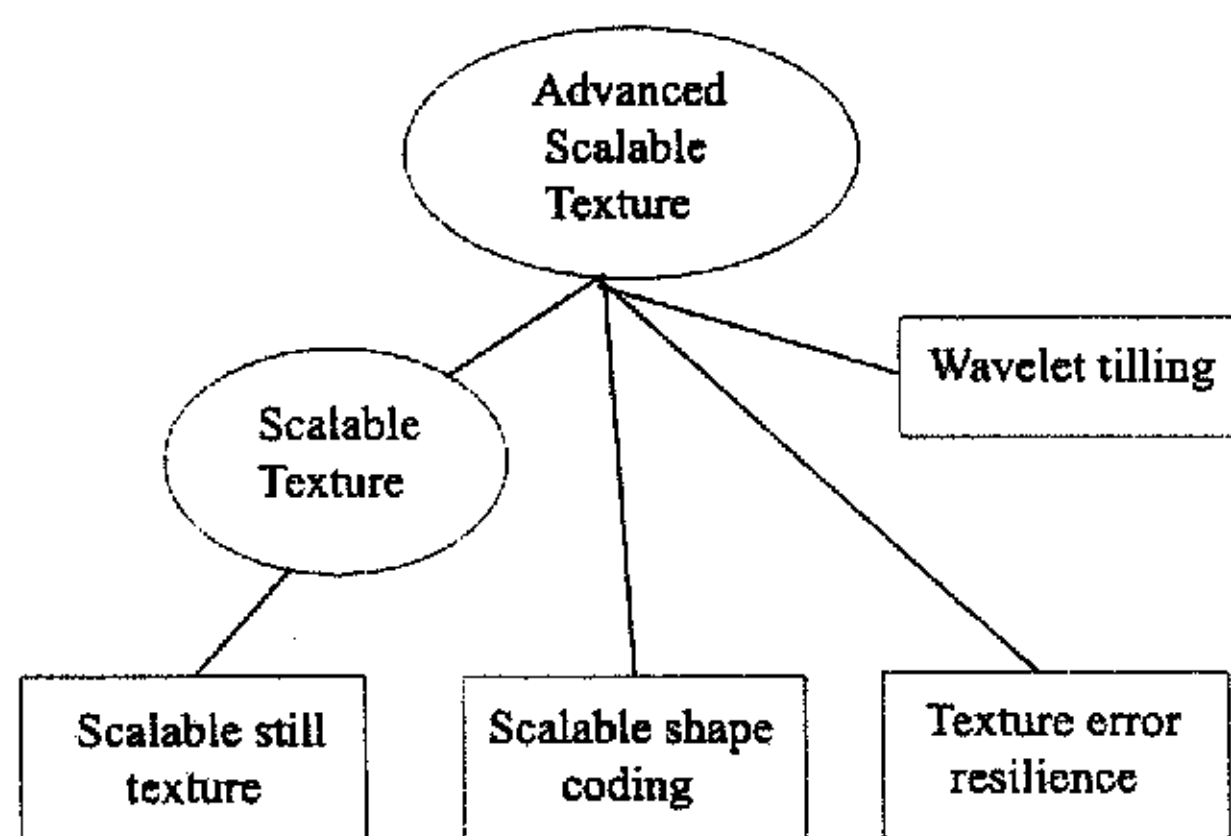


图 2-8 纹理编码的工具和对象示意图

Figure 2-8 Tools and objects for texture coding

在 MPEG-4 标准中, 静态纹理编码主要采用离散小波变换 (DWT)<sup>[11]</sup> 作为主要编码工具。视频图像通过一个 2D 的 DWT 变换, 得到一个 DC 分量 (低频子带), 三个 AC 分量 (高频分量), 然后重复对 DC 分量继续进行变换。这样就得到一个树状结构图 (见图 2-9)。

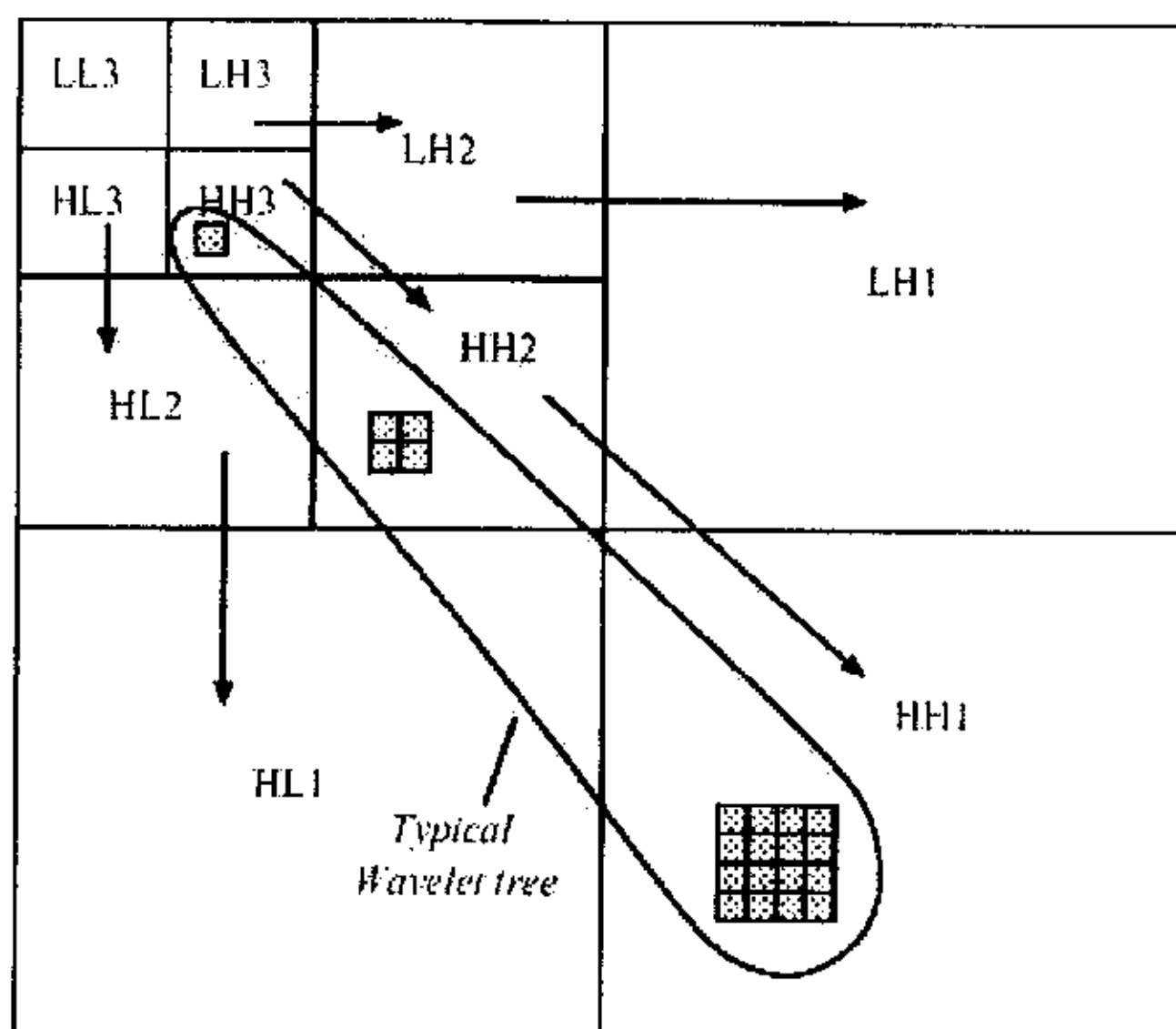


图 2-9 离散小波变换的树结构

Figure2-9 Tree Structure of DWT

图 2-10 演示了静态纹理编码流程, 通过 DWT 变换, DC 分量做量化、预测编码, 最后使用算术熵编码。而 AC 分量则经过量化、扫描以及零树编码, 最后使用熵编码。其中, AC 分量的扫描有两种: 一种是树结构, 最低的子带首先编码, 后面是下一层的子带; 另外一种带结构, 第一层的子带全部编码结束后才进行下一层子带编码。

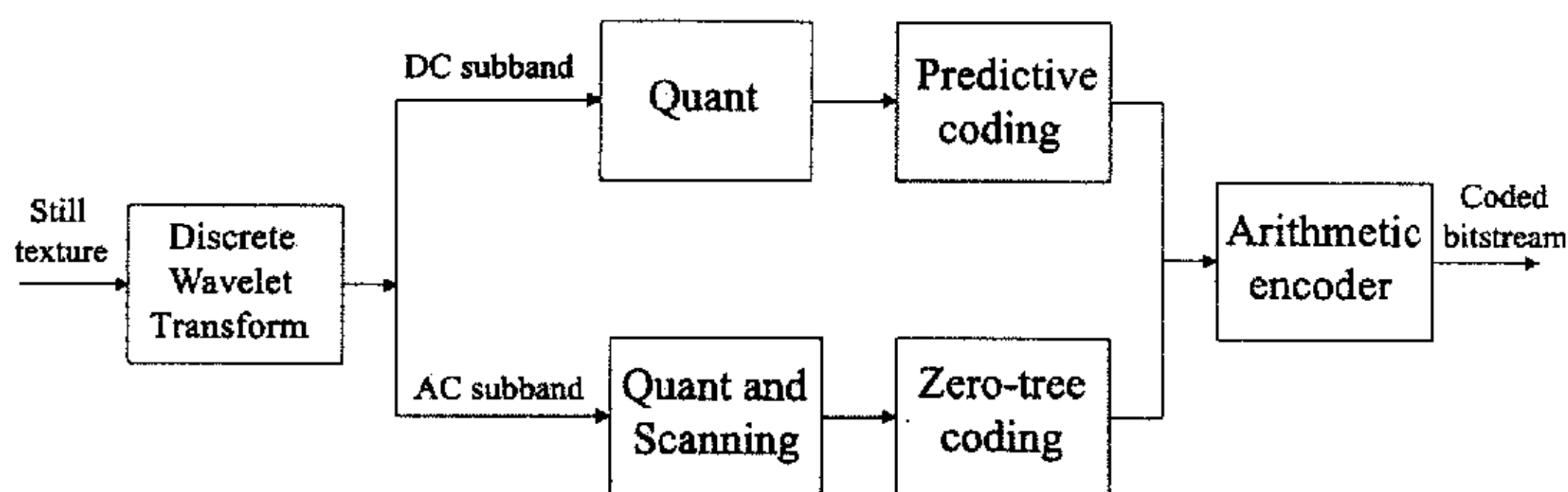


图 2-10 静止纹理的编码过程

Figure2-10 Still Texture encoder block diagram

### 2.2.3 运动信息编码

去除时间冗余信息是视频压缩的主要手段之一。一般来说,运动估计和运动补偿是目前去除帧间时间冗余信息的主要手段。与其他的视频编码压缩方案不同的是,在 MPEG-4 中运动估计和运动补偿的基本单位不是宏块,而是 VOP。P-VOP 和 B-VOP 由于需要参考前后 VOP 的信息,所以运动编码在这两类 VOP 中使用,而 I-VOP 则是被参考的数据源。

对 VOP 进行运动估计/补偿,有多种情况。对于 VOP 内不同的宏块需要采用不同的技术进行运动估计/补偿。如果一个宏块完全位于一个 VOP 内部,则采用  $16 \times 16$  或  $8 \times 8$  的块进行运动估计,运动向量采用  $1/4$  以上像素精度;如果一个宏块只有一部分位于一个 VOP 内部,则采用修正的块匹配技术估计运动向量;如果参考块位于 VOP 边界上时,采用重复填补技术给位于 VOP 外的那些像素制定像素值,然后用来估计运动矢量。至于 VOP 外部的宏块,则不作运动估计。这样,在 VOP 边界上搜索预测所需的候选像素时就有多种选择。

对于常规意义下的运动估计,将在第三章 H.264 技术中做进一步的阐述。在 MPEG-4 标准中,VOP 运动估计不仅支持基本的宏块运动估计,同时还支持高级预测模式。前者允许运动向量指向 VOP 之外,后者则允许每个宏块有 4 个运动向量。

### 2.2.4 Sprite 对象编码

对于一段视频序列,如果背景内容没有场景变换/切换,则每帧图像的背景就是这个大的固定背景的一部分。通常情况下视频图像的背景不会有任何运动,而在视频中体现出来的背景变化则是由于摄像机的运动或者前景物体运动造成的。Sprite 对象是针对背景对象这一特点提出的<sup>[15]</sup>。

通过图像的镶嵌技术将背景内容拼接成为一个大的背景图像,即为 Sprite 图像。Sprite 图像主要用来进行背景 VOP 的直接重建和预测编码。Sprite 图像只需编码一次然后存储在解码端,在随后的每一时刻的背景图像只需要利用传输摄像机相对于背景的运动参数,通过空间变换(如平移、放大、旋转、仿射、透视等)得到。图 2-11 为一个 Sprite 图像编码的实例:左上角是背景 Sprite 图像,通过仿射放大变换,得到相应的背景图像;左下角是前景视频对象,与变换后的背景图

像合成出右边的实际图像。

这样背景图像只需要传输一次即可在接收端根据用户需求合成相应视频,可以大大提高编码效率,增加视频的交互性。但是 Sprite 编码技术最大的缺陷在于 Sprite 图像所需要的传输量太大,容易造成传输延迟等问题。此外,如果背景变化过于复杂,或者背景 VOP 无法明确定义,则没有办法分割图像从而得到正确的 Sprite 背景图像。

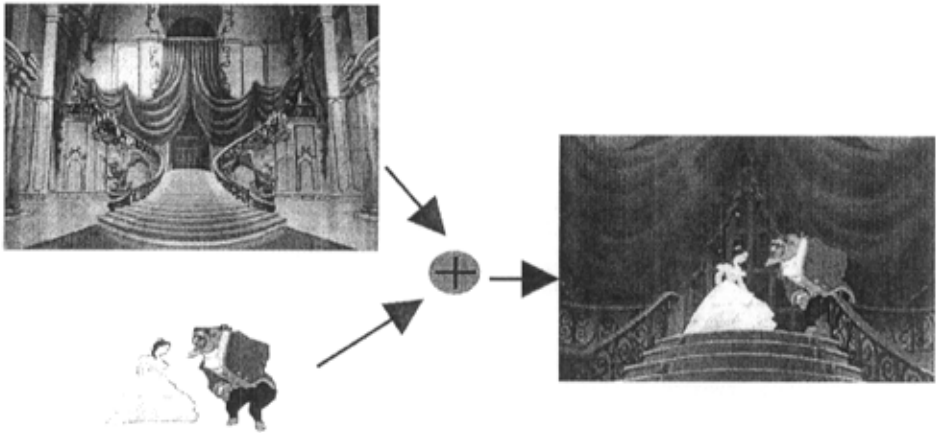


图 2-11 Sprite 编码样例  
Figure2-11 Example of Sprite Coding

### 2.2.5 可分级扩展编码

MPEG-4 标准最大的优势之一就在于针对不同的网络传输带宽,编码不同视频质量的数据流,这样可以满足不同规格的需求。

一般来说,MPEG-4 标准提供了时间和空间两种可分级扩展编码方式。其中,时间可分级扩展编码利用增强层增加 VOP 基本层的时间分辨率,主要通过增加传输和编码中间帧来获得更为准确的 VOP 运动信息;而空间可分级扩展编码则增加 VOP 基本层的空间分辨率,对源图像作下采样传输,然后在解码端通过插值恢复图像。

为此,MPEG-4 标准提供了不同水平应用层次的 Profile 以及相应的 Level 定义。Profile 是编码器提供的工具集的对象子集,内部又有多个级别(Level)来限制计算复杂度。不同的编码质量需求有相应的 Profile 层次。

Profile 的框架层次结构见图 2-12 所示。

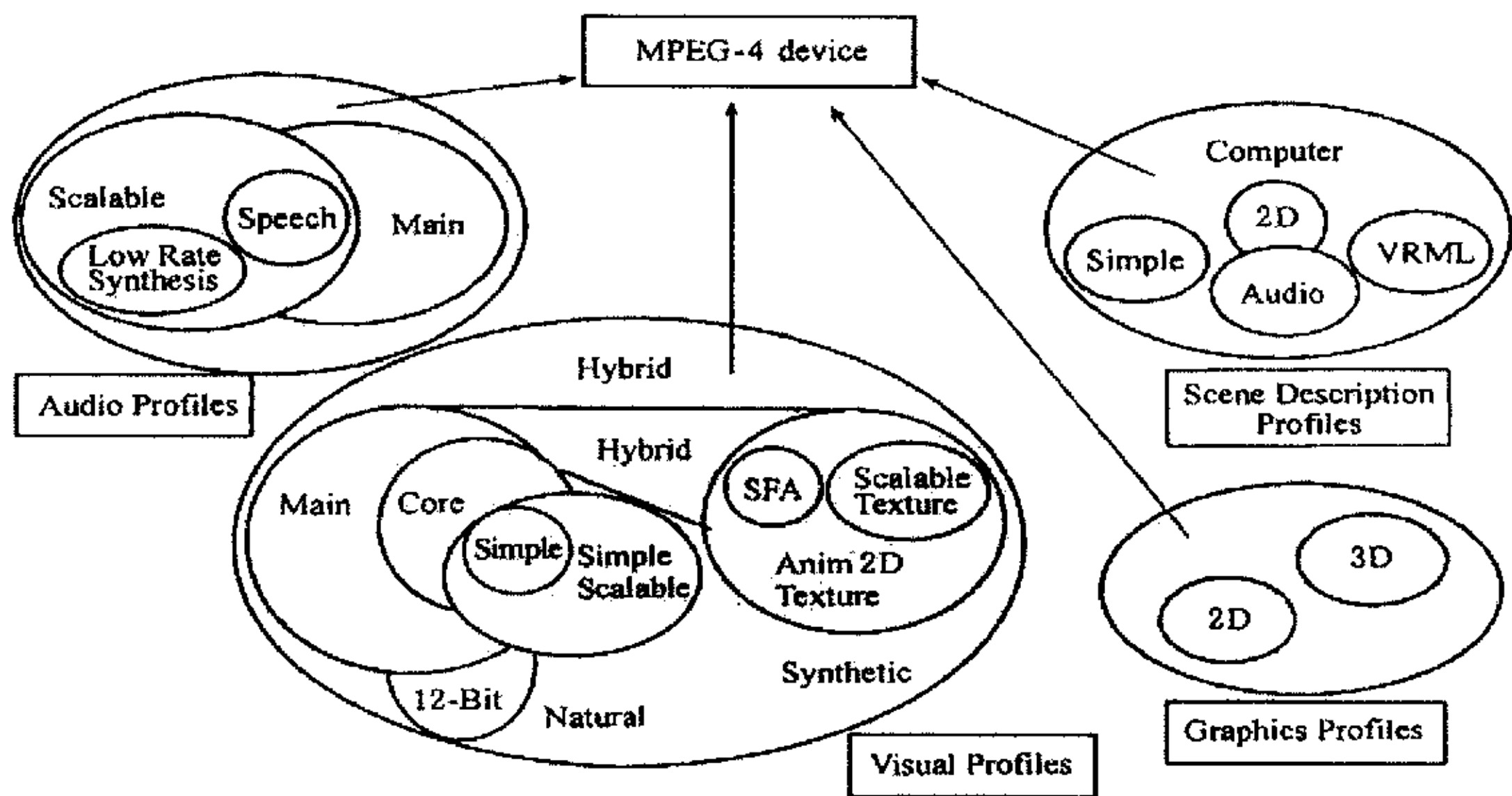


图 2-12 Profile 框架图

Figure2-12 Profile framework demonstration

### 2.2.6 VOP 解码

MPEG-4 编解码与传统的 DPCM 编解码原理一致，但是因为其操作对象是 VOP，因此针对 VOP 特点，编解码框架做了专门的改进与优化。在本节的开始部分已经介绍了 VOP 的解码原理，下面主要介绍解码框架（参见图 2-13）。

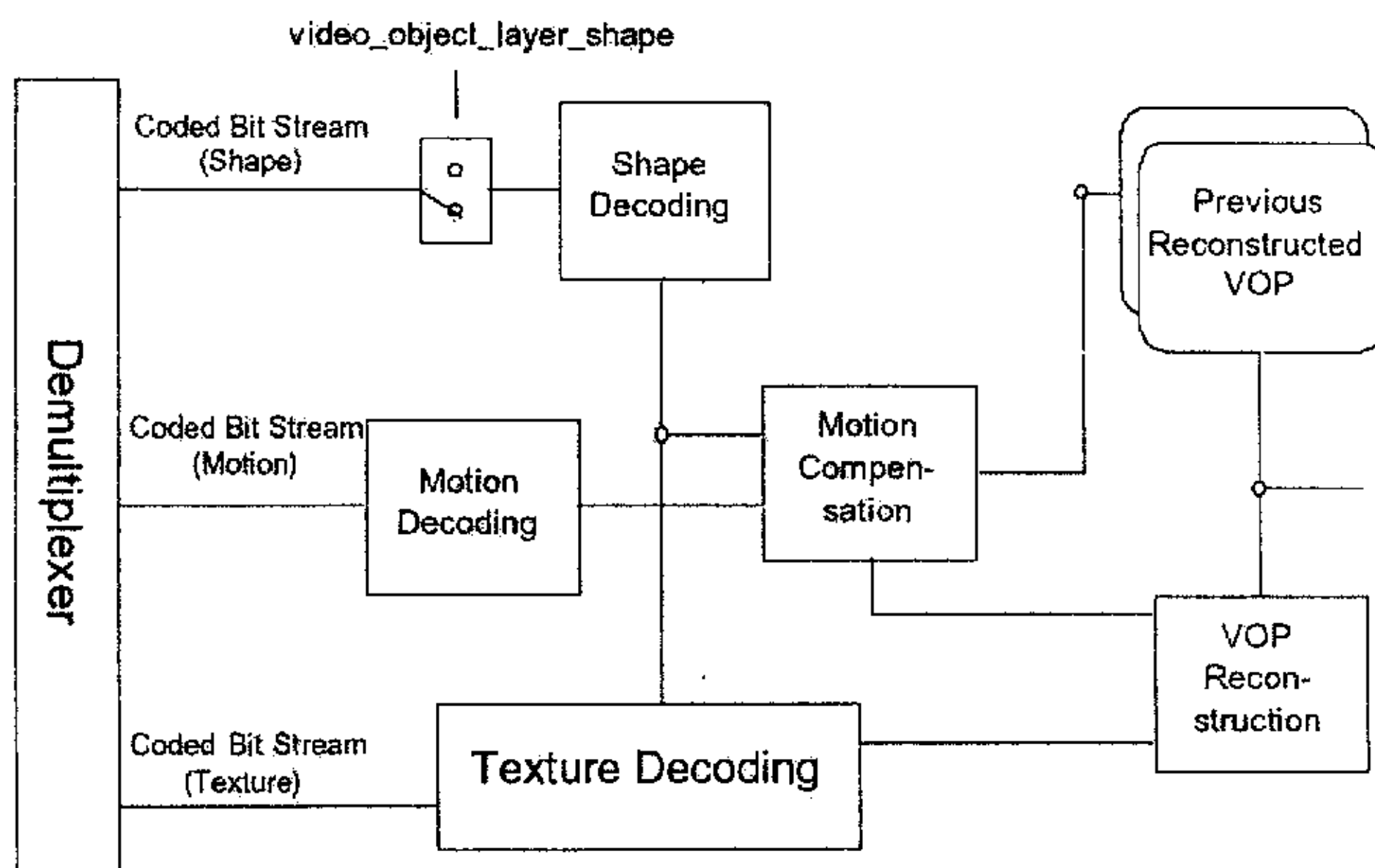


图 2-13 基于 VOP 的解码结构

Figure2-13 Decoder structure based on VOP

因为 MPEG-4 将视频对象分别按照形状、纹理、运动信息等分类分别进行编



码，所以在解码的相应环节也应该分别进行处理。如图 2-13 所示，编码数据流首先解复用得到编码后的形状数据、纹理数据以及运动信息数据，三种数据流分别进入相应的解码模块。这里的解码主要包括反量化和反变换，经过解码后的三种数据流利用前向重构 VOP 做运动补偿后，加上解码后的纹理信息，合成得到当前重构 VOP。

