

第一章 激光影碟机的技术基础

激光影碟机又称激光视盘机，在家用电器消费类领域是继录像机之后的又一代新宠。它集精密机械、激光技术、微电子技术和计算机技术于一身，发展迅速。其中有许多新技术已经渗透到其他工业领域，具有广阔的发展前景，代表了当今信息科学的最新成就，比如，数据压缩技术、纠错技术、调制技术和数据通信技术。本章主要介绍数码激光影碟机的基本概念以及所要用到的一些新技术。

1.1 概述

激光视盘技术（又简称 AV 技术）是一个内容比较广泛的概念，从技术角度讲主要包括两层意思：其一，信息记录过程，即采用激光方式记录声音或图像或一般数据信息，通常把记录以上信息的圆盘形的媒体叫做光盘；其二，信息重放过程，即采用激光方式从光盘上拾取信息，再进行信号的还原，最后送至播放设备，通常把具有信息拾取功能和信号还原能力的设备称为激光影碟机。应该说，光盘和激光影碟机是一对“孪生兄弟”，因此任何一种激光视盘技术的发展过程和普及程度必须依赖这对“孪生兄弟”的“同步成长”。

激光视盘技术从 1972 年诞生至今经历了以下几个发展阶段：

- 从模拟到数字（比如，从 LD 发展到 VCD）；
- 从低密度小容量到高密度大容量（比如，从 VCD 发展到 DVD）；
- 从线性到压缩（例如，从 CD 到 VCD）；
- 从单放到可录放（例如，从 DVD-ROM 到 DVD-RAM）。

1.1.1 激光影碟机的发展历史

光盘和激光影碟机的品种繁多，功能各异，但普通老百姓在日常生活中能够广泛接触到的和拥有的不外乎以下几种：即 LD、CD、CD-ROM、VCD、DVD 和超级 VCD。

1. LD 系统

1972 年，美国 RCA 公司和荷兰飞利浦公司联合开发并发表了 LD(Laser Disc) 激光视盘系统，从此开创了以激光和高密度方式记录信息为标志的光盘时代。1981 年 10 月，经日本先锋公司再度开发后，LD 系统正式大批量进入商品市场。

但是，LD 系统仍然采取模拟处理方式记录音频和视频信号，即用调频和频分复用技术把音频和视频信号调制在各自的频带上，然后再用光刻方式记录在 LD 光盘上。在重放过程中，激光头拾取 LD 光盘上的信息，经 FM 解调后再还原成记录前的模拟的音频和视频信号。

2. CD 系统

1982 年，由日本索尼公司和荷兰飞利浦公司共同开发的 CD 唱机问世，标志数字光盘时代的到来。实际上，CD 是 CD-DA(Compact Disc Digital Audio) 的简称，是数字化技术、编码技术、纠错技术和大规模集成电路飞速发展的必然结果。

CD 系统采用了全新的数字化方式处理音频信号，在信号记录过程中，先将模拟的音频信

号数字化，使之变成 16 位的二进制数；再经过编码处理用光刻方式记录在 CD 光盘上。在重放过程中，激光头拾取 CD 光盘上的数字信息，经过解码和数/模(D/A)转换，使之还原成模拟的音频信号。

3. CD-ROM 系统

1983 年，人们将 CD 系统所具有的优良性能应用于计算机领域，从而开发出 CD-ROM (Compact Disc Read Only Memory) 系统，并掀起了一场印刷业的革命，开辟了无纸印刷的新时代。CD-ROM 主要用于存储文字和数据，其显著特点是容量巨大，是电子出版物的理想媒体。

4. VCD 系统

在 1992 年，经过多方努力，活动图像专家组 MPEG (Moving Picture Experts Group) 正式发布 MPEG-1 标准。该标准是一种全活动图像的压缩标准，也是 VCD (Video CD) 技术的核心。日本索尼公司和荷兰飞利浦公司在 1993 年 3 月推出了基于该压缩标准的卡拉 OK CD。同年 8 月，飞利浦、JVC、索尼和松下等 4 家公司终于一致同意在卡拉 OK CD 的规格基础上加以发展，制定了 VCD 标准，即 VCD1.1 版本。日本 JVC 等公司使卡拉 OK CD 进一步实用化，将其改称为 VCD1.0 版本。VCD 是第一代全部采用数字化技术的激光视盘系统，它不仅包含有一对数字立体声音频信号，还包括一路全活动的数字图像信号。虽然 VCD 系统的图像清晰度赶不上原来的 LD 系统，但它毕竟是从模拟到数字的一次飞跃，而且通过压缩技术可将长达 74 分钟的数字化活动图像和伴音信号刻录在一张直径只有 12cm 的 VCD 盘上。

由于 VCD 盘比 LD 要便宜得多，且与 CD 兼容，再加上我国已经积累了良好的生产 CD 系统的工业基础，因此 VCD 一经问世，便迅速进入千家万户，得到广大普通消费者的青睐。可以说我国 VCD 市场的发展速度非常迅猛，据行家估计 VCD 用户高达 5000 万左右，目前尽管城镇市场已日趋饱和，但在广阔的农村还存在较大的发展空间。

5. DVD 系统

从技术角度来看，VCD 系统的确存在两个明显的缺陷：其一，VCD 的清晰度不高，尚不及 LD 系统；其二，VCD 系统的播放时间不够长，单张 VCD 光盘放不下一部电影。

1993 年 7 月 MPEG 发布了适用于高清晰度的图像和环绕立体声的 MPEG-2 标准，在此基础上提出了图像质量和声音质量均佳的 DVD (Digital Versatile Disc) 系统。作为家庭用的第一代 DVD 影碟机，已于 1996 年 11 月在欧美、日本等一些发达国家开始上市。我国也有不少的“AV 发烧友”或高收入家庭购买了松下公司生产的松下 A300MU 型 DVD 影碟机。但是，由于 DVD 系统从光盘和影碟机两个方面比 VCD 要贵一些，而且要达到理想的播放效果 DVD 系统要求有较高清晰度的新一代彩色电视机和较好的音响设备，因此期望 DVD 系统马上像 VCD 一样迅猛发展和大范围普及是不太现实的。

