

Visual C++

音视频编解码 技术及实践

求是科技 编著

附·源·代·码·光·盘

□ 本书以 Visual C++ 作为开发工具，从实用角度出发，介绍了音视频编解码技术的基础理论、实现方法和实用技巧，并给出具体的工程案例。全书主要内容包括压缩技术基础、JPEG 压缩编解码技术、MPEG-1 压缩编解码技术、MPEG-2 标准、MPEG-4 压缩编解码标准、MPEG-4 实用源代码分析、H.263 视频编解码技术、H.264 视频编解码技术、音视频网络传送技术以及通信实例。



人民邮电出版社
POSTS & TELECOM PRESS

Visual C++ 音视频编解码 技术及实践

■ 求是科技 编著

人民邮电出版社

图书在版编目 (CIP) 数据

Visual C++音视频编解码技术及实践 / 求是科技编著.

—北京: 人民邮电出版社, 2006.6

ISBN 7-115-14813-9

I. V... II. 求... III. C 语言—程序设计 IV. TP312

中国版本图书馆 CIP 数据核字 (2006) 第 051907 号

内 容 提 要

本书以 Visual C++作为开发工具, 从实用角度, 向读者介绍了数字音、视频编解码技术的基础理论、实现方法和实用技巧, 并给出具体的工程案例。

全书主要包括数字视频技术基础, JPEG 编解码技术与实现, MPEG-1 编解码技术与实现, MPEG-2 压缩编码标准, MPEG-4 压缩编码技术与实现, MPEG-4 实用源代码分析, H.263 视频压缩编码技术与实现, H.264 视频压缩编码技术与实现, 音、视频网络传送技术等, 最后给出了一个音、视频通信工程的实例。

本书内容丰富、叙述详细、实用性强, 可供从事数字编码及多媒体开发工作的技术人员阅读。

Visual C++音视频编解码技术及实践

- ◆ 编 著 求是科技
责任编辑 张立科
- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京密云春雷印刷厂印刷
新华书店总店北京发行所经销
- ◆ 开本: 787×1092 1/16
印张: 41
字数: 1 008 千字 2006 年 6 月第 1 版
印数: 1—5 000 册 2006 年 6 月北京第 1 次印刷

ISBN 7-115-14813-9/TP · 5424

定价: 88.00 元 (附光盘)

读者服务热线: (010) 67132692 印装质量热线: (010) 67129223

前 言

随着编码理论和媒体网络应用的发展,近年来诸如 MPEG-2、MPEG-4 和 H.264 等编码标准逐渐为大家所熟悉,基于这些标准的数字视频、音频技术的应用也越来越广泛。

Visual C++是广大工程技术人员最为熟悉的开发工具,并且数字编码技术对代码效率有很高的要求,对系统也有一定的依赖,所以本书选择了 Visual C++作为主要开发工具。

本书的前身《Visual C++实现 MPEG/JPEG 编解码技术》第一版于 2002 年面世以来,深受广大读者的青睐,同时也收到了很多读者的宝贵意见和建议,包括原稿中程序、文字上的不足和对当前流行的音、视频网络传输技术知识的需求等。为此,我们策划了本书。本书在保持原有特色的基础上去除了原书第 3 章和第 8 章的内容,新增加了 H.263、H.264 视频编码标准等新内容,并在最后一章给出了音、视频开发的综合实例。

全书共分 10 章,主要内容如下。

第 1 章 介绍了压缩技术基础。主要包括压缩的目的、必要性和图像压缩的基本思想。

第 2 章 介绍了 JPEG 压缩编解码技术。JPEG 压缩编码技术是全书的基础,本章介绍了许多图像压缩编码技术的基本算法和实现方法,如离散余弦变换、量化、熵编码及 Huffman 编码。另外,还详细介绍了 JPEG 的实现、压缩编码 API 函数的封装以及 API 的调用演示。

第 3 章 介绍了 MPEG 压缩编码技术。本章是视频压缩编码的基础,讲述了 MPEG 视频模型、I 图像帧编码、P 图像帧编码和 B 图像帧编码,同时介绍了 MPEG 音频压缩。

第 4 章 介绍了 MPEG-2 标准。本章从 MPEG-2 标准的特征进行讲述,分别介绍了 MPEG-2 视频和音频的特征,随后介绍了改进后的 MPEG-2 算法。

