

第 3 章

多媒体数据编码标准

在多媒体系统中,多媒体数据编码标准是系统设计、开发和应用的基础。本章主要介绍数字图像和数字音频压缩标准。

3.1 静态图像压缩标准 JPEG

3.1.1 JPEG 标准的主要内容

ISO/IEC 10918 号标准“多灰度连续色调静态图像压缩编码”即 JPEG 标准选定 ADCT 作为静态图像压缩的标准化算法。

本标准有两大分类:第一类方式以 DCT 为基础;第二类方式以二维空间 DPCM 为基础。虽然 DCT 和 FFT 变换类似,是一种包含有量化过程的不能完全复原的非可逆编码,但它可以用较少的变换系数来表示,逆变换还原之后恢复的图像数据与变换前的数据更接近,故作为本标准的基础。另一方面,空间方式虽然压缩率低,但却是一种可完全复原的可逆编码,为了实现此特性,故追加到标准中。

在 DCT 方式中,又分为基本系统和扩展系统两类。基本系统是实现 DCT 编码与解码所需的最小功能集,是必须保证的功能,大多数的应用系统只要用此标准,就能基本上满足要求。扩展系统是为了满足更为广阔领域的应用要求而设置的。另一方面,空间方式对于基本系统和扩展系统来说,被称为独立功能。它们的详细功能如下。

(1) 基本系统。输入图像精度 8 位/像素/色,顺序模式,Huffman 编码(编码表 DC/AC 分别有两个)。

(2) 扩展系统。输入图像精度 12 位/像素/色,累进模式,Huffman 编码(编码表 DC/AC 分别有 4 个),算术编码。

(3) 独立功能。输入图像精度 2~16 位/像素/色,序列模式,Huffman 编码(编码表 4 个),算术编码。

3.1.2 JPEG 静态图像压缩算法

JPEG 定义两种相互独立的基本压缩算法，一种是基于 DCT 的有失真的压缩算法，另一种是基于空间线性预测技术(DPCM)的无失真压缩算法。

1. 基于 DPCM 的无失真编码

为了满足无失真压缩的需要,JPEG 选择一个简单的预测编码,处理过程见图 3.1。

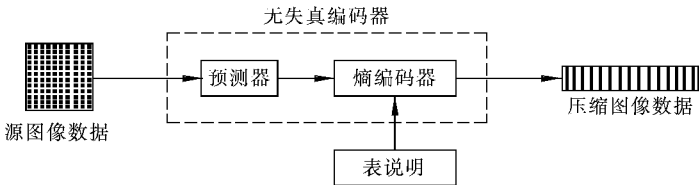


图 3.1 无失真编码器简化框图

这种编码的优点是硬件容易实现,重建图像质量好。缺点是压缩比太低,大约为 2 : 1。其中预测器工作原理是对 X 的预测值 $\bar{X}$,将 $X - \bar{X}$ 进行无失真熵编码。对 $\bar{X}$ 的求法见图 3.2 所给出的预测方式。

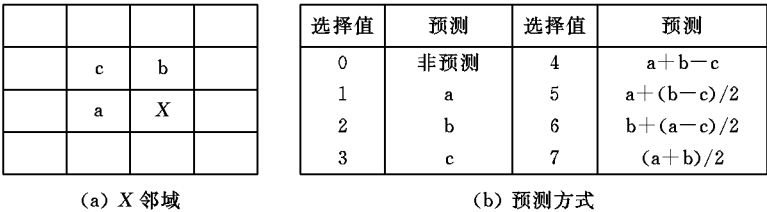


图 3.2 预测器

熵编码器采用 Huffman 编码或算术编码。

2. 基于 DCT 的有失真压缩编码

基于 DCT 压缩编码算法包括两种不同的系统,即基本系统和增强系统。增强系统是基本系统的扩充。基于 DCT 编码器和解码器的工作原理框图分别见图 3.3 和图 3.4。它们表示图像中一个单分量信号(Y,U,V)的压缩和解码过程。