6. 超级 VCD 系统

为了达到尽可能好的播放效果，同时又能让广大消费者接受，在众多 VCD 生产厂家和国外集成电路开发供应商的共同努力下，1998 年上半年开始先后出现两种介于 VCD 和 DVD 之间的新产品 CVD (China Video Disc) 和 SVCD (Super VCD)。

从本质上讲，CVD 和 SVCD 这两种技术没有太大的区别，其开发的出发点也有异曲同工之妙。但是由于这两种技术分别得到国内 VCD 影碟机生产厂家的两个主力联盟和国际上不同的机芯和解码芯片开发商的支持，从而爆发了一场激烈的标准之争。直到 1998 年 9 月 29 日，国家信息产业部颁布了我国“超级 VCD”新标准——《SJ/T11196-1998 超级 VCD 系统技术规范》电子行业标准（此标准从 1998 年 11 月 1 日正式实行），才结束了 CVD 和 SVCD 标准

之争,将 VCD 和 SVCD 统称为超级 VCD。

为了清晰地了解不同类型激光影碟机的特性,读者可从表 1.1 得到较全面的认识。

表 1.1 不同类型激光影碟机的特性

比较项目	VCD	超级 VCD	DVD
视频压缩标准	MPEG-1	MPEG-2	MPEG-2
音频编码	MPEG-1 层 2 单立体声	MPEG-1 层 2, 双立体声 可扩展为 AC-3	MPEG-2 层 2 多声道 AC-3
字幕	低清晰度	高清晰度, 支持 4 种文字	高清晰度, 支持 8 种文字
外径	120mm	120mm	120mm
盘厚	1.2mm	1.2mm	1.2mm(2×0.6mm)
坑长	0.9μm~3.3μm	0.9μm~3.3μm	0.40μm~2.13μm
坑宽	0.4μm	0.4μm	0.4μm
坑深	0.13μm	0.13μm	0.11μm
信息轨间距	1.67μm	1.67μm	0.74μm
光盘容量(每盘可播放时间)	74min	45min	135min(单面单层)
激光波长	780nm	780nm	650(635)nm
数值孔径	0.45μm	0.45μm	0.6μm
调制方式	EFM	EFM	EFM+
纠错方式	CIRC	CIRC	RS-PC
机芯速度	单倍速	双倍速	相当 12 倍速~24 倍速
采样频率	6.75MHz	9.0MHz	13.5MHz
帧频	25Hz(PAL) 29.97Hz(NTSC)	25Hz(PAL) 29.97Hz(NTSC)	25Hz(PAL) 29.97Hz(NTSC)
极限带宽	3.375MHz	4.50MHz	6.75MHz
视频码率	固定码率 1.15Mb/s	可变码率 最小:1.15Mb/s 最大:2.60Mb/s	可变码率 最小:1.15Mb/s 最大:9.80Mb/s
数据码率	固定比特率:1.47Mb/s	可变比特率:0~2.5Mb/s	可变比特率:0~26Mb/s
水平和垂直像素结构	352×288(PAL) 352×240(NTSC)	480×576(PAL) 480×480(NTSC)	704×576(PAL) 704×480(NTSC)
极限水平分辨率	268 线	357 线	535 线
兼容性	VCD,CD	CVD,SVCD VCD,CD	DVD,CVD SVCD,VCD,CD

1.1.2 “3D” 之争

在人们还没有从 VCD 的热潮中回过神来,并期望 DVD 能早日走进普通老百姓的家庭时,新型的激光影碟机 CVD 和 SVCD 又横空出现在我们眼前。从目前的激光影碟机市场发展趋势来看,在我国影碟机市场似将形成“3D 大战”的局面。这里的“3D”是指 VCD 超级 VCD

(包括 CVD 与 SVCD)和 DVD。有人比喻:当前的“3D大战”形同 1700 年前的“魏蜀吴”三国鼎立。尽管蜀国偏安,吴国据险,但两者最后均被实力强大的魏国吃掉。当今的 DVD 就是“魏国”,其一统天下的局面只是时间问题。

虽然超级 VCD 采用较为简单的驱动器,但需要三碟的机芯才能放完一整部电影。因此,超级 VCD 的零售价格也应该在人民币 900 元~1200 元之间,它比 DVD 影碟机约便宜人民币 700 元~800 元。

尽管 DVD 的价格还较高,1998 年全世界只售出 80 万台(美国占 35 万台),但对全世界来说,DVD 市场非常重要,中国 DVD 市场尤其重要。中国 VCD 市场的发展经历了一段辉煌的时期,拥有开发 DVD 市场牢固的技术基础和市场消费潜力。预计 1999 年 DVD 影碟机的销量将与 VCD 影碟机持平,大约在 1000 万台~1200 万台之间。800 万台 DVD 是可以实现的销售量,其中美国、日本和中国各 200 万台,其他地区 200 万台。2000 年,根据市场分析专家的预测,DVD 的产量大约 2000 万台,其中中国约占 30% 而 VCD 将减少到 500 万台。到 2001 年,DVD 将在中国的数字影碟机市场中占主导地位。

1.2 模/数转换技术及其应用

在数码激光视盘系统中,记录在光盘上的信息都是数字信号,所谓数字信号是指在时间上和数值大小上均不连续的离散信号。目前的音频信号源(声音)和视频信号源(图像)都是模拟信号,模拟信号是指在时间和数值大小上均连续变化的信号。因此在把它们刻到光盘上之前,就必须先进行模拟信号到数字信号的转换,简称模/数转换,这种转换过程常称为模拟信号的数字化。

1.2.1 模拟信号的数字化

把完成上述模/数转换过程的器件称为模/数转换器,常用 ADC(Analog to Digital Converter)表示。一次完整的模/数转换过程包括以下 3 个阶段。

- 采样(Sampling):每间隔一定时间对连续变化的模拟信号观测一次,得到的观测值叫做采样值;
- 量化(Quantization):是用一个与采样值最接近的整数来代替采样值的过程,该整数叫做量化结果;
- 编码(Code):是用“0”和“1”这两个数码位的不同组合形式(组合规律)来表示量化结果的过程。

