

## 中文摘要

**摘要:** 随着高清数字视频时代的到来,“高清”已经成为各种数字设备发展的必然趋势,用户对高清和高保真度视频通信的需求日益升高,希望无需对现有网络进行较多改动和升级便能实现更好的视频质量。因此如何在高保真度下,提高视频编码压缩效率,有效去除各种冗余信息,尽可能降低码率以节省带宽成为视频编解码领域研究的热点。

当前,MEPG-2 和 H.264 标准中的频率加权技术,利用不同频段图像注意度不同,采用不同的量化参数提高编码效率。该技术逐帧更新量化矩阵,但没有重点考虑不同图像的内容特征,主观质量和压缩比不够好。本文在 H.264 帧级自适应频率加权的基础上,针对 H.264 宏块级自适应频率加权技术展开研究。

首先,本文分析了 H.264 的结构特点和编解码原理,总结了 H.264 相对于以往视频编码标准的优势,并剖析了产生该优势的 H.264 关键技术。

其次,本文通过对人类视觉系统特性的分析,提炼总结了和视频压缩相关的人类视觉系统敏感度特性,针对 H.264 帧级自适应频率加权算法的缺陷,对其进行了改进,提出了基于人类视觉系统的宏块级自适应频率加权算法。

再次,本文分析了现有感兴趣区域检测技术的优劣势,引出了感兴趣区域的重要性,并吸取以往感兴趣区域定义的优点,提出了以视觉兴趣性为前提的改进地感兴趣区域提取算法,并把该算法应用于宏块级自适应频率加权算法,进一步改进了频率加权矩阵,提出了基于感兴趣区域的宏块级自适应频率加权算法。

最后,对全文工作进行了总结,指出当前工作中的不足,为以后的研究工作指明了努力方向。

**关键词:** H.264 标准; 频率加权; 人类视觉系统; 感兴趣区域;

**分类号:** TN919.8

## ABSTRACT

**ABSTRACT:** With the advent of high-definition (H-D) digital video age, H-D has become the inevitable tendency of the development of a variety of digital devices. The demands for high fidelity with no change and upgrade to existing networks increase greatly. How to raise video coding compression ratio, remove all redundant information and reduce the rate to conserve bandwidth as much as possible with a high fidelity is a very hot research field.

Today, the frequency weighting technique has been used in MPEG-2 and further developed in H.264 High Profile (HP), in which quantization matrix is updated picture by picture without considering the different characteristics of different parts. However, the subjective quality and compression ratio are not so good. Based on the picture-level adaptive frequency weighting of H.264, the Macroblock-Level adaptive frequency weighting algorithms for H.264 are proposed.

First, in the report, the structure and decoding principle of H.264 are analyzed, and then the advantages and key technology are concluded comparing with the previous video coding standard.

Second, the characteristic of human visual system is presented and employed. Considering the shortcoming existing in H.264 picture-level adaptive frequency weighting algorithm, an improved Macroblock-Level adaptive frequency weighting algorithm based on human visual system is proposed in this paper to improve the subjective quality.

Third, by the analysis of the region of interest detecting technique, the importance of the region of interest is used. In order to ensure the efficiency of algorithm, a better approach to detect the region of interest is proposed.

Moreover, the improved region of interest detect algorithm is applied in the Macroblock-Level adaptive frequency weighting, combing with the improved quantization matrix, a Macroblock-level adaptive frequency weighting based on region of interest is presented.

Finally, the whole article is summarized and the inadequacy of the current work is pointed out. Also, the direction for future work is specified at the end of the paper.

**KEYWORDS:** H.264; Human Visual System; Region of Interest; Frequency Weighting

**CLASSNO:** TN919.8

## 致谢

时光如梭，转眼间硕士生活即将结束。能够在现代通信网实验室这个和谐融洽、学习气氛浓厚的大家庭中完成我的学业，不仅是我学业中的收获，更是我人生中的财富。

感谢我的导师李兴华老师，本论文的工作是在李老师的悉心指导下完成的，李兴华老师严谨的治学态度和科学的工作方法给了我极大的帮助和影响。在此衷心感谢三年来李兴华老师对我的关心和指导。

感谢实验室荆老师给予我的指导。在我攻读硕士学位期间，荆老师提供了很多实践机会，并悉心指导我们完成了实验室的科研工作，在学习上和生活中都给予我很大的关心和帮助，在此向荆涛老师表示衷心的感谢。

在实验室工作及撰写论文期间，霍炎博士在专业知识的学习和研究工作中给予了热情帮助，并对我的科研工作和论文提出了许多宝贵意见，在此向他表达我的感激之情。霍炎博士深厚的理论功底、一丝不苟的科研精神，使我受益匪浅。

仲丽媛、郭英慧等同学对我论文中的研究工作给予了大力支持和热情帮助，在此向他们表达我的感激之情。

另外，感谢一直以来支持我、鼓励我的家人，他们的理解和关心使我能够在学校专心完成我的学业。

最后，再一次对大家的帮助表示诚挚的谢意。

# 1 引言

## 1.1 研究背景和意义

随着信息社会和高清数字视频时代的到来,多媒体应用需求越来越大,数字视频技术已广泛应用于通信、计算机、广播电视等领域,带来了会议电视、可视电话及数字电视、媒体存储等一系列应用,“高清”已经成为各种数字设备发展的必然趋势,用户对高清和高保真度视频通信的需求日益升高,希望无需对现有网络进行较多改动和升级便能实现更好的视频质量。为满足高清和高保真度需求,研究者提出了许多技术,包括(1)提高比特深度:像素比特深度由8比特提高到10比特或12比特;(2)扩展色度保真度:由4:2:0提高到4:4:2或4:4:4。上述技术由于视频内容更加丰富,信息量也就更大,网络传输码率增加,无可避免地受到有限带宽资源的制约,不利于网络传输。而在用甚低比特率对视频进行压缩时,往往会丢失一些细节信息,甚至其中部分信息是人们特别感兴趣的。从主观视觉角度来说,这些区域的恢复质量直接影响到重建图像的整体主观感受质量。因此如何对视频编解码算法进行优化,在满足高保真度同时,提高视频编码压缩比,有效去除各种冗余信息,尽可能降低码率以节省带宽成为研究的热点。

视频编码算法的优化无论在实时视频应用中,还是在面向存储的多媒体应用中都有着重要的意义。

在实时视频应用中,更优的视频编码算法能够用较低的输出码率,在较小的带宽上传输同等质量的视频,提高网络资源利用率。

在面向存储的多媒体应用中,同等视频质量下,码率越小,存储编码码流所需要的存储容量越小,进而媒体存储介质的成本愈小。

研究表明,人眼是视频质量的最终评判者,可利用人类视觉系统(Human Visual System, HVS)敏感度特性在不改变主观质量甚至提高主观质量时较大程度对视频序列进行压缩,降低编码比特数<sup>[1]</sup>。

研究还发现,对于图像和视频等多媒体数据中包含的很多信息,人们在观察和理解图像时并不是对所有区域都分配相同的注意力,而是重点关注某些区域,这些区域被称为感兴趣区域(Region of Interest, ROI)。如果计算机能找出这些关键区域对其进行重点分析,而对非感兴趣区域适当调整量化参数,将会在降低编码比特数的同时大大提高图像和视频编码效率和精度。

## 1.2 发展现状

### 1.2.1 视频编码标准发展现状

H.261<sup>[2]</sup>是国际电信联盟(International Telecommunication Union, ITU-T)制定的第一个视频编码标准,是视频压缩编码的里程碑,其采用的压缩算法成为以后其它各种压缩标准的基础和核心;H.261的输出码率是  $p \times 64\text{kbps}$ ,主要应用于综合业务数字网(Integrated Services Digital Network, SDN)、异步传输模式(Asynchronous Transfer Mode, ATM)等宽带信道上声音和图像信息的实时传输,不适合在公共交换电话网(Public Switched Telephone Network, PSTN)和移动通信网等带宽有限的网络上应用<sup>[2]</sup>。为满足低速率视频通信的应用需要,1995年ITU-T推出了H.263视频编码标准,该标准可用于64kbps以下的低码率视频传输。2002年ITU-T又公布了H.263的第二版——H.263+和H.263++,在H.263的基础上增加了一些选项以提高压缩效率或特定功能,具有很强的扩展性,且进一步提高了压缩编码性能<sup>[3]</sup>。

国际标准组织(International Organization for Standardization/International Electrotechnical Commission, ISO/IEC)的活动图像专家组(Moving Picture Expert Group, MPEG),于1991年公布了MPEG-1视频编码标准,该标准主要用于家用VCD的视频压缩。1994年公布的MPEG-2标准,除了用于家用DVD之外,主要用于数字视频广播和高清晰电视。1999年公布的MPEG-4标准,重点强调了多媒体通信的交互性和灵活性<sup>[4]</sup>。

H.264视频编码标准由ITU-T和ISO/IEC成立的联合视频小组(Joint Video Team, JVT)于2003年公布,即MPEG-4的第10部分内容。H.264的压缩效率更好,且具有更有的网络亲和性,主要应用于视频实时通信,数字广播电视和网络流媒体等领域。H.264的推出,是视频编码标准的一次重要进步,与其它现有的视频编码标准相比,在相同的带宽下提供更加优秀的图像质量,具有明显优越性<sup>[5]</sup>。

图 1-1 为视频编码标准发展历程。

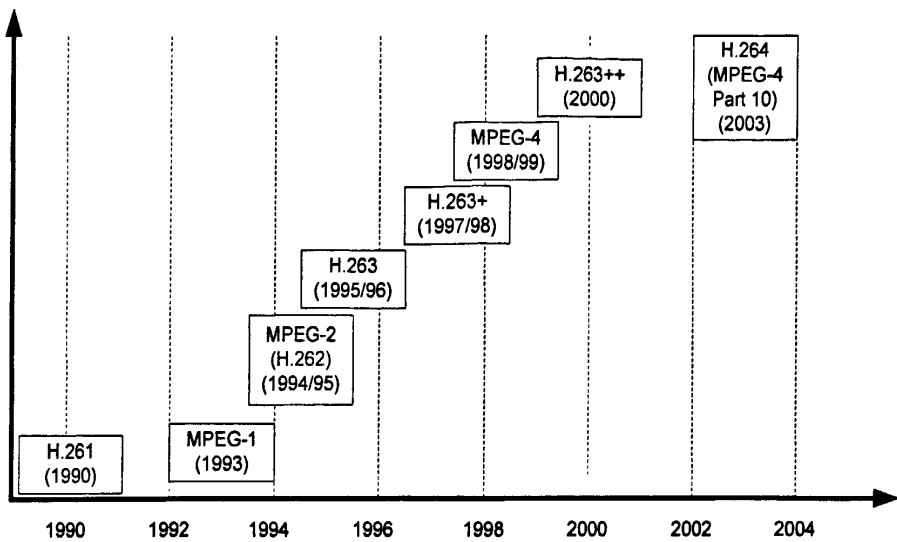


图 1-1 视频编码标准发展历程

Fig. 1-1 Video coding standard development process

### 1.2.2 频率加权技术发展现状

频率加权技术是根据 HVS 频率敏感度特性提出的一种自适应量化加权方法。该方法利用人眼对高频部分失真不敏感特性,对不同频段的 DCT 系数采用不同的量化参数,提高压缩比。

在 JPEG 和 MPEG-2 中应用频率加权技术,高频部分粗略量化,低频部分精细量化,针对帧内和帧间提出两种量化矩阵,对  $8 \times 8$  块进行自适应量化; H.264 高档 (High Profile, HP) 中采用可变大小块编码,利用频率加权技术,采用帧级自适应频率加权算法 (Picture-Level Adaptive Frequency Weighting, PAFW),逐帧更新量化矩阵,分别对  $4 \times 4$  块和  $8 \times 8$  块进行自适应量化。但 JPEG 和 MPEG-2 的频率加权算法中粗略量化的高频部分往往代表视频图像细节,若被粗略量化的细节是人类视觉系统敏感区域,视频主观质量下降。H.264 HP 中的 PAFW 算法,逐帧更新量化矩阵,一帧中所有宏块采用相同量化矩阵,不考虑视频图像不同区域的内容特征,在一定程度上没有突出细节,压缩比不够好。可把帧级自适应频率加权扩展到宏块级,逐宏块更新量化矩阵,满足人类视觉特性同时,提高视频压缩比<sup>[6]</sup>。

### 1.3 本文结构安排

本文在 H.264 国际视频编码标准的基础上,通过对人类视觉系统特性的研究

提出了基于人类视觉系统的宏块级自适应频率加权算法；并进一步分析感兴趣区域检测技术，提出新的感兴趣区域提取算法的同时，将该算法应用于宏块级自适应频率加权算法，提出了基于感兴趣区域的宏块级自适应频率加权算法。图 1-2 为论文结构图。

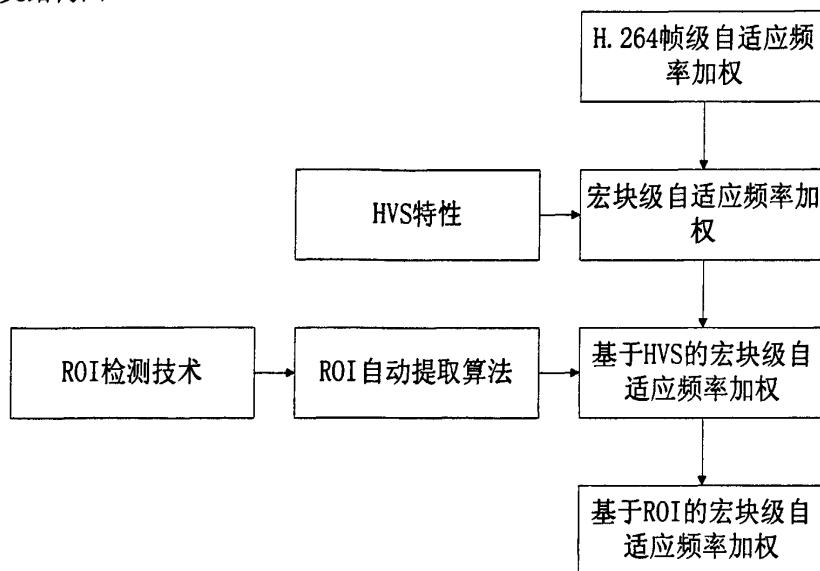


图 1-2 本文研究内容结构图

Fig. 1-2 Research contents

如上图所示，论文主要内容如下：

第一章引言，介绍了视频编码优化算法研究的背景意义和目前视频编码标准和频率加权技术的发展现状；

第二章首先分析了 H.264 视频编码标准的档次结构特点和编解码原理，其次对帧内预测、帧间预测、整数变换和量化等关键技术进行了剖析；

第三章针对现有的基于 H.264 的帧级自适应频率加权算法的缺陷，详细分析了人类视觉系统特性，并在此基础上提出了基于人类视觉系统的宏块级自适应频率加权算法；

第四章对感兴趣区域的定义提取算法进行了研究，在基于信息熵的自适应阈值确定算法基础上，提出了新的感兴趣区域判定算法，并引入到宏块级自适应频率加权算法，进一步提高了视频主观质量；

第五章总结了本文提出的基于人类视觉系统和基于感兴趣区域的宏块级自适应频率加权算法的优势特点和下一阶段研究的方向，并就视频编码优化技术的发展进行展望。

## 2 H.264 国际视频编码标准概述

H.264 标准是一种高性能的视频编解码技术，是一个面向未来 IP 和无线环境下的最新数字视频压缩编码标准。其目标应用涵盖了目前大部分的视频服务，应用较广。因此，本文以 H.264 标准为背景，对视频编码优化算法自适应频率加权技术进行研究。本章将针对 H.264 标准进行分析阐述。

### 2.1 H.264 档次和分层结构

#### 2.1.1 H.264 档次

H.264 规定了三种档次，每个档次支持一组特定的编码功能<sup>[7]</sup>。

(1) 基本档次：不包含以下两点内容，初此之外，包含其他所有内容。这两部分是：

1) B 帧、加权预测、基于上下文的自适应算术编码(Context-based Adaptive Binary Arithmetic Coding, CABAC)、视频图像宏块自适应切换场合帧编码(Macroblock Adaptive Field/Frame Coding, MB-AFF)和场编码；