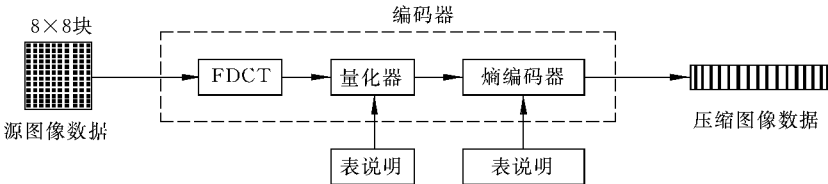


图 3.3 基于 DCT 编码过程

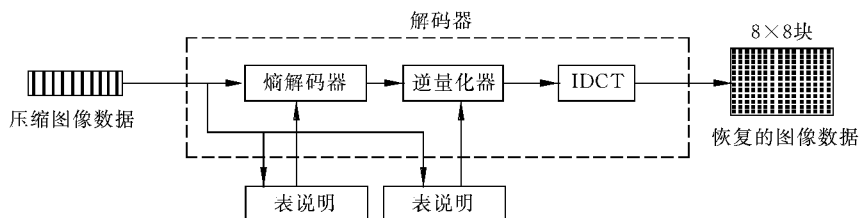


图 3.4 解码过程

(1) 离散余弦变换

① 首先把原始图像顺序分割成 8×8 子块；

② 采样精度为 P 位(二进制),把 $[0, 2^P - 1]$ 范围的无符号数变换成 $[-2^{P-1}, 2^{P-1}]$ 范围的有符号数,作为离散余弦正变换(forward DCT, FDCT)的输入；

③ 在输出端经离散余弦逆变换(inverse DCT, IDCT)后又得到一系列 8×8 子块,需将数值范围 $[-2^{P-1}, 2^{P-1}]$ 变换回 $[0, 2^P - 1]$ 来重构图像。

这里用的 8×8 FDCT 的数学定义为：

$$F(u, v) = \frac{1}{4} C(u) C(v) \left[\sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos\left(\frac{(2x+1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right) \right]$$

8×8 IDCT 的数学定义为：

$$f(x, y) = \frac{1}{4} \left[\sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F(u, v) \cos\left(\frac{(2x+1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right) \right]$$

$$\text{其中: } \begin{cases} C(u), C(v) = \frac{1}{\sqrt{2}} & \text{当 } u, v = 0 \\ C(u), C(v) = 1 & \text{其他} \end{cases}$$

下面的编码针对 FDCT 输出的 64 个基信号的幅值(称为 DCT 系数)来进行。

(2) 量化处理

这一步关键是找最小量化失真(误差)的量化器。为了达到压缩数据的目的,对 DCT 系数 $F(u, v)$ 需作量化处理,量化是一个“多到一”映射,这是信息损失的原因。JPEG 采用线性均匀量化器,定义为对 64 个 DCT 系数除以量化步长,然后四舍五入取整：

$$F^Q(u, v) = \text{Integer Round}[F(u, v)/Q(u, v)]$$

$Q(u, v)$ 是量化器步长,它是量化表的元素。量化表元素随 DCT 系数的位置和彩色分量不同有不同的值,量化表尺寸为 8×8 ,与 64 个变换系数一一对应。这个量化表应由用户规定(JPEG 给出参考值),并作为编码器的一个输入。表中每个元素值为 $1 \sim 255$ 的任意整数,对应 DCT 系数量化器步长。表 3.1 给出了亮度量化表,表 3.2 给出了色度量化表,是 JPEG 给出的参考值。

表 3.1 亮度量化表

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

表 3.2 色度量化表

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

量化的作用是在一定主观保真度图像质量的前提下,丢掉那些对视觉影响不大的信息,通过量化可调节数据压缩比。

(3) DC 系数的编码和 AC 系数的行程编码

64 个变换系数经量化后,坐标 $u=v=0$ 的 $F(0,0)$ 称 DC 系数,是直流分量,即 64 个空域图像采样值的平均值。相邻 8×8 块之间 DC 系数有强相关性。JPEG 对量化后的 DC 系数采用 DPCM 编码,即对 $\text{DIFF}=\text{DC}_i-\text{DC}_{i-1}$ 编码。如图 3.5 所示。