1. 采样过程

在刻录光盘之前,必须每隔一定时间从要记录的声音信号/图像信号等连续变化的模拟量中提取一个观测值,这个观测值也就是人们常说的某一时刻的采样值。经过采样过程之后,连续变化的模拟信号变成了一个个时间上等间距、数值大小分散的离散信号,如图 1.1 所示。其中相邻的两个采样点之间的时间间隔叫做采样周期(又称取样周期),常用 T_s 表示,而反映采样速度快慢程度的量常用采样频率 f_s 来表征,采样周期 T_s 和采样频率 f_s 满足以下关系:

$$f_s = 1/T_s$$

由图 1.1 可见,要想用一系列离散的采样值来体现被采样信号的变化情况,必须使相邻的两次采样过程之间的时间 T_s 尽可能短,即采样频率 f_s 应比较高。形象地说,只有当采样周期

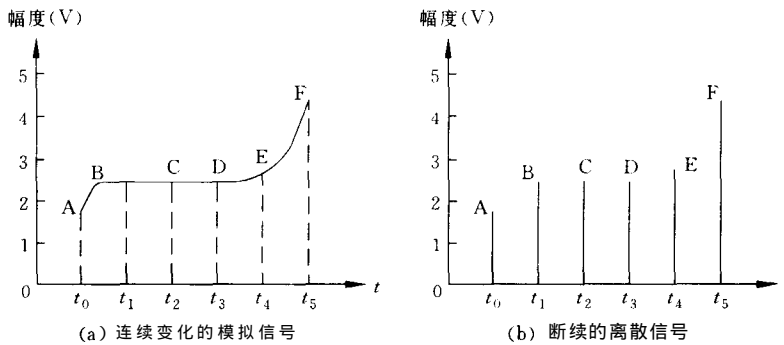


图 1.1 采样过程的示意图

比较小时，图 1.1(b)中的 A~F 各点依次相连后的折线才会和图 1.1(a)中的曲线非常相似（理想状态时相同）。虽然提高采样频率可改进这两条曲线的接近程度，但是在单位时间内所要处理的数据量却加大了。

实际上，从物理实现上考虑采样频率不可能无限制地增大，因为模/数转换器在将模拟信号转换成数字信号时，需要一定的时间，这个时间就是模/数转换时间。显然，当采样周期 T_s 小于模/数转换时间时，无疑得不到期望的采样结果。这一点间接地告诉我们一个道理：对于速度变化很快的信号，要求选择转换时间比较短的模/数转换器。

事实上，一味地提高采样频率，是“弊大于利”的保守思想。对于变化速度比较缓慢的模拟信号，在短时间内反复进行采样是完全没有必要的。比如，由于人体的体温在短时间内（如 1 分钟~2 分钟）是不会有太大变化的，因此医院的护士并没必要每隔 3 分钟~5 分钟为病人测一次体温，道理就这么简单。这个例子告诉我们又一个道理：采样频率的选择必须考虑被采样信号的变化快慢，准确地讲，采样频率的确定必须满足以下采样定理。

所谓采样定理是指：当对连续变化的模拟信号进行采样时，若采样频率高于该信号最高频率 f_{max} （信号变化最快的部分所对应的频率）的 2 倍，则可以在采样值的基础上利用插补技术正确地恢复原信号的波形，否则将会引起频谱混叠（Aliasing），产生混叠噪声（Aliasing Noise），如图 1.2 所示。

2. 量化过程

由于模拟信号在时间上是连续变化的，因此采样值很可能刚好不是整数。为此常采用四舍五入的归整方法用一个与采样值最接近的整数来表示之，这个数据处理过程就叫做量化。把完成量化任务的电路称为量化器（Quantizer），量化器的输入信号是采样值，输出信号是量化结果。

显然，数据经过量化处理之后不可避免地会出现舍入误差，引起信号的失真，产生量化噪声。在数字系统中，量化结果只能用二进制数的整数倍来表示。量化单位 q 自然是二进制数的最低有效位 LSB(Least Significant Bit) 所代表的物理量，量化误差则为 $\pm q/2$ 。

例如，将一个 0~5V 的电压信号（模拟量）转换为 8 位二进制数（0000 0000~1111 1111），则 q 代表的物理量为 $5V/256$ ，它同时反映了量化的基本“台阶”。也就是说，当两个模拟信号的大小之差的绝对值小于 $q/2$ 时，量化的结果会是相同的。比如，采样值 1V 和 1.0098V 经 8 位量化后，其结果均为二进制数 0011 0011（十进制数 51），即不能区分两者的大小。但是，在将 0~5V 的电压信号（模拟量）转换为 16 位二进制数（16 个 0~16 个 1）时，采样值 1V 和 1.0098V

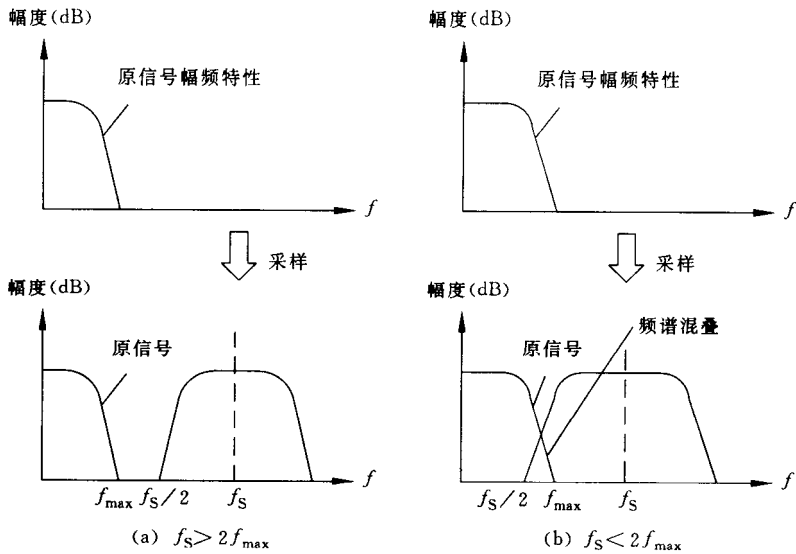


图 1.2 采样引起的频谱变化

经 16 位量化后，其结果分别为二进制数 0011 0011 0011 0011（十进制数 13107）和 0011 0011 1011 0011（十进制数 13235），此时能够较好地区分这两个电压信号。