第 5 章 介绍了 MPEG-4 压缩编码标准。本章首先介绍了 MPEG-4 标准的特征,然后就 MPEG-4 视频编码算法介绍了 MPEG-4 的结构和句法、MPEG-4 系统数据类型描述、形状编码、运动信息编码、纹理编码、sprite 编码、可缩放性以及容错和码率控制。随后分析了基于视频对象平面的编码和 MPEG-4 视频解码算法。最后介绍了 MPEG-4 音频流技术。

第 6 章 是在第 5 章的基础上介绍 MPEG-4 的实用源代码,介绍了代码的编译、MPEG-4 文件处理以及简要的源代码分析。

第 7 章 介绍了 H.263 视频编码标准。本章从 H.263 标准的技术框架入手,详细阐述了该标准所应用的视频模型以及模型中各层的定义与描述;然后根据系统框架详细讲解了 H.263 所采用的运动估计技术、视频帧图像编码技术、码率控制技术以及视频可缩放技术。

第 8 章 介绍了 H.264 视频编码标准。本章着重介绍 H.264 标准的框架模型、NAL、VCL 等新概念,然后重点讲解帧内/帧间预测技术、熵编码技术、码率控制技术、去块效应滤波技术、变换与量化技术等。

第 9 章 介绍了音、视频网络传送技术。本章着重介绍音、视频传输中的关键技术,例如路由选择、传输协议、拥塞控制、数据同步、差错控制、安全传输、QoS 服务质量等。

第 10 章 介绍了两个通信实例。本章主要通过对两个通信实例——音频通信实例和视频通信实例的讲解,帮助读者深入领会、切实掌握音频、视频编码以及网络传送各个方面的知识。

另外附带光盘中有本书涉及到的全部源程序，读者可以参考使用，稍做改动即可应用于实际项目的研发中。

本书主要由求是科技组织编写，其他参加编写工作以及提供帮助的人员有梁国远、林飞、吴杰、刘世杰、翟华等。

由于时间仓促，加上编者水平有限，书中难免存在疏漏和不足之处，请广大读者指正。

编者

2006年5月

目 录

第 1 章 数字视频技术基础	1
1.1 图像压缩的必要性	1
1.2 数字图像压缩的基本思想	2
1.2.1 无损压缩	2
1.2.2 有损压缩	4
1.3 MPEG 技术基础	6
1.4 H.26x 技术基础	9
第 2 章 JPEG 编解码技术及实现	12
2.1 JPEG 压缩标准	12
2.2 JPEG 模型与算法流程	13
2.3 离散余弦变换	15
2.4 量化和之字序列	20
2.4.1 量化原理	20
2.4.2 标量量化	20
2.4.3 矢量量化	21
2.4.4 JPEG 中的量化	21
2.5 熵编码	23
2.5.1 熵	23
2.5.2 熵编码原理	24
2.5.3 JPEG 中的熵编码	24
2.6 Huffman 编码	28
2.6.1 理论基础	28
2.6.2 程序实现	29
2.7 JPEG 文件的格式	41
2.7.1 色度空间	41
2.7.2 JPEG 文件格式	42
2.8 JPEG 压缩编解码的实现	47
2.8.1 实现 JPEG 压缩编解码的 Cjpeg 类	47
2.8.2 JPEG 压缩编解码程序的实现	54
2.9 封装方便 JPEG 与 Bmp 相互转换的 API 函数	62
2.9.1 API 函数的具体封装办法	63
2.9.2 API 函数的 Visual Basic 调用演示	65
2.10 小结	68
第 3 章 MPEG 编解码技术与实现	70
3.1 MPEG-1 标准	70
3.2 MPEG 视频模型	73