## 2.3 本章小结

MPEG-4 标准（视频部分）是一种全新的基于对象的视频编码技术，代表了新一代视频编码技术水平。通过对视频内容进行分析，分割出不同的 VOP，并对 VOP 的形状信息、纹理信息以及运动信息等单独进行编码合成，从而可以有效提高视频质量和编码效率。

MPEG-4 标准最大的优势之一就在于针对不同的网络传输带宽以及不同的用户需求，编码不同视频质量的数据流，这样可以满足不同规格的需求。

MPEG-4 标准提供了前所未有的交互性、灵活性和鲁棒性，并且具有极为高效的编码效率。可以预见在未来的多媒体编码领域中，以及在无线通讯领域中，MPEG-4 都将具有广阔的应用前景。

### 第三章 H. 264 (AVC) 编码技术

H. 264 是国际电信联盟组织和国际标准化组织成立的联合专家组 (JVT) 提出的一项音视频编码标准<sup>[16][17][18]</sup>, 英文名称: Advanced Video Coding (AVC)。该标准从 H. 26L 标准基础上发展起来, 于 2003 年, 正式发布 H. 264 标准文档。凭借其卓越的编码性能和高效的视频编码效率, 不仅为全世界各大视频编码芯片厂商所认同, 而且正式成为了 MPEG-4 编码标准中的一部分。

#### 3.1 H. 264 标准概述

H. 264 标准于 1998 年 1 月份开始草案征集, 1999 年 9 月完成第一个草案, 2001 年 5 月制定了测试模式 TML-8, 2002 年 6 月的 JVT 第 5 次会议通过了 H. 264 的 FCD 版<sup>[19]</sup>。2003 年 3 月正式发布, 同时作为 MPEG-4 标准的第十部分。标准主要侧重于端到端的低比特数据流的实时传输, 如视频会议等。

H. 264 和以前的 H. 26X 系列标准一样, 也是 DPCM 加变换编码的混合编码模式。但它采用了“回归基本”的简洁设计, 不用众多的选项, 却获得了比 H. 263++ 好得多的压缩编码性能; 同时还加强了对各种信道的适应能力, 采用了“网络友好”的结构和语法, 有利于对误码和丢包的处理; 另外, 增加了两层网络结构概念: 视频编码层 (VCL) 和网络抽象层 (NAL), 从而极大拓宽了该技术标准的应用目标范围, 可以满足不同速率、不同解析度以及不同传输 (存储) 场合的需求。

H. 264 标准不仅仅支持 QCIF、CIF 等低分辨率视频格式, 而且还支持 720P、1080I 等高分辨率视频格式。处理的视频帧结构为 YUV 信号数据, 具体的格式主要有 4:2:0 逐行扫描或隔行扫描等多种格式的视频序列。至于预测方案, H. 264 标准主要使用帧内预测和帧间预测两种模式。

H. 264 标准还有多处技术优势, 如统一的 VLC 符号编码, 高精度、多模式的位移估计, 基于  $4 \times 4$  块的自适应 DCT 变换, 分层的编码语法等。这些技术特点使得 H. 264 标准算法具有很高的编码效率, 在相同的重建图像质量下, 能够比 H. 263 节约 50% 左右的码率。H. 264 的码流结构网络适应性强, 增加了差错恢复能力, 能够

很好地适应IP和无线网络的应用。

## 3.2 H.264 编码原理

H.264 标准的编码原理也是以传统的 DPCM 编解码框架为基础的，并在编码的各个环节做了改进。下面，介绍 H.264 的基本框架结构和编码技术特点。

### 3.2.1 H.264 结构

与 MPEG-4 标准类似，H.264 标准定义了三类框架(Profile): Baseline、Main、Extended 三个级别。每一级的 Profile 定义了不同的码字功能以及相应的编码器/解码器。Baseline 级别包括帧内编码、帧间编码、CAVLC 熵编码；Main 级别包括隔行扫描视频、使用 B 帧以及权重预测的帧间编码、CABAC 熵编码；Extended 级别不支持隔行扫描视频和 CABAC 编码，但增加了 SP 和 SI 帧，并提高了容错性能。图 3-1 展示了三类 Profile 的相互关系结构。

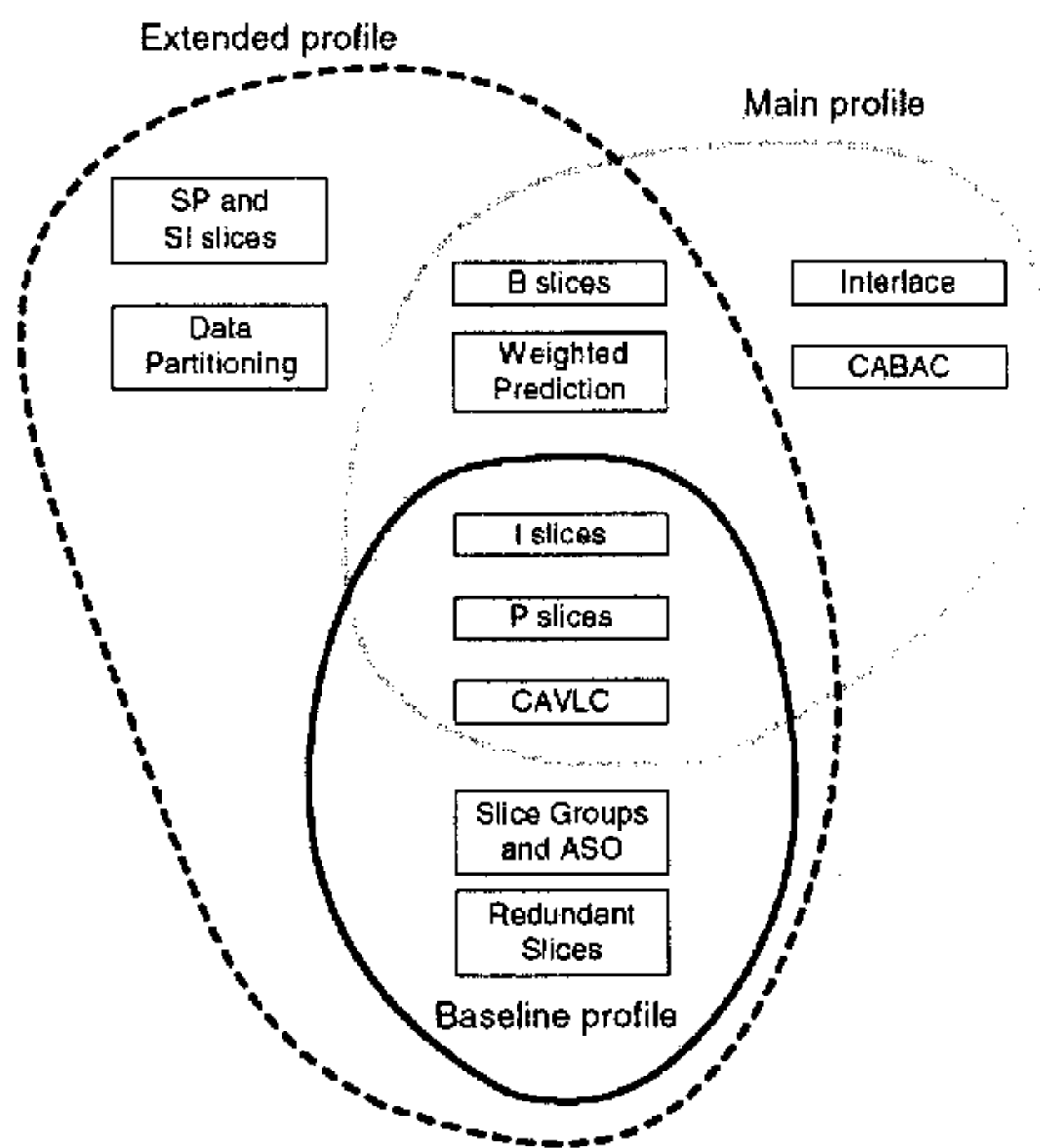


图 3-1 H.264 三类框架关系结构图

Figure3-1 Profiles structure of H.264

从应用层面上讲，Baseline 框架可以提供视频电话、视频会议和无线传输等应用，Main 框架可以提供电视广播和视频存储等应用，至于 Extended 框架则可以提供流媒体一级的应用。

H.264 的算法在概念上可以分为两层：视频编码层（VCL）和网络提取层（NAL）。前者负责高效的视频内容表示，包括基于块的运动补偿混合编码和一些新特性；后者负责按照网络所要求的恰当的方式对数据进行打包和传送，包括组帧、逻辑信道的信令、定时信息的利用或序列结束信号等。在 VCL 和 NAL 之间还定义了一个基于分组方式的接口，打包和相应的信令属于 NAL 的一部分。这样，高编码效率和网络友好性的任务分别由 VCL 和 NAL 来完成。两层的具体结构参见图 3-2。

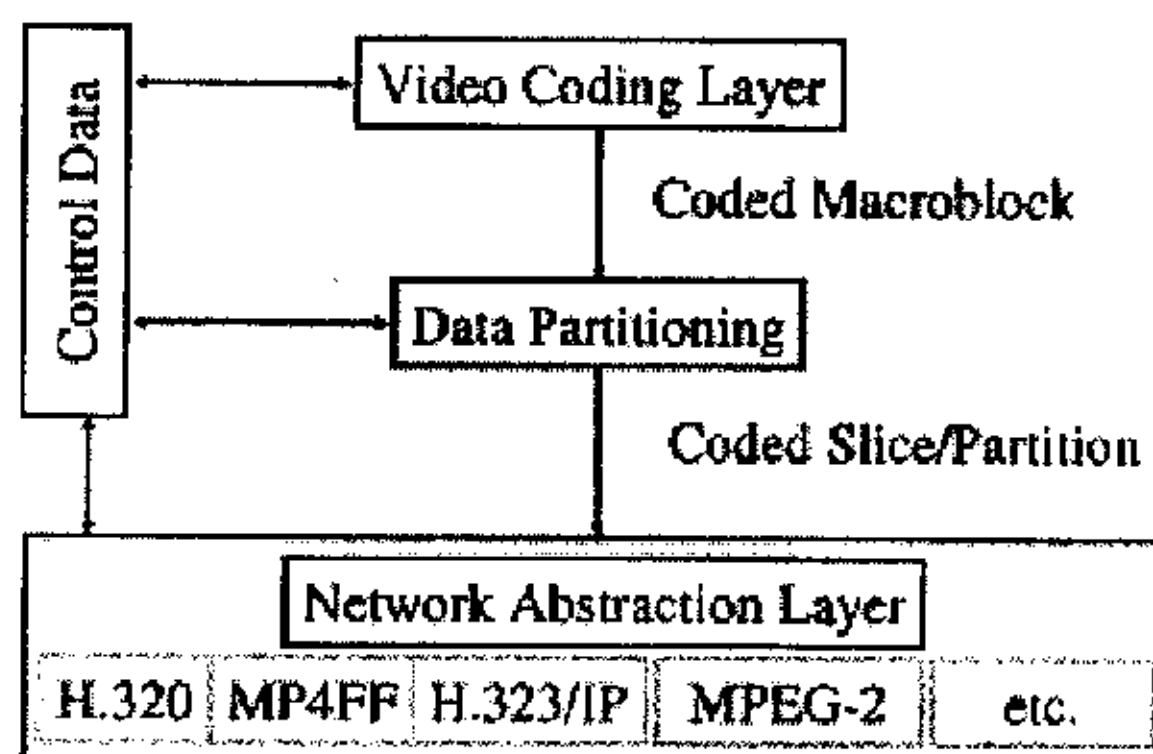


图 3-2 H.264 中 VCL 和 NAL 关系框图(Coder)

Figure3-2 Relationship of NAL and VCL in H.264

另外，H.264 标准还引入了一个全新的编码单位条带（Slice）。条带是由整数个宏块（Macroblock）组成，每帧图像一般包括 1 个或多个 Slice。引入了条带这一编码单位，可以在条带内直接进行编码和纠错，就可以极大提高视频编码容错率和编码效率。图 3-3 是 Slice 的语法结构图。

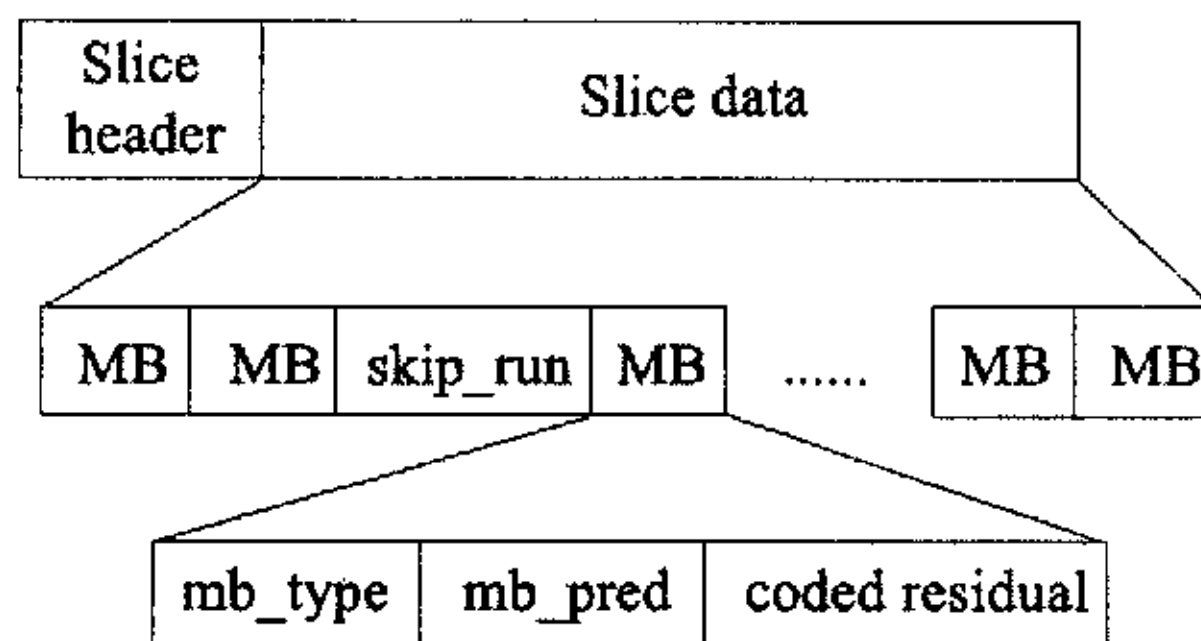


图 3-3 H.264 中的条带语法结构图

Figure3-3 Slice syntax of H.264

这里的宏块（Macroblock）与 H.26X 系列的前述标准定义类似，均为  $16 \times 16$  的亮度块和  $8 \times 8$  的色度块（以 4:2:0 的视频格式为例）。

### 3.2.2 编码技术

H.264 作为新一代视频编码标准, 主要定义编码出来的数据流的语法规则和相应的解码手段, 而对编码器一般不作规定。H.264 的编码框架也是以 DPCM 原理框架为基础的 (参见图 3-4)。

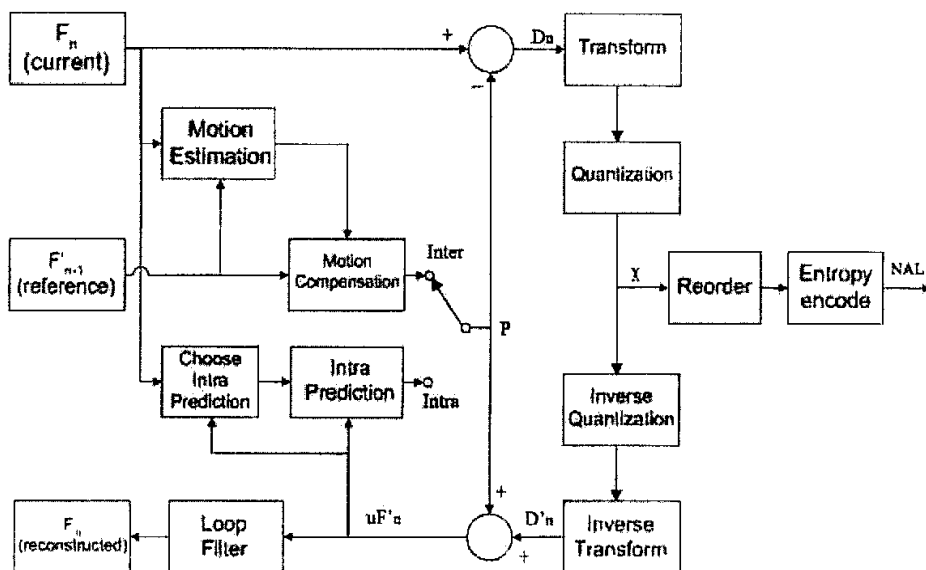


图 3-4 H.264 编码器原理框图(Coder)

Figure3-4 Encoder Block Diagram of H.264

从上图可以看到, H.264 编码器主要包括两个数据流向: 一个是前向编码路径 (从左向右), 一个是反向重构路径 (从右向左)。具体编码过程如下: 如果是帧内编码, 则当前编码帧利用帧内预测方案, 得到相应的像素预测值, 然后对图像残差值做 DCT 变换、量化、熵编码; 如果是帧间编码, 当前帧与前向重构帧进行运动估计得到运动向量, 在此基础上对前向重构帧做运动补偿, 得到的当前重构帧与当前帧做差值, 对差值做 DCT 变换、量化、熵编码; 至于重构帧的生成, 主要依靠对编码出来的数据流做反量化和反 DCT 变换; 最后采用环路滤波消除编码所带来的视频块效应。

### 3.2.3 解码技术

H.264 标准提供的解码器从 NAL 层接收压缩数据流。首先做熵解码, 得到量化系数 X; 然后做反量化和反 DCT 变换。利用数据包的信息头解码得到预测

块,如果是帧内编码,则直接按照所选择的帧内预测模式恢复;如果是帧间编码,则利用前向参考帧做运动补偿。在重构的最后一个环节依然是用来消除块效应的环路滤波。具体解码流程参见图 3-5。

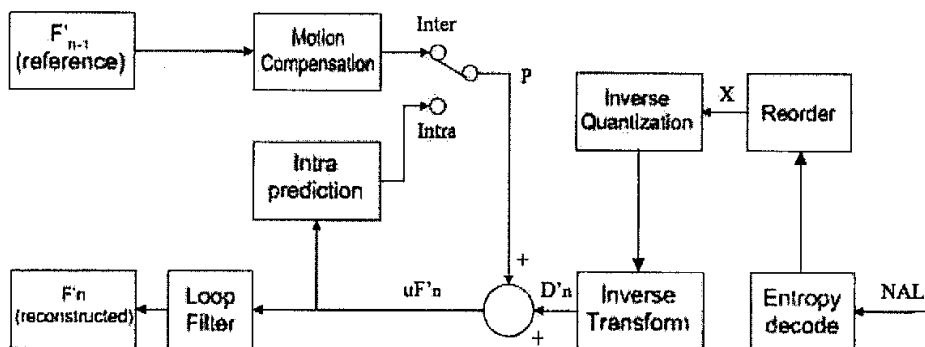


图 3-5 H.264 解码器原理框图(Decoder)

Figure3-5 Decoder Block Diagram of H.264

### 3.2.4 视频编码帧类型

在 H.264 标准中,也定义了 3 类基本图像帧类型: I 帧、P 帧和 B 帧。I 帧为帧内预测帧,一般作为视频序列的首帧,以及每个视频组 (GOP) 的首帧; P 帧为帧间预测帧,主要通过参考前向的 I 帧或者 P 帧进行编码; B 帧为双向预测帧,参考前向/后向的 I 帧或者 P 帧进行编码。由于参考帧的类型和数量不同,所以 B 帧的编码效率最高,但是编码质量最差;而 I 帧反之。

一般来说, 由于编码效率的需要, 需要设定视频序列中每帧的编码类型, 但是由于不同类型的视频帧所需要的参考顺序和参考方向的不同, 所以编码帧的输入顺序和解码帧的输出顺序将有很大不同。图 3-6 演示了不同类型的编码帧之间的参考关系, 箭头方向为参考方向。