其余 63 个交流系数(AC)采用行程编码。从左上方 AC_{01} 开始沿对角线方向做 Z 字形扫描直到 AC_{77} 扫描结束(见图 3.6),这样可增加行程中连续 0 的个数。AC 系数编码的码字用两个字节表示,如图 3.7 所示。

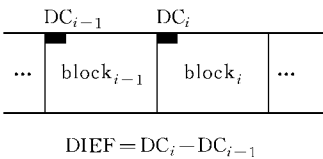


图 3.5 DC 系数差分编码

【例 3.1】 假设 AC 系数扫描结果中包含“...,3,0,0,0,0,0,12,0,0,...”数据,则对它的行程编码的结果为“..., (5,4), (12), ...”,其中 (5,4) 占用 1 个字节存放, (12) 占用 4

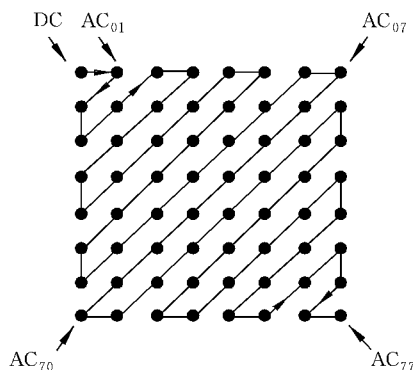


图 3.6 Z字形扫描

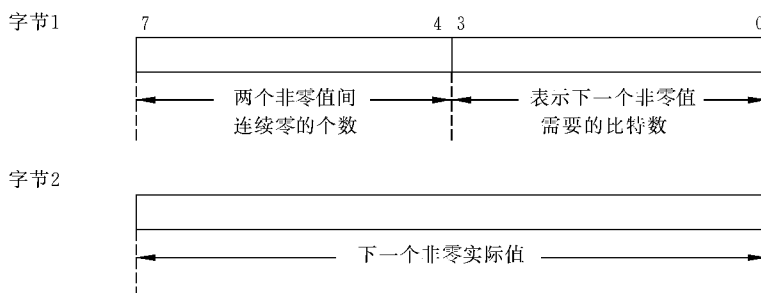


图 3.7 AC系数行程编码码字

位存放。

(4) 熵编码

为了进一步压缩数据,需对 DC 码和 AC 行程编码的码字再做基于统计特性的熵编码。JPEG 建议的熵编码是 Huffman 编码和自适应二进制算术编码。熵编码可分成两步进行,先把 DC 码和行程码转换为中间符号序列,然后给这些符号赋予变长码字。

① 熵编码的中间格式

熵编码的中间格式由两个符号组成:

符号 1: (行程, 尺寸)

符号 2: (幅值)

符号 1 中的第一个信息参数——“行程”表示前后两个非零 AC 系数之间连续 0 的个数,第二个信息参数——“尺寸”是后一个非零的 AC 系数幅值编码所需比特数。行程取值范围为 1~15,超过 15 时用扩展符号 1(15,0)来扩充,63 个 AC 系数最多增加 3 个扩展符号 1。编码结束时用(0,0)表示。“尺寸”取值范围为 0~10。

符号 2 中“幅值”用以表示非零 AC 系数的幅值,范围为 $[-2^{10}, 2^{10}-1]$ (最长 10 比

特),结构形式如表 3.3 所示。

表 3.3 符号 2 结构

尺 寸	幅 值	尺 寸	幅 值
1	-1, 1	6	-63...-32,32...63
2	-3...-2,2...3	7	-127...-64,64...127
3	-7...-4, 4...7	8	-255...-128,128...255
4	-15...-8, 8...15	9	-511...-256,256...511
5	-31...-16,16...31	10	-1023...-512,512...1023

对于直流分量 DC 也有类似于 AC 系数的编码格式：

符号 1: (尺寸)

符号 2: (幅值)

这里,“尺寸”表示 DC 差值的幅值编码所需的比特数,而“幅值”表示 DC 差值的幅值,范围为 $[-2^{11}, 2^{11}-1]$ 。可在表 3.3 中多加一级,幅值尺寸以 1~11 比特表示。