2) 数据分割和 SP/SI 片。

(2) 主要档次：采用帧间编码和帧内编码，支持隔行视频和 CABAC。

(3) 扩展档次：采用数据分割，能够改进误码性能；支持 SP 和 SI 片的有效切换，不支持隔行视频和 CABAC。

由图 2-1 可知，扩展档次不包括主要档次的所有功能，但包括了基本档次的功能，且参数的不同设置对应不同的编码器性能。



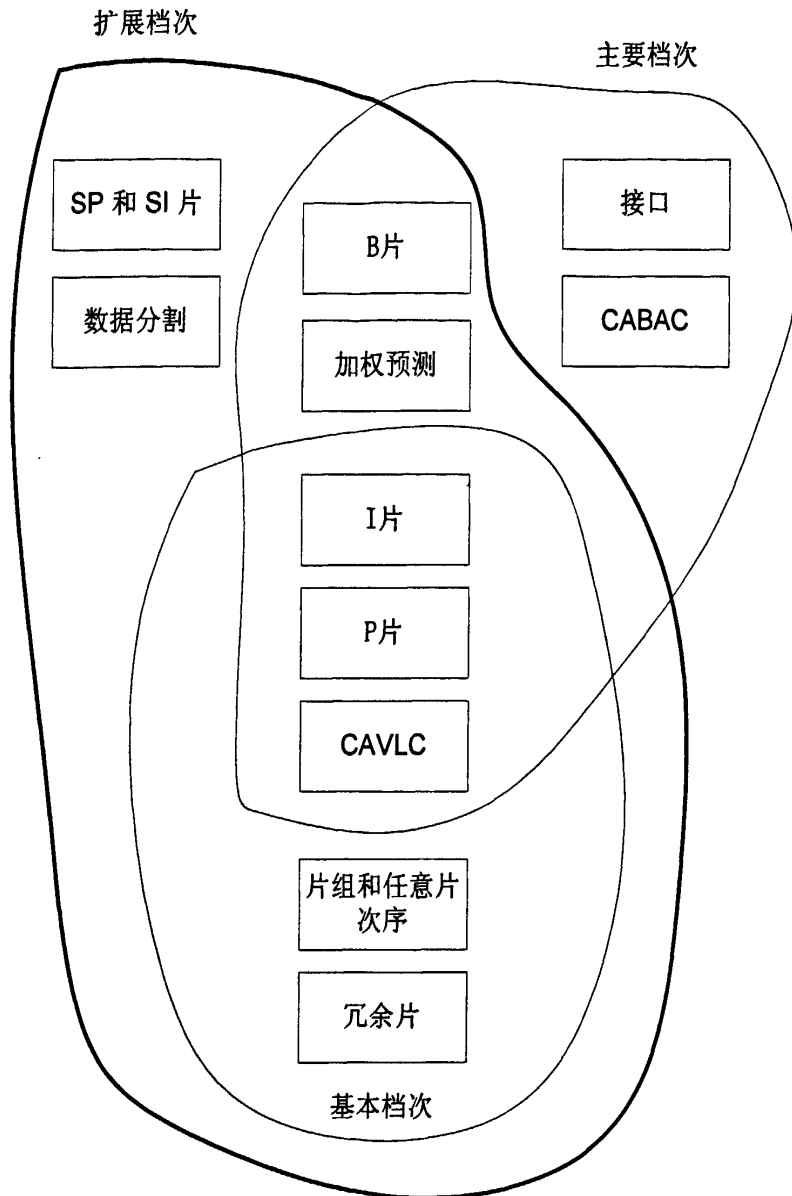


图 2-1 H.264 档次

Fig. 2-1 H.264 level

### 2.1.2 H.264 分层结构

H.264 应用广泛，主要包括视频会议、数字电视、交互媒体、有线电视远程监控、流媒体服务和视频点播等。为解决不同应用中网络传输的差异，H.264 编码算法定义了两层：视频编码层(Video Coding Layer, VCL)和网络适配层(Network Abstraction Layer, NAL)。分层结构如图 2-2 所示。

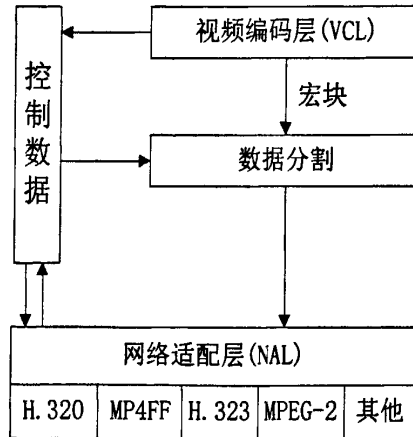


图 2-2 H.264 的分层结构

Fig. 2-2 Hierarchical structure of H.264

### (1) VCL 层

VCL 层<sup>[8]</sup>负责高效的视频内容表示，采用基于块的编码算法，编码算法更加灵活，且加入了一系列新的编码方法来提高编码效率。在 VCL 中，运动补偿的过程采用可变形状和尺寸的块进行预测，运动矢量采用更为精细的像素，支持小数取样内插滤波和两种类型的 S 帧(SP/SI)，且在编码器预测环路采用去方块效应滤波<sup>[15]</sup>。

### (2) NAL 层

NAL 层<sup>[9]</sup>负责使用下层网络的分段格式，以网络所要求的恰当的方式封装 VCL 数据，使得 VCL 层中的内容能够简单有效地应用于各种系统中，提供网络友好性。具体可把 VCL 层的数据封装成 RTP/TP 协议、ISO/MP4 存储格式、H32X 协议和 MPEG-2 系统等传输层。

## 2.2 H.264 编解码原理

H.264 没有明确规定如何实现编解码，而是规定了一个编了码的视频比特流的句法和该比特流的解码方法，它和以前标准中的各功能块在细节上有所不同。由于视频内容的时刻变化和恶劣环境下信道的多变，要求采取相应的自适应技术来满足内容的多样性并对抗信道畸变带来的不良影响。这两方面的多变性带来了自适应压缩技术的复杂性。H.264 就是利用实现的复杂性获得更好的压缩性能。

### 2.2.1 H.264 编码器原理

编码器采用变换和预测的混合编码法，输入的帧或场  $F_n$  以宏块为单位进行编

码，具体过程如图 2-3 所示<sup>[10]</sup>。关键点如下所示：

- (1) 预测编码包括帧内预测和帧间预测两种方式；
- (2) 预测值 PRED 由已编码参考块  $F'_{n-1}$  经运动补偿所得；
- (3) 残差块  $D'_n$  等于预测值 PRED 减去当前块；
- (4)  $D'_n$  经整数变换、量化、熵编码的信息与解码所需信息组成压缩后的码流；
- (5) 参考图像由图像重建功能获得；具体为  $X$  经反量化、反变化得到  $D'_n$  与预测值 PRED 相加，再经环路滤波器滤波去噪，得到具有较高图像质量的重建图像，即为下一帧编码的参考图像。

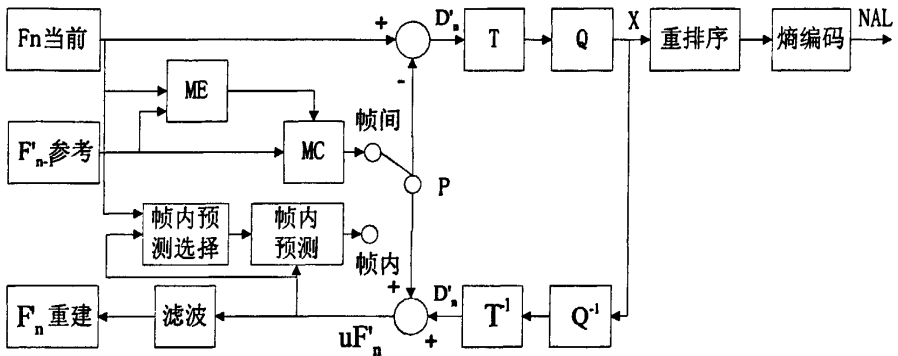


图 2-3 H.264 基本编码框图

Fig. 2-3 Basic encoding block diagram of H.264

## 2.2.2 H.264 解码器原理

由图 2-3 可知，编码器经 NAL 输出 H.264 压缩比特流。根据图 2-4 中的解码器框图，解码流程如下<sup>[10]</sup>：

- 首先，对码流进行熵解码得到量化系数  $X$ ；
- 其次，量化系数经过反量化和反变换得到残差数据  $D'_n$ ；
- 再次，解码器使用从码流中解码得到的头信息创建预测块 PRED，并与  $D'_n$  求和得到图像块的数据  $uF'_n$ ；
- 最后， $uF'_n$  通过去方块滤波得到重建图像的解码块  $F'_n$ 。

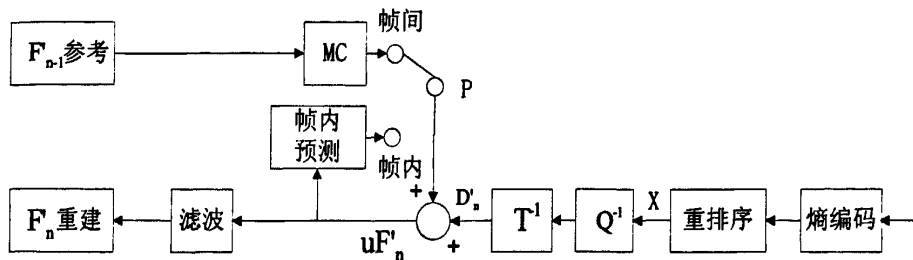


图 2-4 H.264 基本解码框图

Fig. 2-4 Basic decoding block diagram of H.264

## 2.3 H.264 优势和关键技术

### 2.3.1 H.264 优势

H.264 具有以下优势:

(1) 码率低; 在同等图像质量下, H.264 压缩后的数据量只有 MPEG-2 的  $1/2 \sim 1/3$ , 较大程度节省存储容量、信道带宽及用户的下载时间和数据流量收费。

(2) 图像质量高; H.264 能够以较小码率, 在较低带宽上提供高质量的图像传输。

(3) 应用范围广; H.264 的不同 Profile 既可以应用于有严格时延限制的实时通信中, 也可以应用于对时延要求不高的其他应用中。

(4) 较少的编码选项; H.264 力求“回归基本”, 不需要像 H.263 一样设置很多选项, 降低了编码复杂度。

(5) 鲁棒性; H.264 针对分组丢失和比特误码提出了相应的工具, 使得 H.264 在分组交换网和无线网络中传播时具有更强的抗误码性能。

(6) 网络友好性; H.264 中增加了 NAL 层, 负责将编码器的输出码流适配到各种类型的网络中, 从而提供了友好的网络接口。

### 2.3.2 H.264 关键技术

H.264 之所以能够达到以上性能, 是因为采用了以下先进技术。

#### (1) 帧内预测<sup>[11]</sup>

H.264 采用新的帧内预测模式<sup>[11]</sup>, 利用相邻像素的相关性, 通过当前像素块的已编码重建的左边和上边像素进行预测, 最大程度减少图像的空间冗余。亮度块有 9 种  $4 \times 4$  块和 4 种  $16 \times 16$  块的预测模式。

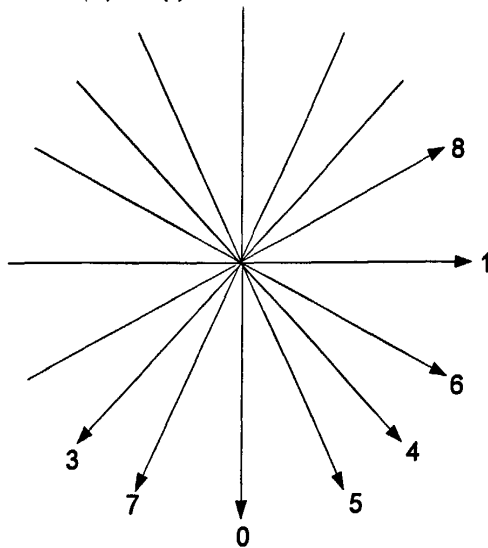
1) 4×4亮度预测模式

除了边缘块特别处置外，对于4×4块，每个像素都可用所在块的左上角的17个像素进行加权来预测，即进行空间预测。如图2-5(a)所示，A、B.....P是已编码像素，a、b.....p为16个待预测的像素点。

4×4帧内预测的基本思路<sup>[12]</sup>是从不同的方向计算、比较块中各个像素之间的亮度差值，即梯度值，选择最小预测误差的方向作为最佳预测方向。图2-5列出了亮度预测的9类不同模式。图2-5(b)箭头标明每种模式的预测方向，表2-1对9种模式进行了描述。

Q	A	B	C	D	E	F	G	H
I	a	b	c	d				
J	e	f	g	h				
K	i	j	k	l				
L	m	n	o	p				
M								
N								
O								
P								

(a) 预测像素点



(b) 预测方向

图 2-5 4×4 亮度预测模式

Fig. 2-5 4×4 luminance prediction modes

表 2-1 预测模式描述

TABLE 2-1 Description of prediction modes

模式	详述
垂直: 模式 0	由 A,B,C,D 垂直推出相应像素值
水平: 模式 1	由 I,J,K,L 水平推出相应像素值
DC: 模式 2	由 A 到 D 及 I 到 L 平均推出相应像素值
下左对角线: 模式 3	由 $45^\circ$ 方向的像素内插得出相应像素值
下右对角线: 模式 4	由 $45^\circ$ 方向的像素内插得出相应像素值
右垂直: 模式 5	由 $26.6^\circ$ 方向的像素内插得出相应像素值
下水平: 模式 6	由 $26.6^\circ$ 方向的像素内插得出相应像素值
左垂直: 模式 7	由 $26.6^\circ$ 方向的像素内插得出相应像素值
上水平: 模式 8	由 $26.6^\circ$ 方向的像素内插得出相应像素值

2)  $16 \times 16$  亮度预测模式

$16 \times 16$  预测有垂直、水平、直流和平面预测四种方式, 如图 2-6 和表 2-2 所示。

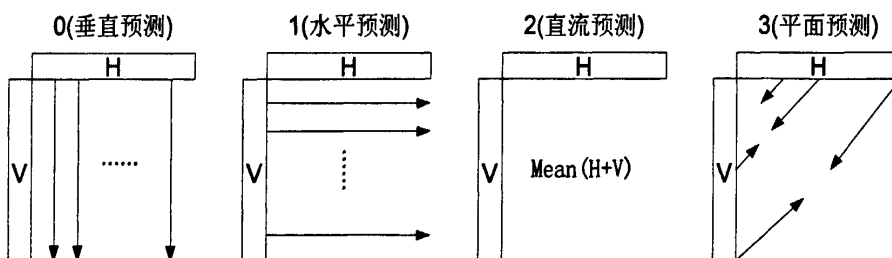
图 2-6  $16 \times 16$  亮度预测模式Fig. 2-6  $16 \times 16$  luminance prediction modes

表 2-2 预测模式描述

TABLE 2-2 Description of prediction modes

模式	详述
垂直: 模式 0	由上边的像素得出相应像素值
水平: 模式 1	由左边的像素得出相应像素值
DC: 模式 2	由上边和左边像素的平均值得出相应像素值
平面: 模式 3	利用线形函数及左边和上边的像素得出相应像素值

(2) 帧间预测<sup>[13]</sup>

H.264 帧间预测编码利用连续帧中的时间冗余来进行运动估计和补偿, 其运动估计和补偿有以下 4 个特性:

## 1) 不同大小和形状的宏块分割

H.264 采用不同大小和形状的宏块分割和亚分割方法<sup>[13]</sup>, 其划分方法有  $16 \times 16$ 、 $16 \times 8$ 、 $8 \times 16$  和  $8 \times 8$  四种。当选用  $8 \times 8$  方式时, 可按照  $8 \times 8$ 、 $8 \times 4$ 、 $4 \times 8$  和  $4 \times 4$  进行亚分割, 如图 2-7 所示<sup>[14]</sup>。由于每个划分区域包含自己的运动矢量, 且和区域选取信息一起通过进行编码传输, 选用区域的大小亦有所不同, 具体如下:

a) 选用小区域: 残差减少, 预测更精确, 但用于表示区域选取和运动向量的数据量增大;

b) 选用大区域: 用于表示区域选取和运动向量的数据量减少, 运动补偿后的残差增大。

总之, 小区域适合表现细节部分, 大区域适合反映同质部分。

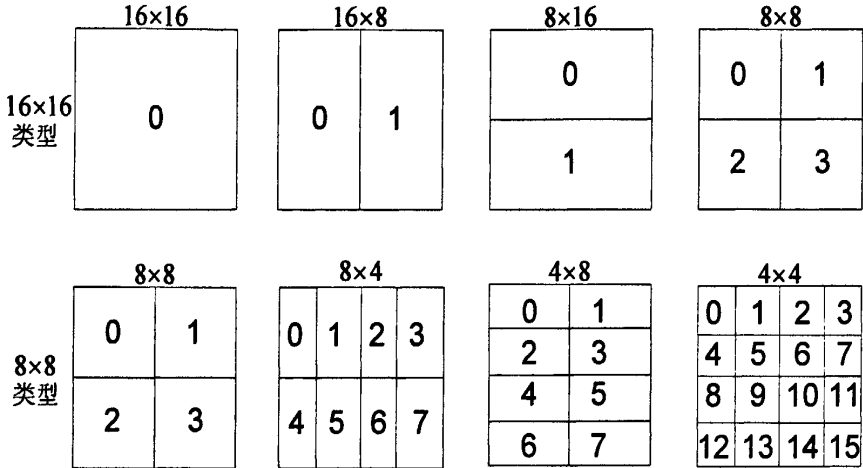


图 2-7 宏块划分方式

Fig. 2-7 Macroblock partition modes

## 2) 高精度的亚像素运动补偿

H.264 中可以采用 1/4 或者 1/8 像素精度的运动估值, 相同精度下, 比半像素精度运动估计后的残差小, 所需的码率更小。

## 3) 多帧预测

H.264 提供可选的多帧预测功能, 在帧间编码时, 可选 5 个不同的参考帧, 提供了更好的纠错性能, 进一步改善视频图像质量。

## (3) 整数变换与量化<sup>[7]</sup>

H.264 采用整数变换, 并将变换和量化中的乘法运算合并起来, 减少运算量的同时, 提高了压缩实时性。H.264 中整数变换及量化具体过程如图 2-8 所示<sup>[16]</sup>。

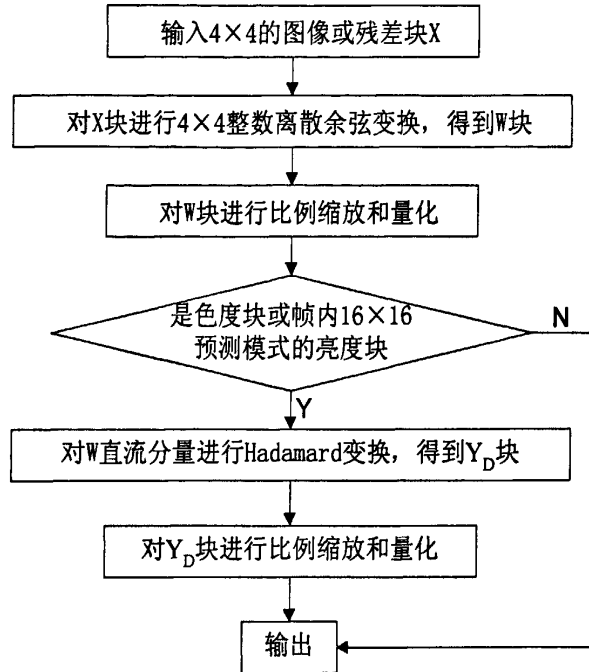


图 2-8 H.264 的整数变换和量化流程图

Fig. 2-8 Integer transform and quantization of H.264

1) 整数变换

在 H.264 标准中，其基本配置根据要编码的残差数据类型使用三种变换<sup>[17]</sup>：

- a) 对(16×16模式中预测的)4×4内部宏块的直流系数亮度分量的变换，变化矩阵如图 2-9(a)；
- b) 对(在任何宏块中) 2×2的直流系数的色度分量的变换，变换矩阵如图 2-9(b)；
- c) 对所有其他的 4×4 块的残差数据的变换，变换矩阵如图 2-9(c)所示。

如果使用可选择的“自适应大小变换模式”，根据运动补偿块大小(4×8、8×4、8×8、16×8等)选择进一步的变换。

$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$
(a) 普通残差 变换矩阵	(b) 色度块直流 变换矩阵	(c) 亮度块直流 变换矩阵

图 2-9 H.264 的三种整数变换矩阵

Fig. 2-9 Integer transform matrices of H.264

2) 量化

H.264 采用标量量化技术，将每个图像样点编码映射成较小的数值。标量量化



器的基本原理如下:

$$FQ = \text{round}\left(\frac{y}{QP}\right) \quad (2-1)$$

其中,  $FQ$  为  $y$  的量化值,  $\text{round}()$  为取整函数,  $y$  为输入样本点编码,  $QP$  为量化步长。反量化为:

$$y' = FQ \cdot QP \quad (2-2)$$

在量化和反量化过程中,  $QP$  的大小能反映编码压缩率及图像精度。由公式(2-2)可知, 若  $QP$  比较大, 则量化值  $FQ$  动态范围较小, 其相应的编码长度较小, 但图像细节信息损失较多; 若  $QP$  较小, 则  $FQ$  动态范围较大, 相应的编码长度也较大, 图像细节信息损失较少。最佳的  $QP$  值能够很好的权衡编码长度和图像精度, 以达到最佳的整体效果。

在 H.264 中, 量化步长  $Q_{step}$  共有 52 个值<sup>[18]</sup>, 即 0-51。如表 2-3 所示。其中  $QP$  为量化参数, 每个  $QP$  对应一个量化步长, 两者的对应关系为:  $QP$  每增加 6,  $Q_{step}$  增加一倍。 $QP$  越小, 量化越精细。对于色度编码, 一般使用与亮度编码同样的量化步长。在 H.264 中, 为了避免在较高量化步长时出现颜色量化人工效应, 色度的  $QP$  最大值大约限制在亮度  $QP$  最大值的 80% 范围内, 亮度  $QP$  的最大值是 51, 则色度  $QP$  的最大值是 39。

表 2-3 H.264 中编解码器的量化步长

TABLE 2-3 Quantization step of H.264 CODEC

QP	Qstep	QP	Qstep	QP	Qstep	QP	Qstep	QP	Qstep
0	0.625	12	2.5	24	10	36	40	48	160
1	0.6875	13	2.75	25	11	37	44	49	176
2	0.8125	14	3.25	26	13	38	52	50	208
3	0.875	15	3.5	27	14	39	56	51	224
4	1	16	4	28	16	40	64		
5	1.125	17	4.5	29	18	41	72		
6	1.25	18	5	30	20	42	80		
7	1.375	19	5.5	31	22	43	88		
8	1.625	20	6.5	32	26	44	104		
9	1.75	21	7	33	28	45	112		
10	2	22	8	34	32	46	128		
11	2.25	23	9	35	36	47	144		

在 H.264 中, 具体量化过程如下所示:

$$Z_{ij} = \text{round} \left( \frac{Y_{ij}}{Qstep} \right) \quad (2-3)$$

其中,  $Z_{ij}$  是输出的量化系数,  $Y_{ij}$  是矩阵  $Y$  中的转换系数,  $Qstep$  是量化步长。  
H.264 量化过程和变换中的乘法运算相结合, 表述如下:

$$Z_{ij} = \text{round} \left( W_{ij} \frac{PF}{Qstep} \right) \quad (2-4)$$

其中,  $W_{ij}$  是矩阵  $W$  中的转换系数,  $PF$  是矩阵  $Ef$  中的元素, 根据样本点在图像块中的位置  $(i, j)$ ,  $PF$  可以是  $a^2$ 、 $b^2/4$  或  $ab/4$ ,  $PF$  值见表 2-4。

表 2-4 不同  $(i, j)$  对应的  $PF$

TABLE 2-4 PF values

位置	PF
(0, 0), (2, 0), (0, 2), (2, 2)	$a^2$
(1, 1), (1, 3), (3, 1), (3, 3)	$b^2/4$
其他	$ab/4$

利用量化步长随量化参数每增 6 而增加一倍的性质, 可进一步简化计算。设:

$$qbits = 15 + \text{floor}(QP/6) \quad (2-5)$$

$$MF = \frac{PF}{Qstep} 2^{qbits} \quad (2-6)$$

则公式(2-4)可以写成:

$$Z_{ij} = \text{round} \left( W_{ij} \frac{MF}{2^{qbits}} \right) \quad (2-7)$$

$MF$  可取整数, 表 2-5 是对应  $QP$  值为 0 到 5 的  $MF$  值, 当  $QP$  大于 5 时,  $qbits$  随  $QP$  每增加 6 而增加 1, 对应的  $MF$  值不变。

表 2-5 H.264 中  $MF$  值

TABLE 2-5 MF values of H.264

QP	位置(0,0), (2,0), (2,2), (0,2)	位置(1,1), (1,3), (3,1), (3,3)	其它位置
0	13107	5243	8066
1	11916	4660	7490
2	10082	4194	6554
3	9362	3647	5825
4	8192	3355	5243
5	7282	2893	4559

由式(2-5)、(2-6)、(2-7)可得具体量化过程为:

$$|Z_{ij}| = (|W_{ij}| \cdot MF + f) \gg qbits \quad (2-8)$$

$$\text{sign}(Z_{ij}) = \text{sign}(W_{ij}) \quad (2-9)$$

其中,  $\text{sign}()$  为符号函数; “ $\gg$ ” 为右移运算, 右移一次完成整数除以 2;  $f$  为偏移量, 它的作用是改善恢复图像的视觉效果, 在参考模型软件的帧内块和帧间块中,  $f$  分别  $2^{qbits/3}$  和  $2^{qbits/6}$ 。

## 2.4 本章小结

本章对 H.264 国际视频编码标准进行了分析概述。

首先,介绍了 H.264 的档次和分层结构,并进一步阐述了各档次的应用范围及 VCL 层和 NAL 层的特点。

其次,给出了 H.264 编解码原理,分别分析了编码器和解码器整体过程。

最后,详细阐述了 H.264 较以往视频编码标准的优势,深刻剖析了产生该优势的 H.264 关键技术。

### 3 基于 HVS 的宏块级自适应频率加权算法研究

#### 3.1 HVS 特性分析

由于频率加权技术来源于 HVS，频率加权策略选择要统筹考虑 HVS 对不同内容的敏感度和编码区域内容特征，以满足人类视觉特性。本文在 HVS 的多通道及视觉阈值特性<sup>[19]</sup>、视觉掩盖效应<sup>[20]</sup>、平滑跟踪眼睛移动效应<sup>[21]</sup>等特性的基础上总结了以下和视频压缩相关的人类视觉系统特性：

- (1) 低频部分有较高敏感度<sup>[22]</sup>；
- (2) 静止区域的运动部分有较高敏感度；
- (3) 静止区域的纹理细节部分和平滑运动区域有较高敏感度；
- (4) 剧烈运动区域敏感度较低。

#### 3.2 基于 HVS 的宏块级自适应频率加权算法

##### 3.2.1 基于 HVS 的 MBAFW 算法框架

H.264 HP 中的 PAFW 算法，量化矩阵逐帧更新。频率加权参数可从图像头信息中获得，一帧中所有宏块均采用相同的频率加权矩阵。但是，视频一帧中图像的内容有很大不同，只用帧级自适应频率加权是不够的。若频率加权扩展到宏块级，采用宏块级自适应频率加权(Macroblock-level Adaptive Frequency Weighting, MBAFW)算法，即同一帧图像的不同区域可根据图像内容特征选择不同的频率加权策略。理论上，宏块级自适应频率加权可达到更好的编码效率。利用 3.1 节 HVS 特性和相邻宏块相关信息，视频图像不同频率部分采用不同的频率加权矩阵和相应的频率加权策略。算法整体框架如下：

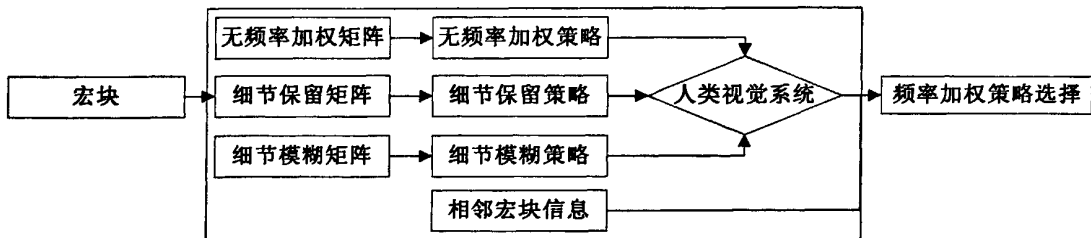


图 3-1 MBAFW 算法整体框架图

Fig. 3-1 Overall diagram of MBAFW algorithm

### 3.2.2 基于 HVS 的 MBAFW 算法关键问题

本章在 H.264 HP 基础上，通过分析 H.264 HP 的具体量化过程及 PAFW 算法特性，总结频率加权策略的核心思想，并用于本章的 MBAFW 算法。

#### (1) 频率加权矩阵和频率加权策略

##### 1) 频率加权矩阵

频率加权<sup>[23]</sup>概念来自于 HVS 的频率敏感特性，此概念相继用于图像编码标准 JPEG、MPEG2，之后扩展到 H.264 HP。在 H.264 HP 中，针对 4×4 和 8×8 两种不同大小的块定义了相应的量化矩阵，量化过程<sup>[24]</sup>如下所示：

$$Xq_{ij} = \text{sign}\{X_{ij}\}(|X_{ij}|A(QP, i, j) + f^{2^{17+QP}}) \gg (17 + QP/6) \quad (3-1)$$

反量化过程如下：

$$Xr_{ij} = Xq_{ij} B(QP, i, j) \ll QP/6 \quad (3-2)$$

其中， $Xq_{ij}$  是量化后的系数， $X_{ij}$  是 DCT 变换系数， $QP$  是量化参数。 $A(QP, i, j)$  和  $B(QP, i, j)$  是量化缩放因子， $f$  是量化取整。缩放因子的计算如下：

$$A(QP, i, j) = WS(i, j) \times \text{Norm}(QP, i, j) \quad (3-3)$$

$$B(QP, i, j) = WS(i, j) \times \text{Norm}(QP, i, j) \quad (3-4)$$

其中， $\text{Norm}(QP, i, j)$  是缩放调节参数， $WS(i, j)$  是量化矩阵。

H.264 HP 中频率加权可通过调节公式(3-3)(3-4)量化矩阵中不同频带的值来实现，且量化矩阵可以是默认的也可以是预定义的。

##### ● 4×4 块

根据 HVS 特性，在 H.264 HP 基础上，本文预定义了三类量化矩阵以实现 MBAFW：无频率加权矩阵(QM<sub>0</sub>)，细节保留矩阵(QM<sub>1</sub>)，细节模糊矩阵(QM<sub>2</sub>)。

无频率加权矩阵如图 1(a)所示，对所有频率系数采用相同的量化参数，默认为 16；

细节保留矩阵如图 1(b)所示，由于高频区域往往代表图像细节部分，细节保留矩阵对高频系数采用较小的量化参数，以达到细节保留的效果；

细节模糊矩阵如图 1(c)所示，和细节保留矩阵相反，细节模糊矩阵对高频系数采用较大的量化参数，以达到更好的压缩比。

$\begin{bmatrix} 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 \end{bmatrix}$	$\begin{bmatrix} 25 & 19 & 13 & 11 \\ 19 & 13 & 11 & 10 \\ 13 & 11 & 10 & 9 \\ 11 & 10 & 9 & 8 \end{bmatrix}$	$\begin{bmatrix} 6 & 13 & 20 & 28 \\ 13 & 20 & 28 & 32 \\ 20 & 28 & 32 & 37 \\ 28 & 32 & 37 & 42 \end{bmatrix}$
(a) 无频率加权矩阵	(b) 细节保留矩阵	(c) 细节模糊矩阵