3. 编码过程

编码就是要把量化结果变成二进制的数码序列，在模 / 数转换过程中，对双极性（有正和负之分）信号常采用以下 3 种编码方法：

- 符号-数值码：和计算机中的原码表示方法相同，用最高位表示数据的极性，为“0”表示正数，为“1”表示负数，其他位表示数据的绝对值大小，如“0001”表示 +1，“1001”表示 -1；
- 偏移二进制码：一种直接的二进制编码，最高位表示数据的极性，为“1”表示正数，为“0”表示负数，其他位表示对满刻度的偏移，如“1111”表示 +7，“0001”表示 -7，“1000”表示 +0 和 -0；
- 二进制的补码：和计算机中的补码表示方法相同，正数的补码 = 正数的原码，负数的补码为符号位不变，原码的其他位按位取反再加 1。

1.2.2 数字化的特点和优势

当今的信息社会可以说已经进入了一个数字化时代，数字化的产品可谓琳琅满目、比比皆是，如移动通信系统、数码相机、数字视听设备和各种计算机控制系统等。

数字化技术具有以下特点：

- 数字信号可以进行有效的压缩，以便快速传输，节省资源，如卫星通信系统；
- 数字信号存储方便，便于携带和复制，如数码相机中的照片可直接移植到计算机中；
- 数字信息容易加密，安全性好，如数字手机不易被“码机”；
- 在数字信息中可以很容易地加入纠错信息，信号不易失落，如 VCD 的纠错功能；
- 数字信息便于混合和分离，如多媒体电脑；
- 数字信号可在大规模数字集成电路上进行快速处理，硬件成本较低，如股票接收机；

- 数字信号的抗干扰能力很强，不易受其他信号源的影响，如数字电视发射系统。

1.2.3 音频信号的数字化

1. 采样频率的选择

人耳朵能听到的声音信号的频率范围是 20Hz~20kHz，可闻的上限频率是 20kHz，因此采样频率应在 40kHz 以上。为了避免高于 20kHz 的高频信号对采样过程的影响，在采样前需要对声音信号进行滤波。考虑到滤波器对输入的声音信号在 20kHz 处有 10% 的衰减作用，因此用 22kHz 的 2 倍(44kHz)作为声音信号的采样频率比较合理。为了使 VCD 的软件与 CD 兼容，故将音频信号的采样频率选为 44.1kHz。

2. 量化位数的确定

前述的例子说明：当 ADC 的位数足够多时，舍入误差将很小，可以认为数字信号非常逼近采样信号。也就是说，量化的位数越多，量化误差就越小，数字化处理的精度也就越高。但是，随着量化位数的增加，ADC 的价格将比较昂贵，数据量也将随之加大，数字化的技术复杂性和设备成本会成倍增长。

在数字激光视盘系统中，声音信号的量化位数一般选为 16 位，理论上的动态范围可达

$$20\lg 2^{16} = 96(\text{dB})$$

此时，实际的听觉效果已经很不错。

3. 数据传输的码率

在 VCD 系统中，音频信号的采样频率选为 44.1kHz，量化位数是 16 位。对立体声音响信号（左右两声道）进行数字化，每秒钟要传送的数据位数达

$$44.1 \times 10^3 \times 16 \times 2 = 1.41 \times 10^6 (\text{b/s}) = 1.41 (\text{Mb/s})$$

即音频信号的码率为 1.41 Mb/s。

1.2.4 视频信号的数字化

从采样过程来看，视频信号的数字化和音频信号有不少的共性。但是，从实现数字化的具体技术来讲，前者要复杂得多。一般地说，视频信号的数字化有全信号数字化和分量数字化两种不同的形式。前者对全电视信号直接进行数字化；后者对亮度信号 Y、色差信号 $C_R (=R-Y)$ 和 $C_B (=B-Y)$ 分别进行数字化，然后利用时分复用原理进行再处理。在数字激光视盘系统中，普遍采用分量数字化技术。

分量数字化主要有以下两个好处：

- 亮度和色差分开处理，相互间不存在干扰，可提高图像重放的质量；
- 有利于将常用的 PAL 制式和 NTSC 制式彩色电视统一起来。

1. 采样频率的选择

基本的电视制式有正交平衡调幅 NTSC 制式和逐行倒相正交平衡调幅 PAL 制式两种，这两种制式都是按隔行扫描方式进行工作的。一幅图像即有水平扫描，又有垂直扫描。当对图像进行采样时，很自然会存在一个采样点的分布问题。一般把由采样点构成的排列结构图，称为采样结构。如果前后两帧（即两幅图像）的采样点在空间位置上完全一样的话，则可给数字处理电路带来很多方便，有利于提高图像重放的质量，甚至于很容易实现特技处理。视频信号的两种典型采样结构如图 1.3 所示。