② 可变长度熵编码

将 63 个 AC 系数表示为符号 1 和符号 2 序列,其中行程长度超过 15 时,有多个符号 1;块结束(EOB)时仅有一个符号 1(0,0)。可变长度熵编码就是对上述序列进行变长编码。

对 DC 系数和 AC 系数中的符号 1 采用 Huffman 表中的变长码编码(VLC),这里 Huffman 变长码表必须作为 JPEG 编码器输入。符号 2 用码字长度在表 2.4 中给出的变长整数 VLI 码编码。VLI 是变长码,但不是 Huffman 码。VLI 的长度存放在 VLC 中, JPEG 提供 VLI 码字表供用户使用。

基于 DCT 的基本系统的编码器,只能存储两套不同的 Huffman 码表(每套 Huffman 码表中有一个 DC 表和一个 AC 表)。表 3.4 和表 3.5 分别给出直流系数 DC 的亮度分量典型的 Huffman 表和 DC 的色度分量典型的 Huffman 表。

典型的直流分量 Huffman 表说明如下。一般以 16 字节说明亮度 DC 系数表的码字长度：

X'00 01 05 01 01 01 01 01 01 00 00 00 00 00 00 00'

长度表中第 i 个元素值,表示长度为 i 的 Huffman 码个数。紧跟 16 字节之后是说明亮度表分类的一组值：

X'00 01 02 03 04 05 06 07 08 09 0A 0B'

同样方法,以 16 字节说明色度 DC 系数表的码字长度表：

X'00 03 01 01 01 01 01 01 01 01 01 00 00 00 00 00'



表 3.4 亮度 DC 系数表

分类	码长	码字	分类	码长	码字
0	2	00	6	4	1110
1	3	010	7	5	11110
2	3	011	8	6	111110
3	3	100	9	7	1111110
4	3	101	10	8	11111110
5	3	110	11	9	111111110

表 3.5 色度 DC 系数表

分类	码长	码字	分类	码长	码字
0	2	00	6	6	111110
1	2	01	7	7	1111110
2	2	10	8	8	11111110
3	3	110	9	9	111111110
4	4	1110	10	10	1111111110
5	5	11110	11	11	11111111110

紧跟 16 字节之后是说明色度表分类的一组值：

X'00 01 02 03 04 05 06 07 08 09 0A 0B'

同时,JPEG 也给出了交流系数 AC 的亮度分量典型的 Huffman 表和 AC 系数色度分量典型的 Huffman 表。这里不再详细说明,可查阅有关参考文献。

(5) 压缩比和图像质量

基于 DCT 的 JPEG 压缩算法,对中等复杂程度压缩效果的评价见表 3.6。

表 3.6 JPEG 压缩效果评价

压缩效果(比特/像素)	质 量	压缩效果(比特/像素)	质 量
0.25~0.50	中~好	0.75~1.5	极好
0.50~0.75	好~很好	1.2~2.0	与原始图像分不出来

3. 基于 DCT 的累进操作方式编码

基于 DCT 的顺序操作方式的编码过程是一次扫描完成的,基于 DCT 的累进操作方式编码方法基本与顺序方式一致,不同的是,累进方式中每个图像分量的编码要经过多次扫描才完成。第一次扫描只进行一次粗糙图像的扫描压缩,以相对于总的传输时间快得多的时间传输粗糙图像,并重建一帧质量较低的可识别图像,在随后的扫描中再对图像做较细的压缩,这时只传递增加的信息,可重建一幅质量提高一些的图像。这样不断累进,



直到满意的图像为止。

为了实现累进操作方式,需在量化器的输出与熵编码的输入之间,增加一个足以存储量化后 DCT 系数的缓冲区,对缓冲区中存储的 DCT 系数多次扫描,分批编码。

有以下两种累进方式。

(1) 频谱选择法。一次扫描中,只对 64 个 DCT 变换系数中某些频带的系数进行编码、传送,在随后的扫描中,对其他频带编码、传送,直到全部系数传送完毕为止。如选择分组为 $(0,1,2), (3,4,5), \dots, (61,62,63)$ 。