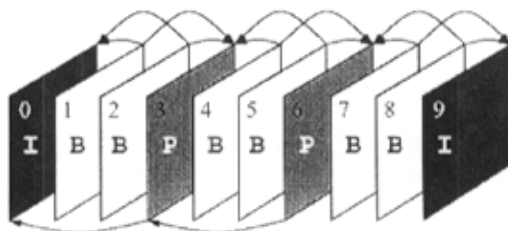


图 3-6 视频序列的每帧参考顺序示意图

**Figure3-6 Frame referenced sequence of H.264**



如果给定一段视频序列，且规定不同类型的编码帧的输入顺序为 I、B、B、P、B、B、P……（参见图 3-6），则必须按照被参考的帧先编码、参考帧后编码的顺序，输出视频顺序为：I、P、B、B、P、B、B……。在视频最后输出阶段，还需要对解码输出的视频编码帧按照原始输入顺序重新排列，才可以得到正确的解码视频流。

为了提高视频数据流的网络传输质量，满足不同规格的需求，H.264 标准针对不同的应用框架（Profile）提出了不同的编码帧类型。在 Extended 框架下，提出了两类全新的视频编码帧——SP 帧和 SI 帧。

如果要从一个视频码流到另外一个视频码流的转换，由于 P 帧和 B 帧的参考关系，只能转到目标码流的 I 帧。为了避免目标码流的 I 帧重复设置，出现了 SP 帧和 SI 帧。SP 帧类似于 P 帧，允许在参考帧不同的情况下重构出相同的图像，这样可以实现码流切换、快进、实时访问等功能。而 SI 帧类似于 I 帧，采用帧内编码，可以在不需要参考帧的情况下完全相同的重构出另外一幅 SP 帧。图 3-7 即为码流转换中的 SP 帧示意图，码流转换一般为不同码率下的同一内容的视频流。这里 AB2 即为 SP 帧，实现了从 A1 到 B3 的直接转换。

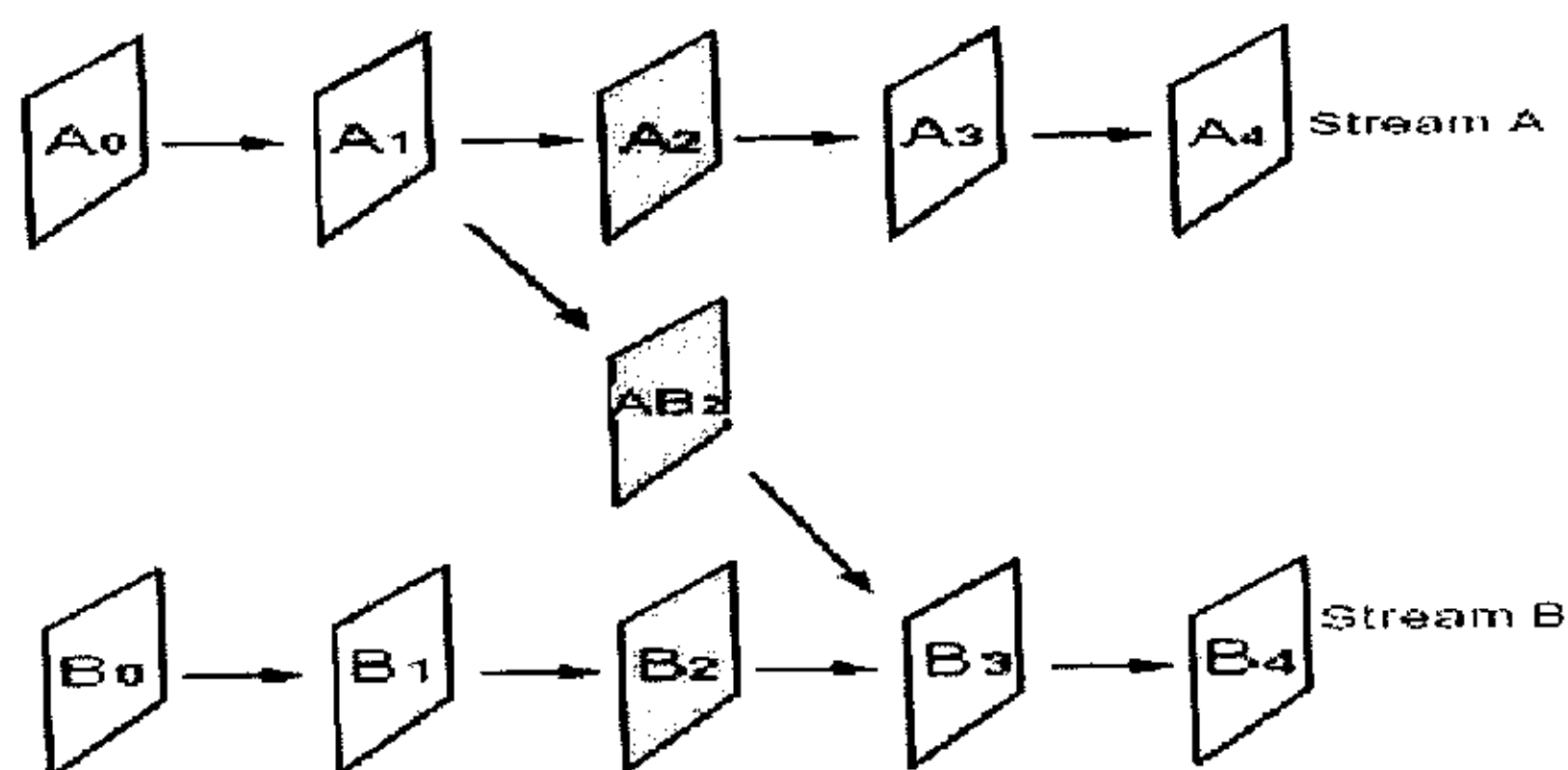


图 3-7 SP 帧的码流转换示意图  
Figure3-7 Switching streams using SP-slices

### 3.3 本章小结

H.264 标准具有广阔的应用前景，如实时视频通信、互联网视频传输、视频流媒体服务、异构网上的多点通信、压缩视频存储、视频数据库等。

H.264 建议的技术特点可以归纳为三个方面，首先是注重实用，采用成熟的

技术，追求更高的编码效率，简洁的表现形式；其次是注重对移动和 IP 网络的适应，采用分层技术，从形式上将编码和信道隔离开来，实质上是在源编码器算法中更多地考虑到信道的特点；最后是在 DPCM 混合编码器的基本框架下，对其主要关键部件都做了重大改进，如多模式运动估计、帧内预测、多帧预测、统一 VLC、 $4 \times 4$  二维整数变换等。

不过，H. 264 标准优越编码性能的代价是计算复杂度的大大增加，据估计，编码的计算复杂度大约相当于 H. 263 标准的 3 倍，解码复杂度大约相当于 H. 263 标准的 2 倍。

迄今为止，H. 264 因其更高的压缩比，更好的信道适应性，必将在数字视频的通信或存储领域得到越来越广泛的应用。

在随后的章节里，在现有的 H. 264 标准技术基础上，将分别介绍块匹配算法、帧内预测以及环路滤波等三个方面的创新思路，所做的实验也是在 H. 264 标准提供的软件技术平台上实现。

## 第四章 基于内容的简单化块匹配算法

在视频编码与压缩技术中，块匹配算法（BMA）作为一种应用十分广泛的算法，主要用于对运动向量的估计和补偿等计算<sup>[20][21]</sup>。此外，在MPEG-4标准中的特征点跟踪也有着广泛的应用。传统的块匹配算法具有较高的计算复杂度、匹配结果的不确定性和随机性等缺点，导致算法在实际中的应用受到严重的限制。

本章将就MPEG-4标准中的人脸特征点（FAP、FDP）跟踪为应用技术角度，对现有的传统块匹配算法做优化与简化，针对匹配效果做进一步的改进。

### 4.1 BMA 概述

块匹配算法的主要手段是衡量两幅图像目标块与匹配块的相关程度。通过图像目标块与匹配块的相关性参数的计算，寻找与当前像素块最匹配的目标点。图4-1演示了块匹配算法的基本原理，通过相关窗口在搜索窗口的移动匹配，找到最佳匹配点。但是由于在计算过程中，目标块（相关窗口）大小涉及到了像素相关性的相互影响问题，其大小选择将直接影响匹配结果，所以针对不同的图像应该选择不同大小的相关窗口进行计算。在搜索计算过程中，由于涉及到窗口内每个像素的匹配计算过程，所以计算复杂度一直是BMA发展的瓶颈之一。

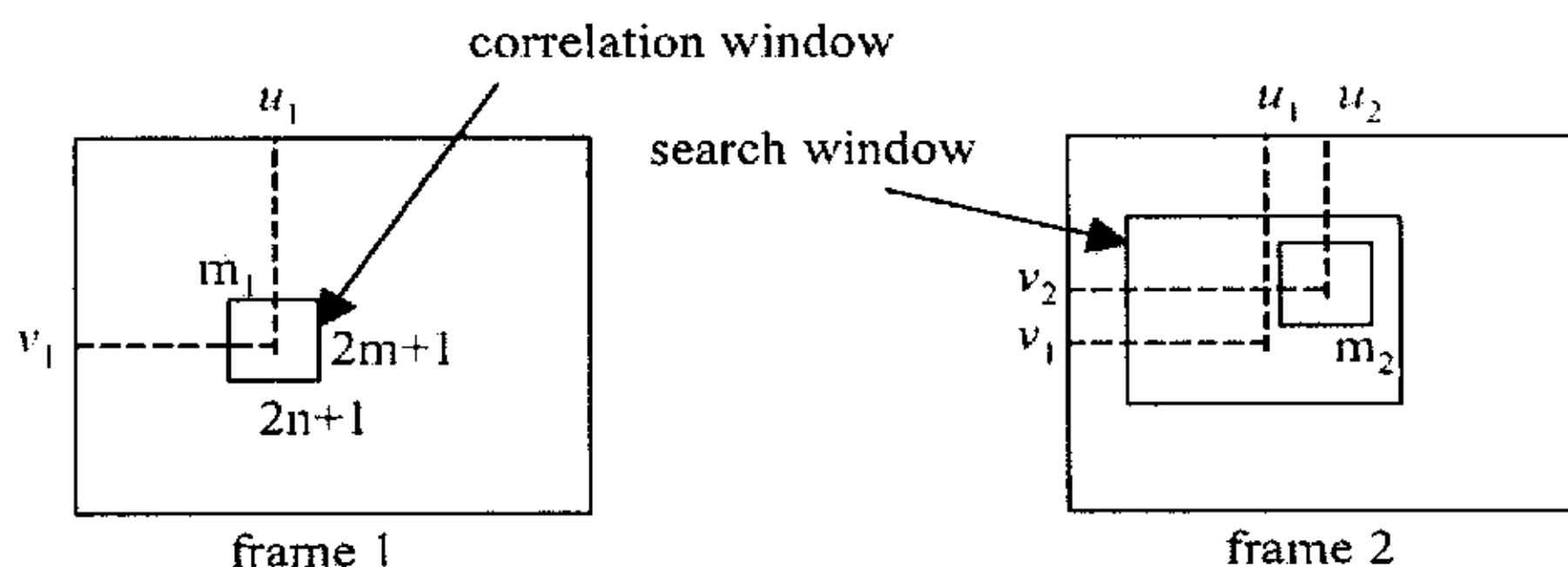


图 4-1 块匹配计算原理示意图

Figure4-1 Block-matching between two consecutive frames

目前，针对块匹配算法计算复杂度高的特点，利用现代搜索策略，对传统算法作了很多的改进，主要有：三步搜索（3SS）<sup>[22]</sup>、对数搜索<sup>[23]</sup>以及交叉搜索算法<sup>[24]</sup>等。这些算法基本上可以分为三大类。第一类是通过减少搜索区域来提高计

算效率的算法；第二类是通过减少匹配时间的策略算法<sup>[25]</sup>；还有一类就是两者相结合的算法<sup>[26]</sup>。目前，大多数改进算法主要是基于第一类算法，但有时为了更加明显地提高计算速度同时兼顾第二类策略。为了大幅度地提高计算效率，这些改进算法最大的弊端就是把全局最优变成了局部最优，这样就造成了匹配错误率和匹配结果不稳定性增加。此外，几乎所有的块匹配算法都没有基于内容的优化方案，这样针对不同内容的视频流来说，如视频光照条件不同、视频对象运动速度不同等情况下，匹配效果会有很大差异。

由于具体应用场合的限定，基于特定内容的块匹配改进算法将有更大的应用前景。因此，许多改进算法针对所处理的特定内容，基于匹配目标的几何特点和图像特性，加入了相应的几何位置约束以及像素之间相互关系的限制，同时还考虑特定视频内容的相互限制关系，从而大幅减小搜索区域，提高了匹配搜索的准确度，并降低了计算复杂度<sup>[27][28]</sup>。但是，这一做法的缺点在于仅仅限制于特定的视频内容，应用范围十分局限。

## 4.2 基于内容的简单化块匹配算法

块匹配算法高昂的计算代价主要是来自于频繁的像素值叠加计算。影响相关系数结果的另外一个重要因子是搜索窗口和相关窗口的大小  $m, n$ （参见图 4-1）。不同的图像，如果采用固定  $m, n$  将直接影响相关系数反映匹配块和目标块的相似程度的准确性。实际上，对于不同的匹配块，应该采用不同大小的  $m, n$ 。

一般来说，块匹配计算有多种计算方案，最为简便的是 SAD（见公式 4-1）。该算法主要用于视频编码中运动向量的计算等方面。但是，SAD 计算具有简便性的同时，却没有考虑搜索匹配块内部的像素分布，因此匹配结果可能会产生错误。如果使用相关系数计算公式<sup>[29]</sup>，则可以把匹配块的像素内容考虑进来，但是计算复杂度将随之大幅增加（参见公式 (4-2) (4-3) (4-4)）。该算法主要用于 MPEG-4 标准视频编码中的特征点跟踪和检测等方面。

$$SAD(x, y) = \sum_{v=y-n}^{v=y+n} \sum_{u=x-m}^{u=x+m} |I_j(u, v) - I_{j-1}(u, v)| \quad (4-1)$$

$$score(m_1, m_2) = \frac{\sum_{i=-n}^n \sum_{j=-m}^m ([I_1(u_1 + i, v_1 + j) - \overline{I_1(u_1, v_1)}] \times [I_2(u_2 + i, v_2 + j) - \overline{I_2(u_2, v_2)}])}{(2n+1)(2m+1)\sqrt{\delta^2(I_1) \times \delta^2(I_2)}} \quad (4-2)$$

$$\text{其中, } \overline{I_k(u, v)} = \frac{\sum_{i=-n}^n \sum_{j=-m}^m I_k(u+i, v+j)}{(2n+1)(2m+1)}, (k=1,2) \quad (4-3)$$

$$\delta(I_k) = \sqrt{\frac{\sum_{i=-n}^n \sum_{j=-m}^m I_k^2(u+i, v+j)}{(2n+1)(2m+1)} - \overline{I_k(u, v)}^2}, (k=1,2) \quad (4-4)$$

为了减小计算复杂度,同时保证匹配效果,本文提出了预处理反馈环和采样匹配等方案。下面将分别讨论具体技术细节。

### 4.2.1 预处理反馈环

在块匹配算法中,相关窗口的大小直接影响了匹配结果。通常,两幅图像的相似程度大,则应该计算主要的相关像素而不考虑过多的周边像素,否则周边的噪音可能会影响相关系数的准确度量,所以应该选用较小的相关窗口。反之,如果两幅图像的运动差异较大,则应该选用较大的相关窗口,这样可以尽量多的考虑到周边像素,从而获取最为相似的目标点。

针对块匹配算法这一特点,添加了一个预处理反馈环,主要目的是衡量两幅图像的内容相关性,获得两幅图像之间的相似度,依此确定下一步精细块匹配算法所使用的相关窗口的大小。具体流程如图 4-2 所示:

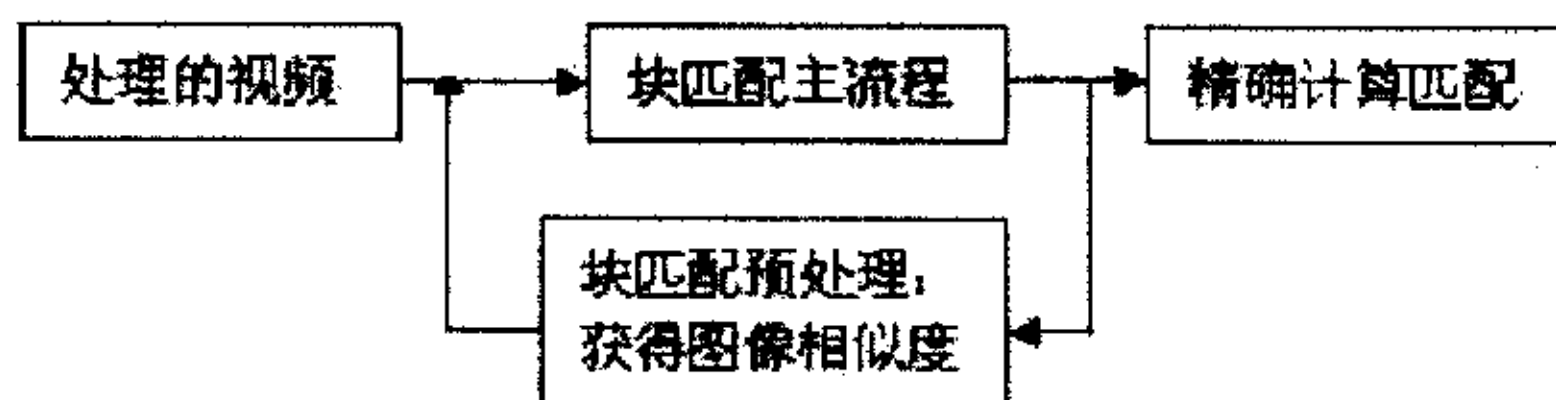


图 4-2 预处理反馈环算法流程

Figure4-2 Flow of Pre-process Feedback

为了减小计算复杂度,相关窗口应该尽量小,但是小到一定阈值以后,就有可能根本改变目标块的相关性。与此同时,为了尽量准确地获取像素之间的相互关系,需要扩大相关窗口,但这样又会因为图像噪音和本身不相关的像素的加入而导致根本改变目标块的相关性。所以应该判断目标图像的相似程度,得到图像内容的大致相关关系,从而确定合适的相关窗口的大小,做到有目的的计算与匹配。本文的预处理反馈环评估图像目标块的相关性,将评估结果与统计阈值进行比较,以此决定下一步精确匹配所用相关窗口的大小,从而实现相关系数的准确

计算。一般来说,评估办法是计算当前目标块和原始块的综合误差和平均绝对误差(见公式(4-5)、(4-6))。其中,统计阈值是经过对大量视频数据的统计分析求得,具有普适性意义。通常,两个目标块的差异越小,则相关窗口大小应该减小;反之,相关窗口应该相应增大。

$$e = \left| \sum_{i=0}^m \sum_{j=0}^n P_{ij} - \sum_{i=0}^m \sum_{j=0}^n Q_{ij} \right| \quad (4-5)$$

$$SAD = \sum_{i=0}^m \sum_{j=0}^n |P_{ij} - Q_{ij}| \quad (4-6)$$

(其中, P 代表目标原始块, Q 代表目标匹配块,  $P_{ij}$  和  $Q_{ij}$  分别代表  $(i,j)$  位置的像素值)

此外,在取得相关窗口的大小之前,还需要做一个精确调整的过程。因为第一次的预处理仅对两幅图像之间作了一个粗略的计算和估计,没有考虑其他影响因素,诸如图片本身的自我相关性等。所以,预处理反馈环还需要对目标块本身的相关性作一个评价。这里采用均方差(MSE)作为评价目标块本身相关性的主要指标(见公式(4-7))。通常, MSE 的大小反应了目标块本身的像素差异程度, MSE 越大,说明目标块的差异程度越明显,那么相关窗口的大小应该相应减小。

$$MSE = \sum_{i=0}^m \sum_{j=0}^n (P_{ij} - \bar{P})^2 \quad (4-7)$$

(其中  $P_{ij}$  代表  $(i,j)$  位置的像素,  $\bar{P}$  代表目标块 P 的像素值均值)。

#### 4.2.2 镂空块的匹配计算

影响块匹配搜索效果的主要因素是相关窗口的大小。相关窗口如果选择较大,则匹配的准确率提高,但相应的计算复杂度也随之升高;反之,如果窗口小,则匹配效果差,计算复杂度较低。

从图 4-3 为相同搜索视频图像下,不同大小的搜索窗口的匹配结果,其中(a)为  $17 \times 17$  的相关窗口搜索结果, (b)为  $5 \times 5$  的相关窗口搜索结果。可以看出,尽管(a)的搜索结果比(b)的搜索结果更加准确,但是(a)的计算复杂度远远高于(b)的计算复杂度。因此,应该尽量选择大的搜索窗口做匹配计算才可以得到最优的匹配结果,但是由此带来的计算复杂度高昂的代价需要做进一步的处理。





(a)  $17 \times 17$  相关窗口匹配结果 (b)  $5 \times 5$  相关窗口匹配结果

图 4-3 不同大小的相关窗口的搜索结果

Figure4-3 The search results compare of different correlation windows

针对这一情况, 本文提出了镂空相关窗口的匹配方案<sup>[30]</sup>。如图 4-4(a) 所示, 对相关窗口进行隔点像素采样, 即当前帧每 4 个像素为一组, 选取其中一个像素, 组成一个新的相关窗口进行匹配计算。由于每 4 个相邻像素之间的像素值差距较小, 所以选取的采样像素可以基本代表其他像素; 同时由于隔点采样保证了像素之间的相对位置关系, 没有影响像素之间的相关性, 所以依然可以反映图像之间的相互关系。经过像素采样后的匹配计算, 计算量比原先减小约 75%。

图 4-4(b) 为采用了  $17 \times 17$  镂空相关窗口的搜索结果, 可以看出, 在计算复杂度约下降 75% 的前提下, 准确率没有明显影响。



(a)  $17 \times 17$  镂空窗口 (b)  $17 \times 17$  镂空相关窗口搜索结果

图 4-4 采样镂空相关窗口示意图及搜索结果

Figure4-4 sampled correlation window and search result

### 4.2.3 试验结果

按上述算法思想, 对 256 灰度视频和 RGB 彩色视频分别进行了试验, 结果表明效果良好, 不仅在算法复杂度上有大幅降低, 而且匹配的准确率没有明显降低。具体测试及结果如下所示: 图 4-5 为局部图像 (人物嘴部) 变化较大的情况, 由于不涉及所要跟踪的眼睛和鼻子部分, 所以算法依然建议采用小相关窗口。图