3.2.1	MPEG 视频压缩	73
3.2.2	宏块 (Macro Block)	74
3.2.3	运动补偿预测	75
3.2.4	MPEG 帧图像的类型	76
3.3	为 I 图像帧编码	79
3.3.1	算法原理	79
3.3.2	实现代码	79
3.4	为 P 图像帧编码	90
3.4.1	算法原理	90
3.4.2	实现代码	93
3.5	为 B 图像帧编码	101
3.5.1	算法原理	101
3.5.2	实现代码	102
3.6	MPEG 音频压缩	112
3.6.1	听觉系统的感知特性	113
3.6.2	声音编码	116
3.6.3	编码层	118
3.6.4	MPEG 音频层 1/2/3 的帧头	140
3.6.5	MPEG 音频编码的性能	143
第 4 章	MPEG-2 压缩编码标准	144
4.1	MPEG-2 标准特性	144
4.1.1	MPEG-2 视频特性	146
4.1.2	MPEG-2 音频特性	148
4.2	改进后的 MPEG-2 算法	149
4.2.1	视频算法	149
4.2.2	音频算法	172
第 5 章	MPEG-4 压缩编码技术与实现	194
5.1	MPEG-4 标准特性	194
5.1.1	AV 对象 (AVO)	194
5.1.2	MPEG-4 标准的构成	196
5.1.3	MPEG-4 的应用前景	199
5.2	MPEG-4 视频编码算法概述	200
5.2.1	MPEG-4 的结构和句法	200
5.2.2	MPEG-4 系统数据类型描述	203
5.2.3	形状编码	207
5.2.4	运动信息编码	207
5.2.5	纹理编码	217
5.2.6	sprite 编码	238
5.2.7	可缩放性 (Scalability)	240
5.2.8	容错和码率控制	242

5.2.9 小结	247
5.3 基于视频对象平面的编码	248
5.3.1 视频对象平面的编码结构	248
5.3.2 视频对象平面编码的代码分析	249
5.4 MPEG-4 视频解码算法概述	259
5.4.1 媒体对象的组成和流式数据的传输	259
5.4.2 MPEG-4 系统流	262
5.4.3 MPEG-4 视频流解码	267
5.4.4 MPEG-4 解码器代码分析	277
5.4.5 小结	282
5.5 MPEG-4 音频流	284
5.5.1 自然声音	284
5.5.2 参数编码器	285
5.5.3 码激励线性预测编解码器	287
5.5.4 TwinVQ 与 AAC 简介	288
5.5.5 结构化音频	288
5.5.6 MPEG-4 音频编码编程介绍	289
5.5.7 小结	303
第 6 章 MPEG-4 实用源代码分析	304
6.1 如何完成工程的编译和安装	304
6.2 MPEG-4 文件处理	306
6.2.1 MPEG-4 视频	306
6.2.2 MPEG-4 音频	307
6.2.3 多路技术	307
6.2.4 MPEG-4 的编码	308
6.2.5 MPEG-4 的播放	314
6.3 MPEG-4 编码器源代码分析	315
6.3.1 源代码的组织结构	315
6.3.2 代码功能分析与讲解	316
6.4 小结	333
第 7 章 H.263 视频压缩编码技术与实现	334
7.1 H.263 视频压缩编码标准	334
7.1.1 H.263 简介	334
7.1.2 H.263 标准的构成	338
7.1.3 H.263 的应用前景	340
7.2 视频模型的描述	341
7.2.1 视频图像层	341
7.2.2 GOB 层	347
7.2.3 编码宏块层	348
7.2.4 编码块层	350

7.3	H.263 视频压缩编码算法及实现	351
7.3.1	H.263 基本数据结构描述	351
7.3.2	运动估计技术	355
7.3.3	高级变长编码技术	380
7.3.4	容错与码率控制技术	382
7.3.5	视频可缩放性	384
7.3.6	视频编码器设计	387
7.3.7	小结	401
7.4	H.263 视频解码技术	402
7.4.1	运动补偿技术	402
7.4.2	系数解码	407
7.4.3	编码块重构	412
7.5	系统编译与使用	412
7.5.1	编码器系统编译与使用	412
7.5.2	解码器系统编译与使用	413
第 8 章	H.264 视频压缩编码技术与实现	415
8.1	H.264 视频编码标准	415
8.1.1	H.264 简介	415
8.1.2	H.264 标准的构成	418
8.1.3	H.264 的应用前景	422
8.2	H.264 视频编码算法概述	423
8.2.1	基本概念	424
8.2.2	H.264 编码器数据类型描述	431
8.2.3	帧内预测技术	438
8.2.4	帧间预测技术	449
8.2.5	熵编码技术	462
8.2.6	去块效应的循环滤波器技术	462
8.2.7	假想参考解码器模型	467
8.3	总结	468
第 9 章	音、视频网络传送技术	469
9.1	路由选择	469
9.1.1	基本概念	470
9.1.2	单播路由	471
9.1.3	组播技术	473
9.2	传输协议	485
9.2.1	基本概念	485
9.2.2	传输低层协议	486
9.2.3	传输高层协议	500
9.3	拥塞控制	502
9.3.1	基本概念	502