在图 1.3(a) 中，圆圈表示亮度信号的采样点，黑色三角形表示色差信号的采样点。在图

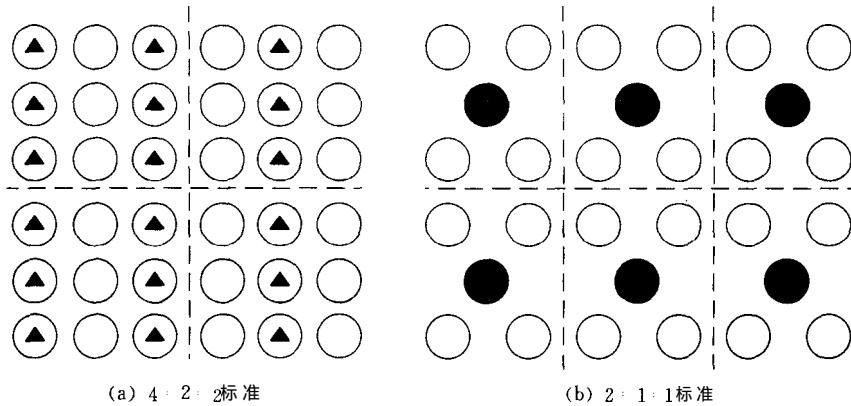


图 1.3 视频信号的采样结构示意图

1.3(b) 中，空白圆圈表示亮度信号的采样点，黑色圆圈表示色差信号的采样点。

若希望达到上述目的，则要求视频信号的采样频率是行频的整数倍，以便扫描线与像素同步。由于 PAL 制式的行频是 15 625Hz 而 NTSC 制式的行频却为 15 734.266Hz 因此，能被这两个行频整除的采样频率必须满足以下关系

$$x \times 15\,734.266 = y \times 15\,625$$

其中 x, y 应取最小的正整数，由上式可以求得

$$x = 143, y = 144$$

故，满足上述要求的最小采样频率应是 2.25MHz。

(1) 亮度信号的采样频率 f_s

在对亮度信号进行数字化处理时，采样频率的选择必须考虑以下因素：

根据采样定理，采样频率至少应是被采样信号上限频率的 2 倍（考虑采样前的低通滤波器在 f_{\max} 处会带来 10% 的衰减，通常选 2.2 倍）。PAL 制式的 625 行扫描方式要求有 5.8MHz~6MHz 的带宽，NTSC 制式的 525 行扫描方式要求有 5.6MHz 的带宽。

为了在采样后使混叠噪声足够小，要求采样频率是信号带宽的 2.2 倍~2.7 倍。

采样频率 f_s 必须是行频的整数倍，以利于行间、场间和帧间的信号处理。

为了兼容 PAL 制式和 NTSC 制式，两者应采用相同的采样频率 f_s 。

根据上述条件，亮度信号的采样频率 f_s 选定为

$$f_s = 6 \times 2.25 = 13.5(\text{MHz})$$

(2) 色差信号 C_R 和 C_B 的采样频率

为了获得满意的彩色图像，色差信号的带宽应达 2MHz。采用与亮度信号采样频率相似的确定方法，可将色差信号 C_R 和 C_B 的采样频率确定为 6.5MHz。该频率是亮度信号的采样频率的一半，且正好是 PAL 制式行频的 432 倍、NTSC 制式行频的 429 倍。

2. 量化位数的确定

对于单极性的电视信号，其信噪比（信噪比是指信号峰值与量化噪声的均方根之比）可由下式确定：

$$S/N = 6n + 10.8 (\text{dB})$$

式中， n 为量化位数。在数字激光视盘系统中，视频信号的量化位数一般为 8 位，其信噪比可达 58.8dB。当然，量化位数越多，信噪比自然比较高。但由于视频信号的采样频率达几兆赫，因此

量化位数每增加一位，就会带来巨大的数据量。

3. 数据传输的码率

视频信号的数字化有 4 : 2 : 2 和 2 : 1 : 1 两种标准，分别适用于 DVD 和 VCD 详细的参数见表 1.2。前者的采样频率为

$$Y : C_R : C_B = 13.5\text{MHz} : 6.75\text{MHz} : 6.75\text{MHz}$$

后者的采样频率为

$$Y : C_R : C_B = 6.75\text{MHz} : 3.375\text{MHz} : 3.375\text{MHz}$$

表 1.2 两种 MPEG 视频标准的图像格式参数

图像格式 参数名称	4 : 2 : 2 标准(适用于 DVD)		2 : 1 : 1 标准(适用于 VCD)	
	PAL 制式	NTSC 制式	PAL 制式	NTSC 制式
每帧行数	625	525	625	525
场频	50Hz	60Hz	50Hz	60Hz
帧频	25Hz	30Hz	25Hz	30Hz
行频	15625Hz	15734.266Hz	15625Hz	15734.266Hz
亮度 Y 的采样频率	13.5kHz	13.5kHz	6.75kHz	6.75kHz
色差 C _B 的采样频率	6.75kHz	6.75kHz	3.375kHz	3.375kHz
色差 C _R 的采样频率	6.75kHz	6.75kHz	3.375	3.375kHz
一行的总像素个数	864	858	432	429
一行的有效像素个数	720(704)	720(704)	352	352
有效行数	576	480	288	240
每帧的亮度像素个数	720(704)×576	720(704)×480	352×288	352×240
每帧的色差 C _B 像素个数	720(704)×576/4	720(704)×480/4	352×288/4	352×240/4
每帧的色差 C _R 像素个数	720(704)×576/4	720(704)×480/4	352×288/4	352×240/4
每帧的总像素个数	622080(608256)	518400(506880)	152064	126720
数据传输的码率	124.416(120.512) Mb/s	124.416(121.6512) Mb/s	30.4128 Mb/s	30.4128 Mb/s

(1) DVD 系统的码率

图像的分割

一般把一幅完整的图像称为一帧，一帧 PAL 制式和 NTSC 制式的电视信号分别由 625 行和 525 行组成，对应的帧频分别为 25Hz 和 30Hz。即在 PAL 制式中，每秒传送 25 帧图像，在 NTSC 制式中，每秒传送 30 帧图像(注在电影中每秒传送 24 帧图像)。

为了对图像进行数字化处理，必须对每帧图像进行分割。在数字激光视盘系统中，一帧图像的分割方法如图 1.4 所示，详细的分割步骤如下：

- 将一帧图像横向切成若干条 (PAL 制式 :18 条 ;NTSC 制式 :15 条)，每一条称为一片 (Slice)；
- 每一片再纵向切成 22 块 (PAL 制式和 NTSC 制式)，称之为宏块 Macro Block) 也称大块；
- 将每个宏块分 3 层，一层亮度 Y，两层色差 C_B 和 C_R)；
- 人的视觉要求亮度比色度有更高的清晰度，故将宏块的亮度部分再细分成 4 个块；