4-6 为整体图像差异较大的情况，算法自动建议采用较大的相关窗口。图 4-7 为彩色图像主体偏移较大的情况，算法自动建议采用较大的相关窗口；图 4-8 为彩色图像差异较小的情况，算法自动建议采用较小的相关窗口。

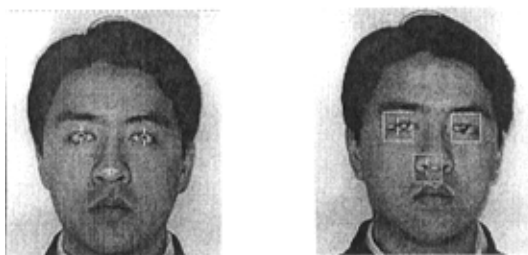


(a) 原始标定图像

(b) 目标匹配图像

图 4-5 灰度视频匹配结果（嘴部差异较大）

Figure4-5 Match results of Gray video (local large difference)



(a) 原始标定图像

(b) 目标匹配图像

图 4-6 灰度视频匹配结果（图像整体差异较大）

Figure4-6 Match results of Gray video (Large difference)



(a) 原始标定图像

(b) 目标匹配图像

图 4-7 彩色视频匹配结果（图像主体偏移较大）

Figure4-7 Match results of Color video (Large difference)

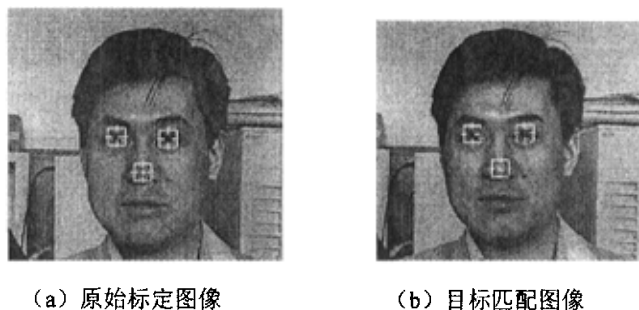


图 4-8 彩色视频匹配结果 (图像差异较小)

Figure4-8 Match results of Color video (Small difference)

从以上试验结果可以发现,在图像差异大的情况下,依然有较好的匹配效果。由于采用了像素采样镂空相关窗口进行匹配,所以计算量有明显下降(约下降75%),但是由于隔点采样的像素之间的相对位置没有改变,所以目标块的相关值不会发生质的变化。使用了 HSV 色彩空间中的亮度分量投影,RGB 彩色视频的匹配结果依然很好。

### 4.3 本章小结

在 MPEG-4 标准视频编码中,人脸特征点(FAP、FDP)的提取与跟踪属于基于内容/基于对象的视频编码应用范畴。

由于其视频内容的特殊性,可以对块匹配算法做很多改进工作,例如添加匹配点的各种限制条件等手段。前期的工作包括对特定检测内容的几何位置约束限制、窗口边缘的梯度值范围限制等,虽然效果比较显著,但是同时也严重限制了视频内容的适用范围。

本章提出了一种基于内容的自适应块匹配改良算法。首先用一个反馈环对视频进行预处理。经过两帧图像的误差计算,和预定统计阈值进行比较,以此来评估图像的相似程度,以此得到初步的建议窗口大小;然后利用 MSE 来评价目标图像的自身相似性,精确调整所建议的窗口大小;最后进行精确的块匹配匹配。所增加的预处理反馈环尽管提高了匹配效果,但同时也增加了计算复杂度。

为了有效降低计算复杂度,对相关窗口进行像素采样,即每 4 个像素采样一个像素作为匹配样本。因为相邻 4 个像素之间的差异很小,因此样本像素可以代替其它三个像素参加匹配计算。与此同时,采样不改变像素与像素之间的相对位

置关系，所以计算得到的匹配结果不会改变。通过采用镂空相关窗口进行计算匹配，使计算量降低了约 75%。

当累积多帧图像连续进行块匹配后，算法的匹配误差将逐渐加大，因此必须对累积误差进行修正消除。一般来说，主要通过对视频内容的分析，提取特征匹配点的几何约束关系和像素值等特征进行匹配结果的二次调整。实践证明，改良算法的实验效果良好，在匹配速度和匹配结果上均有不同程度的提高与改善。

目前该算法存在的主要问题是预处理反馈环提供的阈值参数针对不同的视频内容将有不同的范围，应提高阈值选择的自学习能力，实现阈值参数的自适应调整，才可以进一步提高预处理反馈环的适用范围。由于相关窗口大小对匹配计算的结果十分敏感，因此阈值的精确选择将对匹配结果有明显影响。所以，一旦出现相关窗口大小定义误差，将导致匹配结果错误。

下一步的工作可以集中到相关窗口大小的精确调整，以及整个算法计算复杂度的进一步降低。实现针对不同种类视频内容所对应的阈值参数的自适应调整学习，以及匹配精度的自适应调整等方面。

## 第五章 基于块内相邻像素的帧内预测技术

视频编码压缩技术的本质就是最大限度的去除视频图像内的冗余信息。一般来说图像的冗余数据主要可以分为三大类：时空冗余、统计冗余和心理冗余<sup>[31]</sup>。其中，时空冗余主要指图像帧内和帧间的数据冗余；而统计冗余是指图像数据的编码冗余；至于心理冗余则是指人类视觉系统（HSV）的感知部分。

帧内预测技术的主要原理是消除图像的空间冗余数据，利用当前帧内的已知像素值预测其它像素值。在 H. 264 视频编解码标准中，主要对视频序列中的 I 帧和 P 帧做帧内像素值预测。

目前，H. 264 标准采纳的帧内预测有传统的 9 种预测模式，分别由 Nokia 公司和 Real 公司专利保护。国内外先后就此提出了多种改进的帧内预测方案，包括预测模式的增加，RDO 计算的简略等。

本章将提出一种基于块内相邻像素的帧内预测技术，在传统的 9 种预测模式基础上，面向块内像素，增加 4 种帧内预测模式。试验表明效果显著。

### 5.1 帧内预测技术概述

帧内预测技术作为视频编解码中的重要一环，属于视频编码流程中的第一个处理阶段，主要目的是去除图像内部的空间冗余信息。由于视频图像的空间冗余数据占据了视频内容当前帧主要数据的绝大部分，所以能否有效去除该冗余数据将直接关系到帧内预测技术的编码效率。

在 MPEG-4 标准和 H. 264 标准中，对帧内预测方案做了详尽的技术规定；此外，基于静态图像的编码标准 JPEG 系列也广泛使用了帧内预测技术<sup>[32][33]</sup>。

帧内预测技术已经成为视频编解码技术研究的热点关键问题。目前，国内外针对这一技术开展的研究创新工作主要可以分为计算复杂度和空间复杂度的降低以及预测模式的创新等两个方面，具体有下面两类。

由于传统的 9 种帧内预测模式需要计算每个预测模式的 RD cost，所以计算复杂度较高。为了降低计算复杂度，可以通过对当前需要计算 RD cost 的模式做预测效果分级，按照预测效果来决定是否进一步计算 RD cost<sup>[34][35]</sup>。

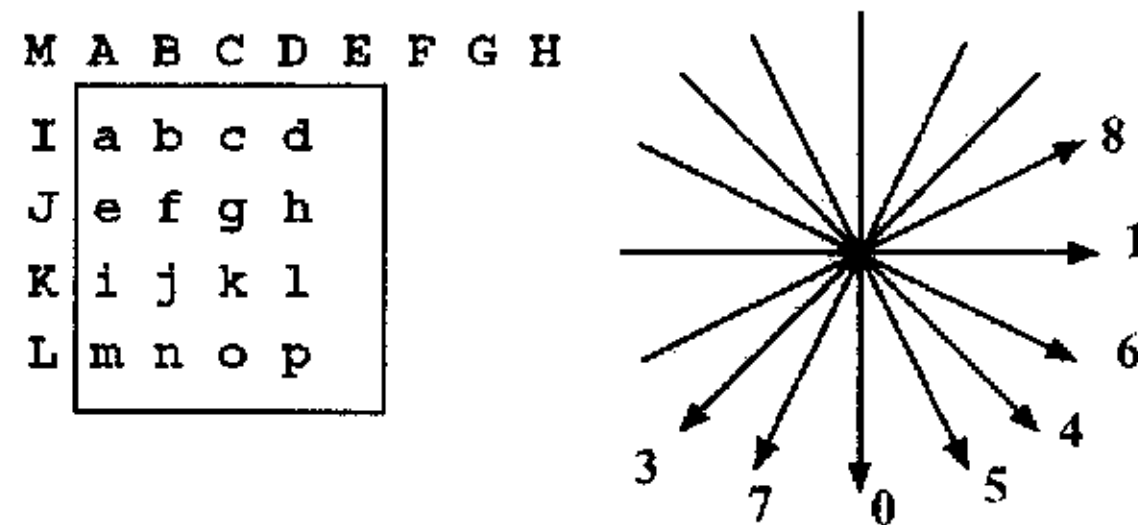
至于预测模式的创新，则主要是对传统的9种预测模式进行改进，具体包括预测方向和预测模式的改进<sup>[36]</sup>、预测模式的删减<sup>[37]</sup>、预测顺序的改变以及预测参考值的选择<sup>[38][39]</sup>等方面。

## 5.2 帧内预测技术原理

本节将以 H.264 技术标准为基础，介绍帧内预测的技术细节与编码原理。

帧内预测位于整个视频编码框架的第一部分，因为第一帧视频图像做编码时，没有任何参考帧可用，只能做帧内预测编码。

图 5-1 是目前 H.264 标准采用的帧内预测模式示意图，其中，(a)为当前像素块的预测位置排列表，(b)为帧内预测方向模式。其中，像素 a~p 的像素值通过相邻块的像素 A~M 来预测<sup>[40]</sup>。



(a) 当前像素块的位置结构 (b) 帧内预测模式方向

图 5-1 用于帧内预测的像素排列表及其预测方向

Figure 5-1 Identification of samples used for Intra prediction with its directions

预测的模式一共为 9 种，图中没有标出的模式 2 (DC 模式) 是一种无方向的均值预测模式。9 种预测模式具体参见表 5-1。

表 5-1 4×4 亮度像素块帧内预测模式名称列表

Table 5-1 List of 4×4 Luma Block Intra Prediction Name

4×4 亮度块帧内预测模式编号	4×4 亮度块帧内预测模式名称
0	Intra_4x4_Vertical
1	Intra_4x4_Horizontal
2	Intra_4x4_DC
3	Intra_4x4_Diagonal_Down_Left
4	Intra_4x4_Diagonal_Down_Right
5	Intra_4x4_Vertical_Right
6	Intra_4x4_Horizontal_Down
7	Intra_4x4_Vertical_Left
8	Intra_4x4_Horizontal_Up

至于  $16 \times 16$  块的帧内预测, 则与  $4 \times 4$  块的 9 种模式类似, 但预测模式种类为 4 种, 具体为: 垂直预测、水平预测、DC 预测以及平面预测等模式。由于采用 4:2:0 的视频序列格式, 色度块是亮度块的一半, 所以色度块的预测办法类似于亮度  $16 \times 16$  块的预测办法, 也是 4 种预测模式。

帧内预测的具体办法是依靠已经得知的像素值通过既定的插值计算公式得到需要预测的像素值。

### 5.3 基于块内相邻像素的帧内预测技术

目前传统的 9 种帧内预测模式, 依据的是像素块之间的相关性。因为相邻块之间有较强的数据相关性, 所以可以利用像素 A~M 来预测像素 a~p 的值。但实际上, 同一个块内各像素之间的相关性明显大于相邻块之间的相关性, 所以如果可以利用这一更为明显的相关数据的话, 必然可以得到更加准确的预测值。

为此, 本文提出了利用块内像素的帧内预测技术, 在这一方案中使用了块内的相关数据来做插值预测<sup>[41]</sup>, 下面将讨论其具体技术细节。

图 5-2 是  $4 \times 4$  像素块的像素标号排列表, 像素  $P_{00} \sim P_{33}$  为当前需要预测的块内像素值。由于首次预测像素值的时候, 块内的各个像素值尚属未知, 所以无法直接使用块内相关性进行帧内预测。为了利用块内像素的相关性, 把图 5-2 中的 16 个像素按照物理位置分布进行划分, 分为第一批预测像素与后期预测像素。当预测到当前像素块的时候, 首先利用传统的 9 种预测模式计算当前块的第一批像素, 之后对于其他像素则利用块内已经预测得到的像素值进行累积插值预测。这样就可以有效利用块内像素的相关性。

$P_{00}$	$P_{01}$	$P_{02}$	$P_{03}$
$P_{10}$	$P_{11}$	$P_{12}$	$P_{13}$
$P_{20}$	$P_{21}$	$P_{22}$	$P_{23}$
$P_{30}$	$P_{31}$	$P_{32}$	$P_{33}$

图 5-2 交互式帧内预测模式像素表

Figure5-2 Interactive Intra prediction mode pixels table

至于块内像素如何判定是否属于第一批像素, 主要通过对视频内容进行分析

统计, 依靠对不同物理位置的定义, 提出了 4 种模式。这 4 种模式按照预测方向定义, 分别划分了 4 类不同的第一批像素与其他像素分类。

### 5.3.1 L 型扩展预测模式

由于当前块和相邻块具有一定的相关性, 所以首先利用传统的 9 种预测模式来预测当前块的边缘 7 个像素 (参见图 5-3 中的第一个块的 7 个灰色像素), 然后利用已经得到像素值的 7 个像素来预测块内剩余的 9 个像素, 具体步骤和预测顺序参见图 5-3 中的第二个块、第三个块以及第四个块的灰色像素。

为了得到更加准确的预测值, 可以不单单利用第一批的 7 个像素预测其余像素, 而重复利用上一步所得到的当前所有可用像素值进行累积预测, 这样可以得到更多的预测候选值参与下一步的 RD Cost 计算, 从而为编码效率的进一步优化奠定基础。

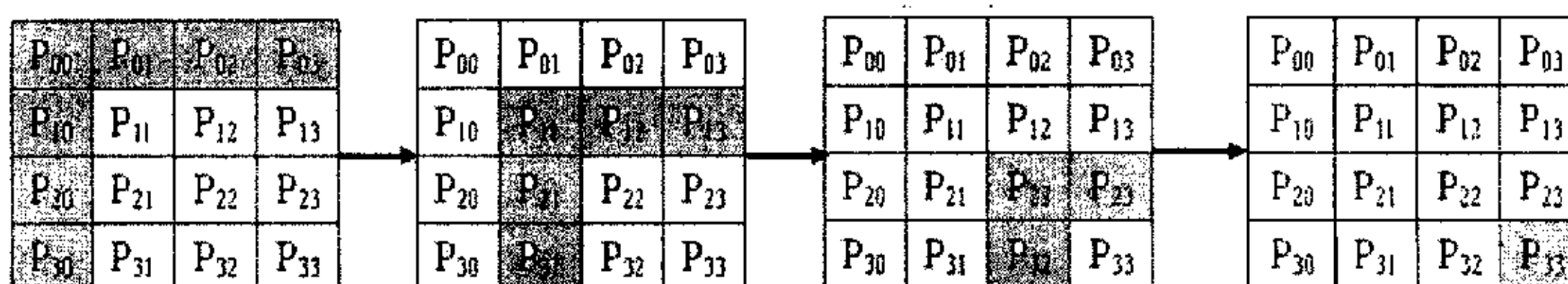


图 5-3 L 型扩展帧内预测模式

Figure5-3 L style Extended predict mode

至于预测所使用的插值计算公式, 可以有多种选择。本文使用了根据像素之间的物理距离和插值权重挂钩的办法来做插值计算, 与被预测像素的物理位置越接近的像素, 其分配的权重就越大; 反之就越小。具体计算见公式 5-1:

$$P_i = A_i \times P_{00} + B_i \times P_{01} + C_i \times P_{02} + D_i \times P_{03} + E_i \times P_{10} + F_i \times P_{20} + G_i \times P_{30} \quad (5-1)$$

其中,  $A_i \sim G_i$  为像素值权重值。对于  $P_{11}$  来说, 由于  $P_{00}$ 、 $P_{01}$ 、 $P_{10}$  的物理位置与  $P_{11}$  最为接近, 所以权值  $A_{11}$ 、 $B_{11}$ 、 $E_{11}$  相对较大, 而其余权值参数相对较小。对于  $P_{13}$  来说, 由于  $P_{02}$ 、 $P_{03}$  的物理位置与  $P_{13}$  最为接近, 所以权值  $C_{13}$ 、 $D_{13}$  相对较大, 而其余权值参数相对较小。

预测到达第二阶段后, 由于原先的第一批 7 个像素可用, 同时像素  $P_{11}$ 、 $P_{12}$ 、 $P_{13}$ 、 $P_{21}$ 、 $P_{31}$  可用, 所以预测剩余的 4 个像素时, 可以同时使用像素  $P_{11}$ 、 $P_{12}$ 、 $P_{13}$ 、 $P_{21}$ 、 $P_{31}$  进行累积插值计算, 具体参见公式 5-2 与公式 5-3。



$$P_i = H_i \times P_{11} + I_i \times P_{12} + J_i \times P_{13} + K_i \times P_{21} + L_i \times P_{31} \quad (5-2)$$

$$P_i = M_i \times P_{22} + N_i \times P_{23} + O_i \times P_{32} \quad (5-3)$$

### 5.3.2 右下斜线预测模式

右下斜线预测模式主要针对视频图像内容的斜纹理做预测。与 L 型扩展预测模式类似，首先根据方向划分出首批预测的 4 个像素  $P_{00}$ 、 $P_{11}$ 、 $P_{22}$ 、 $P_{33}$ ，这 4 个像素按照传统的 9 种预测模式进行预测（参见图 5-4）；其余像素按照右下斜线方向，分别划分为三批像素： $(P_{01}$ 、 $P_{10}$ 、 $P_{12}$ 、 $P_{21}$ 、 $P_{23}$ 、 $P_{32})$ 、 $(P_{02}$ 、 $P_{13}$ 、 $P_{20}$ 、 $P_{31})$ 、 $(P_{03}$ 、 $P_{30})$ 。累积插值的规则同 5.3.1 所述，也是按照物理位置关系确定插值权重系数大小。具体计算方法见公式 5-4、5-5、5-6。

$$P_i = A_i \times P_{00} + B_i \times P_{11} + C_i \times P_{22} + D_i \times P_{33} \quad (5-4)$$

$$P_i = E_i \times P_{01} + F_i \times P_{12} + G_i \times P_{23} + H_i \times P_{10} + I_i \times P_{21} + J_i \times P_{32} \quad (5-5)$$

$$P_i = K_i \times P_{02} + L_i \times P_{13} + M_i \times P_{20} + N_i \times P_{31} \quad (5-6)$$

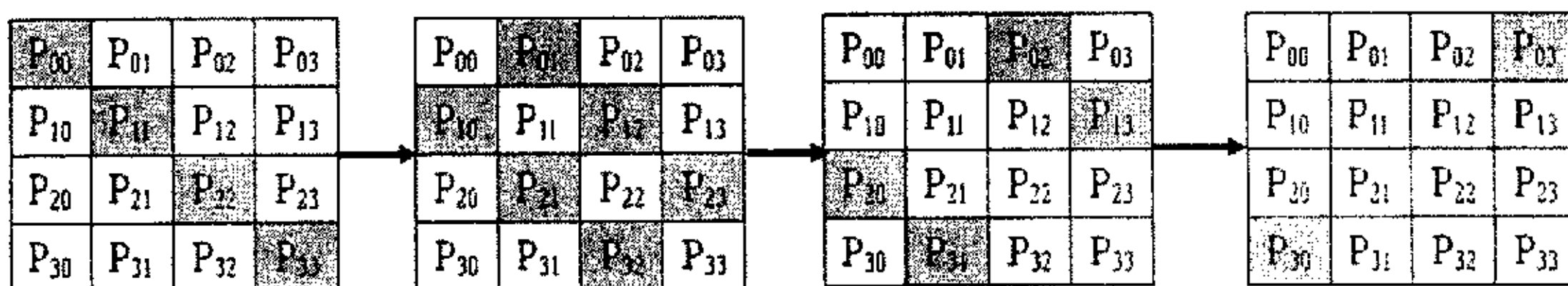


图 5-4 右下斜线帧内预测模式

Figure 5-4 Right-Down Diagonal predict mode

### 5.3.3 左下斜线预测模式

左下斜线预测模式也是针对视频图像内容的斜纹理做预测。与右下斜线预测模式类似，首先根据方向划分出首批预测的 4 个像素  $P_{03}$ 、 $P_{12}$ 、 $P_{21}$ 、 $P_{30}$ ，这 4 个像素按照传统的 9 种预测模式进行预测（参见图 5-5）；其余像素按照左下斜线方向，分别划分为三批像素： $(P_{02}$ 、 $P_{11}$ 、 $P_{13}$ 、 $P_{20}$ 、 $P_{22}$ 、 $P_{31})$ 、 $(P_{01}$ 、 $P_{10}$ 、 $P_{23}$ 、 $P_{32})$ 、 $(P_{00}$ 、 $P_{33})$ 。累积插值的规则以及计算公式同 5.3.2 所述。插值计算公式参见公式 5-7、5-8、5-9。

$$P_i = A_i \times P_{03} + B_i \times P_{12} + C_i \times P_{21} + D_i \times P_{30} \quad (5-7)$$

$$P_i = E_i \times P_{02} + F_i \times P_{11} + G_i \times P_{20} + H_i \times P_{13} + I_i \times P_{22} + J_i \times P_{31} \quad (5-8)$$

$$P_i = K_i \times P_{01} + L_i \times P_{10} + M_i \times P_{23} + N_i \times P_{32} \quad (5-9)$$