9.3.2 拥塞控制策略	503
9.3.3 控制算法性能评价	509
9.4 数据同步	510
9.4.1 基本概念	510
9.4.2 同步描述	511
9.4.3 同步机制	513
9.5 差错控制	515
9.5.1 基本概念	515
9.5.2 编码方式	516
9.5.3 基本协议	518
9.5.4 编程实现	519
9.6 安全传输	521
9.6.1 基本概念	521
9.6.2 安全体系	521
9.6.3 安全措施	522
9.6.4 编程实现加-解密	525
9.7 QoS 服务质量	537
9.7.1 基本概念	537
9.7.2 QoS 规范	538
9.7.3 QoS 机制	539
9.7.4 QoS 体系	540
9.7.5 编程实现 QoS	542
9.8 小结	572
第 10 章 音频、视频通信工程实例	573
10.1 音频通信实例	573
10.1.1 设计概要	573
10.1.2 关键技术	574
10.1.3 关键代码	575
10.1.4 系统改进	614
10.2 视频通信实例	614
10.2.1 设计概要	614
10.2.2 关键技术	614
10.2.3 关键代码	621
10.2.4 系统改进	642
参考文献	644

第 1 章 数字视频技术基础

20 世纪 90 年代,多媒体技术随着人类进入信息时代而得到了迅猛发展。大信息量的音、视频数据被广泛使用,如常见的 VCD、DVD 等都是将大量的音、视频数据经过处理后供用户使用的。音、视频相关技术在一些行业领域也得到了很好的应用,如保安监控系统中对现场进行音、视频的录制,以便存档、查看、取证等。

由于音、视频数据量非常大,若未经处理,其存放、网络传送都存在很大问题。为此,需要将它们进行压缩(如 VCD 盘的内容就是经过压缩处理的),待用户使用时进行解压缩即可(如 VCD 播放机就是将光盘内容解压缩后进行音、视频播放的)。

为了压缩音、视频数据,出现了各种不同的压缩算法,并都取得了一定成效。但被普遍认可、最为权威的还得是由 ISO 组织颁布的 MPEG 技术以及由 ITU-T 组织颁布的 H.26x 技术(常见的是 H.263 和 H.264 技术)。本章将主要介绍数字视频压缩相关技术的原理和实现方法,并结合静态图像压缩标准,介绍目前流行的 JPEG 图像压缩标准技术。

1.1 图像压缩的必要性

在过去的 10 年里,需要存储、传输和处理的信息的数量成指数级的增加。这主要是两项技术的发展,一是多媒体系统在众多领域广泛应用,过去计算机只能处理数字和文本,现在已能处理声音、图像、电影;另一项是 Internet 的大发展,它使信息可以被许多人共享。这两项技术综合作用产生了所谓的 World Wide Web,一个交互的、多媒体的、基于超文本的信息网络。

首先,硬件方面的革新使这些发展成为可能。CPU、磁盘、传输信道的性能都迅猛发展。但是,从下面的事实可以看出,压缩技术仍有一段路要走。

- 存储一幅中等大小的图像,比如说 512×512 pixel, 24 位真彩色,需要 0.75MB 的空间。
- 一幅分辨率为 $12 \mu\text{m}$ 的 35mm 数字照片需要 18MB 的空间。
- 一秒钟的 NTSC 彩色视频需要 23MB 的空间。

由此可见,无论是从经济还是技术的角度,光靠硬件的支持并不能满足人们的需要。在本章中要讲的压缩技术,将提供一种解决方案。信息压缩的动机是很明显的。如果能用一种压缩了的形式来表示信息,将会有如下的好处:

- 节省存储空间;
- 节省 CPU 时间;
- 节省传输时间。

大部分信息都具有高度的相关性,或者说,它们本身包含着冗余,因此,没有信息丢失的压缩是可能的。压缩中最关键的问题是能够在原始数据和压缩数据之间快速地切换。

1.2 数字图像压缩的基本思想

压缩机制通常可以分为有损压缩和无损压缩两种。无损压缩通常被用于文本文件的压缩中,其中没有信息丢失的数据的精确重建是让人们关心的问题。

对于有损压缩,允许压缩后产生质量上的误差。有损压缩机制的优点是可以得到比无损压缩高得多的压缩比,但是它只能用近似的数据代替原始数据,而这种相近数据又是容易被压缩的。举例来说,如果图像与原图像在视觉上没有区别,那么它就可以看作是原图像的近似图像。