图 3-2 4×4 块频率加权矩阵

Fig. 3-2 Frequency weighting matrices for 4×4

● 8×8 块

在 H.264 HP 中，8×8 块默认量化矩阵中各元素的值均为 16，本文算法中的无频率加权矩阵采用标准默认矩阵，如图 3-3(a)所示；细节保留矩阵和细节模糊矩阵采用文献[24]在多次试验基础上提出的最优加权矩阵，如图 3-3(b)。

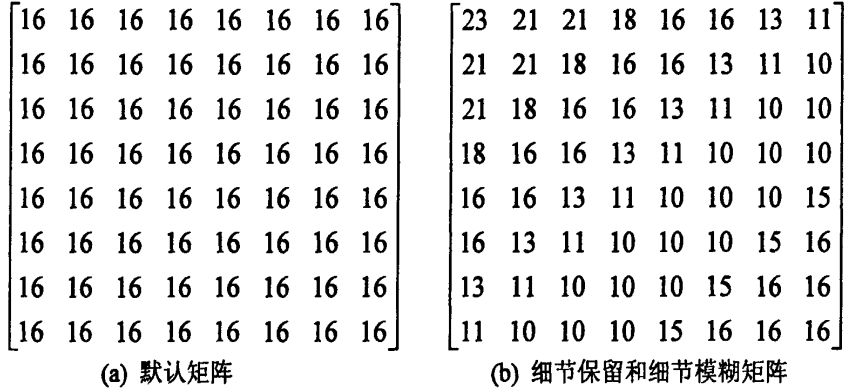


图 3-3 8×8 块频率加权矩阵

Fig. 3-3 Frequency weighting matrices for 8×8

2) 频率加权策略

在 JM10.1 中，量化过程如下：

$$Xq_{ij} = (|X_{ij}| \times scale(i, j) + f) \gg 15 + floor(QP/6) \quad (3-5)$$

其中，f 为偏移量，它的作用是改善恢复图像的视觉效果， $floor()$  为取整函数（其输出为不大于输入实数的最大整数），QP 为量化参数， $scale(i, j)$  为缩放因子，不同的频率加权矩阵对应不同缩放因子。本文针对图 4 和图 5 提出的频率加权矩阵定义了三种宏块级频率加权策略。频率加权过程如下：

策略 0：无频率加权策略，所有频率系数均采用相同的量化参数，其中  $MF(i, j)$  为 H.264 中每个 QP (0-51) 所对应的值， $QM_0$  为无频率加权矩阵；

$$scale(i, j) = (MF(i, j) \gg 4) / QM_0 \quad (3-6)$$

策略 1：细节保留策略，保留更多高频系数和图像细节部分，具体过程如(13)所示，式中  $QM_1$  为细节保留矩阵；

$$scale(i, j) = (MF(i, j) \gg 4) / QM_1 \quad (3-7)$$

策略 2：细节模糊策略，更多的高频系数被粗略量化为 0，细节被模糊，具体过程如(13)所示，其中  $QM_2$  为细节模糊矩阵；

$$scale(i, j) = (MF(i, j) \gg 4) / QM_2 \quad (3-8)$$

由上可知，策略 1 用更多的比特达到较高的视频质量；策略 2 在不损坏视频主观质量的情况下降低了比特率。

(2) 相邻宏块信息的提取

1) 相邻宏块定义

在 H.264 HP 中，相邻宏块包括编码宏块左方、上方、左上方、右上方的宏块，具体位置如下，其中，E 为当前编码宏块，A、B、C、D 为相邻宏块。

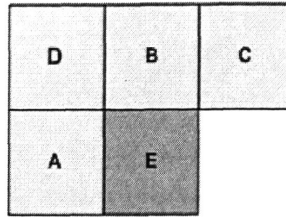


图 3-4 相邻宏块定义

Fig. 3-4 Neighboring macroblock definition

### 2) 相邻宏块相关信息

传统的图像分析研究中，若利用 HVS 特性，常常需要从图像数据中提取一些额外的数据特征。数据特征提取过程会大大增加计算复杂度，对当前快速图像处理的应用造成很大的障碍。实际上，在编码过程中，图像的边信息，如相邻宏块的划分方式、预测模式以及运动矢量都是已知的，且从某种程度上反映了当前编码区域特点。本文提出的算法依据相邻宏块的划分方式和预测模式来确定选用哪种频率加权策略。

H.264 帧内帧间预测模式如表 3-1 所示：

表 3-1 H.264 帧内帧间预测模式

TABLE 3-1 Intra and inter prediction modes of H.264

H.264 帧内预测模式		H.264 帧间预测模式
亮度分量	色度分量	
Intra_4×4	Intra_8×8	P16×16, P16×8, P8×16, P8×8, PS8×8, PS8×4, PS4×8, PS4×4
Intra_16×16		

表 3-1 中，Intra\_4×4 预测模式提供 9 种可选的方式，Intra\_16×16 预测模式和 Intra\_8×8 预测模式提供 4 种可选的方式；H.264 帧间预测支持 7 种大小和形状的块：16×16、16×8、8×16、8×8、8×4、4×8、4×4，根据采用块大小的不同，编码模式分为表 1 所示种类。

### 3.2.3 基于 HVS 的 MBAFW 算法频率加权策略选择规则

基于 HVS 特性，定义了频率加权策略选择规则。编码时，根据相邻宏块信息唯一确定频率加权策略。具体步骤如下：

步骤 1：如果相邻宏块 ABCD 有一个不可得，则当前编码区域为图像边缘部分，依据 HVS 特性，人眼对图像轮廓或边缘的量化误差不敏感，采用策略 2；

步骤 2: 判断当前编码帧的类型, 针对 I、P、B 采用不同的策略选择算法。

(1) I 帧

对于帧内编码, 细节比较丰富的区域, 各相邻宏块采用不同的预测模式; 有相同特征的区域通常采用相同的预测模式。由于 I 帧对整个视频序列的主观质量起着至关重要的作用, 更多 I 帧宏块采用细节保留策略。对于 I 帧, 在没有更好的预测模式下, 通常采用 Intra\_4x4 模式, 采用该模式的区域一般含有丰富的细节。I 帧的策略选择规则如下:

- 1) 如果相邻宏块 ABCD 中有一个是 Intra\_4x4 模式, 则可预知当前编码区域含有较丰富细节, 采用策略 1;
- 2) 否则如果 A 和 D 或 B 和 C 有相同的预测模式, 则可预知该区域具有较相似的内容特征, 细节不丰富, 采用策略 2;
- 3) 否则如果 A 和 D 或 B 和 D 是 I\_16x16 模式, 由于采用较大的划分方式, 细节不丰富或运动比较缓慢, 采用策略 0;
- 4) 否则, 采用策略 1。

流程如图 3-5。

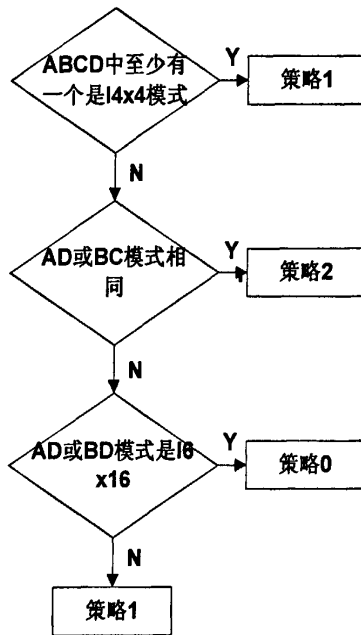


图 3-5 I 帧策略选择算法流程图

Fig. 3-5 Flow chart of strategy selection algorithm for I frame

(2) P 帧

对于帧间编码, 宏块的划分大小可以表示编码区域的特点。细节比较丰富的区域通常采用较小的划分方式; 运动缓慢的背景区域采用较大的划分方式。对于 P 帧, 策略 0 是默认策略。如果某宏块采用帧内宏块模式或帧间 4x4 宏块模式, 则



该宏块很可能是运动较剧烈的部分；若采用 P\_SKIP 模式或 P\_16×16 模式，则该宏块通常是背景区域或较平坦区域。考虑到背景区域的细节较敏感，这些区域通常不采用细节模糊策略。P 帧的策略选择规则如下：

- 1) 如果相邻的四个宏块有一个是帧内宏块模式，采用策略 1；
- 2) 否则如果相邻宏块有一个是 inter\_4×4 模式，采用策略 1；
- 3) 否则如果 A 和 D 或 B 和 D 是 P\_SKIP 模式，采用策略 0；
- 4) 否则如果 A 和 D 或 B 和 D 是 P\_16×16 模式，采用策略 2；
- 5) 否则，采用策略 0。

流程如图 3-6。

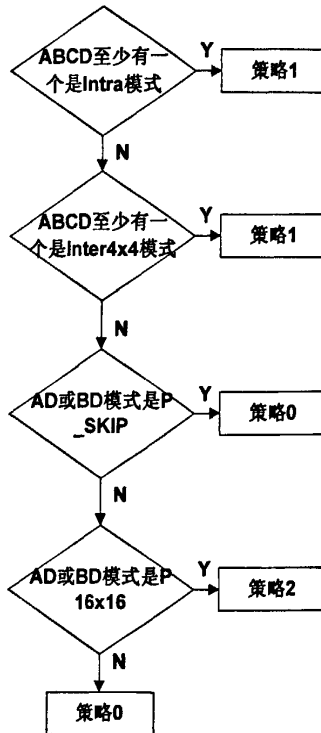


图 3-6 P 帧策略选择算法流程图

Fig. 3-6 Flow chart of strategy selection algorithm for P frame

### (3) B 帧

对于 B 帧编码，大部分宏块采用 B\_SKIP 模式。B 帧的策略选择规则如下：

- 1) 如果 A 和 D 或 B 和 D 是 B\_SKIP 模式，采用策略 0；
- 2) 否则如果 A 和 D 或 B 和 D 是 P\_16×16 模式，采用策略 2；
- 3) 否则，采用策略 1。

流程如图 3-7。

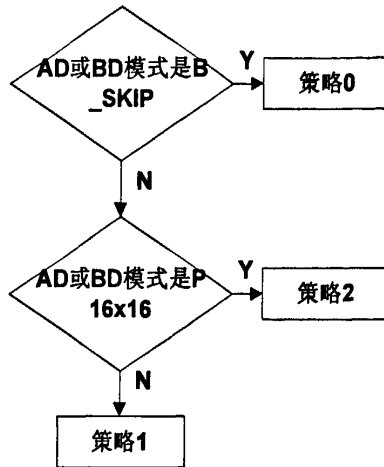


图 3-7 B 帧策略选择算法流程图

Fig. 3-7 Flow chart of strategy selection algorithm for B frame

### 3.3 仿真结果和分析

本章提出的 MBAFW 算法测试的软件基础是 H.264 HP JM10.1, 该实验允许率失真优化以及去块效应滤波器的应用, 参考帧为 2, 搜索范围设定为 32, 测试视频序列为 QCIF 格式的 City、Salesman 和 Bridge, 视频序列结构为 IBBP。本文分别对帧级频率加权、宏块级频率加权、无频率加权从客观质量和主观质量两个方面进行了对比, 并对宏块级频率加权算法复杂度进行了分析。

#### 3.3.1 仿真结果算法复杂度分析

算法复杂度包括时间复杂度和空间复杂度, 本文对宏块级自适应频率加权算法的时间复杂度进行分析。时间复杂度是度量算法执行的时间长短; 一般情况下, 算法的基本操作重复执行的次数是模块  $n$  的某一个函数  $f(n)$ , 因此, 算法时间复杂度记作:  $T(n) = O(f(n))$ ; 随着模块  $n$  的增大, 算法执行时间的增长率和  $f(n)$  的增长率成正比, 所以  $f(n)$  越小, 算法的时间复杂度越低, 算法的效率越高。MBAFW 算法是量化矩阵策略选择的过程。该算法基本操作的语句执行次数为一阶常数, 经上机测试, 该算法的执行时间较帧级自适应频率加权算法几乎相同。结果表明, 该算法没有引入太多的时间复杂度。MBAFW 算法和 PAFW 算法执行时间如下表所示:

表 3-2 MBAFW 算法和 PAFW 算法平均每帧编码时间比较 (秒/帧)

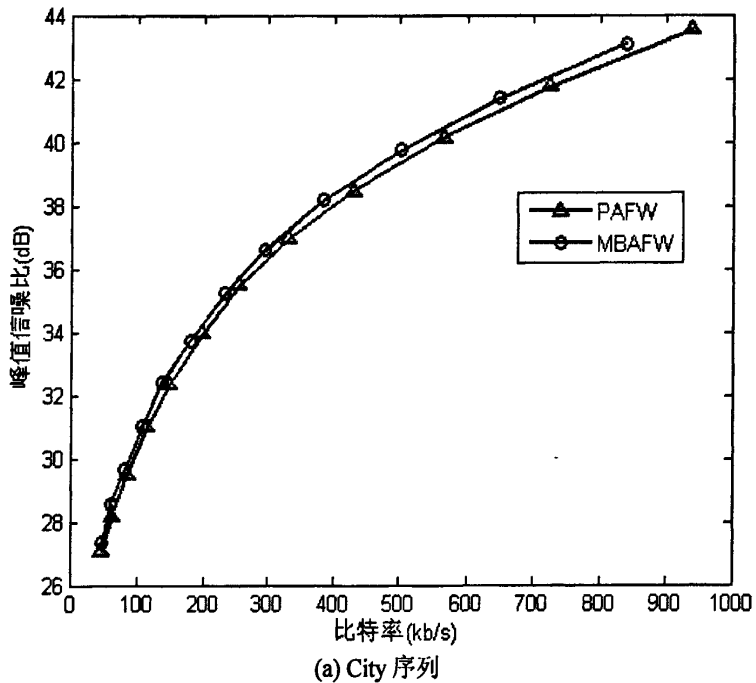
TABLE 3-2 Execution time of the MBAFW and the PAFW

视频序列	MBAFW	PAFW
City(QCIF)	6.574	6.527
Salesman(QCIF)	6.489	6.471
Football(CIF)	27.797	27.737

### 3.3.2 仿真结果客观质量分析

#### (1) PSNR 比较

图 3-8 分别为 City、Salesman 视频序列在 PAFW 算法和 MBAFW 算法下的对比。实验结果表明, 相同比特率情况下, MBAFW 算法的平均峰值信噪比更大; 峰值信噪比相同情况下, MBAFW 算法比 PAFW 算法的比特率低, 可见 MBAFW 算法有更好的峰值信噪比。



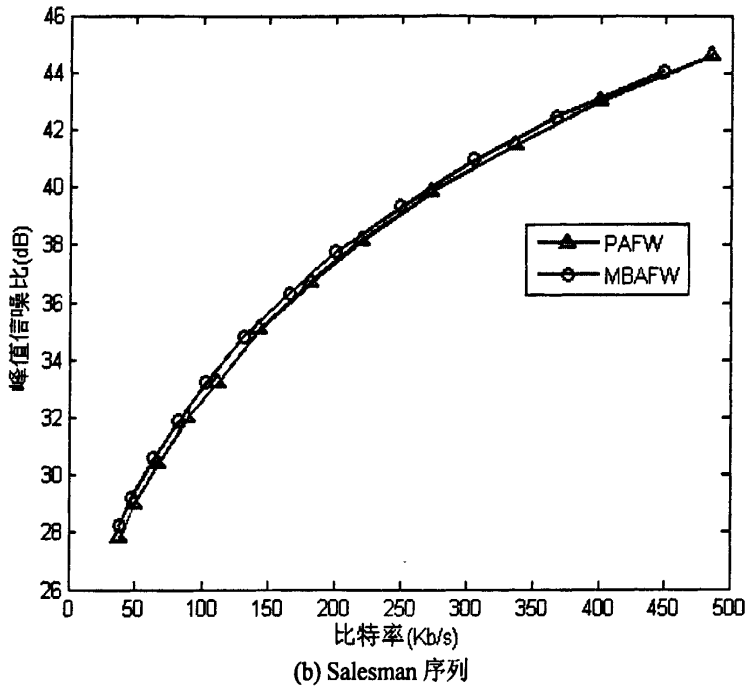


图 3-8 MBAFW 算法和 PAFW 算法的峰值信噪比

Fig. 3-8 The Curves of PSNR versus bitrate for City and Salesman

(2) 比特率比较

一般, 量化参数相同时, 视频序列具有相似的主观质量。和无频率加权(NOFW)相比, MBAFW 算法可大大降低比特率。在图 3-9 中, 当量化参数为 18 时, 该算法较 NOFW 算法降低 11% 的比特率。从下图还可发现, 相同量化参数下, MBAFW 的比特率比 PAFW 稍高。这是因为 MBAFW 算法用更多的比特来进行细节保留。实际上, 若能很好地权衡细节保留参数和细节模糊参数, MBAFW 算法可更大程度降低比特率, 甚至比 PAFW 算法的比特率低。