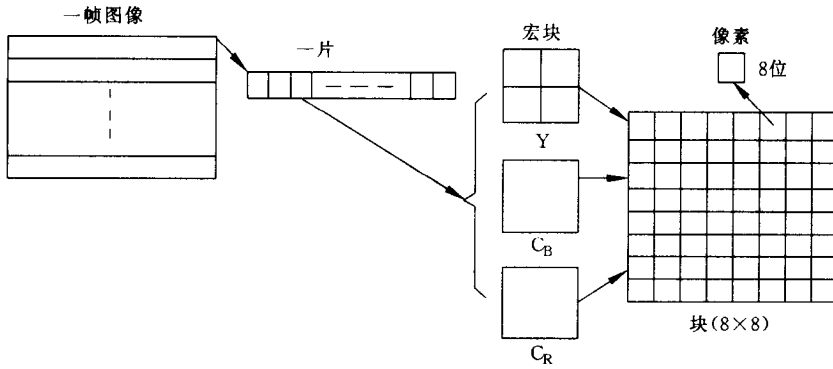


图 1.4 图像分割方法的示意图

● 最后将前述的 6 个块(Block)分成 64 个(8×8)像素小块。

综上所述，图像的分割方法可归纳为：连续图像 → 1 帧图像 → 切成片 → 切成宏块 → 切成块 → 分割成像素小块。在 MPEG 标准中，宏块是视频信号压缩处理的基本单元，视频信号的数字化实际上是对像素小块而言的。每个像素用一个 8 位二进制数表示，因此每个采样点的亮度具有 $256(2^8=256)$ 级灰度，由此可见图像明暗变化的层次是足够的。

基本参数的确定

像素是构成图像的最小单元，且像素越多，图像的清晰度越高，但数据量将会加大。图像的分割原则是既要达到比较满意的清晰度，又要保证目前的数字电路有足够的处理能力去处理这些数据。

在 4:2:2 标准中，亮度 Y 的采样频率为 13.5MHz (采样时间为 $0.074\mu\text{s}$) 即每秒钟有 13.5 兆个采样点 (简称样点)。用采样频率除以行频就可得到每一行亮度信号的样点数，即

$$\text{PAL 制式：每一行亮度信号的样点数} = 13.5\text{MHz} \div 15.625\text{kHz} = 864$$

$$\text{NTSC 制式：每一行亮度信号的样点数} = 13.5\text{MHz} \div 15.734\text{266kHz} = 858$$

通常，把一行中表示图像部分的称为有效行，把行同步头的前沿称为 1 行的起点 O_H 。PAL 制式的 O_H 到有效行的起始点间隔 $9.778\mu\text{s}$ 占用 132 个样点的时间。有效行的结束点到下一行的起始点间隔 $0.889\mu\text{s}$ 占用 12 个样点的时间。因此，PAL 制式的有效样点数 (从第 133 个样点开始到第 852 个样点为止) 为

$$864 - 132 - 12 = 720$$

对于 NTSC 制式，有效行的样点从第 123 个样点开始到第 842 个样点结束，有效的样点数也是 720 个点。

在 525 行的 NTSC 制式中，采用隔行扫描方式，场频为 60Hz，两场为一帧。场的始端是垂直同步信号，脉宽为 22.5 行 (在此区间内是不传送视频信号的，谓之垂直消隐)，因此在一帧图像信号中有效行数实际只有

$$525 - 2 \times 22.5 = 480$$

同样的道理，在 625 行的 PAL 制式中，一帧图像信号的有效行数实际只有

$$625 - 2 \times 24.5 = 576$$

经过上述分析和计算，不难得到在每帧图像中亮度信号的样点数 (像素个数)，即

$$\text{PAL 制式：} 576 \text{ 行} \times 720 \text{ 个样点/行} = 414\,720 \text{ 个样点}$$

$$\text{NTSC 制式：} 480 \text{ 行} \times 720 \text{ 个样点/行} = 345\,600 \text{ 个样点}$$

由于每个色差信号在整个宏块中只占亮度信号的 1/4，因此两个色差所具有的像素个数应是亮度的 1/2，即为 207 360 个样点(PAL 制式)和 172 800 个样点(NTSC 制式)。

最后根据亮度信号的样点数和色差信号的样点数可计算出一帧图像的总样点数，即

$$\text{PAL 制式: } 414\,720 + 207\,360 = 622\,080 \quad (\text{个样点})$$

$$\text{NTSC 制式: } 345\,600 + 172\,800 = 518\,400 \quad (\text{个样点})$$

根据图像分割原理，每个样点(像素)用 8 位二进制数表示，传送图像信号的数据码率应为：

$$\text{PAL 制式: } 8\text{bit/样点} \times 622\,080 \text{ 样点/帧} \times 25 \text{ 帧/秒} = 124.416(\text{Mb/s})$$

$$\text{NTSC 制式: } 8\text{bit/样点} \times 518\,400 \text{ 样点/帧} \times 30 \text{ 帧/秒} = 124.416(\text{Mb/s})$$

值得注意的是：在 DVD 系统中还选用“每行 704 个样点数”的图像格式，对应有另外一种数据传输的码率，详见表 1.2。

(2)VCD系统的码率

VCD 系统采用 2:1:1 标准，其采样频率是 DVD 的一半。由图 1.3 的采样结构可知，VCD 的有效行数和有效行的样点数都只有 DVD 的 1/2。用和 DVD 相似的参数确定方法，可以求得以下数据。

在一帧图像中，亮度信号和色差信号的总样点个数为

$$\text{PAL 制式: } 352 \text{ 个样点/行} \times 288 \text{ 行} + 176 \text{ 个样点/行} \times 144 \text{ 行} \times 2 = 152\,064 \text{ 个样点}$$

$$\text{NTSC 制式: } 352 \text{ 个样点/行} \times 240 \text{ 行} + 176 \text{ 个样点/行} \times 120 \text{ 行} \times 2 = 126\,720 \text{ 个样点}$$