任何压缩机制的目的都是除去数据中存在的相关性。所谓相关性,就是根据给出的一部分数据来判断出其相邻的数据。

实际中存在着很多种数据相关性,这里给出常见的几种。

- 空间相关性: 可以根据图像中某一点的像素值推断出其相邻点的像素值。
- 频率相关性: 一个信号的傅立叶变换通常是光滑的,这意味着可以根据某一部分的频率来推断其相邻部分的频率。
- 时间相关性: 在数字视频中,在时间上相邻两帧图像的大部分像素的值变化很小。

有损压缩的标准过程是变换编码即用了一个和原来不同的数学基来表示数据,其数据的相关性能够显露出来或被拆开。在这种新的基下,大部分的系数都接近零,可以忽略,于是可以将余下的信息存储在一个较小的数据包中。压缩是通过数据变换,把域值以下的系数置零,对非零的数据进行无损的编码来实现。

1.2.1 无损压缩

1. 行程编码(Run-Length)技术

行程编码是相对简单的一种编码,是指在一行扫描的像素中,比较相邻像素的幅度(如亮度)。当幅度有一个显著变化时,就说有一个行程存在。像素幅度的连续长度和终点位置标记是其重要参数。随终点位置标记方法不同,行程编码可分为两类。

(1) 行程终点编码: 行程的终点位置由扫描行的起始点算起至行程终点位置时的像素数确定。

(2) 行程长度编码: 某行程的终点位置由它距前一终点的相对距离来确定。

行程终点编码分为两种。

- 线性码法(A码)根据不同行程长度赋予不同码字。大行程码字长,小行程码字短。
- 对数码法(B码)的码字长度与行程长度的对数成正比。

行程编码适用于二值图像。行程编码的编码效率不如 Huffman 编码方法高,但它的码字结构相对简单,故在许多情况下都采用此法。

2. Huffman 编码压缩

在无损压缩的编码方法中, Huffman 编码方法是一种较有效的编码方法。Huffman 编码是一种长度不均匀的、平均码率可以接近信息源熵值的一种编码。该编码对于出现概率大的

信息采用短字长的码,对于出现概率小的信号用长字长的码,以达到缩短平均码长,从而实现数据的压缩目的。Huffman 编码采用以下方法进行:首先将信源符号按其出现概率的大小顺序排列,然后把出现概率最小的两个符号的概率值相加,得到一个新的概率。第二步,把这个新概率看成是一个新符号的概率,和其他符号再按概率大小排列,再把最后两个概率相加。重复上述做法,直到最后只剩下两个符号的概率为止。完成以上概率相加作顺序排列后,再反过来逐步向前进行编码,每一步有两个分支,各赋予一个二进制码,可以对概率大的赋码元 0,对概率小的赋码元 1,亦可对概率大的赋码元 1,对概率小的赋码元 0。Huffman 在变字长编码方法中是最佳的,其码字平均长度很接近信息符号的熵值。Huffman 编码的最高压缩效率可以达到 8:1,但是在一般实施过程中,很难达到这种压缩比例。若图像文件中存在某个拥有长行程的字节值时,使用行程编码压缩算法可能更好。

3. 字典压缩方法

目前广泛采用的字典压缩算法包括两种类型。一种是在数据压缩过程中,寻找当前等待进行压缩处理的数据串中是否在已经处理过的数据串中出现过,如果确实曾经出现过,则利用指向已经进行处理数据串的指针代替当前等待进行压缩的数据串,此时,字典是隐式的,它用曾经处理过的数据描述。这类字典压缩算法都是基于 Abraham 与 Jakob Ziv,于 1977 提出并发表的 LZ77 算法,该算法提出后,Storer 与 Szymanski 于 1982 年对其进行了改进,并提出相应的 LZSS 算法,LZSS 算法成为现在实践中广泛使用的该类算法的基础。

另外一种字典压缩算法是为输入数据创建一个短语字典,如果在当前等待进行压缩的数据流中发现字典中已经存在相应的短语,则利用该短语在字典中的相应索引值取代原始数据,这种类型的算法基于 Lempel 与 Ziv,并在 1978 年提出发表的 LZ78 算法。后来该压缩算法由 Sperry 公司的研究员 Welch 于 1984 年在硬件设计过程中改进,并用于高性能磁盘控制器的设计,同时,由 Lempel 和 Ziv 在实际工作中实现,LZW 编码也由此而得名。LZW 压缩算法现在已经派生出几种流行的算法格式,不仅可以用于文字数据的压缩,而且也可以成功地用于某些图像数据的压缩处理。现在,WINZIP、GZIP 和 PKZIP 等压缩软件广泛使用 LZ77 压缩算法进行数据压缩。