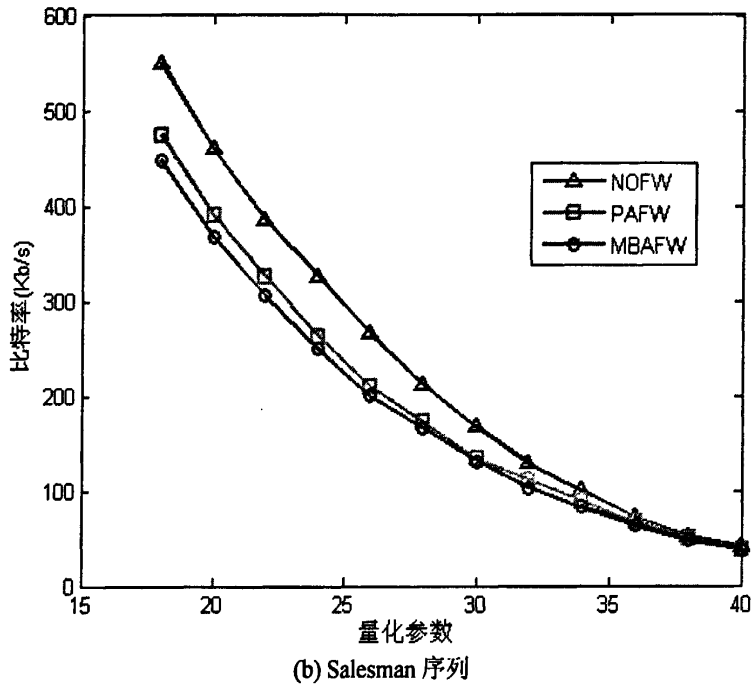
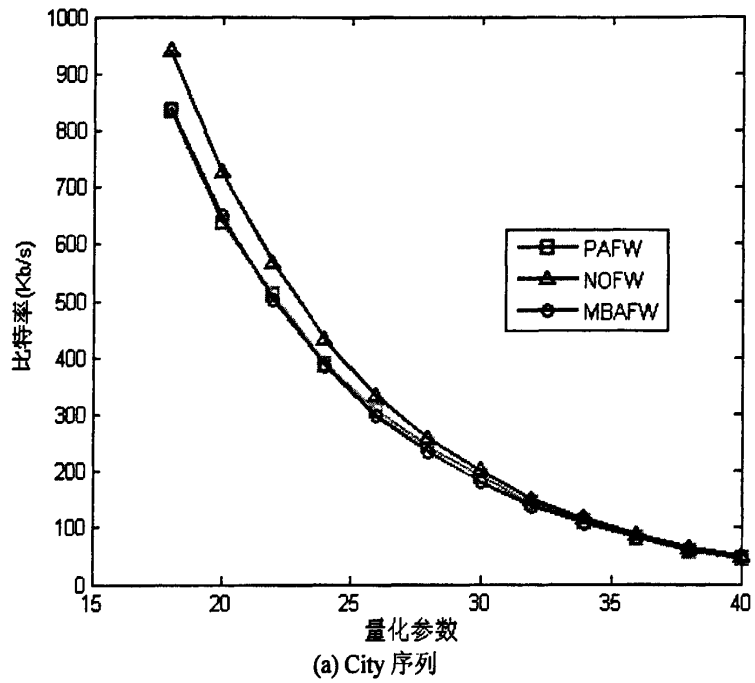


图 3-9 MBAFW 算法和 PAFW 算法的比特率值

Fig. 3-9 Bitrate comparison for City and Salesman

### 3.3.3 仿真结果主观质量分析

图 3-10 为 City 序列在比特率基本相同时，经 PAFW 算法和 MBAFW 算法重

建图像对比。其中, (a)原始图像; (b)量化参数为 35, 比特率为 15.6kb 时, PAFW 算法的重建图像; (c)量化参数为 34, 比特率为 15.3kb 时, MBAFW 算法的重建图像。实验结果显示, PAFW 算法和 MBAFW 算法所达到的主观质量几乎相同, 但细节部分相比, MBAFW 算法比 PAFW 算法更优越。

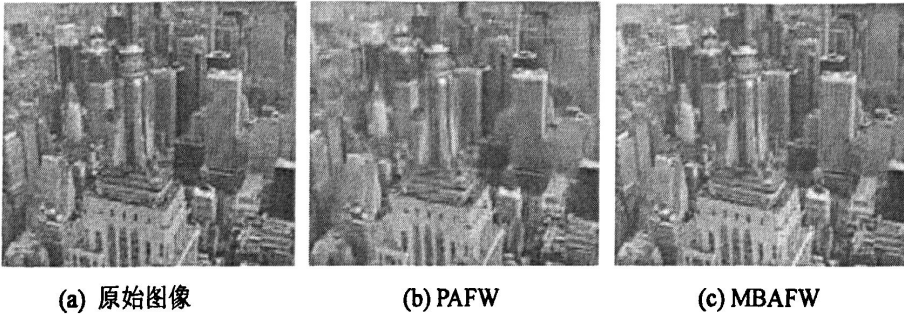


图 3-10 City 序列重建图像的主观质量比较

Fig. 3-10 Reconstructed pictures of PAFW and MBAFW for city

下图 3-11 为 Salesman 序列经 PAFW 算法和 MBAFW 算法重建图像对比。(a)原始图像; (b)量化参数为 34, 比特率为 20.6kb 时, PAFW 算法的重建图像; (c) 量化参数为 35, 比特率为 20.1kb 时, MBAFW 算法的重建图像。实验结果显示, PAFW 算法和 MBAFW 算法所达到的主观质量几乎相同, 细节部分相比, MBAFW 算法比 PAFW 算法更优越。

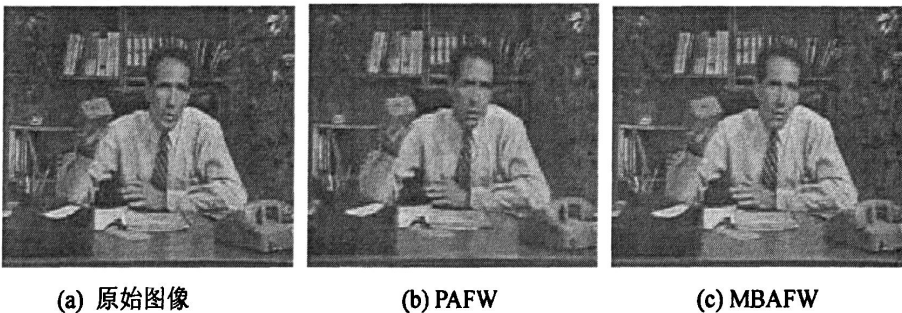


图 3-11 Salesman 序列重建图像的主观质量比较

Fig. 3-11 Reconstructed pictures of PAFW and MBAFW for salesman

下图 3-12 为 Bridge 序列经 PAFW 算法和 MBAFW 算法重建图像对比。(a)原始图像; (b)量化参数为 40, 比特率为 20.6kb 时, PAFW 算法的重建图像; (c)量化参数为 40, 比特率为 20.1kb 时, MBAFW 算法的重建图像。实验结果显示, PAFW 算法和 MBAFW 算法所达到的主观质量几乎相同, 细节部分相比, MBAFW 算法比 PAFW 算法更优越。

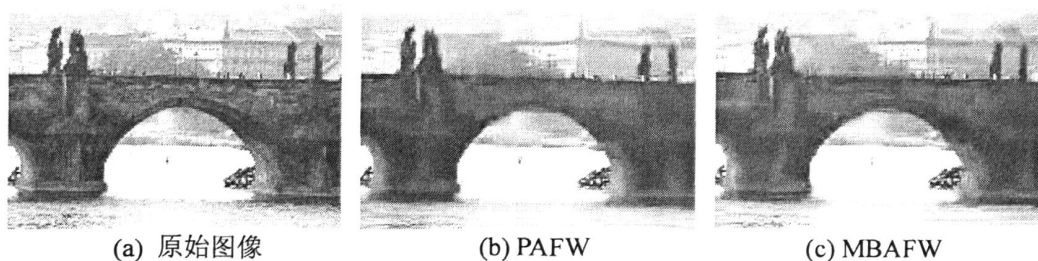


图 3-12 Bridge 序列重建图像的主观质量比较

Fig. 3-12 Reconstructed pictures of PAFW and MBAFW for bridge

### 3.4 本章小结

本章以 H.264 国际视频编码标准为背景，主要对基于 HVS 的自适应频率加权技术展开研究。

首先，在 HVS 的多通道视觉阈值、视觉掩盖和平滑跟踪眼睛移动效应特性的基础上，提炼总结了和视频压缩相关的 HVS 敏感度特性。

其次，在 H.264 HP 的 PAFW 算法的基础上，结合 HVS 敏感度特性，利用视频图像内容特征和编码区域相邻宏块信息，提出了 MBAFW 算法。

该算法针对 I、P、B 帧定义了对应不同量化矩阵的自适应频率加权策略，实现了逐宏块更新量化矩阵，较好地满足了 HVS 特性，对于重建视频序列，在降低几乎相同比特率的情况下，可达到更好的主观质量；相同主观质量时，该算法的压缩效率更好。

## 4 基于 ROI 的宏块级自适应频率加权算法研究

### 4.1 现有 ROI 检测技术分析

随着科学技术的发展,人类的信息载体越来越多样化,图像作为重要的信息载体,一方面给我们信息的传递带来了极大方便,但另一方面其与日俱增的庞大的数据量,给图像的人工分析带来了很大困难。研究者们发现,在图像和视频信息的实际应用中,人们往往只关注图像的某些特定区域,而对特定区域之外的其他部分不太在意。也就是说,由于观察者的主观应用目的不同,使得图像的不同区域为观察者提供信息的比例也不同。这种用户主观选定的影像子区域即为 ROI<sup>[26]</sup>。

近年来,国内外研究者提出了许多 ROI 检测技术,总体上可分为人工检测方法和计算机对图像的自动检测方法。人工检测方法主要包括基于交互的 ROI 检测算法,自动检测算法主要包括<sup>[25]</sup>:

- (1) 基于灰度变化的 ROI 检测技术;
- (2) 基于颜色和熵的检测算法;
- (3) 基于视觉特征的 ROI 检测算法;
- (4) 基于区域生长的 ROI 检测算法。

#### 4.1.1 基于交互的 ROI 检测算法分析

人机交互根据用户对不同图像不同区域有不同要求的特性,由用户选择图像的若干区域作为 ROI 的方法<sup>[26]</sup>。该方法以用户为中心,使永和在检索过程中的作用得到充分发挥,不仅能准备捕获用户的检索目的,检索起来也更加简单高效。该方法适合于医学图像,卫星等内容比较单一的图像库中。

但随着信息社会的发展,表征信息的数据量越来越大,面对海量数据的压力,人工分析就显得力不从心。因此当前的 ROI 检测技术的目的是让计算机快速自动地完成 ROI 提取和索引,ROI 的自动提取方法成为当前研究的热点。

#### 4.1.2 基于灰度变化的 ROI 检测算法分析

Sebe 等人认为图像中灰度变化大的地方是图像的重要部分。小波系数反映了图像在不同尺度上的变化。大尺度下小波系数的绝对值大表明对应的图像区域有



较大的灰度变化。Sebe 等使用非标准的 Haar 小波和 Daubechies4 小波变换<sup>[28]</sup>选取小波系数大(绝对值)的点作为感兴趣点,称这些点为显著点。

该方法优点是:感兴趣点能比较完整的表示图像的内容,而不会像拐点检测器<sup>[27]</sup>只集在纹理区域。

该方法缺点是:只利用图像的灰度信息,与人的观察结果仍有一定的差距。

### 4.1.3 基于颜色的 ROI 检测算法分析

在 ROI 检测中,可利用图像的物理特性提取 ROI 来降低检测的复杂度。在图像的物理特性中,颜色是最基本也是最重要的特征,利用颜色对图像进行 ROI 检测是一种有效的方法。

文献 29 提出了基于颜色信息的检测方法<sup>[29][31]</sup>,把图像中颜色差异较大的区域作为 ROI 区域,该算法中,颜色由 R、G、B 分量表示。

(1) 在水平或垂直方向对图像进行分割<sup>[30]</sup>,如图 4-1 所示;

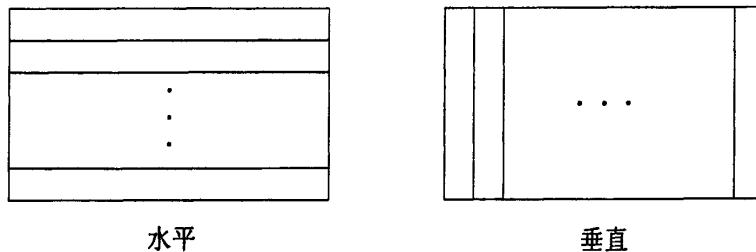


图 4-1 分块模型

Fig. 4-1 Sectional pattern

(2) 计算每个分割后的小区域的像素个数:

$$N_{pixel} = \frac{M \times N}{L} \quad (4-1)$$

其中, M 是图像的高度, N 是图像的宽度, L 是小区域的个数。

(3) 计算每个小区域所有像素的颜色平均值:

$$C_{RA} = \sum_{i=1}^{N_{pixel}} C(i) \quad (4-2)$$

其中,  $C_{RA}$  表示每个区域的颜色平均值,  $C(i)$  表示每个像素的 R、G、B 值;

(4) 根据颜色差异大小选取一半区域作为 ROI, 具体选法如下。

1) 计算各区域之间的颜色差异  $D_{RA}$ , 如下所示:

$$D_{RA} = \left| C_{RA}(i) - \frac{\sum_{j=1}^L C_{RA}(j)}{L} \right| \quad i=1, \dots, L \quad (4-3)$$

2) 按照  $D_{RA}$  的大小选出  $D_{RA}$  最大的  $L/2$  个区域作为 ROI。

$D_{RA}$  的值越大表示此区域的颜色平均值与其他区域的颜色平均值差别越大, 即此区域与其他区域颜色的对比反差越大。

#### 4.1.4 基于视觉特性的 ROI 检测算法分析

人是图像信息的最终评判者, 好的 ROI 检测应当以用户评价为标准; 因此, 越来越多的研究者开始从以图像为核心的分析方法转变到以人类感觉为核心的分析方法, 并把这些研究应用到 ROI 检测技术中, 提出了许多以视觉机制为基础的 ROI 检测方法。下面总结了影响视觉注意的几方面因素, 并在此基础上分析了基于视觉特性的 ROI 检测算法<sup>[32]</sup>。

##### (1) 影响人类视觉注意的因素

目前已经确认的影响视觉注意力的因素主要包括<sup>[32]</sup>:

- 1) 对比度: 与相邻区域具有较高对比度的区域更能吸引人类视觉的注意;
- 2) 形状和运动: 在平等的其他因素下, 狭长形状的区域和运动部分更容易引起人们的视觉注意;
- 3) 位置: 人们在观察图像时, 通常首先都会把注意力定格在图像中央的 25% 的区域范围之内;
- 4) 前景/背景: 人们通常更关注前景对象, 而忽略背景的具体内容;
- 5) 人物: 人的脸部和肢体活动更容易引起人们的注意。

##### (2) 基于视觉注意力的 ROI 检测算法分析

由以上影响人类视觉注意的因素可知, 多方面的合力作用产生了对 ROI 的认知。由于缺乏足够的数据库, 并不能确定各因素在此过程中的具体权重。因此, 该算法的前提假设为: 所有因素之间是相互独立的个体, 互不影响。算法主要步骤如下<sup>[33]</sup>:

- 1) 分割原始图像灰度图像, 得到若干子区域  $Region(i)$ ;
- 2) 计算每个子区域不同因素的重要性, 计算公式分别为:
  - a) 形状因子计算如下:

$$W_2(R_i) = \frac{[Border(R_i)]^p}{Area(R_i)} \quad (4-4)$$

其中  $Border(R_i)$  表示区域边界长度，以像素个数计算。常量  $p=1.75$ ；

b) 位置因子计算如下：

$$W_3(R_i) = \frac{Center(R_i)}{Area(R_i)} \quad (4-5)$$

其中  $Center(R_i)$  表示区域  $R_i$  中处于中央位置的个数，中心像素点 25% 范围内定义为中央位置。

c) 图像中背景区域的检测办法是通过分析区域中有多少比例的像素分布于图像的边界之上。前景/背景因子计算如下：

$$W_4(R_i) = 1 - \min\left(\frac{RBorder(R_i)}{ImageBorder(R_i)}, 1\right) \quad (4-6)$$

其中  $Rborder(R_i)$  表示区域  $R_i$  与图像边界接壤的像素个数， $ImageBorder(R_i)$  表示图像边界的像素个数，计算如下：

$$ImageBorder(R_i) = 2 \times (width + height) - 4 \quad (4-7)$$

其中  $width$  和  $height$  分别表示图像的宽度和高度。

3) 对以上形状因子、位置因子和前景/背景因子采用等权方法进行综合，得到最终的影响权重，如下所示：

$$W(R_i) = \sum_{i=1}^4 W_i^2 \quad (4-8)$$

4) 将 3) 的计算结果归一化。

#### 4.1.5 基于区域生长的 ROI 检测算法分析

区域生长是将具有相似性质的像素集中起来构成一个区域。

首先，选定一个种子点或者一个种子区域作为生长的起始点；

然后，在种子附近搜索与种子点相似特征度满足生长准则的像素，并与种子点所在区域合并，作为新的种子点；

接着，继续以上搜索与合并过程，直到再没有满足条件的像素可以包括进来，一个区域就生长而成了。

因此，区域生长法需要解决以下关键问题：

- 1) 选取种子：选择或确定一组能正确代表所需区域的种子像素；
- 2) 确定门限：确定在生长过程中能将相邻像素包括进来的准则；
- 3) 停止条件：确定让生长过程停止的条件或规则。

上述基于视觉特性的 ROI 检测算法和基于区域生长的 ROI 检测算法可精确提取 ROI，但由于其较高的复杂度要求，并不适用于实时视频编码系统。其次，基

于宏块的 H.264 视频编码标准并不需要高精度的 ROI 检测。

## 4.2 改进的 ROI 提取算法

### 4.2.1 ROI 的定义

图像的物理特征中亮度直接影响着人们对图像产生的印象并且易于提起，但人眼在实际观察景物时得到的亮度感觉并不完全由实际亮度决定，而是取决于相对亮度的变化。本文提出的算法以相对亮度变化的大小作为评判显著度的标准，并以显著度进一步定义 ROI，显著度越大，人眼对该部分的兴趣性越大，反之，兴趣性越小。

由于 H.264 是基于宏块的视频编码标准，本文以宏块为单位，用宏块的相对平均亮度值度量当前编码宏块的显著度，具体过程如下：

(1) 计算每个宏块的平均亮度值：

$$\overline{L_{MB}(i)} = \frac{\sum_{j=0}^{N_{pixel}} L(j)}{N_{pixel}} \quad (4-9)$$

其中  $\overline{L_{MB}(i)}$  为第  $i$  宏块中所有像素点的亮度平均值， $N_{pixel}$  为单个宏块的像素总数，即  $N_{pixel} = 256$ ， $L(j)$  为第  $j$  像素点的亮度值。

(2) 对于任何一个宏块，与其相邻区域(上、下、左、右位置的宏块)的对比度计算如下：

$$W(i) = \overline{L_{MB}(i)} - \overline{L_{MB}(i\_next)} \quad (4-10)$$

其中  $W(i)$  为亮度对比度， $\overline{L_{MB}(i\_next)}$  为所有相邻宏块的平均亮度值。 $W(i)$  值越大，表明该宏块的显著度越大。

### 4.2.2 ROI 的阈值确定算法

不同的图像内容不相同，因此固定的阈值是不合适的。最佳的阈值确定应当能够自动的随着图像的具体内容而自适应地调整。本文采用基于信息熵的自适应阈值确定方法<sup>[34]</sup>，逐帧更新 ROI 判定阈值：

设图像中所有宏块的显著度取值范围为  $[0, M]$ ，显著度为  $t$  的宏块共有  $n_t$  个，其中  $t \in [0, M]$ ，给定一个阈值  $T$ ，大于  $T$  是感兴趣宏块，小于  $T$  为非感兴趣宏块。感兴趣宏块和非感兴趣宏块的概率分布认为是独立的，本文使用基于信息熵的自适应阈值确定方法来区分感兴趣宏块和非感兴趣宏块的概率，分别计算为：

$$P_n(t) = \frac{n_t}{\sum_{i=0}^T n_i}, 0 \leq t \leq T \quad (4-11)$$

$$P_y(t) = \frac{n_t}{\sum_{i=T+1}^M n_i}, T+1 \leq t \leq M \quad (4-12)$$

其中,  $P_n(t)$  和  $P_y(t)$  分别表示非感兴趣宏块和感兴趣宏块的概率, 两种宏块的信息熵定义为:

$$E_n(T) = -\sum_{t=0}^T P_n(t) \lg P_n(t) \quad (4-13)$$

$$E_y(T) = -\sum_{t=T+1}^M P_y(t) \lg P_y(t) \quad (4-14)$$

则最佳阈值  $T's$  为:

$$T's = \arg \max \{E_n(T) + E_y(T)\}, T = 0, 1, \dots, M \quad (4-15)$$

基于信息熵的自适应阈值确定方法流程图如下所示:

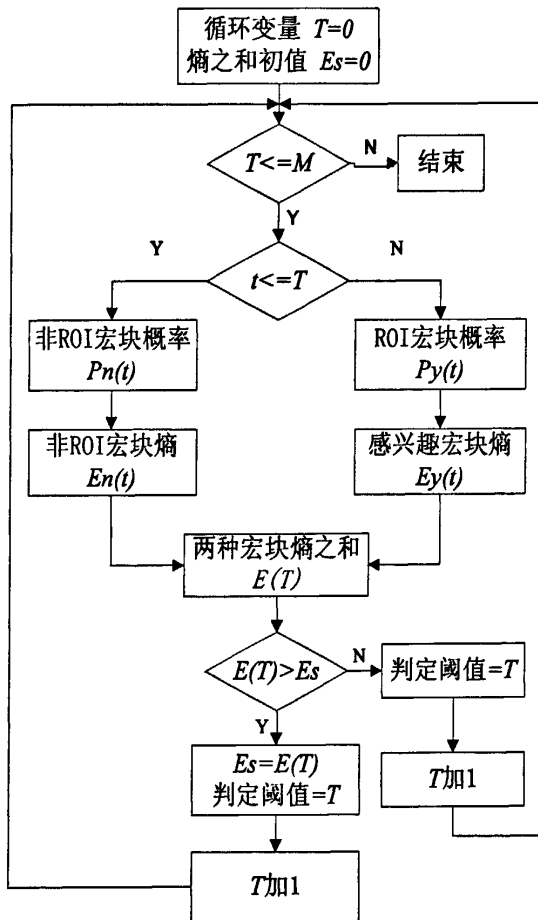


图 4-2 基于信息熵的自适应阈值确定算法流程图

Fig. 4-2 Entropy-based threshold determination algorithm

### 4.3 基于 ROI 的 MBAFW 算法

MBAFW 算法根据相邻宏块的划分方式和预测模式为不同宏块选择不同的频率加权策略,但没有考虑当前编码区域的内容特征,频率加权策略没有得到很好的权衡;本章提出基于 ROI 的宏块级自适应频率加权(ROI-based Macroblock Adaptive Frequency Weighting, ROI-based MBAFW)算法,通过对当前编码宏块 ROI 的判定,结合相邻宏块信息,选择频率频率加权策略,理论上,可较大程度提高视频主观质量。

算法整体框架如下:

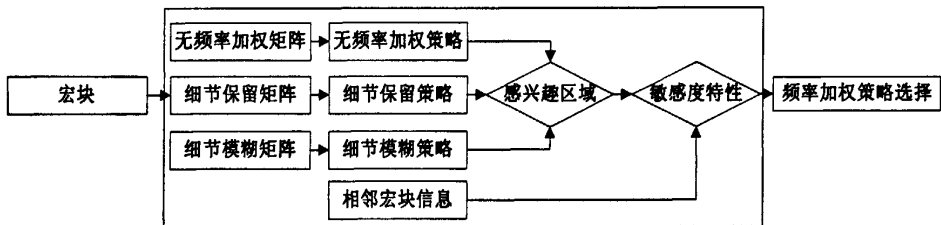


图 4-3 ROI-based MBAFW 算法框图

Fig. 4-3 Framework of the ROI-based MBAFW

#### 4.3.1 感兴趣宏块判定

本章采用 4.2 节提出的改进的感兴趣区域提取算法,对当前编码宏块进行判定。具体过程如下:

(1) 根据式(4-16)循环读取每帧各宏块的平均亮度值  $\overline{L_{MB}(i)}$ , 并依据式(4-17)计算当前宏块的亮度对比度  $W(i)$ , 以此值判定当前宏块的显著度; 该值越大, 显著度越高, 反之亦然;

(2) 根据 4.2.2 节基于信息熵的自适应阈值确定方法, 计算当前编码帧的感兴趣区域判定阈值  $T's$ , 如果满足  $W(i) \geq T's$ , 则当前宏块为感兴趣宏块, 反之为非感兴趣宏块。

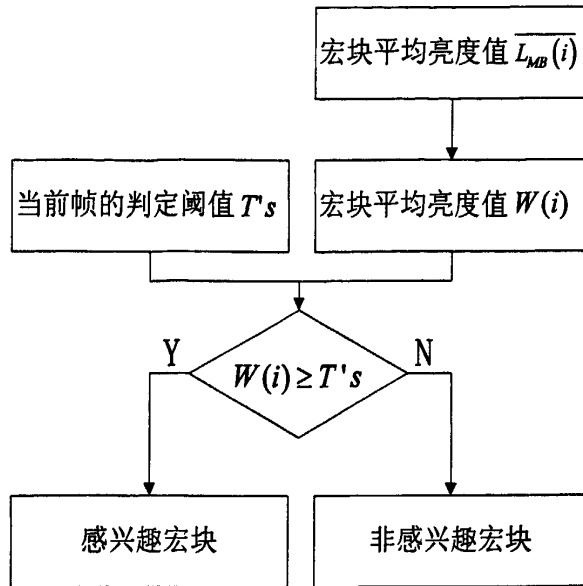


图 4-4 感兴趣宏块判定

Fig. 4-4 ROI macroblock determination

### 4.3.2 优化的频率加权矩阵

频率加权矩阵在文献[24]的基础上，经过多次仿真试验，最优量化矩阵如下：

#### (1) 4×4 块

无频率加权矩阵如图 4-5(a)所示，对所有频率系数采用相同的量化参数，在 H.264 HP 中默认为 16；

基于文献[24]的细节保留矩阵，本文细节保留矩阵如图 4-5(b)所示，由于高频区域往往代表图像细节部分，细节保留矩阵对高频系数采用较小的量化参数，以达到细节保留的效果，低频系数采用 H.264 HP 中的默认值 16；

细节模糊矩阵采用文献[24]提出的最优频率加权矩阵，如图 4-5(c)所示，和细节保留矩阵相反，细节模糊矩阵对高频系数采用较大的量化参数，以达到更好的压缩比。

$\begin{bmatrix} 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 \end{bmatrix}$	$\begin{bmatrix} 16 & 16 & 13 & 11 \\ 16 & 13 & 11 & 10 \\ 13 & 11 & 10 & 9 \\ 11 & 10 & 9 & 8 \end{bmatrix}$	$\begin{bmatrix} 6 & 13 & 20 & 28 \\ 13 & 20 & 28 & 32 \\ 20 & 28 & 32 & 37 \\ 28 & 32 & 37 & 42 \end{bmatrix}$
(a)默认矩阵	(b)细节保留矩阵	(c)细节模糊矩阵

图 4-5 4×4 块频率加权矩阵

Fig. 4-5 Frequency weighting matrices for 4×4

(2)  $8 \times 8$  块

在 H.264 HP 中,  $8 \times 8$  块默认量化矩阵各元素的值为 16, 本章算法中的无频率加权矩阵和细节模糊矩阵均采用标准默认矩阵, 如图 4-6(a)所示; 细节保留矩阵在文献[24]基础上, 通过多次试验得出最优细节保留矩阵如图 4-6(b)所示。

$$\begin{array}{c} \left[ \begin{array}{cccccccc} 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \end{array} \right] \quad \left[ \begin{array}{cccccccc} 16 & 16 & 16 & 16 & 16 & 13 & 13 & 11 \\ 16 & 16 & 16 & 16 & 13 & 13 & 11 & 10 \\ 16 & 16 & 16 & 13 & 13 & 11 & 10 & 10 \\ 16 & 16 & 13 & 13 & 11 & 10 & 10 & 9 \\ 16 & 13 & 13 & 11 & 10 & 10 & 9 & 9 \\ 13 & 13 & 11 & 10 & 10 & 9 & 9 & 9 \\ 13 & 11 & 10 & 10 & 9 & 9 & 9 & 8 \\ 11 & 10 & 10 & 9 & 9 & 9 & 8 & 8 \end{array} \right] \\ \text{(a) 无频率加权和细节模糊矩阵} \qquad \qquad \qquad \text{(b) 细节保留矩阵} \end{array}$$

图 4-6  $8 \times 8$  块频率加权矩阵Fig. 4-6 Frequency weighting matrices for  $8 \times 8$ 

## 4.3.3 基于 ROI 的 MBAFW 算法详述

该算法把各宏块 ROI 标识用于频率加权策略选择规则, 结合各相邻宏块信息, 更好的进行频率加权策略选择, 具体过程如下:

(1) 如果相邻宏块 ABCD 有一个不可得, 且当前编码宏块为非 ROI, 依据 HVS 感兴趣性和人眼对图像轮廓或边缘的量化误差不敏感特性, 采用策略 0, 过程如式(3-6);

(2) 判断当前编码帧的类型, 针对 I、P、B 采用不同的策略选择规则。

## ● I 帧

1) 如果相邻宏块 ABCD 中有一个是 Intra\_4x4 模式, 则可预知当前编码区域含有较丰富细节, 采用策略 1, 过程如式(3-7);

2) 否则如果 A 和 D 或 B 和 C 有相同的预测模式, 且当前编码宏块为非 ROI, 则可预知该区域具有较相似的内容特征, 细节不丰富, 不为人眼所重视, 采用策略 2, 过程如式(3-8);

3) 否则如果 A 和 D 或 B 和 D 是 I\_16x16 模式, 由于采用较大的划分方式, 细节相对不丰富或运动比较缓慢, 采用策略 0, 过程如式(3-6);

4) 否则, 采用策略 1, 过程如式(3-7)。

算法流程如图 4-7。



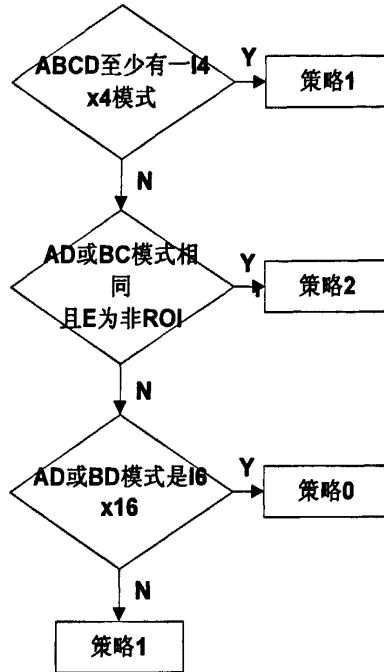


图 4-7 I 帧策略选择算法流程图

Fig. 4-7 Flow chart of the strategy selection algorithm for I frame

●P 帧

- 1) 如果当前编码宏块为感兴趣宏块，采用策略 1，过程如式(3-7);
  - 2) 否则如果相邻的四个宏块有一个是帧内宏块模式，采用策略 1，过程如式(3-7);
  - 3) 否则如果相邻宏块有一个是 inter\_4×4 模式，采用策略 1，过程如式(3-7);
  - 4) 否则如果 A 和 D 或 B 和 D 是 P\_SKIP 模式，采用策略 0，过程如式(3-6);
  - 5) 否则如果 A 和 D 或 B 和 D 是 P\_16×16 模式，且当前编码宏块为非 ROI，采用策略 2，过程如式(3-8);
  - 6) 否则，采用策略 0，过程如式(3-6)。
- 算法流程如图 4-8。