在 VCD 系统中，传送图像信号的数据码率应为：

$$\text{PAL 制式: } 8\text{bit/样点} \times 152\,064 \text{ 样点/帧} \times 25 \text{ 帧/秒} = 30.4128(\text{Mb/s})$$

$$\text{NTSC 制式: } 8\text{bit/样点} \times 126\,720 \text{ 样点/帧} \times 30 \text{ 帧/秒} = 30.4128(\text{Mb/s})$$

1.3 压缩和解压缩原理

在介绍压缩/解压缩原理之前，先来讨论一下为何要对数据进行压缩。我们已经知道，刻录在数码光盘上的是数字信号，这些信号来源于模拟信号的采样和量化。比如，现在要按照 2:1:1 标准的 PAL 制式编码方式记录一部 120 分钟的电影节目，且只考虑图像部分，也不采取纠错措施，其数据总量达

$$152\,064 \text{ 样点/帧} \times 25 \text{ 帧/秒} \times 60 \text{ 秒/分} \times 120 \text{ 分} = 27\,371.52 \text{ 兆样点}$$

即使采用 8 位量化(一个样点占一个字节)，此时也要用 42 张 CD-ROM 光盘(其存储容量达 650MB)才能装得下如此之多的数据，由此可见数据压缩的现实意义。

1988 年，国际标准化组织 ISO(International Standardization Organization)和国际电工委员会 IEC(International Electro-technical Commission)成立了运动图像专家组，简称 MPEG(Moving Picture Experts Group)，专门负责制定运动图像及其伴音的编码标准。MPEG 下设 3 个委员会，即 MPEG 系统委员会(MPEG System)、MPEG 视频委员会(MPEG Video)和 MPEG 音频委员会(MPEG Audio)这 3 个委员会分别制定对应的 MPEG 子标准即 MPEG 系统标准、MPEG 视频标准和 MPEG 音频标准。现在有些人将 MPEG 标准称为运动图像压缩标准，这显然是不恰当的。

由于 MPEG 制定的编码标准实质上是要解决信号的压缩和解压缩问题，因此由该专家组讨论通过的编码标准统称为 MPEG 压缩技术。目前已经实用化的两个压缩标准是 MPEG-1

和 MPEG-2 分别适用于 VCD、超级 VCD 和 DVD。这两种技术的最大区别在于：

- 数据传输的码率不同，MPEG-1 的码率固定为 1.5Mb/s，MPEG-2 的码率的推荐值在 4Mb/s~15Mb/s 之间；
- 场的结构不同，MPEG-1 采用逐行扫描，在 1 秒钟内只有 25/30 场画面（PAL 制式 / NTSC 制式），而 MPEG-2 采用隔行扫描，在 1 秒钟内有 50/60 场画面（PAL 制式 / NTSC 制式）；
- 音频的兼容性不同，MPEG-2 音频兼容 MPEG-1 立体声、MPEG-2 环绕立体声和杜比 AC-3 环绕声。

1.3.1 视频压缩编码和解压缩原理

所谓视频压缩编码是指，在数字图像信息处理中，采用适当的压缩编码技术识别和剔除数据冗余，以降低数据传输的码率。压缩后的数据经解码器处理之后可以形成一幅“新”图像。衡量视频压缩技术是否科学，关键看以下两个方面：

- 用最少的数据代码“表达”被压缩的原图像信息，且失真很小；
- 人眼一般感觉不到新图像与原图像有什么明显的区别。

1. 视频压缩原理

视频压缩技术的基本工作原理是：利用信息论、人类视觉生理学和心理学以及景物特征分析等方面的知识，消除以下 3 种数据冗余，即像素间冗余、视觉信息冗余和编码冗余。在激光视盘系统中，通过帧内压缩和帧间压缩两种方法来消除图像信息中的冗余信息，以达到数据压缩的目的。

(1) 帧内压缩技术

通过观察、分析之后发现，人眼的视觉生理学和心理学特性允许压缩编码的新图像相对于原图像存在一定的失真。因为一般人受分辨力的限制对这些失真部分可能并不敏感，比如，对于一幅（一帧）完整的正面人头像（原图像），观众对其面部各个部位的重视程度和感观结果往往是完全不一样的。一般情况下，眼部和唇部的线条对面部表情和特征的影响比较大，比较容易得到重视；而额头和面颊等部分的轮廓变化较少，灰度层次差别不明显，往往不易引起观众的注意。也可以说，这幅图像的各个部位在观众记忆中所占的信息量是不同的，一段时间后，一个有绘画基础的观众可通过回忆在纸上“再现”一幅与原图像差不多的画像。因此，为了使新图像能够比较真实地反映原图像，需要用较多的数据“表达”前者，而后者所用的数据可以适当少一些。相对于整幅图像用相同数据量来“表示”的情况而言，实际“记住”的信息量少多了，可以达到压缩图像信息的目的。

从本质上讲，上述信息压缩方法利用了相邻像素的相关性，这里的相关性是指图像比较规则、细节变化比较平缓的图像，其相邻像素之间的关系比较密切，即从其中一个像素的值可推测出另一个像素的值，也就是说存在信息冗余。由于这种压缩是在一帧图像中完成的，故称帧内压缩，显然它减少的是空间冗余信息。

帧内压缩主要采用了以下 3 种技术：

- 离散余弦变换 DCT (Discrete Cosine Transform)；
- 自适应量化；
- 变步长编码 VLC (Variable Length Coding)。