4. 算术压缩方法

算术压缩方法与 Huffman 压缩方法相似,都是利用比较短的代码取代图像数据中出现比较频繁的数据,而利用比较长的代码取代图像数据中使用频率比较低的数据,从而达到数据压缩的目的。它同时又采用了 LZW 压缩算法的思想,不仅压缩数据值,而且压缩值序列,从而可以达到更加理想的压缩比例,尤其适合于由相同的重复序列组成的图像文件。但是算术压缩算法的实现比较复杂,其基本思想是将每个不同的序列按照出现频率映像到 0 和 1 之间的相应数字区域内,该区域表示成可以改变精度的二进制小数,其中出现频率低的数据由精度高的小数表示。算术压缩算法中两个基本的要素为源数据出现的频率以及其对应的编码区间。其中,源数据的出现频率决定该算法的压缩效果,同时也决定编码过程中源数据对应的区间范围,而编码区间则决定算术压缩算法最终的输出数据。

算术压缩算法可以大幅度地减小文件长度,甚至可以达到 100:1 的压缩比例,在 JBIG 与 JPEG 等图像文件格式的数据压缩处理步骤中占有很重要的地位。针对不同的图像文件,其压缩比例主要是与源文件的数据分布及其所采用标准模式的精度有关,同时,阻碍算术压缩

算法广泛推广的原因不仅是因为算法的复杂性，而且其受到几项 IBM 的专利保护，从而导致了算法许可证的不确定性。

1.2.2 有损压缩

1. 预测编码方法

如果已知图像一个像素离散值，利用其相邻像素的相关性，预测它的下一个像素（水平方向或垂直方向）的可能值，求其两者差，再量化、编码，这种方法称为预测编码方法，简称预测法（DPCM 法）。

预测编码方法计算简单，若采用 Huffman 编码技术，压缩比为 2:1~4:1 仍有满意的效果。但本方法是基于差值信号的统计特性基础上发展起来的，它还有如下缺点。

- 对黑白灰度有突变的点，会有较大的预测误差，使重建图像的边缘模糊，分辨率降低。
- 对图像亮度值变化缓慢区域，其差值信号应为零，但因预测值偏大而使重建图像产生噪音。

针对预测编码方法的缺点，其改进的方法是自适应预测编码方法。

2. 变换编码

图像变换编码系统框图如图 1-1 所示。

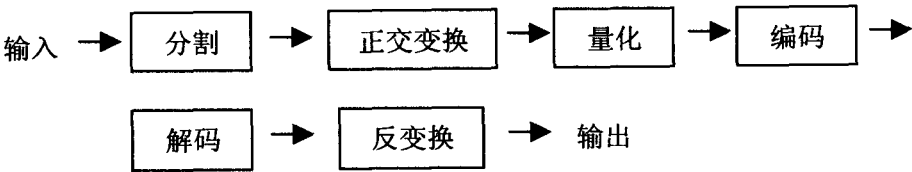


图 1-1 图像变换编码系统框图

图像经过正交变换后，能够实现图像数据压缩的物理本质在于经过多维坐标系中的适当的坐标旋转和变换，能够把散布在各个坐标轴上的原始图像数据，在新的坐标系中集中到少数坐标轴上，因而有可能用较少的编码比特数来表示一幅子图像，实现图像的压缩编码。

从数学上来看，可用于图像压缩编码的正交变换的方法除 Fourier 变换、Walsh-Hadamard 变换外，还有正弦变换、余弦变换、斜变换、哈尔变换、K-L 变换等。不同的变换有不同的压缩效果（压缩比和重建图像质量）。数学上已证明，采用均方差最小准则，K-L 变换（即离散信号的 Hotelling 变换）具有最佳变换性质，余弦变换次之，几种正交变换方法由优到劣排列顺序如下：

K-L ———> DCT ———> ST ———> DFT ———> WT

虽然 K-L 变换最优，但其实现起来需要很大的计算量，因而常用 DCT 来替代。变换的效率还与子图的大小有关，一般说来子图增大能提高编码效率，但也会使计算量增大而给实现带来困难。同时图像的相关性一般局限在 20 个邻近像素之内。因此，目前通常采用的子图