(2) 按位逼近法。沿着 DCT 量化系数有效位(表示系数精度的位数)方向分段累进编码。如第一次扫描只取最高有效位的 n 位编码、传送,然后对其余位进行编码、传送。如扫描分段依次为 7654 位,3 位, $\dots$, 0 位。

4. 基于 DCT 的分层操作方式

分层方式是将一幅原始图像的空间分辨率,分成多个分辨率进行“锥形”的编码方法,水平(垂直)方向分辨率的下降以 2 的倍数因子改变,如图 3.8 所示。过程如下:

① 把原始图像空间分辨率降低。

② 对已降低分辨率的图像采用基于 DCT 的顺序方式、累进方式或无失真预测编码中的任何一种编码方法进行编码。

③ 对低分辨率的图像解码,重建图像,使用插值滤波器,对它插值,恢复图像的水平和垂直分辨率。

④ 把分辨率已升高的图像作为原始图像的预测值,对它们的差值采用基于 DCT 的顺序方式、累进方式或用无失真方式进行编码。

⑤ 重复③、④直到图像达到完整的分辨率编码。

在必须使用低分辨率的设备来存取或观察高分辨率图像的应用中,这种方式非常有效。

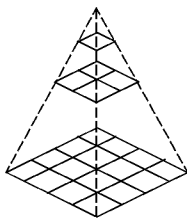


图 3.8 分层操作方式

3.1.3 JPEG 2000 简介

多媒体应用的快速发展,尤其是基于 Internet 的多媒体应用,给图像编码提出了新的要求。为此,2000 年 12 月公布的新的 JPEG 2000 标准(ISO 15444),其目标是在高压压缩的情况下,如何保证图像传输的质量。

JPEG 采用以 DCT 变换为主的分块编码方式,DCT 变换考查整个时域过程的频域特征或整个频域过程的时域特征。JPEG 2000 作为 JPEG 升级版,采用以小波变换为主的多分辨率编码方式。小波变换对时域或频域的考查都采用局部的方式,小波在信号分析中对高频成分采用由粗到细渐进的时空域上的采样间隔,所以能够像自动调焦一样看清远近不同的景物,并放大任意细节,是构造图像多分辨率的有效方法。



JPEG 2000 统一了面向静态图像和二值图像的编码方式,是既支持低比率压缩又支持高比率压缩的通用编码方式。该算法主要特点如下。

(1) 高压缩率(低比特速率)。与 JPEG 相比,可修复约 30% 的速率失真特性。也就是说,在新的算法下,JPEG 和 JPEG 2000 在压缩率相同的情况下,JPEG 2000 的信噪比将提高 30% 左右。

(2) 无损压缩。预测编码作为对图像进行无损编码的成熟方法被集成在 JPEG 2000 中,使它能实现无损压缩。

(3) 渐进传输。JPEG 2000 可实现以空间清晰度和信噪比为首的各种各样的可调节性,从而实现渐进传输,即具有“渐现”特性,这是 JPEG 2000 的一个极其重要的特征。它先传输图像的轮廓,然后逐步传输数据,不断提高图像质量,让图像由朦胧到清晰显示,而不必像现在的 JPEG 一样,由上到下慢慢显示。

(4) 感兴趣区域压缩。JPEG 2000 支持所谓的“感兴趣区域”(region of interest, ROI)特性,可以任意指定图像上感兴趣区域的压缩质量,还可以选择指定的部分先解压缩。这样就可以很方便地对图像感兴趣的部分采用低压缩比以得到较好的压缩效果,对其他部分采用高压缩比以节省存储空间。

从多方测试的结果来看,JPEG 2000 的压缩效果更优秀,特别在要求高压缩比的场合下表现更加突出。在中度与低度的压缩比率下,传统的 JPEG 表现得非常出色。但是在较高的压缩比率之下,传统的 JPEG 压缩方式就不太令人满意。用 JPEG 压缩的图像,在压缩比较高的情况下,有明显的马赛克现象(比如出现较大色斑或颜色信息丢失),但是用 JPEG 2000 压缩的图像效果就能得到保证。即使在很高的压缩比下,图像的内容也很容易辨别。另外,JPEG 2000 有两个方面重要特征得到广泛关注。