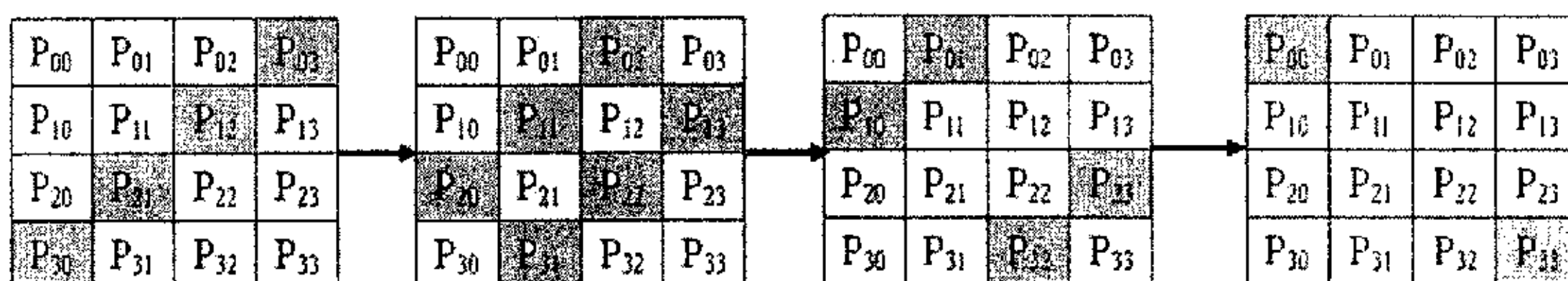


图 5-5 左下斜线帧内预测模式

Figure5-5 Left-Down Diagonal predict mode

### 5.3.4 中心扩散预测模式

由于有的视频图像具有中心扩散放射图案, 所以考虑添加了中心扩散预测模式。在该种预测模式下, 第一批预测的像素为  $P_{11}$ 、 $P_{12}$ 、 $P_{21}$ 、 $P_{22}$ ; 之后按照边缘扩展模式依次进行预测 (参见图 5-6)。具体计算方案见公式 5-10、5-11。

$$P_i = A_i \times P_{11} + B_i \times P_{12} + C_i \times P_{21} + D_i \times P_{22} \quad (5-10)$$

$$P_i = E_i \times P_{01} + F_i \times P_{02} + G_i \times P_{31} + H_i \times P_{32} + I_i \times P_{10} + J_i \times P_{20} + K_i \times P_{13} + L_i \times P_{23} \quad (5-11)$$

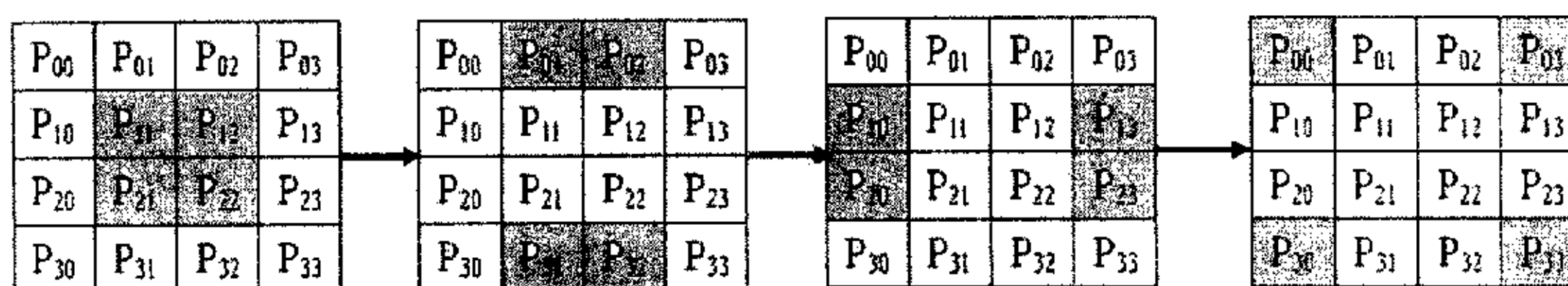


图 5-6 中心扩散帧内预测模式

Figure5-6 Center Originated Extended predict mode

### 5.3.5 反馈评价环

4 种帧内预测模式利用了块内像素之间的相关性, 从而获得了更加准确的像素预测值, 但同时也不可避免的带来了高昂的计算复杂度, 主要耗费在 4 种模式 RD Cost 的计算上。为了降低计算复杂度, 需要避免计算大量的 RD Cost。

为了实现更加高效的帧内预测，添加了一个反馈评价环。通过计算块内相关性预先决定所采用的预测模式。反馈评价环早在码率控制研究课题<sup>[42]</sup>中就有广泛应用，利用反馈环可以评估当前所采取的方案预期效果，同时可以依据评估结果选择下一步的选择模式。具体步骤如图 5-7 所示。

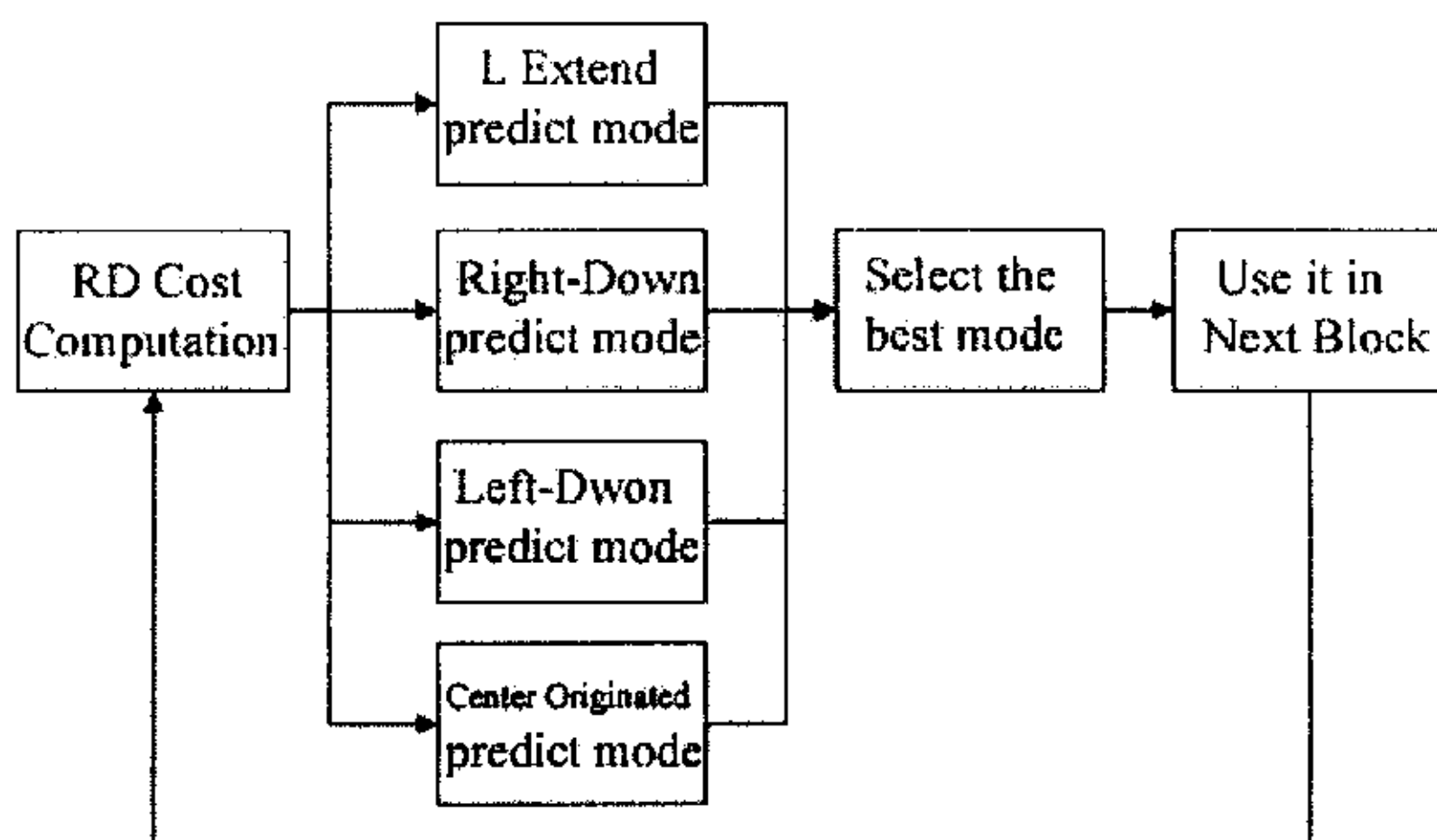


图 5-7 交互式帧内预测反馈环

Figure5-7 Feedback circle for interactive intra prediction

因为当前  $4 \times 4$  的像素块根据 RD Cost 计算确定了所选择的预测模式，根据图像视频的连贯性，相邻像素块选择当前预测模式的概率应该相对较大，所以该反馈环直接将当前块所选择的预测模式传给下一个块，在下一个块做 RD Cost 的时候，直接计算上一块所用的预测模式的 RD Cost，如果低于预定阈值，则直接采用该模式，而无需计算其他三种预测模式，从而可以大幅降低计算复杂度。这里的既定阈值需要对不同视频内容做统计分析后得到。

## 5.4 试验结果

在 5.3 节中提出了基于块内像素相关性的 4 种帧内预测模式。由于利用块内更为紧密的相关数据，所以帧内预测效果增益比较明显。与此同时，采用了反馈评价环，利用前一像素块所采用的预测模式来预测当前块，避免计算全部 4 类预测模式的 RD Cost，从而计算复杂度有较大降低。

针对上述方案，在 JVT 5.0C 软件平台上做了实验。测试条件与结果如下：

- Frame Rate: 30Hz
- Frame number for video coding: 20

- Sequence mode of Frame: I I I I I I...
- The scope of search:  $\pm 16$
- Value of QP: 18, 20, 22, 24

表 5-2 高清晰视频序列 Night.yuv 测试结果 (1280×720)

Table 5-2 Test result of Night.yuv (1280×720)

		Original Scheme	Improved Scheme
QP=18	PSNR(Y)(db)	45.53	45.51
	Bit-Rate(kbps)	48278.98	43508.08
QP=20	PSNR(Y)(db)	43.92	43.91
	Bit-Rate(kbps)	40652.83	37063.67
QP=22	PSNR(Y)(db)	42.51	42.35
	Bit-Rate(kbps)	34122.40	31322.16
QP=24	PSNR(Y)(db)	41.01	40.74
	Bit-Rate(kbps)	28047.79	25852.06

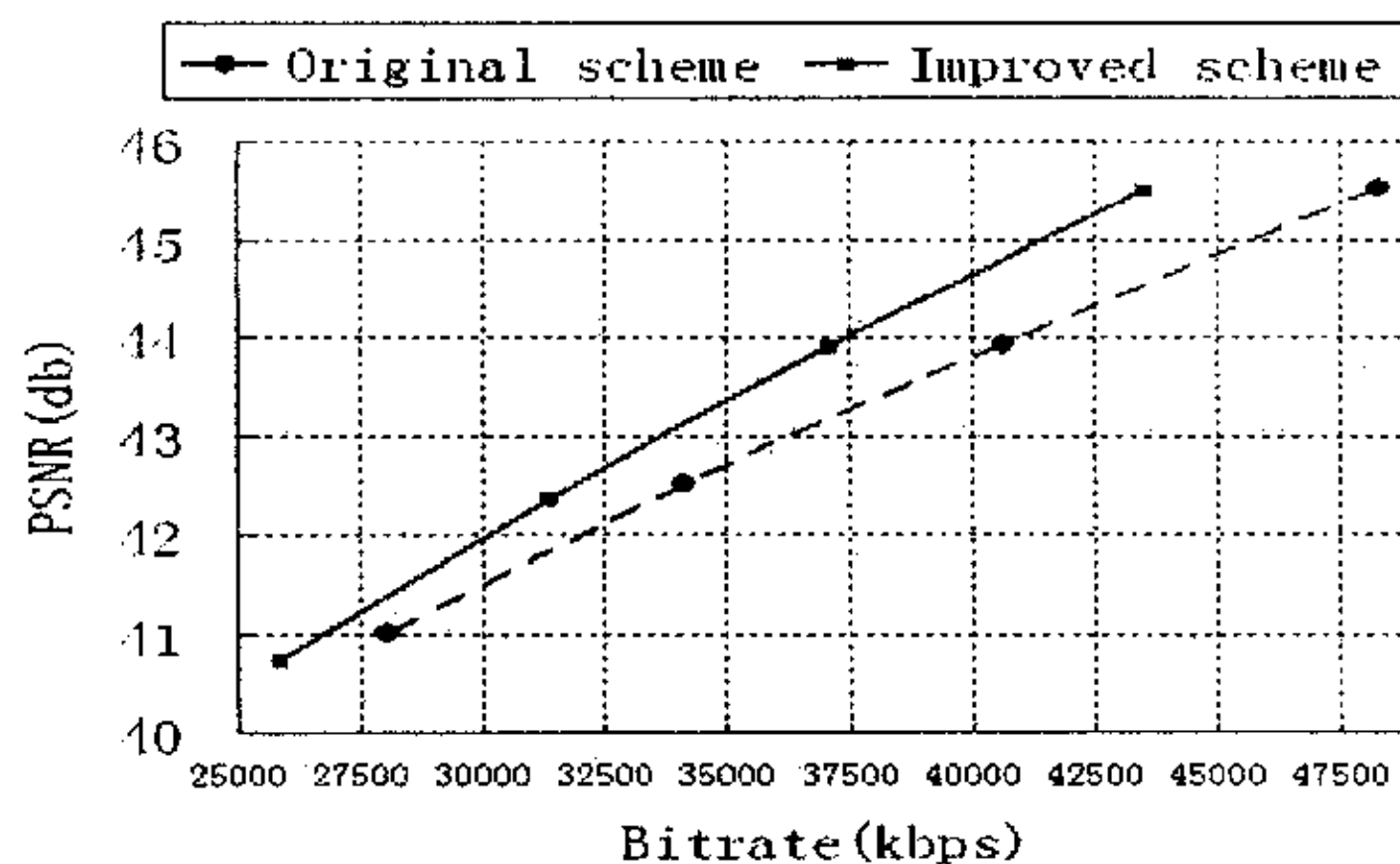


图 5-8 高清晰视频序列 Night.yuv 测试 RD 曲线图

Figure 5-8 The RD curve of Night.yuv

表 5-3 高清晰视频序列 Jets.yuv 测试结果 (1280×720)

Table 5-3 Test result of Jets.yuv (1280×720)

		Original Scheme	Improved Scheme
QP=18	PSNR(Y)(db)	44.73	45.06.
	Bit-Rate(kbps)	23260.16	22198.66
QP=20	PSNR(Y)(db)	43.53	43.83
	Bit-Rate(kbps)	16657.66	16106.23
QP=22	PSNR(Y)(db)	42.68	42.79
	Bit-Rate(kbps)	11975.93	11631.82
QP=24	PSNR(Y)(db)	41.86	41.85
	Bit-Rate(kbps)	8683.19	8441.14

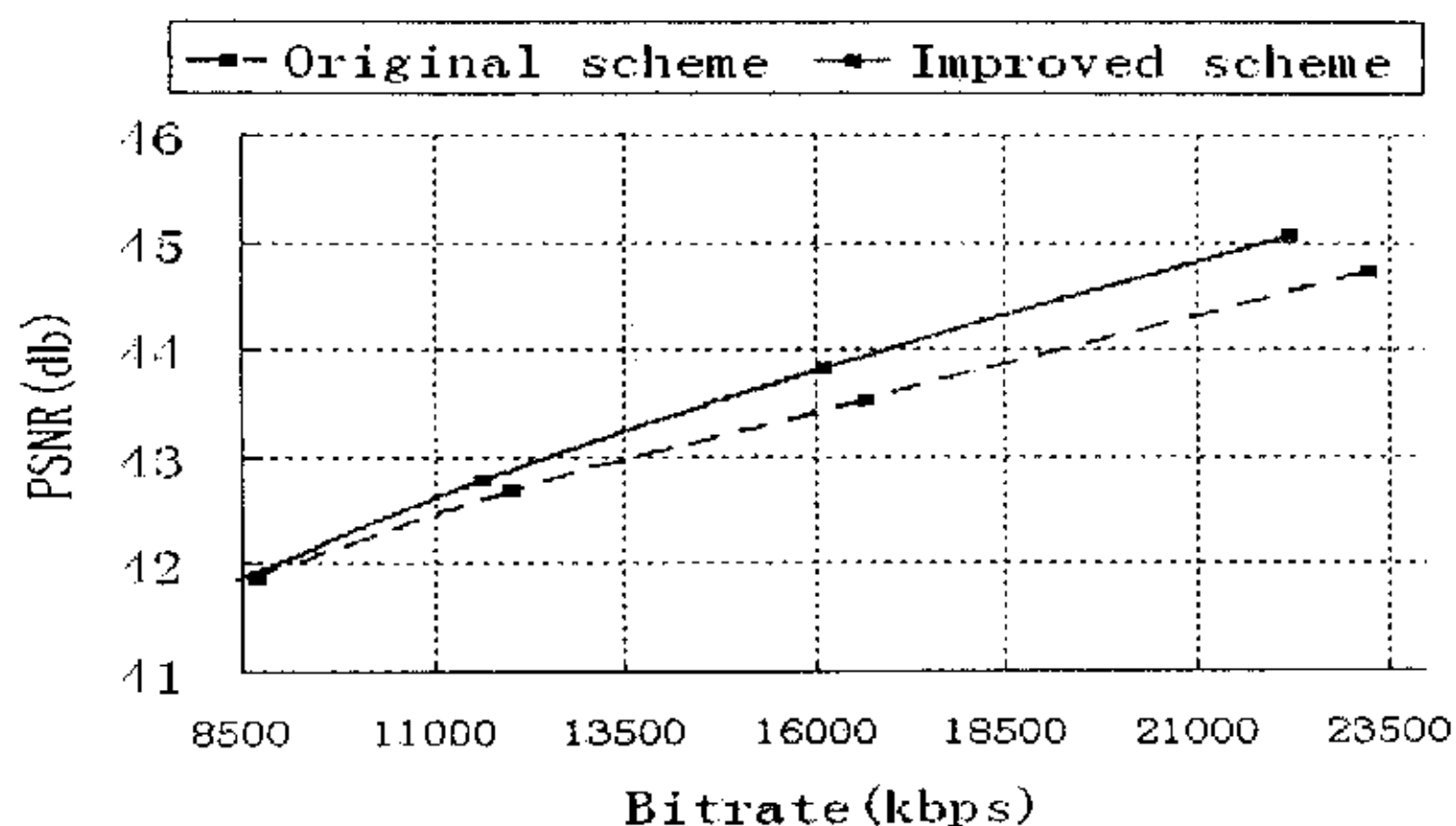


图 5-9 高清晰视频序列 Jets.yuv 测试 RD 曲线图

Figure 5-9 The RD curve of Jets.yuv

表 5-4 高分辨率静态图片 7150.Raw 测试结果 (1024×1024)

Table 5-4 Test result of picture 7150.raw (1024×1024)

		Original Scheme	Improved Scheme
QP=18	PSNR(Y)(db)	45.09	45.37
	Bit-Rate(kbps)	38734.32	38513.04
QP=20	PSNR(Y)(db)	43.67	43.93
	Bit-Rate(kbps)	31799.04	31592.40
QP=22	PSNR(Y)(db)	42.45	42.55
	Bit-Rate(kbps)	26120.40	25913.52
QP=24	PSNR(Y)(db)	41.16	41.17
	Bit-Rate(kbps)	21074.16	20880.72

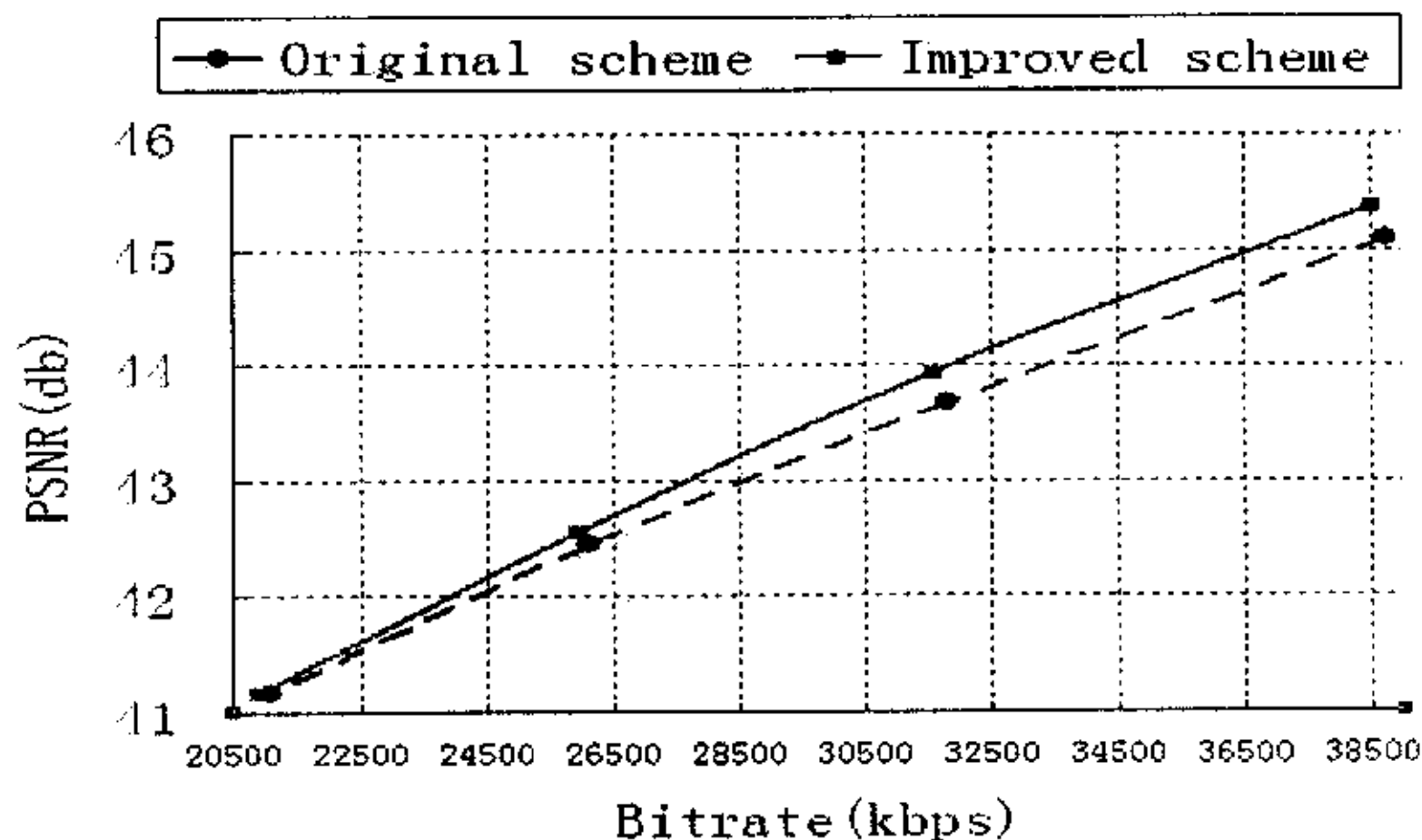


图 5-10 高分辨率静态图片 7150.Raw 测试 RD 曲线图

Figure 5-10 The RD curve of Picture 7150.raw

针对上述帧内预测算法, 主要对两个高清晰视频序列 night.yuv 和 jets.yuv

做了实验，格式均为  $1280 \times 720$ ；此外还对一个高分辨率的静态航拍图片做了实验，格式为  $1024 \times 1024$ 。在 RD 曲线图中，实线为本文提出的预测算法编码曲线，虚线为 H.264 技术 JVT 5.0C 版本的编码曲线。从图中可以看到，QP 值较小的时候，本文提出的预测算法增益比较明显。