像大小为 $8 \times 8\text{pixel}$ 或 $16 \times 16\text{pixel}$ 。

3. 金字塔编码

由 Burt 和 Adelson 提出的金字塔算法,是把原图像分解成许多不同的分辨率的子图像,并把高分辨率(尺寸较大的)的子图像放在下层,把低分辨率(尺寸较小)的子图像放在上层,从而构成了一个金字塔,借助于拉普拉斯金字塔,对图像的每一层分别量化、编码,并对视觉不敏感的层粗化,用较少的码字编码,从而达到压缩的目的。

图像的金字塔表示方法也是计算机视觉中常用的一种多分辨率表示法,利用金字塔表示法,能分析图像中不同大小的物体,例如高分辨率的下层可用于分析细节,低分辨率的上层可用于分析较大的物体,同时,通过对低分辨率、尺寸较小的上层进行分析所得到的信息还能用来指导对高分辨率、尺寸较大的下层进行分析,从而大大简化了分析和计算。

图像的金字塔编码有其合理的一面,金字塔编码能大致上做到在对数频率上以半倍频形式分解频段,并且金字塔编码本身符合逐级浮现的图像编码要求,高层子图显示全图概貌,在解码过程中能由上到下,逐层添加细节,直到较低层的子图满足要求。然而这种塔形分解并不理想,它使数据量增大了,并且没有体现方向性,因此,目前主要集中在性能更加优越的图像分解表达方法上,如小波变换,但多分辨率表达和复原图像的思想已经体现在各种图像编码的标准中。

4. 子带编码

子带编码最早用于语音编码,Wood 和 O'Neil 在 1986 年首先把子带编码应用于图像编码。子带编码先将原图用若干数字滤波器(分解滤波器)分解成不同频率成分的分量,再对这些分量进行亚抽样,形成子带图像,最后对不同的子带图像分别用与其相匹配的方法进行编码,在接受端,将解码后的子带图像补零、放大,并经合成滤波器的内插,将各子带信号相加,进行图像复原。

与 DCT(离散余弦变换)编码相比,子带编码的最大优点是复原图像无方块效应,因此得到了广泛的研究,是一种有潜力的图像编码方法。

各子带图像的统计特性是根据人眼对它们的敏感程度而定的,可以根据信息论和人眼的视觉特性,用不同的方法对各子带编码,其要点可以归结为以下两条。

- LL 分量最重要,应该较精确地编码,一般可用 DPCM、DCT 等方法,量化也要精确一些。
- 高频分量重要性差一些,不适于用 DCT 编码,由于用 DPCM 不比 PCM 好多少,所以通常直接用 PCM 编码,另外,为了降低码率及噪声的影响,高频分量的量化在零附近有个死区。

5. 矢量量化编码

矢量量化在图像编码中要优于标量量化,目前得到较为广泛的研究应用,主要用于空间域矢量量化、预测矢量量化、变量矢量量化、混合矢量量化,以及子带矢量量化、神经网络矢量量化等技术。

6. 图像压缩的国际标准

• JPEG

JPEG (Joint Photographic Experts Group) 是联合影像专家小组的缩写。JPEG 是为静止的灰度 (色彩) 连续变化的图像编码压缩而制定的标准。JPEG 定义了两种压缩算法, 一种是离散余弦变换 (DCT), 另一种是空间预测。DCT 算法偏重于图像的视觉效果, 用该算法得到输出图像的视觉效果相当好, 它与原图像的视觉效果几乎一样。空间预测采用了脉冲编码调制技术, 它偏重于无失真编码或基本无失真编码。

• JPEG2000

JPEG2000 压缩编码标准在保证失真率低和主观图像质量优于现有标准的条件下, 能够提供对图像的低码率的压缩。并且在速率畸变的情况下, 主观图像质量在性能上优于现行的 JPEG 标准。JPEG2000 标准是和 JPEG 标准兼容的, 但是它不是来代替它们的。JPEG2000 标准使用了基于小波的先进压缩技术, 期望在至少 10 年的时间里占据标准图像压缩领域的领先地位, 并且以它在高端应用和成像设备上表现出的优秀性能为基础, 开拓图像压缩尚未涉足的市场。