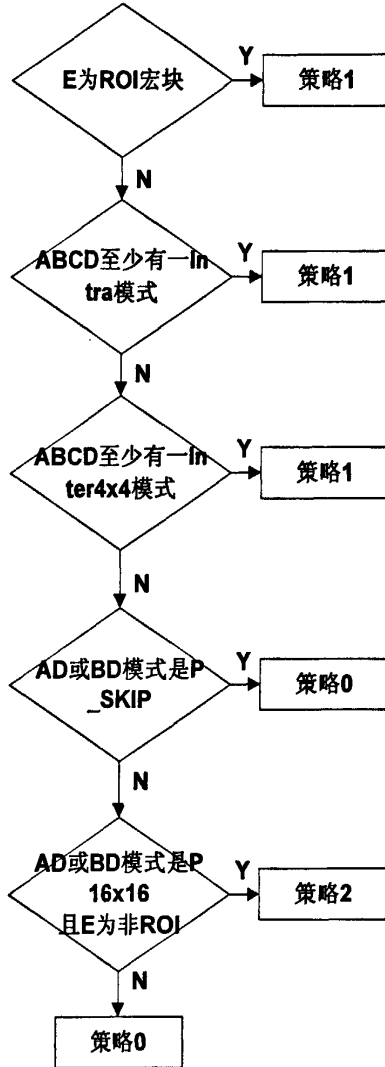


图 4-8 P 帧策略选择算法流程图

Fig. 4-8 Flow chart of the strategy selection algorithm for P frame

●B 帧

- 1) 如果 A 和 D 或 B 和 D 是 B\_SKIP 模式，采用策略 0，过程如式(3-6)；
  - 2) 否则如果 A 和 D 或 B 和 D 是 P\_16x16 模式，且当前编码宏块为非 ROI，采用策略 2，过程如式(3-8)；
  - 3) 否则，采用策略 1，过程如式(3-7)。
- 流程如图 4-9。

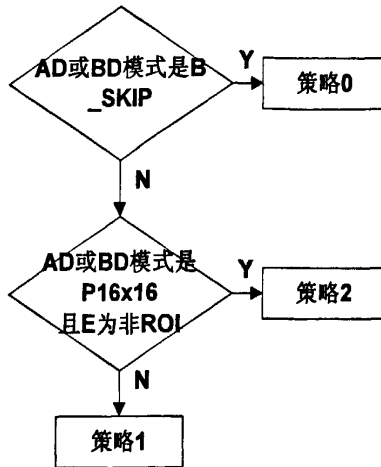


图 4-9 B 帧策略选择算法流程图

Fig. 4-9 Flow chart of the strategy selection algorithm for B frame

#### 4.3.4 仿真结果和分析

本章提出的 ROI-based MBAFW 算法测试的软件基础是 H.264 HP JM10.1, 该实验允许率失真优化以及去块效应滤波器的应用, 参考帧为 2, 搜索范围设定为 32, 测试视频序列为 QCIF 格式的 City、Mobile 和 Salesman, 视频序列结构为 IBBP。本文分别对 MBAFW 算法和 ROI-based MBAFW 算法从客观质量和主观质量两个方面进行了对比, 并对 ROI-based MBAFW 算法复杂度进行了分析。

##### (1) 仿真结果算法复杂度分析

算法复杂度包括时间复杂度和空间复杂度, 本文对 ROI-based MBAFW 算法的时间复杂度进行分析。时间复杂度是度量算法执行的时间长短, 一般情况下, 算法的基本操作重复执行的次数是模块  $n$  的某一个函数  $f(n)$ , 因此, 算法时间复杂度记作:  $T(n) = O(f(n))$ ; 随着模块  $n$  的增大, 算法执行时间的增长率和  $f(n)$  的增长率成正比, 所以  $f(n)$  越小, 算法的时间复杂度越低, 算法的效率越高。ROI-based MBAFW 算法是感兴趣区域判定和量化矩阵策略选择的过程。该算法基本操作的语句执行次数为一阶常数, 经上机测试, 该算法的执行时间较 MBAFW 算法几乎相同。结果表明, 该算法没有引入太多的时间复杂度。ROI-based MBAFW 算法和 MBAFW 算法执行时间如下表所示:

表 4-1 MBAFW 算法和 ROI-based MBAFW 算法平均每帧编码时间比较 (秒/帧)

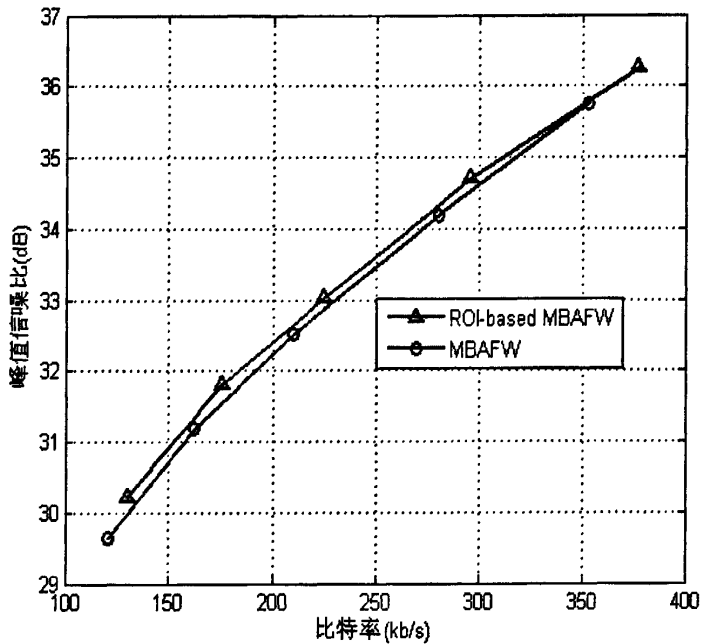
TABLE 4-1 Execution time of the MBAFW and the PAFW

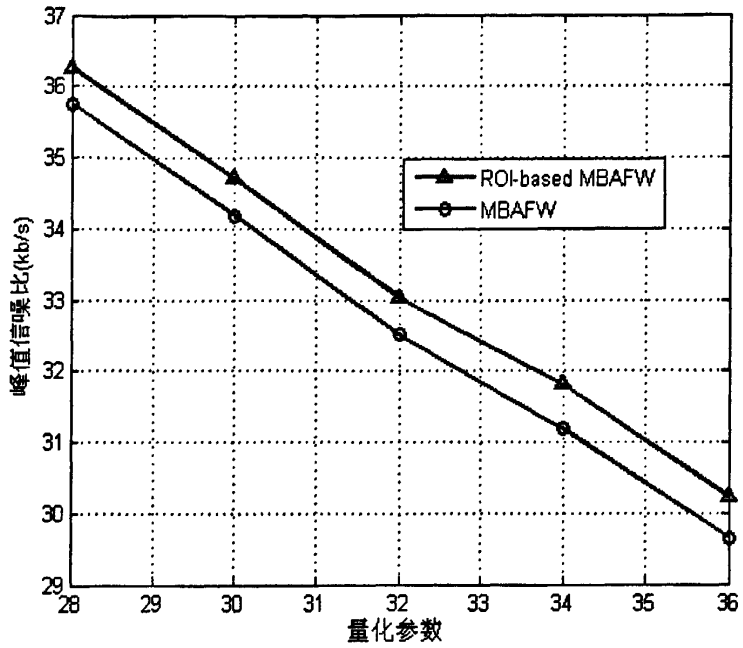
视频序列	MBAFW	ROI-based MBAFW
City(QCIF)	7.348	7.668
Salesman(QCIF)	5.969	6.136
Mobile(QCIF)	8.144	7.855

(2) 仿真结果客观质量分析

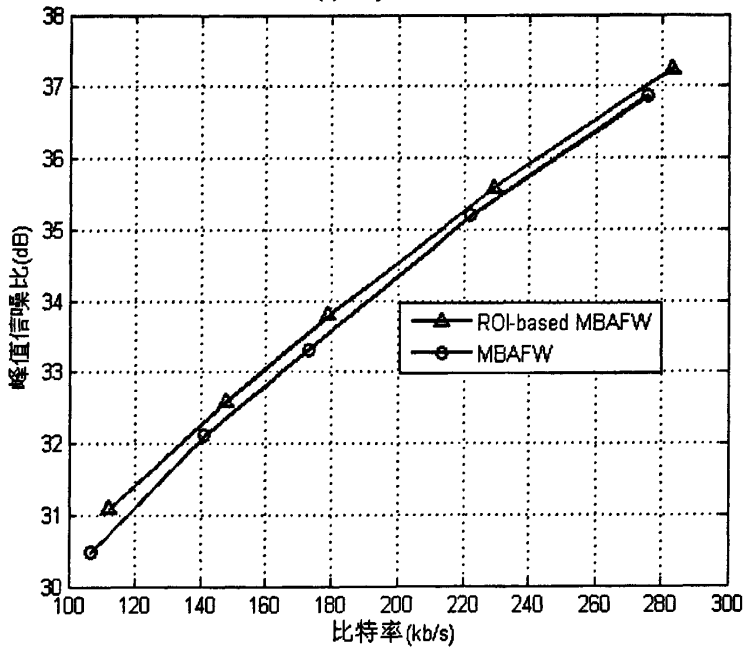
1) PSNR 比较

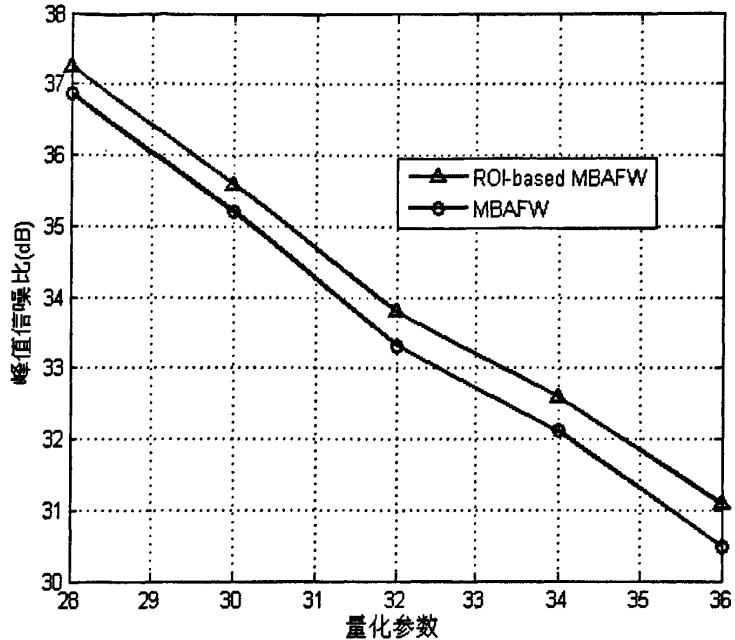
图 4-10 分别为 City 和 Salesman 视频序列在 MBAFW 算法和 ROI-based MBAFW 算法下的对比。实验结果表明，相同比特率或相同量化参数情况下，ROI-based MBAFW 算法的平均峰值信噪比更大，针对 city 序列，当量化参数为 34 时，峰值信噪比提高 0.63dB。峰值信噪比相同情况下，ROI-based MBAFW 算法比 MBAFW 算法的比特率低，有更好的压缩效果。





(a) City 序列





(b) Salesman 序列

图 4-10 ROI-based MBAFW 算法和 MBAFW 算法的峰值信噪比

Fig. 4-10 The Curves of PSNR versus bitrate and QP for City and Salesman

## 2) 比特率比较

一般, 量化参数相同时, 视频序列具有相似的主观质量。和无频率加权(NOFW)相比, ROI-based MBAFW 算法可大大降低比特率。在图 4-11 中, 当量化参数为 28 时, 该算法较 NOFW 算法降低 9%~10% 的比特率。从下图还可发现, 相同量化参数下, ROI-based MBAFW 算法的比特率比 MBAFW 稍高。这是因为 ROI-based MBAFW 算法用更多的比特来进行感兴趣部分和细节部分的保留。

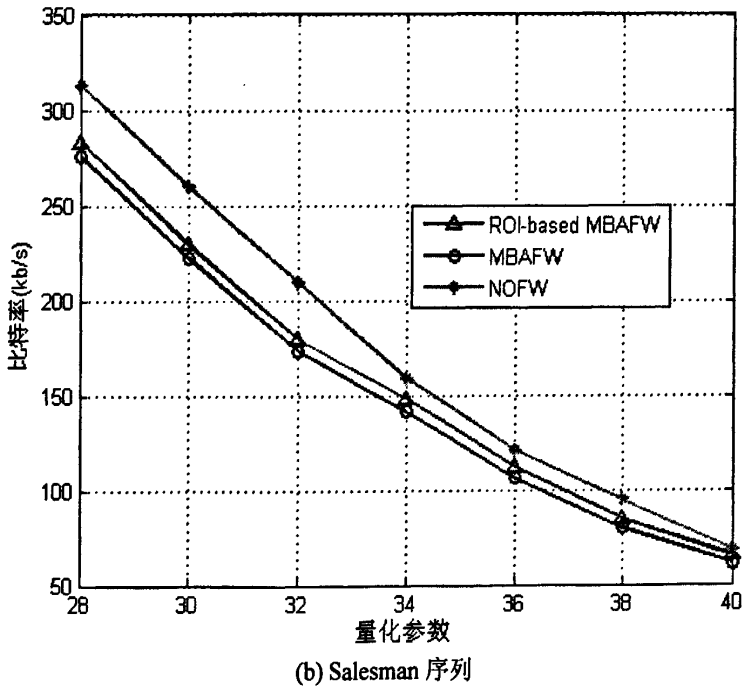
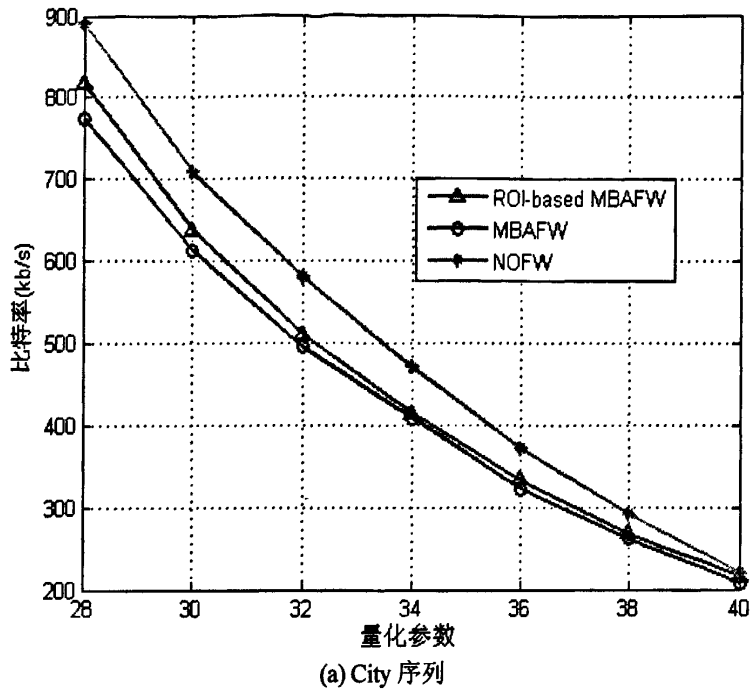