## 5.5 本章小结

利用帧内相关数据预测像素值是帧内预测技术的原理所在。目前传统的 9 种帧内预测模式主要利用了相邻像素块之间的相关性进行预测。但实际上，块内像素之间的相关性远远大于相邻块之间的相关性，所以利用块内像素做帧内预测将可以得到更加准确的预测像素值。

本文提出了 4 种基于块内相关性的帧内预测模式（L 型扩展、右下斜线、左下斜线、中心扩展），分别针对视频图像内容的不同纹理结构划分了预测像素的计算顺序。由于新添加了 4 种预测模式，必然带来额外的计算复杂度。为了减小计算复杂度，添加了反馈评价环。由于相邻块之间的纹理结构不会有很大差异，而且具有一定的相关性，所以可以依据上一个像素块采用的预测模式，决定当前块采用的预测模式，这样可以避免计算额外模式的 RD Cost。如果当前选择模式的 RD Cost 超出既定阈值，则重新计算全部 4 种模式的 RD Cost。

最后，对高清晰视频序列和高分辨率静态图像做了实验测试。结果表明，QP 值较小时，编码效率有明显增益。

下一步的工作可以集中在针对不同内容的视频图像，制定更加细致的预测模式，从而可以得到更高的编码效率。此外，由于预测模式的 RD Cost 计算需要耗费大量的计算时间，所以如何大幅减小计算复杂度仍然是本课题研究的主要技术难点之一。

## 第六章 基于内容的自适应环滤波技术

环路滤波技术在视频编码中起到了十分关键的作用。由于视频图像需要以 $4\times 4$ 、 $16\times 16$ 的像素块为基本单位进行编码,这样势必造成由于条块分割带来的块效应。就信号处理的范畴来说,块效应属于信号噪声。为了消除噪声信号、提高视频质量,视频编码使用了滤波技术。

本章将提出一种基于内容的自适应环路滤波技术,针对不同视频内容,分析相邻块之间的相关性,从而实施不同类型、不同强度的滤波方案。

### 6.1 滤波技术概述

早在20世纪40年代,就对平稳随机信号建立了维纳滤波理论。根据有用信号和干扰噪声的统计特性(自相关函数或功率谱),以线性最小均方误差估计准则所设计的最佳滤波器,称为维纳滤波器。该滤波器可以最大限度的滤除干扰噪声,提取有用信号。60年代初,由于空间技术的发展,出现了卡尔曼滤波理论。它利用状态变量模型对非平稳、多输入多输出随机序列做最优估计<sup>[43]</sup>。

Widrow B.在1967年提出的自适应滤波理论,可使自适应滤波系数的参数自动调整而达到最佳状态,而且在设计时,只需要很少的或根本不需要任何关于信号与噪声的先验统计知识<sup>[44]</sup>。由于该滤波器的优良性能,近十年来,自适应滤波理论和方法得到了迅速发展。

在视频图像编码领域,滤波技术主要用来消除以块为单位的编码所带来的块效应。由于这一类噪音的大小和视频图像信号内容本身及其编码方式有很大关系,所以H.264标准提出了根据视频帧的编码类型与编码方式来决定相应的滤波强度和滤波范围。

目前,针对应用于视频图像编码中的滤波技术主要分为两类:后滤波和环路滤波,其中环路滤波是视频编码的主要应用技术。由于滤波本身属于信号处理学科范畴,因此国内外学者研究的有关课题内容主要集中在滤波器的改进和滤波器应用范围与应用种类的探索方面。

在历次的JVT技术会议中,对H.264视频编码标准中的环路滤波部分都有技



术改进。具体有滤波器tap的变化和滤波位移的变化<sup>[45]</sup>、滤波强度的适用范围的改变、滤波器的适用范围和适用条件的变换<sup>[46]</sup>、滤波器计算复杂度的减小<sup>[47]</sup>等。

由于自适应信号处理已经成为了国际前沿技术，因此在视频编码中也逐渐用到了自适应的滤波技术<sup>[48]</sup>。

## 6.2 环路滤波原理

在视频图像编码中所应用的滤波技术主要用来去除图像的块效应。图像的块效应是因为视频编码以宏块、块为处理单元，这样势必将本来有机的图像整体分割为一个一个的像素块，必然造成块边界的明显存在。图6-1为滤波效果图，左边(a)为未进行滤波的编码重构图，右边(b)为滤波后的编码重构图。从图中可以明显看到，人物脸部光照部分的块效应及其处理效果。

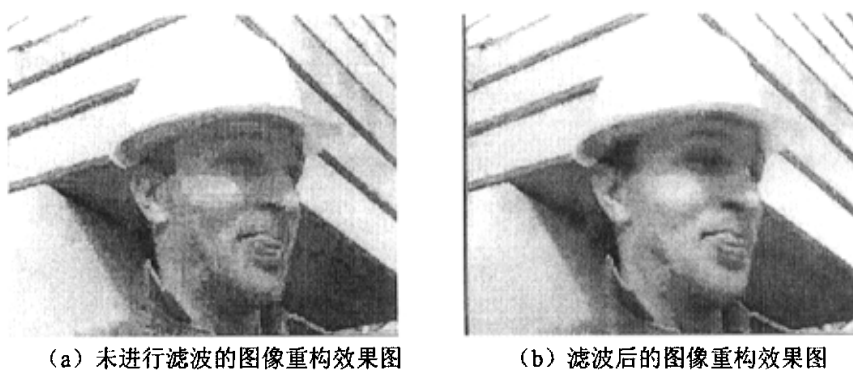


图 6-1 图像滤波效果图

Figure 6-1 Filter effect demonstration of image

### 6.2.1 滤波技术分类

为了消除块效应，可以有两种滤波方式：后滤波与环滤波。图6-2为后滤波的编码流程，图6-3为环滤波的编码流程。灰色框为滤波功能块，由两图可以看到两类滤波的区别就在于滤波器的位置及其实施时间。

后滤波仅仅在解码端的视频序列输出之前进行滤波，具有计算简单、不改变传输码流以及不参与编解码过程等优势，但其缺点在于不能提高编码效率；而环滤波则参与了编码环节，在重构帧的时候做滤波，参与运动估计/补偿环



节, 能够有效提高编码效率, 但同时也带来了计算复杂度升高的代价并改变了编解码过程。实际上, 由于环滤波不仅改变了视频输出的效果, 而且也改变了编码中间数据, 因此具有更高的编码效率<sup>[49][50]</sup>。

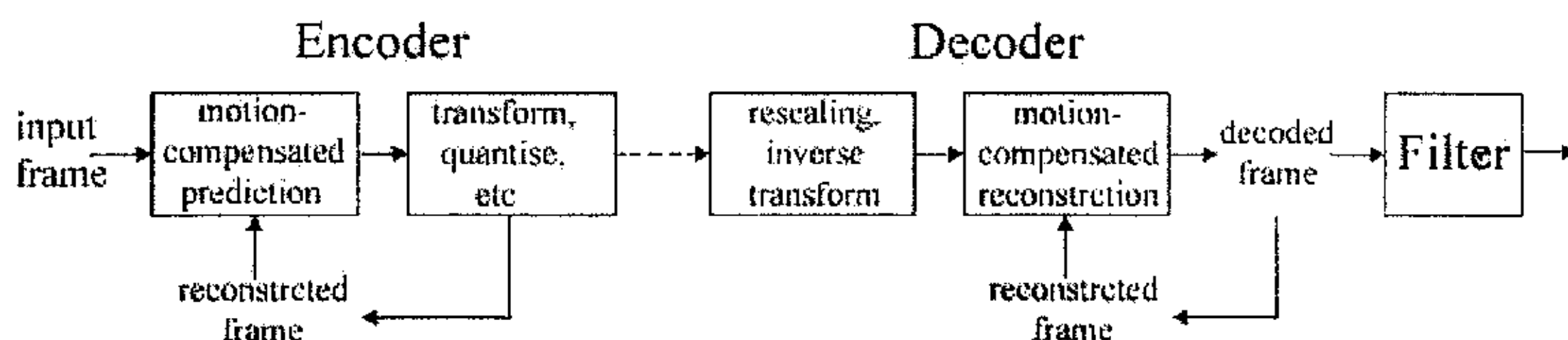


图 6-2 后滤波实现流程图

Figure 6-2 Post-filter implementation

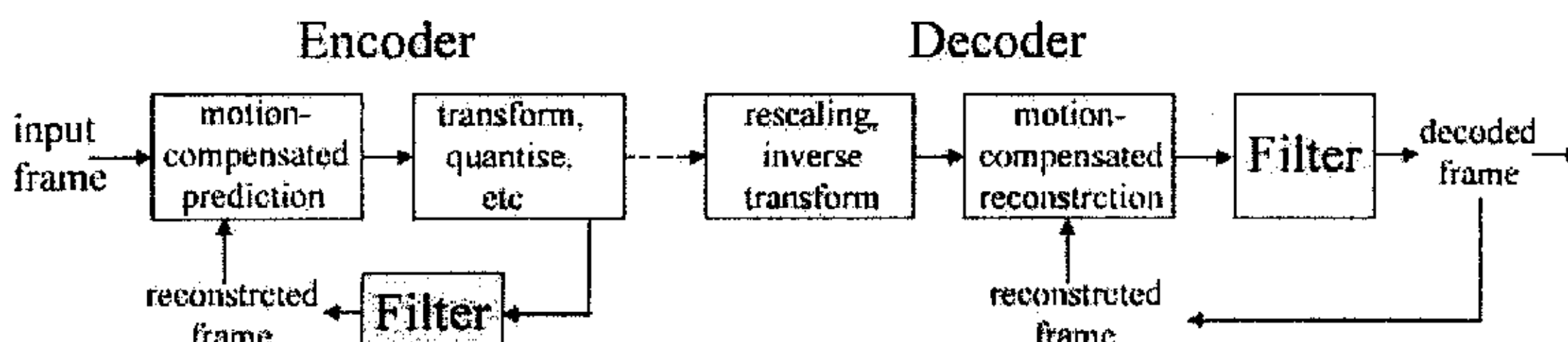


图 6-3 环滤波实现流程图

Figure 6-3 Loop-filter implementation

### 6.2.2 环路滤波条件判定

目前, H. 264标准采用了环路滤波技术。图6-4为 $16 \times 16$ 宏块编码的分割示意图, 需要进行滤波处理的边界有横向和纵向两种。

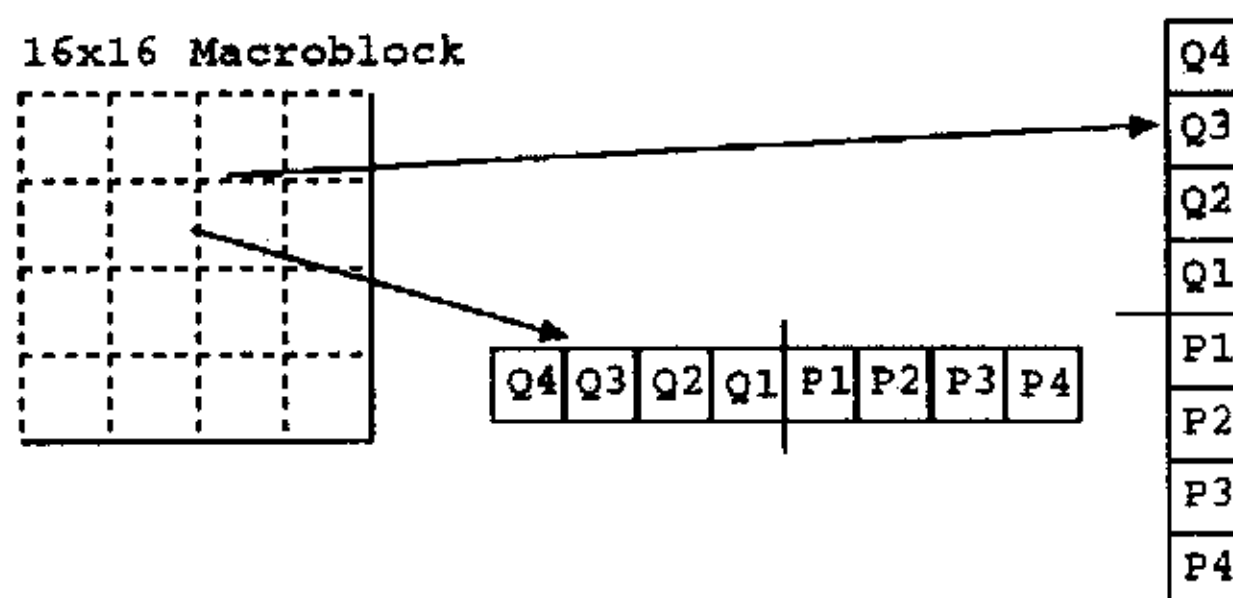


图 6-4 宏块滤波边界示意图

Figure 6-4 Edge of filter in Macroblock

在边界做滤波之前, 需要判定边界的块效应严重程度。一般来说, 当前视频帧如果按照I帧方式进行编码的时候, 块效应是最为严重的; 其次, 判断块

边界点的相邻像素值的差值，如果差值较大，则说明块效应严重。

除图像边界及条带的边界之外，宏块的所有边界都应进行滤波。此处宏块边界定义为宏块内部各个 $8\times 8$ 块以及 $4\times 4$ 块的边界，以及当前宏块与相邻宏块的上边界和左边界。图6-5是视频格式为4:2:0的序列滤波边界示意图，其中黑色边界是需要滤波的边界。

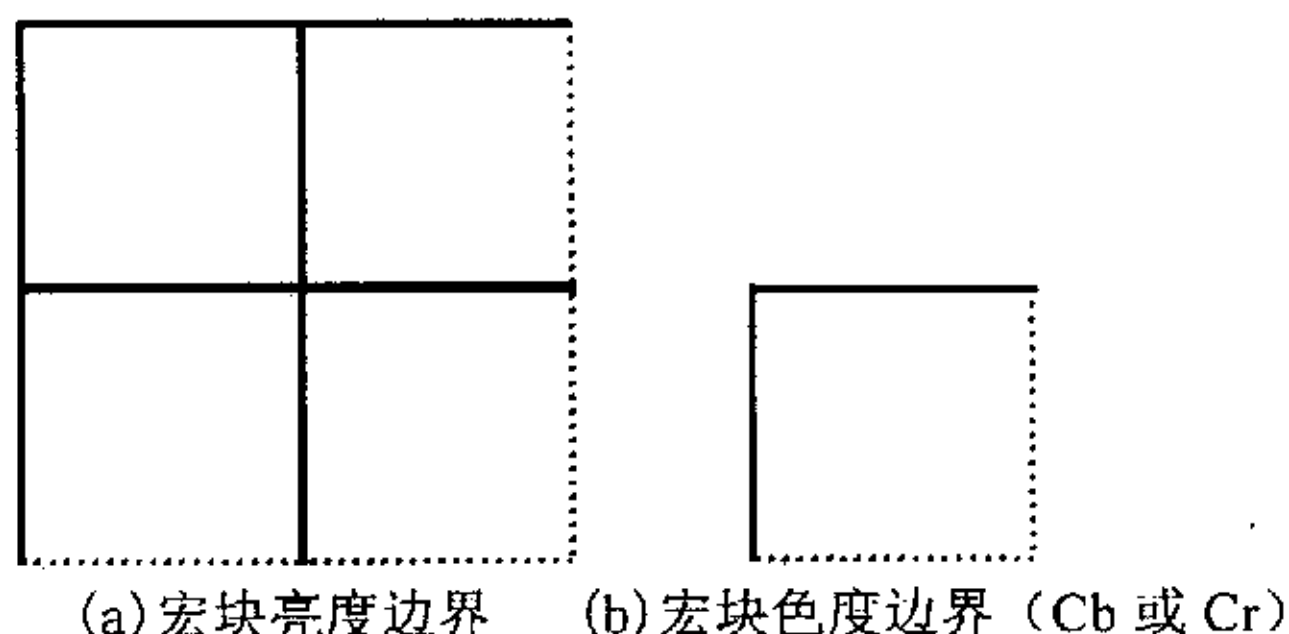


图 6-5 4:2:0 视频格式的滤波边界示意图（黑线为滤波边界）

图 6-5 The edge of filter in 4:2:0 YUV sequence (Marked Black)

为了确定不同的滤波强度，对每种边界定义了一个边界滤波强度  $B_s$ ，色度块的边界滤波强度用对应位置亮度块边界的  $B_s$  代替，如图 6-5 所示（以 4:2:0 格式为例）。如果  $B_s$  等于 0 则不要对边界滤波，否则根据局部样本值的特性和  $B_s$  值对边界滤波。

### 6.2.3 环路滤波处理流程

环路滤波以宏块为单位，按照光栅扫描顺序依次处理。图像中每个宏块的滤波过程如下（以国内视频编解码标准 AVS 为例）：

由于 YUV 视频序列有亮度和色度两类数据分量，因此需要对它们分别做环路滤波（参见图 6-5）。对于每个宏块的 4 个边界，首先从左到右对垂直边界滤波，然后从上到下对水平边界滤波。由于不同类型的边界需要做不同强度的滤波，因此滤波之前需要判定各个边界的滤波强度。图 6-6 为滤波强度的判断流程，滤波强度分为不同等级 0~2：如果当前图像帧为帧内预测模式，则滤波强度为 2；如果是 P 帧或者 B 帧，当相邻块的运动向量差值大于 4 或者参考帧不同，则相应的滤波强度为 1，否则为 0（即不做滤波）。

对块边界划分出滤波强度以后，实施相应的滤波器。一般来说，滤波插值公式是 4-tap 或 5-tap，以当前需要滤波的像素点为中心，向周边像素扩展。

插值公式中各个元素的权重值依靠与中心像素的物理位置关系界定，或使用既定滤波系数。

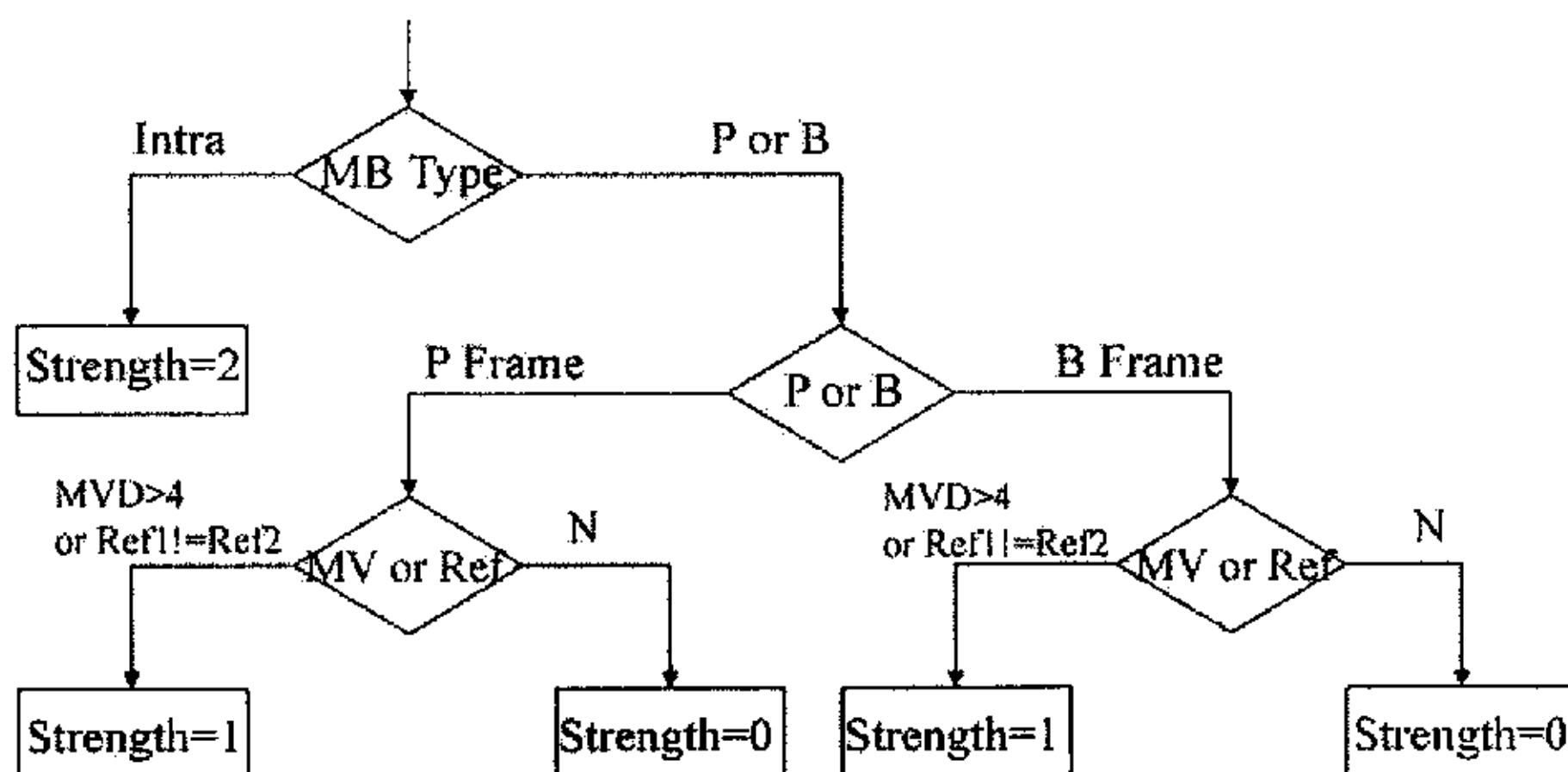


图 6-6 滤波强度判断流程图

Figure 6-6 Filter Strength judgment Flow

### 6.3 基于内容的自适应环滤波

目前，H.264 标准的环路滤波技术使用两类固定系数的滤波器，对不同的边界滤波强度，实施相应的滤波。滤波强度仅仅通过边界像素差值和相邻块的运动向量差值以及参考帧差异等三种因素判定。实际上，影响滤波强度的还有相邻块之间的相关度、像素块的像素排列结构、运动向量以及 QP 值等因素。