无论哪种压缩算法, 其压缩效率都与原始图像数据的分布特点密切相关, 没有哪一种压缩算法的效率针对任何图像数据而言, 能够高于其他算法。而且一般情况下, 压缩效率比较高的算法, 其具体的演算过程则相对要比较复杂, 从而需要更长的时间进行转换编码操作。这使得多种算法的并存成为可能和必要。

1.3 MPEG 技术基础

1. 发展历史

MPEG 是动态图像专家组 (Moving Pictures Experts Group) 的英文缩写, 该专家组始建于 1988 年, 专门负责制定视频和音频标准, 其成员均为音、视频和系统领域的技术专家。由于此小组提出并制定了 ISO/IEC1172 压缩编码标准, 使得 MPEG 扬名于世。现今所泛指 MPEG-X 版本, 是指一组由国际电信联盟 (ITU, International Telecommunications Union) 和国际标准化组织 (ISO, International Standards Organization) 制定并发布的音、视频数据的压缩标准。

MPEG 的成员最初打算开发 MPEG-1~MPEG-4 4 个版本, 以适用于不同带宽和数字影像质量的要求。后来由于 MPEG-3 被放弃, 所以只保留了 MPEG-1、MPEG-2 和 MPEG-4 3 个版本。后来又提出了 MPEG-7。

2. 功能特点

总体来说, MPEG 在 3 个方面优于其他压缩/解压缩方案。

- 兼容性好, 主要是因为它在一开始就被作为一个国际化的标准来研究制定。
- 能够达到更好的压缩比, 最高可达 200:1。
- 在提供高压缩比的同时, 数据损失造成的音、视频失真很小。

下面具体介绍3个MPEG版本各自的特点。

(1) MPEG-1

于1992年为工业级标准而设计,适用于不同带宽的设备,如CD-ROM、Video-CD等。针对SIF标准的分辨率(NTSC制为 352×240 pixel、PAL制为 352×288 pixel)的图像进行压缩时,可达如下性能。

- 传输速率达1.5Mbit/s,每秒播放30帧(NTSC制式)或25帧(PAL制式)。
- 具有CD(指激光唱盘)音质,质量级别基本与VHS相当。

MPEG的编码速率最高可达4~5Mbit/s,但随着速率的提高,其解码后的图像质量有所下降。

说明: NTSC是National Television Systems Committee的缩写,即全国电视制式委员会;PAL是Phase Alternation Line的缩写,逐行倒相制式。VHS是Video Home System的缩写,即家用录像系统。

MPEG-1也被用于数字电话网络上的视频传输,如非对称数字用户线(ADSL, Asymmetrical Digital Subscriber Line)、视频点播(VOD, Video-On-Demand)以及教育网络等。同时,MPEG-1也可被用做记录媒体或是在Internet上传输音频。

(2) MPEG-2

MPEG制定于1994年,其目标是为了保障高级工业标准的图像质量以及更高的传输率。下面是MPEG-2主要性能。

- 传输率为3~10Mbit/s。
- 在NTSC制式下的分辨率可达 720×486 pixel。
- 提供广播级的视像和CD级的音质。
- 音频编码可提供左右中及两个环绕声道,以及一个加重低音声道和多达7个伴音声道(这是DVD可有8种语言配音的原因)。
- 可提供一个较大范围的压缩比,以适应不同画面质量、存储容量以及带宽的要求。

MPEG-2表现出色,已能适用于高清晰度电视(HDTV, High Definition TV),从而使MPEG-3(原本为HDTV设计的)未发布就被放弃了。

MPEG-2除了做为DVD的指定标准外,还可为广播、有线电视网、电缆网络以及卫星直播(Direct Broadcast Satellite)提供广播级的数字视频。

由于MPEG-2在设计时的巧妙处理,使得大多数MPEG-2解码器也可播放MPEG-1格式的数据,如VCD。

对于电视用户来说,由于现有电视机分辨率的限制,MPEG-2所带来的高清晰度画面质量(如DVD画面)在电视上效果并不明显,倒是其音频特性(如加重低音,多伴音声道等)更引人注目。

(3) MPEG-4

MPEG的专家们还在为MPEG-4的制定而努力工作。此标准主要应用于视频电话(Video Phone)、视频电子邮件(Video E-mail)和电子新闻(Electronic News)等,其性能如下。

- 传输速率要求较低,为64kbit/s~4.8kbit/s。
- 分辨率为 176×144 pixel。