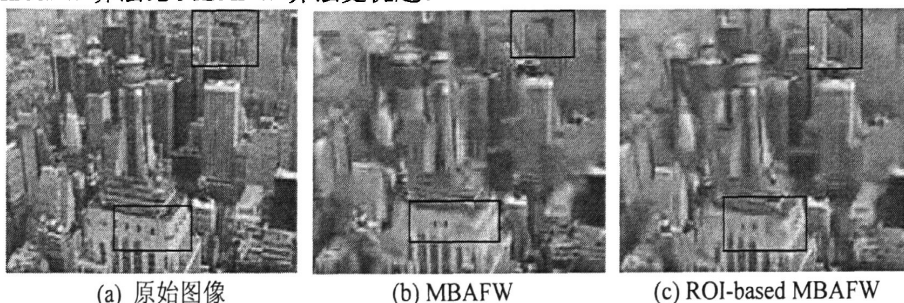
图 4-11 ROI-based MBAFW 算法和 MBAFW 算法比特率

Fig. 4-11 Bitrate comparison for City and Salesman

(3) 仿真结果主观质量分析

图 4-12 为 City 序列在量化参数为 40 时, 经 MBAFW 算法和 ROI-based MBAFW 算法重建图像对比。其中, (a) 原始图像, (b) MBAFW 算法的重建图像; (c) ROI-based MBAFW 算法的重建图像。实验结果显示, MBAFW 算法和 ROI-based MBAFW 算

法所达到的主观质量几乎相同，但细节部分相比，如下图方框中所示，ROI-based MBAFW 算法比 MBAFW 算法更优越。



(a) 原始图像

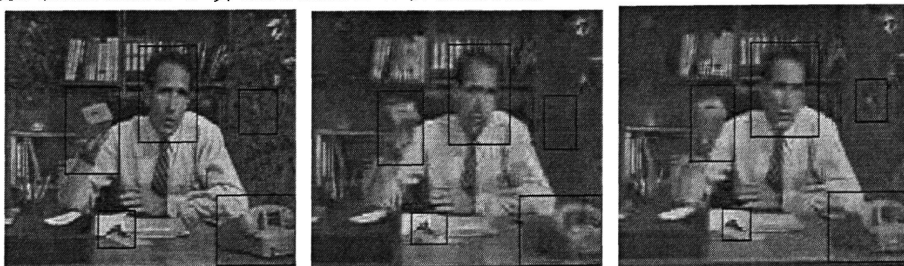
(b) MBAFW

(c) ROI-based MBAFW

图 4-12 City 序列重建图像的主观质量比较

Fig. 4-12 Reconstructed pictures of MBAFW and ROI-based MBAFW for city

图 4-13 为 Salesman 序列在量化参数为 40 时，经 MBAFW 算法和 ROI-based MBAFW 算法重建图像对比。其中，(a)原始图像，(b)MBAFW 算法的重建图像；(c)ROI-based MBAFW 算法的重建图像。实验结果显示，MBAFW 算法和 ROI-based MBAFW 算法所达到的主观质量几乎相同，但细节部分相比，如下图方框中所示，ROI-based MBAFW 算法比 MBAFW 算法更优越。



(a) 原始图像

(b) MBAFW

(c) ROI-based MBAFW

图 4-13 Salesman 序列重建图像的主观质量比较

Fig. 4-13 Reconstructed pictures of MBAFW and ROI-based MBAFW for salesman

图 4-14 为 Mobile 序列在量化参数为 40 时，经 MBAFW 算法和 ROI-based MBAFW 算法重建图像对比。其中，(a)原始图像，(b)MBAFW 算法的重建图像；(c)ROI-based MBAFW 算法的重建图像。实验结果显示，MBAFW 算法和 ROI-based MBAFW 算法所达到的主观质量几乎相同，但细节部分相比，如下图方框中所示，ROI-based MBAFW 算法比 MBAFW 算法更优越。



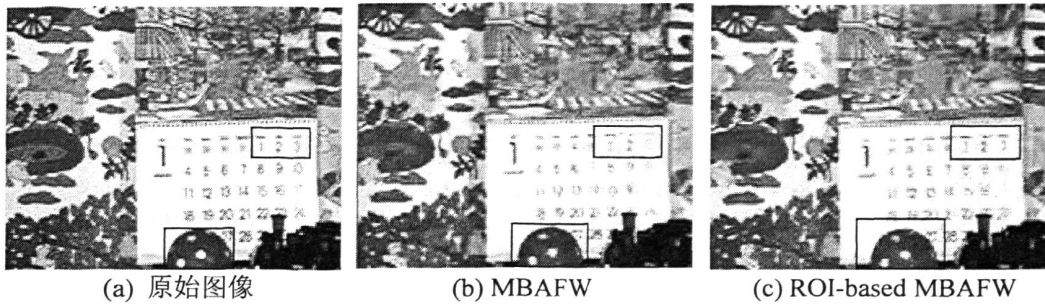


图 4-14 Mobile 序列重建图像的主观质量比较

Fig. 4-14 Reconstructed pictures of MBAFW and ROI-based MBAFW for mobile

#### 4.4 本章小结

本章在 HVS-based MBAFW 算法基础上，主要对 ROI-based MBAFW 算法进行了研究。

首先，分析了现有 ROI 检测技术，指出了其在当前视频编码应用中的不足。

其次，以视觉兴趣性为前提，提出了新的基于亮度对比度的 ROI 定义方法和基于信息熵的自适应阈值确定算法，并在此基础上提出了改进的 ROI 提取算法。

最后，结合本章提出的 ROI 提取算法，优化频率加权量化矩阵，进一步改进 MBAFW 算法，提出了 ROI-based MBAFW 算法。

该算法统筹考虑敏感度特性和视觉兴趣性，把当前编码宏块的 ROI 标识融入策略选择规则，实现逐宏块更新量化矩阵的同时，较好地保留了 ROI，进一步提高了视频主观质量。和 MBAFW 算法相比，主观质量保持不变时，编码比特数大大降低。

## 5 总结和展望

### 5.1 总结

本文针对 H.264 宏块级自适应频率加权算法进行了研究。本文主要工作如下：

1. 系统研究了视频编码标准发展历程，把 H.264 国际视频编码标准与以往视频编码标准进行了对比，剖析了 H.264 的编解码原理及先进的关键技术，并在此基础上总结了 H.264 的特点优势及应用前景。

2. 研究了现在的频率加权技术在 MPEG-2 和 H.264 标准中的应用，归纳总结了频率加权技术在 H.264 标准中的实现过程，并对 H.264 帧级自适应频率加权算法进行了研究。

3. 分析提炼了人类视觉系统特性，在 H.264 帧级自适应频率加权算法的基础上，结合当前宏块相邻宏块信息，提出了基于 HVS 的宏块级自适应频率加权算法。该算法把帧级自适应频率加权扩展到宏块级，一帧中不同宏块可选择不同的量化矩阵，降低编码比特率的同时，提高了视频重建序列的主观质量。

4. 通过对当前 ROI 检测技术的分析，本文总结了人们较为感兴趣的图像特征，即人类的视觉兴趣性。利用人类视觉兴趣性，本文提出了改进的感兴趣区域定义算法、感兴趣区域阈值确定算法和感兴趣区域提取算法。

5. 针对宏块级频率加权算法没有考虑当前编码宏块兴趣性的问题，本文在第三章所提算法的基础上，进一步提出了基于感兴趣区域的宏块级自适应频率加权算法。该算法通过改进的感兴趣区域提取方法对当前宏块进行感兴趣区域判定，并进行感兴趣区域标识，把该标识融入频率加权策略选择规则，更加全面严密地进行频率加权策略的选择。通过该算法进行优化的 H.264 标准有更好的编码效率，重建视频序列质量进一步改善。

### 5.2 展望

对于本文进一步的相关研究和改进工作建议如下：

1. 本文研究的宏观背景是 H.264 国际视频编码标准，该标准以宏块为单位进行编码，因此本文提出的算法并不能应用于所有当前流行的编码标准。下一步尽可能把多种当前流行的编码标准进行对比，并把频率加权技术灵活应用于多种视频编码标准。

2. 频率加权技术的概念来源于 HVS，而 HVS 是比较复杂的模型，本文总结

的 HVS 特性并不全面, 后续阶段课题可对 HVS 进行更加深入的研究, 以更好地服务于视频编码算法的优化。

3. 本文提出的两种视频编码优化技术均作用于量化模块, 应用范围比较局限, 可在此基础上针对运动估计和运动补偿模块、码率控制算法模块、VCL 编码模块和预处理进行视频编码算法的优化。

## 参考文献

- [1] 霍炎, 荆涛, 张德珍, 冯玉珉. 基于人眼视觉系统的误差分布反馈量化策略. 信号处理. 2009. 第 25 期. pp. 537-542
- [2] Tino von Roden. H.261 and MPEG1-A Comparison. IEEE Conference on Computer and Communications. 1996. pp. 65-71
- [3] Berna Erol, Michael Gallant. H.263+: video coding at low bit rates. IEEE Trans. on Circuits and Systems for Video Technology. 1998. vol. 8. pp. 849-866
- [4] Sikora. MPEG digital video-coding standards. IEEE Trans. on Signal Processing. 1997. vol.14. pp. 82-100
- [5] D. Lefol, D. Bull, N. Canagrajah, F. Rovati. Performance Evaluation of Transcoding Algorithm for H.264. 2006. 52(1). pp. 215-222
- [6] T. Suzuki, K. Sato, Y. Yagasaki. Weighting matrix for JVT codec. Joint Video Team of ISO/IEC MPEG & ITU-T VCEG JVT-C053, Fairfax, USA, 2002
- [7] 毕厚杰. 新一代视频压缩编码标准——H.264/AVC. 北京. 人民邮电出版社. 2004. pp. 99-100
- [8] Thomas Wiegand, Gary J. Sullivan. Overview of the H.264/AVC Video Coding Standard. IEEE Trans. on Circuits and Systems for Video Technology. 2003. 13(7). pp. 1009-1126
- [9] Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC, in Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG050, 2003
- [10] S. Wenger. H.264/AVC over IP. IEEE Trans. on Circuits and System for Video Technology. 2003. vol. 13. pp. 645-656
- [11] Hari Kalva. The H.264 Video Coding Standard. IEEE Trans. on Multimedia. 2006. 13(4). pp. 86-90
- [12] T. Wiegand et al. Overview of the H.264/AVC Video Coding Standard. IEEE Trans. on Circuits and Systems for Video Technology. 2003. 13(7). pp. 560-576
- [13] Heiko Schwarz, Detlev Marpe, Thomas Wiegand. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Trans. on Circuits and Systems for Video Technology. 2007. 17(9). pp. 1103-1120
- [14] X.Jing, L.-P. Chau. Fast approach for H.264 inter mode decision. Electronics Letters. 2004. 14(7). pp. 379-388
- [15] P. List, A. Joch, J. Lainema, G. Bjøntegaard, M. Karczewicz. Adaptive deblocking filter. IEEE Trans. on Circuits and System for Video Technology. 2003. vol.13. pp. 614-619
- [16] H. Malvar, A. Hallapuro, M. Karczewicz, L. Kerofsky. Low-Complexity transform and quantization in H.264/AVC. IEEE Trans. on Circuits and System for Video Technology. 2003. vol. 13, pp. 598-603
- [17] Mathias Wien. Variable Block-Size Transforms for H.264/AVC. IEEE Trans. on Circuits and Systems for Video Technology. 2003. 13(7). pp.604 - 607
- [18] F. Pan et al. An intra-frame quantization optimization algorithm for very low bit-rate video

- coding. Inst. for Infocomm Research, Tech. Rep, June 2002
- [19] Bo Wang, Zhibing Wang, Yupeng Liao, Xinggang Lin. HVS-based structural similarity for image quality assessment. IEEE International Conference on Signal Processing. 2008. pp. 1104-1107
- [20] Weisi Lin, Li Dong, Ping Xue. Visual distortion gauge based on discrimination of noticeable contrast changes. IEEE Trans. on Circuits and Systems for Video Technology. 2005. vol.7. pp.900-908
- [21] Claudio M Privitera, Lawrence W.Stark. Algorithms for Defining Visual Regions of Interest Comparison with Eye Fixation. IEEE Trans. on PAMI, 2000. 22(9). pp. 970-980
- [22] C. J. van den Branden Lambrecht, O.Verscheure. Perceptual quality measure using a spatio-temporal model of the human visual system. in Proc. SPIE. 1996. vol. 2668, pp.450 - 461
- [23] HaiBing Yin. Adaptive quantization in perceptual MPEG video encoders. In Proc. PCS2006. Beijing, China. 2006. pp. 13-14
- [24] Jianwen Chen, Jianhua Zheng, Shun Mei and Yun He. Macroblock-Level Adaptive Frequency Weighting for Perceptual Video Coding. IEEE International Conference on Multimedia and Expo. 2007. pp. 304-307
- [25] Z. Lu, et al. Perceptual Region-of-interest (ROI) based Scalable Video Coding. JVT-O056, Busan, KR, 2005. pp. 16-22
- [26] Z Lu, et al. CE8: ROI-based Scalable Video Coding. JVT-O308, Busan, KR, 2005. pp. 16-22
- [27] 张鹏等. 静态图像中的感兴趣区域检测技术. 中国图像图形学报. 2005
- [28] Se Yoon Jeong, Kyuheon Kim, Byung Tae Chun. Entropy and color correlation features for image indexing. IEEE Conference Proceedings. Japan. 1999. vol. 2. pp. 895-899
- [29] Shoji Tanaka, Masayuki Inoue, Michitoshi Ishiwaka. A method for extracting and analyzing "Kansei" factors from pictures. IEEE First Workshop on Multimedia Signal Processing. 1997. pp.251-256
- [30] Toshiyuki I waoka, Hideyuki Kobayahai, Shunji Ota. Digital safari guidebook with image retrieval. IEEE International Conference on Multimedia Computing and Systems. 1999. vol. 2. pp. 1011-1012
- [31] 陆伟, 倪林. 利用颜色和熵提取感兴趣区域的感性图像检索. 中国图象图形学报. 2006. 第4期. pp. 493-494
- [32] Villegas p, Marichal x, Salcedo A. Objective evaluation of segmentation masks in video sequences. WIAMIS99[C]. 1999. pp. 85-88
- [33] 廖斌, 许刚, 刘春颖. 基于高级视觉特性的感兴趣区域判定算法. 计算机应用. 2006. 第12期. pp. 2842-2847
- [34] 王艳娟, 陈晓红, 邹丽. 图像感兴趣区域自动提取算法. 科学技术与工程. 2007. 第12期. pp. 2868-2870

## 作者简历

- 教育经历

2008.9-2010.7      通信与信息系统专业      硕士      北京交通大学

2004.9-2008.7      网络工程      学士      河北师范大学

- 研究课题

2008.10-2010.7 国家自然科学基金项目“无线宽带网络联合信道率失真模型相关技术研究”

- 发表论文

牛青青, 李兴华, 荆涛, 仲丽媛.“基于人类视觉系统的宏块级自适应频率加权算法”, 信号处理, 2010年26卷第8期

仲丽媛, 荆涛, 牛青青, 霍炎.“针对 LDPC 编码的无线视频传输性能分析”, 铁路计算机应用, 2010年19卷第07期

## 学位论文数据集

表 1.1: 数据集页

关键词*	密级*	中图分类号*	UDC	论文资助
H.264 标准, 频率加权, 人类视觉系统, 感兴趣区域		TN919.8	621.39	
学位授予单位名称*		学位授予单位代码*	学位类别*	学位级别*
北京交通大学		10004	工学	硕士
论文题名*		并列题名		论文语种*
H.264 宏块级自适应频率加权算法的研究				汉语
作者姓名*	牛青青		学号*	08120108
培养单位名称*		培养单位代码*	培养单位地址	邮编
北京交通大学		10004	北京市海淀区西直门外上园村 3 号	100044
学科专业*		研究方向*	学制*	学位授予年*
通信与信息系统		多媒体通信	2 年半	2010 年
论文提交日期*				
导师姓名*	李兴华		职称*	副教授
评阅人	答辩委员会主席*		答辩委员会成员	
	肖人毅			
电子版论文提交格式 文本() 图像() 视频() 音频() 多媒体() 其他() 推荐格式: application/msword; application/pdf				
电子版论文出版(发布者)		电子版论文出版(发布)地		权限声明
论文总页数*	53			
共 33 项, 其中带*为必填数据, 为 22 项。				