不同视频内容需要使用不同的滤波器，例如图像边缘需要实施较轻的滤波效果，而图像的物体中央部分由于没有更多的细节信息，则适合使用较重的滤波效果。如果滤波器使用不当，则会造成图像噪音无法消除，甚至于图像细节信息的丢失。因此，针对不同视频内容实施不同滤波器，才可以有效消除块效应，并保留足够的图像细节信息，从而得到最佳的滤波效果。

此外，由于自适应滤波已经成为当前信号处理的最新技术，而且滤波效果十分显著，因此考虑在视频编码中实现该类型的环路滤波器。

在对视频内容进行分析后，根据视频编码原理，可以发现影响滤波强度和滤波范围的因素主要有以下几种：像素块的量化参数值（QP）、像素块的平均运动向量（MV）、相邻块的相关系数以及像素块自身的均方差（MSE）。因此，为了得到自适应滤波的实施条件，需要在传统的固定滤波器中加入以上 4 种影

响因子，实现自适应的环路滤波器<sup>[51]</sup>（参见图 6-7）。

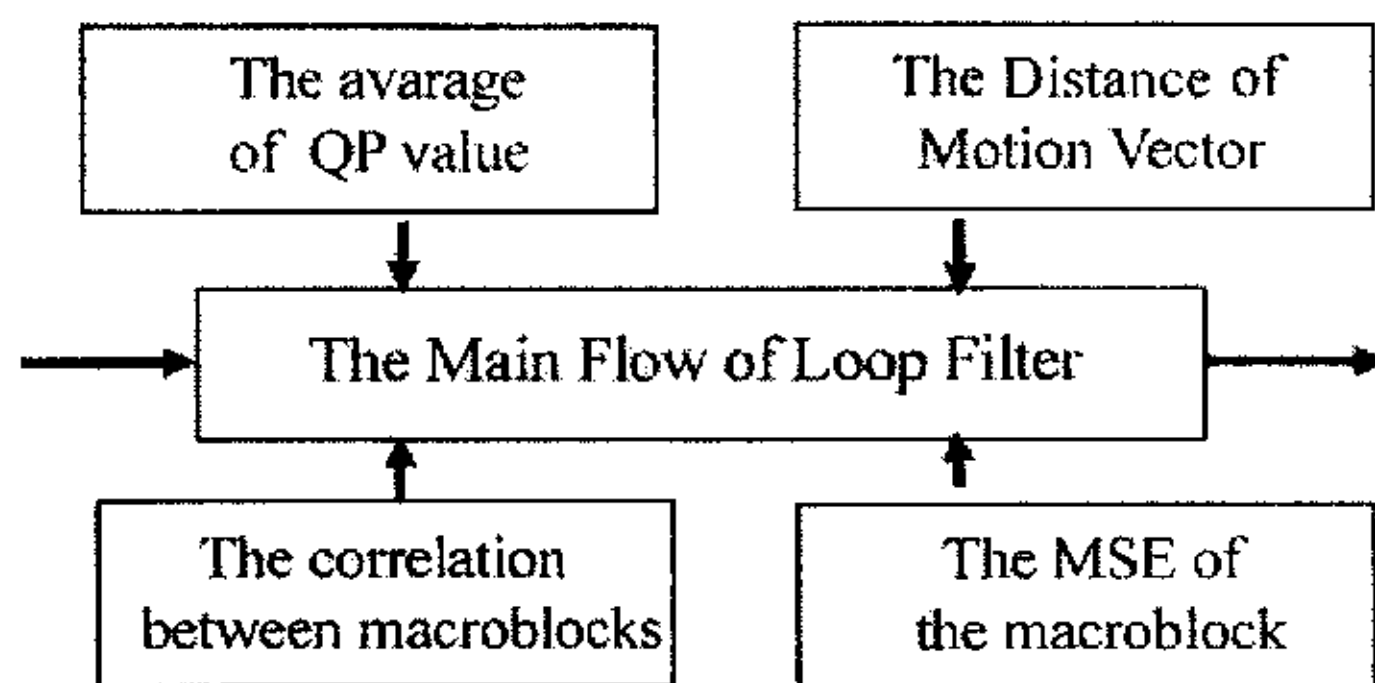


图 6-7 自适应滤波器影响因子结构图

Figure 6-7 Factors structure of Adaptive Filter

### 6.3.1 块内与块间的相关因子

对于  $16 \times 16$  的宏块和  $4 \times 4$  的像素块来说，不仅相邻块之间的关系会影响到滤波强度，而且块内自身的相关性也会影响滤波效果。因此为了实现自适应环滤波，需要添加块内和块间两类相关影响因子。

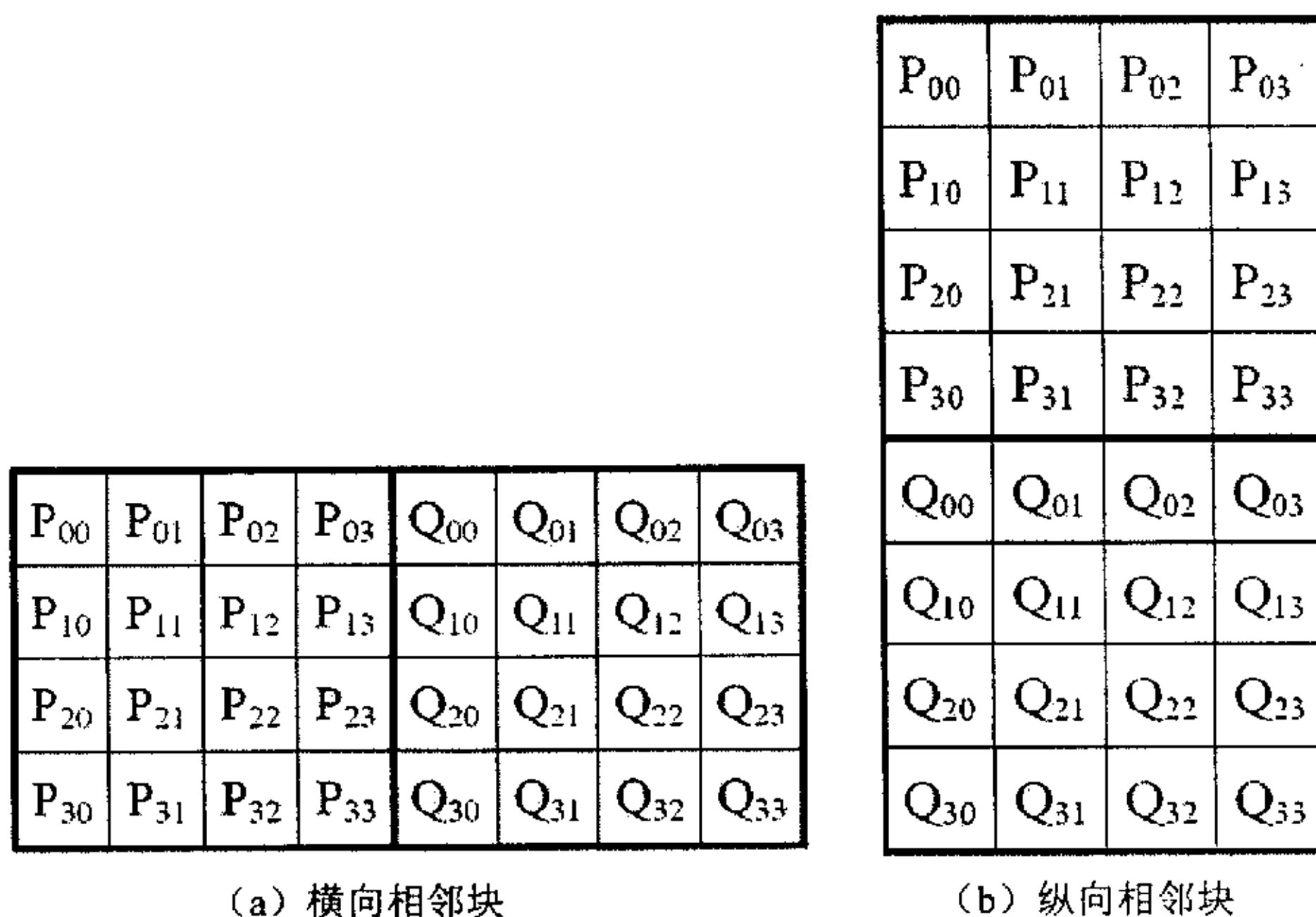


图 6-8 相邻块的像素结构图

Figure 6-8 The sketch map of adjacent blocks (transverse and lengthways)

相邻块之间的位置示意图 6-8。实际上，相邻块之间的相关性反映了视频内容联系的密切程度。如果相关性大，则相邻两块可能在同一个 VOP 内部，也就是说可以采用较大范围的强滤波，而不会丢失重要的细节信息。反之，若

相关性较小，则有可能两块处于图像对象边缘，则需要小范围的轻滤波，才可以消除块效应的同时，保留足够多的图像细节信息。

相邻块之间的相互关系通过公式（6-1）计算得到。

$$correlation = \frac{\sum_{i=0}^3 \sum_{j=0}^3 ((P_{ij} - \bar{P}) \times (Q_{ij} - \bar{Q}))}{\frac{1}{16} \sum_{i=0}^3 \sum_{j=0}^3 (P_{ij} - Q_{ij})^2} \quad (\text{其中, } \bar{P} = \frac{\sum_{i=0}^3 \sum_{j=0}^3 P_{ij}}{16}, \bar{Q} = \frac{\sum_{i=0}^3 \sum_{j=0}^3 Q_{ij}}{16}) \quad (6-1)$$

如果图像细节很多的话，则有可能块内的像素间的差异较大，在这种情况下，滤波范围必须相应缩小，才可以避免将块内的细节信息模糊化。为此，需要在滤波器中添加块内相关数据。

MSE 可以衡量块内像素之间的差异度，所以块内自身的相关性，可以利用 MSE 计算得到（见公式 6-2）。如果 MSE 值较小，则说明块内像素差异度较小，可以使用大范围滤波器；反之，则应该使用小范围滤波器。

$$MSE = \frac{\sum_{i=0}^3 \sum_{j=0}^3 (P_{ij} - \bar{P})^2}{16} \quad (\text{其中, } \bar{P} = \frac{\sum_{i=0}^3 \sum_{j=0}^3 P_{ij}}{16}) \quad (6-2)$$

### 6.3.2 相邻块的差值因子

上一节所讨论的两个相关因子主要用来决定滤波的 tap 长度范围。至于滤波的强度，则依靠下面所讨论的差值因子来改变。

差值因子主要用来衡量块效应的严重程度，而差值可以用像素值差值或运动向量差值以及相邻块的 QP 差值来代表。像素值的差值最为直接的反映了块效应的严重程度，因为相邻两块如果块效应比较明显的话，则在边界处有明显的像素值差异。实际上，传统滤波的主要依据数据就是边界点的像素差值。而运动向量的差值则反映了相邻块之间的运动协调程度，如果相邻块之间的运动向量差值较大，则说明块边界的块效应较为明显。至于相邻块的量化参数（QP）的差值，则反映了相邻块之间的量化步长差异，同时间接反映了边界的块效应严重程度，如果相邻块之间的 QP 差值较大，则说明块边界的块效应较为明显。

一般来说，块效应严重的边界需要实施强滤波，反之，则实施弱滤波。为了实现滤波强度的自适应调整，需要将块边界像素差的均值（参见公式 6-3）、运动向量差值以及 QP 差值加入到滤波器的插值公式中。

$$Distance = \frac{\sum_{i=0}^3 \sum_{j=0}^3 (P_{ij} - Q_{ij})^2}{16} \quad (6-3)$$

由于传统的滤波器中已经考虑了边界点的像素差值，因此在本文提出的自适应滤波器主要考虑另外两种差值影响因子。

### 6.3.3 自适应的环路滤波器

为了得到一个效果良好的自适应环路滤波器，需要将前面详细讨论的 4 个影响因子，添加到滤波插值公式中，才可以对滤波强度和滤波范围产生适当的影响。

经过对不同类型视频的内容测试与统计，将前面所提到的 4 个影响因子做线性组合，得到一个滤波公式的可变影响因子（参见公式 6-4），作为滤波插值公式的一项。

$$AdaptiveFactor = correlation \times A + QP \times B + MV \times C + MSE \times D \quad (6-4)$$

此外，在滤波器的适用条件限制中，主要依靠两个相关因子来确定滤波的范围，依靠两个差值因子来确定滤波的强度。因此，在滤波器实施判定条件中，额外需要进行判别。具体判定流程参见图 6-9。

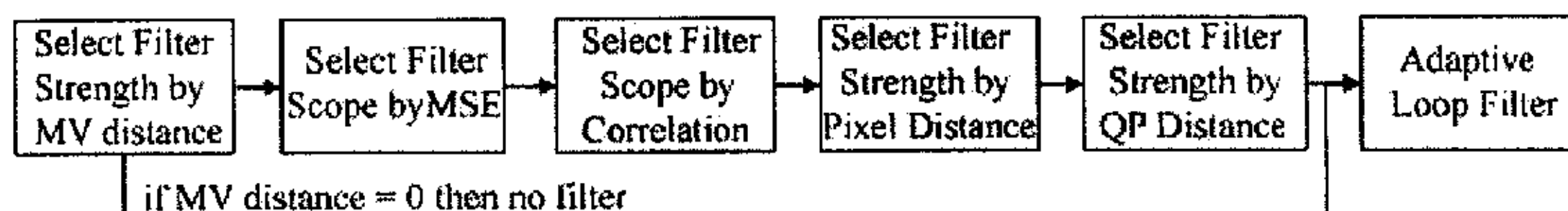


图 6-9 自适应环路滤波实施流程图

Figure 6-9 The Flow of adaptive Loop Filter

## 6.4 实验结果

在 6.3 节中提出了基于视频内容的自适应环路滤波技术。由于利用了相邻块之间的相关性、像素块本身的差异度、边界点的像素值差值、相邻块的运动向量差值以及量化参数的差值等一系列影响因子，所以提出的自适应环路滤波器能够针对不同视频内容实施相应的滤波，从而达到最佳视频处理效果。

针对上述技术方案，在 JVT 6.0 软件平台上做了实验。具体测试条件如下：



- Frame Rate: 30Hz
- Frame number for coding: 30
- Sequence mode of Frame: IPPP...
- The number of reference frame: 2
- The scope of search:  $\pm 32$
- Value of QP: 30, 32, 34, 36

表 6-1 视频序列 Foreman\_A\_00\_0.yuv 测试结果 ( $176 \times 144$ )Table 6-1 Test result of Foreman\_A\_00\_0.yuv ( $176 \times 144$ )

		Original Scheme	Improved Scheme
QP=30	PSNR(Y)(db)	34.98	35.02
	Bit-Rate(kbps)	170.58	169.61
QP=32	PSNR(Y)(db)	33.67	33.68
	Bit-Rate(kbps)	134.07	133.96
QP=34	PSNR(Y)(db)	32.45	32.50
	Bit-Rate(kbps)	108.30	106.37
QP=36	PSNR(Y)(db)	31.03	31.11
	Bit-Rate(kbps)	84.10	84.41

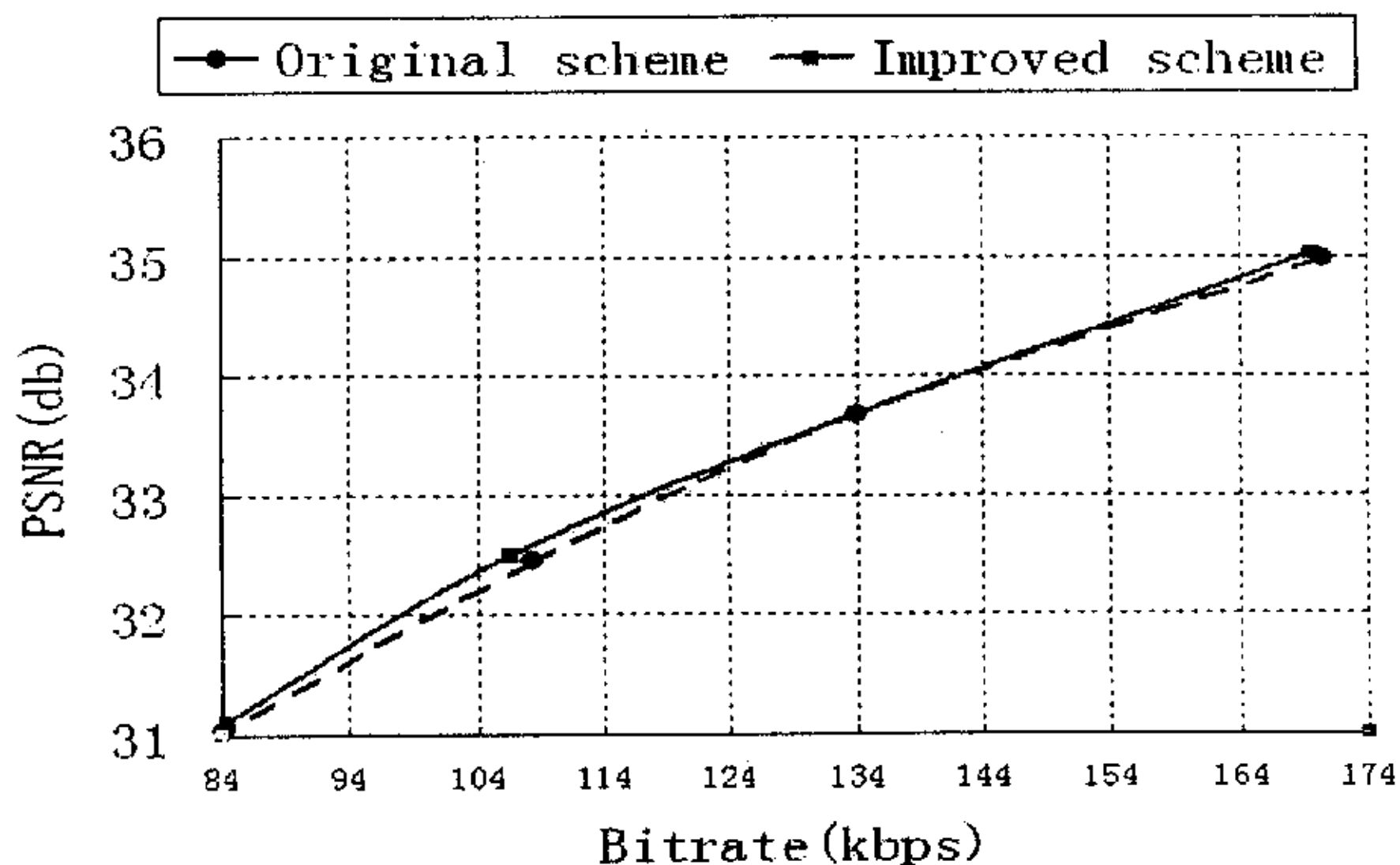


图 6-10 QCIF 视频序列 Foreman\_A\_00\_0.yuv 测试 RD 曲线图

Figure 6-10 The RD curve of Foreman\_A\_00\_0.yuv

表 6-2 视频序列 News\_A\_00\_0.yuv 测试结果 (176×144)

Table 6-2 Test result of News\_A\_00\_0.yuv (176×144)

		Original Scheme	Improved Scheme
QP=30	PSNR(Y)(db)	33.00	33.00
	Bit-Rate(kbps)	226.13	224.03
QP=32	PSNR(Y)(db)	31.48	31.48
	Bit-Rate(kbps)	165.62	165.55
QP=34	PSNR(Y)(db)	30.07	30.09
	Bit-Rate(kbps)	125.14	125.65
QP=36	PSNR(Y)(db)	28.62	28.63
	Bit-Rate(kbps)	91.38	90.84