(1) 纠错能力很强。在文件传输过程中,JPEG 2000 有恢复丢失的数据包的能力。

(2) 可以指定最后文件尺寸大小。在用户定义文件尺寸的情况下,JPEG 2000 保证再现较高图像质量的能力。

这适合目前带宽受到限制的 Web 系统和无线网络上传输图像,应用前景很广。

3.2 运动图像压缩标准 MPEG

3.2.1 MPEG 标准简介

MPEG 标准是面向运动图像压缩的一个系列标准。最初 MPEG 专家组的工作项目是 3 个,即在 1.5Mbps,10Mbps,40Mbps 传输速率下对图像编码,分别命名为 MPEG-1, MPEG-2, MPEG-3。1992 年,MPEG-2 适用范围扩大到 HDTV,能支持 MPEG-3 的所有功能,因而 MPEG-3 被取消。同时为了满足不同的应用要求,MPEG 又将陆续增加其他



一些标准,如 MPEG-4、MPEG-7 和 MPEG-21。本节先介绍 MPEG-1、MPEG-2 和 MPEG-4 的主要内容,然后在其他章节介绍 MPEG-7 和 MPEG-21 的主要思想。

MPEG-1 标准名称为“用于大约高达 1.5Mbps 速率的数字存储媒体的运动图像及其伴音编码”(coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbps),作为 ISO/IEC 11172 号建议于 1992 年通过。该标准分为 4 个部分:①MPEG-1 系统(11172-1),定义音频、视频及有关数据的同步;②MPEG-1 视频(11172-2),定义视频数据的编码和重建图像所需的解码过程,其处理的是 SIF(standard interchange format)格式,即 NTSC 制式为 352 像素×240 行/帧×30 帧/秒,PAL 制式为 352 像素×288 行/帧×25 帧/秒;③MPEG-1 音频(11172-3),定义音频数据的编码和解码;④一致性测试(11172-4)。另外,MPEG-1 标准还提供了软件模拟的技术报告(11172-5)。

MPEG-2 标准名称为“运动图像及其伴音信息的通用编码”(generic coding of moving pictures and associated audio),作为 ISO/IEC 13818 号建议于 1994 年发布。该标准分为 10 个部分:①MPEG-2 系统(13818-1),规定音频、视频及有关数据的同步;②MPEG-2 视频(13818-2),规定视频数据的编码和解码,支持多种格式;③MPEG-2 音频(13818-3),规定音频数据的编码和解码;④MPEG-2 一致性测试(13818-4);⑤MPEG-2 软件模拟(13818-5);⑥MPEG-2 数字存储媒体命令和控制(DSM-CC)扩展协议(13818-6),用于管理 MPEG-1 和 MPEG-2 的数据流,使数据流既可在单机上运行,又可在异构网络环境下运行;⑦MPEG-2 高级声音编码(AAC,13818-7),是多声道声音编码算法标准,这个标准除后向兼容 MPEG-1 音频标准外,还有非后向兼容的声音标准;⑧MPEG-2 系统解码器实时接口扩展标准(13818-9),它用来适应来自网络的传输数据流;⑨MPEG-2 DSM-CC 一致性扩展测试(13818-10);⑩MPEG-2 高级声音编码标准修正版。MPEG-2 Part8(13818-8)原计划用于采样精度为 10B 的视频图像编码,但由于工业界兴趣不大而暂停开发。

MPEG-4 标准名称为“甚低速率视听编码”(very-low bitrate audio-visual coding),1998 年作为 ISO/IEC 14496 号标准草案发布的 MPEG-4 文件有 5 个部分:①MPEG-4 系统(14496-1);②MPEG-4 视频(14496-2);③MPEG-4 音频(14496-3);④MPEG-4 一致性测试(14496-4);⑤MPEG-4 参考软件(14496-5);⑥MPEGF-4 传输多媒体集成框架(DMIF)。

3.2.2 MPEG-1 系统

MPEG-1 标准没有规定编码器和解码器的体系结构或实现方法,但对它们提出了功能和性能上的要求。一个典型的 MPEG-1 编解码器原型如图 3.9 所示。