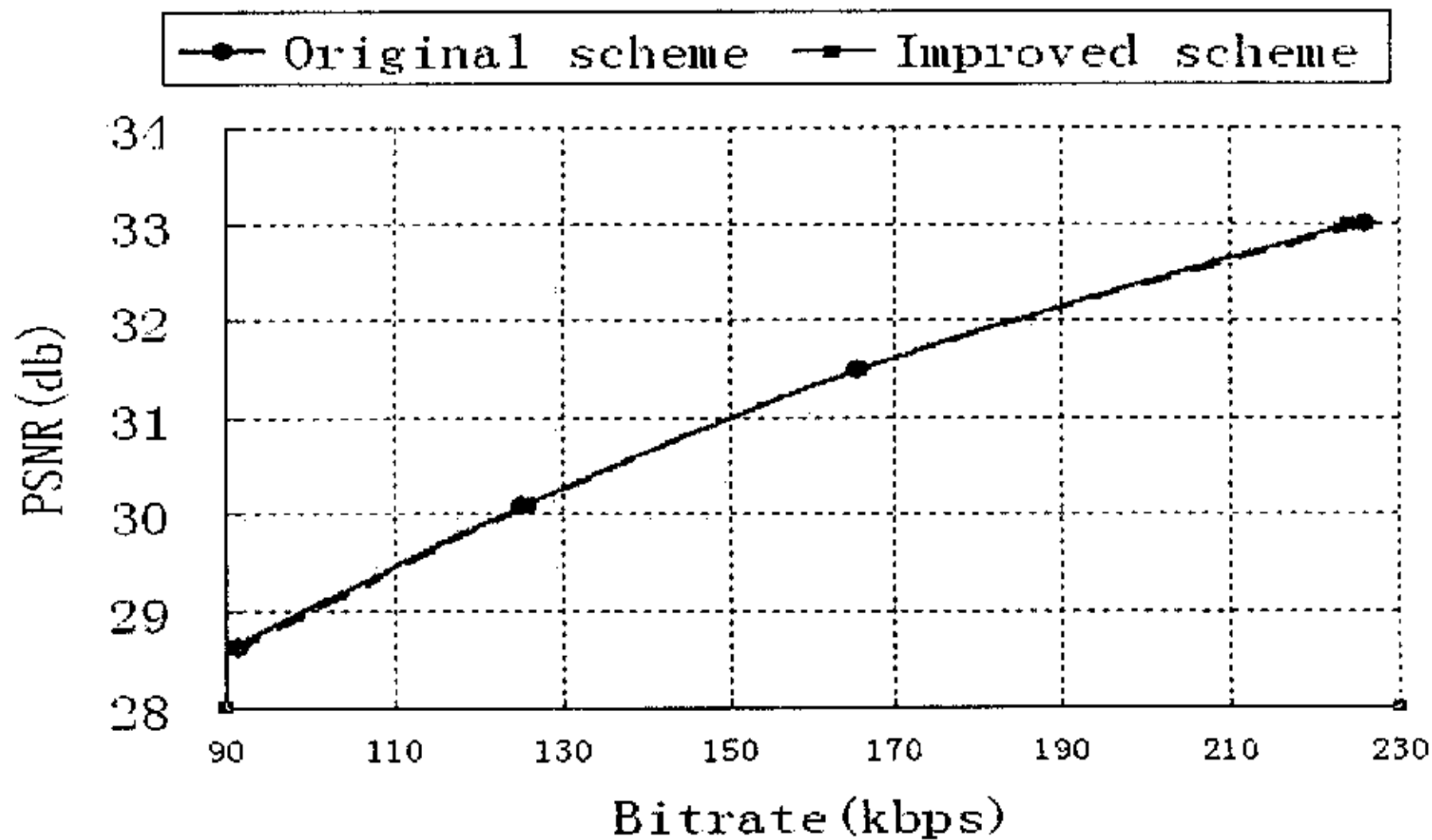


图 6-11 QCIF 视频序列 News\_A\_00\_0.yuv 测试 RD 曲线图

Figure 6-11 The RD curve of News\_A\_00\_0.yuv

针对上述自适应环路滤波算法, 对两个常用的 QCIF 格式的视频序列 Foreman.yuv 和 News.yuv 做了实验, 格式均为 176×144。由于滤波后的重构图像帧参与了后续的编码循环, 因此编码码率有所改善。在 RD 曲线图中, 实线为本文提出的滤波方案编码曲线, 虚线为 H.264 技术 JVT 6.0 版本的编码曲线。由图可知, QP 值较大的时候, 本文提出的滤波算法效果比较好, RD 增益较为明显。

## 6.5 本章小结

在信号处理领域中, 传统的滤波技术从维纳滤波器发展到了卡尔曼滤波



器。不论从实施效果还是计算复杂度等方面，均有了十足的进步。到了六十年代后期，提出了自适应滤波的概念，由于其有针对性的滤波，灵活的实施范围，使得它在短短十几年内发展迅速，已经成为当前国际上信号处理的前沿技术。

在视频编码压缩中，图像处理的基本单元是宏块/块。在宏块/块的分割中，就不可避免的将本来有机构成的图像整体分为不同的像素块。以宏块/块为单位的编码必然会产生块边界的噪音。为了消除这种块效应，滤波技术被引入视频编码领域。

一般来说，图像滤波有两类：后滤波与环路滤波。前者仅仅在解码端提交用户之前，做滤波处理；而后者则参与到编码环节中，为更加准确的运动估计做进一步的改善。在 H.264 标准中使用了环路滤波技术，分别针对不同的滤波强度实施相应强度的滤波。但是，JVT 提供的滤波器的缺点在于滤波系数固定，不能针对不同视频内容做有效的改变。

一个好的滤波器不但可以有效消除块效应，而且还能够保留足够的图像细节信息。为了实现高效的自适应滤波，经过对视频内容的分析与统计，提取了 4 个影响因子。它们分别是相邻块之间的相关度、块内像素的差异度、相邻块运动向量差值、相邻块 QP 差值。加上原有滤波器所考虑的边界点像素值差值后，线性组合成为一个影响因子，在经过对滤波范围和滤波强度的判定之后，实现基于视频内容的自适应滤波。试验表明，整个技术方案效果良好。

本文提出的自适应滤波器尽管滤波效果相对较好，但同时也具有计算复杂度较高等缺点。这样给整个视频编码软件的硬件移植带来了巨大障碍，所以今后的工作可以集中在滤波器计算复杂度的减小等方面。此外，如何更加有效的综合与组织 4 类影响因子，从而可以得到效果更加优良的自适应环滤波也是下一步的研究重点之一。

## 结 论

随着宽带网络的快速发展, 基于互联网的流媒体传输的应用越来越广泛。高效的视频编解码技术是实现流媒体传输的关键, 编码性能的好坏将大大影响流媒体服务质量。

本文主要研究了基于 H. 264 标准的数字视频编解码技术。首先对目前国际标准 MPEG-4 和 H. 264 进行了概括性的原理阐述, 然后在此基础上, 讨论了自己的创新研究工作, 具体分为三个部分: 块匹配算法的简化研究、帧内预测新模式的研究以及自适应环路滤波技术的研究。

在块匹配算法研究中, 首先针对传统算法搜索结果不稳定的特点, 提出了预处理反馈环, 利用预处理技术判定搜索块的相互关系, 确定合适大小的搜索窗口。然后计算当前像素块自身的差异度 (MSE), 进一步精确调整窗口大小。通过预处理, 可以直接确定搜索范围, 避免匹配误差。为了消除由此带来的计算复杂度, 提出了像素采样匹配, 每 4 个像素中选取一个像素参与 BMA 计算, 由于相邻像素之间的差异很小, 所以不会影响匹配结果, 但计算复杂度却因此而降低约 75%。最后, 为了消除 BMA 带来的累积误差, 提出了利用基于视频内容的特征点几何位置约束关系消除累积误差, 做二次校正。

在帧内预测模式的研究中, 分析比较了相邻块之间的相关性和块内像素之间的相关性的关系, 提出了利用块内相关性进行帧内预测。通过对视频内容纹理的分析与统计, 提出了 L 型扩展、右下斜线、左下斜线、中心扩展等 4 种方向预测模式。由于块内像素之间的相关性远远大于相邻块之间的相关性, 因此利用块内相邻像素进行预测可以得到更为准确的像素预测值。预测办法和 RD cost 计算的原则都与原先的 9 种预测模式相同。此外, 为了降低由此带来的计算复杂度, 提出了反馈评价环, 利用相邻块预测模式的延续性, 确定当前块的帧内预测模式, 从而可以避免对所有 4 种模式的 RD Cost 计算。

在自适应环路滤波的研究中, 针对目前 H. 264 标准所使用环路滤波器为固定系数的缺点, 经过对视频内容的分析与统计, 提取了两类影响因子: 相关因子和差值因子。它们分别是相邻块之间的相关度、块内像素的差异度, 相邻块运动向

量差值、相邻块 QP 差值、边界点像素值差值等。对以上影响因子做线性组合成为一个综合影响因子添加到滤波插值公式中,在经过对滤波范围和滤波强度的判定之后,实现自适应滤波。

本文提出的创新思路均在 JVT 提供的基于 H. 264 标准的公用软件平台上进行了实验。实验结果表明,编码效率和计算复杂度均在不同程度上有所提高。

尽管视频编解码技术在基本原理框架方面目前已经发展比较成熟了,但就具体各个环节的技术方案来说,仍然有很多可以改进和完善之处。本文所研究课题的后续工作中,可以从以下几个方面着手。

编解码框架实现复杂度的减小。目前基于 H. 264 标准的视频编解码软件实施的计算复杂度和空间复杂度依然比较高,必然导致了高额的硬件实现成本,因此如果能够进一步减小时间/空间复杂度,不但可以相应提高编码时间和编码效率,而且能够实现低成本硬件移植。

编解码效率的进一步提高。对现有编解码方案进行详细的分析,寻找技术缺陷和不足,从而改进现有技术;对视频内容进行分析统计,寻找视频图像中的各类蕴含信息,对不同种类的信息做相应水平的编码,实现高性能的基于内容的自适应视频图像编解码方案。延拓 MPEG-4 标准所提到的基于 VOP 的编码方案,细化不同 VOP 的编码技术,实现有针对性的基于 VOP 的自适应编码技术。

国际上,ISO/IEC 和 ITU-T 成立的联合视频工作组(Joint Video Team, JVT)已经推出了新一代视频编码标准 H. 264,并且一直在发展完善过程中。此类关键技术包含了国外各大公司的很多关键性的技术发明专利,必然会成为视频编解码芯片生产研发的技术壁垒。为了保护我国国内视频编解码芯片生产厂商的权益,必须提出中国自己的更为高效的视频编解码标准用来抗衡,不仅可以据此生产更为优良的视频编解码产品,同时还可以避免国外昂贵的专利与版权费用。因此,研究并提出一个高效的数字视频编码方案对视频编解码技术知识产权的保护,意义重大。

## 参 考 文 献

- 1 M.barnsley. Fractal image compression. Notices of AMS. Vol. 43. No.6. 1996: 657-662.
- 2 沈兰荪. 图像编码与异步传输. 人民邮电出版社. 1998:10~15
- 3 ISO/IEC 15938. Information technology-multimedia content description interface (MPEG-7), 2002.
- 4 ISO/IEC 21000. Information technology-multimedia framework (MPEG-21), 2003.
- 5 Emerging H.26L Standard: Overview and TMS320C64x Digital Media Platform Implementation (White Paper) 1999.
- 6 ISO/IEC 14496-10 and ITU-T Rec. H.264 Advanced Video Coding. 2003.
- 7 GB/T XXXXX.1—YYYY中华人民共和国国家标准 信息技术 先进音视频编码 第一部分: 系统 国家质量监督检验检疫总局. 2003.11
- 8 GB/T XXXXX.2—YYYY中华人民共和国国家标准 信息技术 先进音视频编码 第二部分: 视频 国家质量监督检验检疫总局. 2003.11
- 9 Ralf Steinmetz, Klara Nahrstedt. Multimedia Fundamentals (Volume I) Media Coding and Content Processing. 电子工业出版社. 2003:65~75
- 10 Yao Wang, Jorn Ostermann, Ya-Qin Zhang. Video Processing and Communications. Publishing House of Electronics Industry. 2003:76~80
- 11 Iain E. G. Richardson. H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia. John Wiley & Sons, Ltd. 2003: 42~63
- 12 Rob Koenen. ISO/IEC JTC1/SC29/WG11 N3536 MPEG-4 Overview (V1.5 –Beijing version) July 2000
- 13 ISO/IEC JTC 1/SC 29/WG 11 N4350. INFORMATION TECHNOLOGY – CODING OF AUDIO-VISUAL OBJECTS – Part 2: Visual Sydney July 2001
- 14 I. Daubechies. The wavelet transform time-frequency localization and signal analysis, IEEE Trans. Inf. Theory 36. 1990:961-1005
- 15 高文, 吴枫. MPEG4编码的现状和研究. 计算机研究与发展 Vol.36, No.6 1999:643~652
- 16 Gary J. Sullivan and Thomas Wiegand. Rate-distortion optimization for video compression. IEEE SIGNAL PROCESSING MAGAZINE Nov.1998:74~90
- 17 The Emerging JVT/H.26L Video Coding Standards, Proceeding of IBC 2002, Amsterdam, NL, Sept. 2002
- 18 田栋, 姚志恒, 沈兰荪 H.263的发展及应用. 测控技术 第20卷第5期 2001:5~12
- 19 ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6 ) Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC) JVT-G050r1
- 20 W. Li and E. Salari, Successive elimination algorithm for motion estimation. IEEE Transaction on Image Processing.4 (1). 1995:105~107

- 21 Y.C.Lin and S.C.Tai. Fast full-search block matching algorithm for motion-compensated video compression, IEEE Transaction on Communications, 45(5). 1997:527~531
- 22 T.Koga, K.Linuma, A.Hirano, Y.Iijima, and T. Ishiguro. Motion-compensated interframe coding for video conferencing [A], Proceedings of the National Telecommunications Conference[C], New York. vol.4. 1981:G5.3.1~G5.3.5.
- 23 Jaswant R. Jain and Anil K. Jain. Displacement measurement and its application in interframe image coding [J]. IEEE Trans. on Communications, vol. COM-29, no.12, 1981:1799~1808
- 24 M.Ghanbari. The cross-search algorithm for motion estimation[J], IEEE Trans. on Communications, vol.38, no.7, 1990:950~953
- 25 Steven L. Kilthau, Mark S. Drew, and Torsten Moller. Full Search Content independent Block Matching Based on the fast fourier Transform, ICIP02,102~105
- 26 Yong-Sheng Chen, et al. Fast Block Matching Algorithm Based on the Winner-Update Strategy [J]. IEEE Transactions on Image Processing, vol.10, no.8, 2001:1212~1222
- 27 王雁来, 尹宝才, 孔德慧等. 块匹配算法在基于MPEG-4的人脸特征提取中的改良与应用[J], 北京工业大学学报(自然科学版). 第29卷, 第1期, 2003(3):98~102
- 28 Ke Wang, YanLai Wang, Yin BaoCai, et al. Face Pose Estimation with a Knowledge based Perspective Model. IEEE International Conference on Neural Networks & Signal Processing Nanjing, December 12-15, 2003 (ICNNSP'03) 2003(12):1131~1134
- 29 Zhengyou Zhang, Richid Deriche, Olivier Faugeras, Quang-Tuan Luong. A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. Artificial Intelligence, 78, 1995:87~119
- 30 王雁来, 尹宝才, 程可等. 基于内容的自适应块匹配算法. 中国图形图像学报(A). 第8卷, 第1期特刊, 2003(10):481~485
- 31 Paul A. Wintz. Transform Picture Coding. Proceedings of the IEEE, vol. 60, no.7, 1972(7): 809~820
- 32 M.Rabbani and R.Joshi. An overview of the JPEG2000 still image compression standard. Signal Processing: Image Communication, vol.17, no.1, 2002(1): 3~48
- 33 Coding of Still Pictures (JBIG, JPEG) of ISO/IEC JTC 1/SC 29/WG 1(ITU-T SG 16) N 2412 The JPEG-2000 Still Image Compression Standard 2002-12-25
- 34 Bojun Meng, Oscar C. Au, Chi-Wah Wong, et al. Efficient Intra-Prediction Mode Selection for 4×4 Blocks in H.264. ICME2003 2003:521~524
- 35 Bojun Meng, Oscar C. Au. Fast Intra-Prediction Mode Selection for 4×4 Blocks in H.264. ICME2003 2003:389~392
- 36 Woo-Shik Kim,Dae-Sung Cho,and Kyung-Wook Kim. Intra Prediction Enhancements. JVT-E051. 5th Meeting: Geneva, CH, 9-17 October
- 37 Minhua Zhou. Scalable Intra Prediction. JVT-C033. 3rd Meeting: Fairfax, Virginia, USA, 6-10 March, 2002
- 38 Zhong Xue,Sheng Mei Shen. An Automatic Mode Decision Method for Intra



- Frame Coding and Decoding. VCEG-O48. 15th Meeting: Pattaya, Thailand, 2001(12): 3~7
- 39 Siu-Leong Yu and Christos Chrysafis. New Intra Prediction using Intra-Macroblock Motion Compensation. JVT-C151. 3rd Meeting: Fairfax, Virginia, USA, May 6-10, 2002
  - 40 ITU-T Rec. H.264/ISO/IEC 11496-10. "Advanced Video Coding", Final Committee Draft, Document JVT-F100, December 2002.
  - 41 Wang YanLai, Yin BaoCai, Ji PengFei, et al. Interactive Intra Prediction Mode in Video Coding. IEEE International Symposium on Intelligent Signal Processing and Communication Systems, December 7- 10, 2003. Awaji Island, Japan (ISPACS 2003). 2003(12):520~524
  - 42 J.Ribas-Corbera and S.Lei. Rate control for low-delay video communications. ITU Study Group 16, Video Coding Experts Group, *Document Q15-A-20*, Portland, June 97
  - 43 何振亚. 自适应信号处理, 科学出版社, 2002:5~8
  - 44 S.Haykin, Adaptive Theory, Prentice-Hall, Inc. 1986:18~26
  - 45 Greg Conklin, Neelesh Gokhale. Dithering 5-tap Filter for Inloop Deblocking, JVT-C056, (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 3rd Meeting: Fairfax, Virginia, USA, 6-10 May, 2002
  - 46 Shijun Sun. Loop Filter with Skip Mode, VCEG-M20.doc, VCEG 13th meeting, Austin, U. S., April 2-4, 2001
  - 47 Guy Côté, Lowell Winger, Michael Gallant. Lower complexity deblocking filter with in-place filtering, VCEG-O39. VCEG. 15th Meeting: Pattaya, Thailand, 4-6 Dec., 2001
  - 48 Thomas Wedi. Adaptive Interpolation Filter for H.26L. VCEG-N28. VCEG. 14th Meeting: Santa Barbara, CA, USA, 24-27 Sep., 2001
  - 49 Khac K. Pang, Thiow K. Tan. Optimum Loop Filter in Hybrid Coders, IEEE CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. Vol 4. No.2. 1994:158~167
  - 50 B.Deknuydt, L.Van Eycken et.al. A deblocking scheme suitable for use inside the codec loop, IEEE 0-7803-1771-6/94. 1994:201~206
  - 51 Wang YanLai, Yin BaoCai, Kong DeHui. Adaptive Video Coding in Loop Filter Based on Content. IEEE International Conference on Neural Networks & Signal Processing Nanjing, December 12-15, 2003(ICNNSP'03). 2003(12):1189~1192.

## 攻读硕士学位期间发表的论文

- 1 王雁来, 尹宝才, 孔德慧等. 块匹配算法在基于MPEG-4的人脸特征提取中的改良与应用. 北京工业大学学报(自然科学版). 2003年3月, 第29卷, 第1期, 98~102.
- 2 Wang YanLai, Yin BaoCai, Kong DeHui. Adaptive Video Coding in Loop Filter Based on Content. IEEE International Conference on Neural Networks & Signal Processing Nanjing, December 12-15, 2003 (ICNNSP'03) IEEE catalog number : 023EX64; ISBN :0-7803-7702-8. 2003(12):1189~1192.
- 3 Ke Wang, YanLai Wang, Yin BaoCai, et al. Face Pose Estimation with a Knowledge based Perspective Model. IEEE International Conference on Neural Networks & Signal Processing Nanjing, December 12-15, 2003 (ICNNSP'03) IEEE catalog number : 023EX64; ISBN :0-7803-7702-8. 2003(12):1131~1134.
- 4 王雁来, 尹宝才, 程可等. 基于内容的自适应块匹配算法. 中国图形图像学报(A). 第8卷, 第1期特刊, 2003(10):481~485.
- 5 Wang YanLai, Yin BaoCai, Ji PengFei, et al. Interactive Intra Prediction Mode in Video Coding. IEEE International Symposium on Intelligent Signal Processing and Communication Systems, December 7- 10, 2003. Awaji Yumebutai Awaji Island, Japan (ISPACS 2003). 2003(12):520~524.
- 6 Wang Ke, Wang YanLai, Yin BaoCai, et al. Face Feature Tracking and The Head Pose Estimation. Journal of Beijing University Technology. (Accepted)
- 7 Wang YanLai, Yin BaoCai, Kong DeHui. Adaptive Object Originated Loop Filter in Video Coding. Chinese Journal of Electronics. (Checking)
- 8 尹宝才, 王雁来, 孔德慧等. 基于内容的视频编码方法. 中国国家发明专利. 03122943. 3.
- 9 孔德慧, 王雁来, 尹宝才等. 一种用于视频编码的简单化环路滤波方法. 中国国家发明专利. 200410000671. 4.

## 致 谢

三年前，一个偶然的事件使我来到了北工大计算机学院，继续我的硕士研究生生活。在这期间的学习和生活中，我得到了尹宝才老师、孔德慧老师、孙艳丰老师的关心和教导。尹老师民主的学术作风给实验室带来十分宽松的学习气氛。正是在这种平和的学术气氛下，使我可以尽心研究我喜欢的课题项目，并有机会出国参加国际学术交流。而孔老师和孙老师的平易近人和谦逊和蔼，则使得我无时无刻感受到她们的关怀。

感谢国家 863 计划项目资助以及北京工业大学研究生科技基金的资助，使我能够顺利完成子课题的研究。

感谢共创联盟软件的刘明宝博士，没有他的帮助，我无法继续研究生生涯。

感谢孔德慧老师对本论文的内容充实以及文字润色等的修改意见。

感谢中科院计算所高文老师和贺玉文老师，他们耐心的帮助和指导，助我答疑解惑，给我学习信心。

感谢微软亚洲研究院的吴枫研究员，他渊博的知识让我受益匪浅。没有他的帮助，没有他给我的压力，我无法达到既定的科研目标。

感谢微软亚洲研究院的林宙辰研究员，他给予了我微软亚洲研究院的实习机会。通过宝贵的 4 个月的实习，我逐渐明白了做科学研究的基本方法。

感谢中科院计算所的吕岩博士、马思伟博士和北京工业大学电控学院的薛金柱博士对我学业的帮助和指导。

感谢重点实验室的王珂博士、邹自强师兄、胡永利博士、张楠博士、王文东同学、赵三强同学、谷春亮同学、程世铨同学、刘文韬同学、李丽岩同学、季鹏飞同学、孙磊同学、岳文颖同学，他们在平时的学习和生活中都曾给过我真诚的帮助和建议。还要谢谢实验室其他同学以及所有关心我帮助我的朋友，感谢他们对我的鼓励和帮助。

最后，感谢远在家乡的亲人和朋友，尤其是我已经故去的父亲！他的挂念和关心一直是我前进步伐的动力，也使我在人生重要的三年中交出了较为满意的一张答卷。



基于H. 264的高清晰视频编码技术研究

作者：[王雁来](#)  
学位授予单位：[北京工业大学](#)  
被引用次数：4次

引证文献(4条)

1. 陈亚菲, 陈生潭, 董元平 对H. 264/AVC 中自适应滤波技术的评测[期刊论文]•[电子技术应用](#) 2006(12)
2. 苏海斌 [基于AVS-M的视频流媒体P2P传输方法研究](#)[学位论文]硕士 2006
3. 朱东巍 H. 264中去块效应滤波器的研究及硬件设计[学位论文]硕士 2006
4. 陈昱东 [3G网络视频流媒体的P2P模式传输方法研究](#)[学位论文]硕士 2006

本文链接：[http://d.g.wanfangdata.com.cn/Thesis\\_Y612755.aspx](http://d.g.wanfangdata.com.cn/Thesis_Y612755.aspx)