离散余弦变换 DCT 编码实际上是一种坐标变换编码方法，具体的编码过程由 DCT 变换、

频率系数量化、系数排序和熵编码等 4 个步骤组成。

① DCT 变换

DCT 变换把图像分割得到的 8×8 数据块 (64 个像素) 映射成频率域的系数, 如图 1.5 所示。频率系数的计算公式为

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

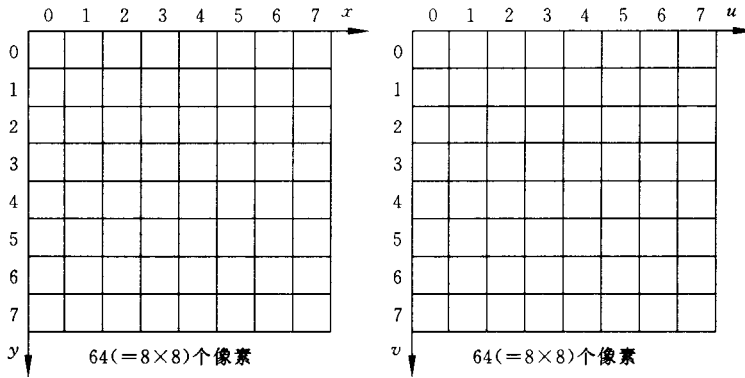


图 1.5 离散余弦变换 DCT 的几何示意图

式中 $f(x, y)$ 为待变换的像素数据, $F(u, v)$ 表示变换后的频率系数, 变换因子 $C(u)$ 和 $C(v)$ ($u = v = 0, 1, \dots, 7$) 满足

- 当 $u = v = 0$ 时 $C(u) = C(v) = 2/\sqrt{2}$;
- 当 $u \neq 0$ 时 $C(u) = 1$;
- 当 $v \neq 0$ 时 $C(v) = 1$ 。

根据频率系数计算公式不难看到: $F(0, 0)$ 等于 64 个像素的代数和除以 8, 相当于直流分量; 随着 u, v 的增加, 相应系数分别代表水平方向和垂直方向的频率分量。

假如某幅图像的一个亮度像素的采样结果如表 1.3 所示, 为了方便起见将像素值用十进制数表示。利用频率系数计算公式, 可求得每个像素值的频率系数, 最后得到经 DCT 变换后的频率系数表, 如表 1.4 所示。

表 1.3 一个亮度块的像素 (例子)

$y \backslash x$	0	1	2	3	4	5	6	7
0	26	22	18	20	24	31	42	47
1	27	24	21	21	23	29	38	44
2	27	25	23	22	22	27	36	43
3	32	27	19	20	23	29	37	43
4	35	28	17	19	24	30	39	44
5	26	24	21	22	23	30	40	45
6	21	22	25	24	23	31	42	45
7	25	24	24	26	27	33	42	45

由表 1.4 可见，左上角部分的数值较大，右下角部分的数值较小。这是因为左上角反映的是亮度信号的直流和低频部分，右下角表示亮度的高次谐波。亮度像素块变化不大，则低频分量较大；反之高频分量加大，低频分量减小。如果频率系数的数值较大，则说明亮度起伏较大，该区域的图像轮廓较细致；反之则说明亮度变化较平缓。

表 1.4 亮度像素的 DCT 频率系数 (例子)

$v \backslash u$	0	1	2	3	4	5	6	7
0	232.3	49.0	41.7	-2.0	2.3	2.7	-2.6	-0.9
1	-4.5	2.2	4.8	2.1	2.3	-1.7	-0.0	0.1
2	1.9	-8.9	-6.4	-4.1	-2.2	2.1	0.7	0.9
3	-0.5	-4.2	3.3	2.5	0.1	0.2	-0.9	-0.0
4	3.3	0.7	3.7	6.3	2.2	-1.1	-1.5	-0.4
5	-1.6	-2.6	-1.5	-2.7	-1.6	0.9	0.6	0.2
6	0.0	-0.4	-0.7	-1.0	0.6	-0.1	-0.1	-0.1
7	0.2	0.2	0.5	-1.0	-0.2	0.3	-0.1	0.4

频率系数量化

量化的目的是在确保图像质量不会明显下降的前提下，舍弃一些对视觉效果影响不大的次要信息，达到压缩数据的目的。事实上，人眼对图像量化误差的敏感程度存在“掩盖效应”，即画面变化越快，量化误差的影响越不容易被觉察，因此对高频部分（即空间快速运动部分）的量化系数可选大一些，这就是所谓的自适应量化器，其量化步长见表 1.5。

由表 1.5 可见，对低频部分的频率系数进行“精”量化，量化误差小；对高频部分的频率系数进行“粗”量化，量化误差大，这样一来就可用较少的数据来表达该亮度信息。

经过自适应量化后就得到亮度像素的 DCT 量化系数，如表 1.6 所示。

表 1.5 自适应量化器的步长

$v \backslash u$	0	1	2	3	4	5	6	7
0	8	16	19	22	26	27	29	34
1	16	16	22	24	27	29	34	37
2	19	22	26	27	29	34	34	38
3	22	22	26	27	29	34	37	40
4	22	26	27	29	32	35	40	48
5	26	27	29	32	35	40	48	58
6	26	27	29	34	38	46	56	69
7	27	29	35	38	46	56	69	83

表 1.6 亮度像素的 DCT 系数 (例子)

$v \backslash u$	0	1	2	3	4	5	6	7
0	29	3	2	0	0	0	0	0
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

系数排序

由表 1.6 可见, 经过量化后的 DCT 系数通常会在右下角出现许多数值为零的系数。在图像复用编码器中用如表 1.7 所示的“之”字形扫描方法把二维矩阵排列的数据转换成一维数列, 使绝大部分的零元素集中在一起。经重新排序后, 可得到以下数据序列:

29	3	0	0	0	2	"0"×58
----	---	---	---	---	---	--------

表 1.7 “之”字形读数的顺序

$v \backslash u$	0	1	2	3	4	5	6	7
0	0	1	5	6	14	15	27	28
1	2	4	7	13	16	26	29	42
2	3	8	12	17	25	30	41	43
3	9	11	18	24	31	40	44	53
4	10	19	23	32	39	45	52	54
5	20	22	33	38	46	51	55	60
6	21	34	37	47	50	56	59	61
7	35	36	48	49	57	58	62	63

熵编码

熵编码是 DCT 编码的最后一步, 它是一种基于量化系数统计特性的无失真编码方法。常用的熵编码有游程长度编码 RLC(Run Length Coding) 和霍夫曼(Huffman) 编码。

在激光视盘系统中普遍采用霍夫曼编码, 霍夫曼编码是 Huffman 于 1952 年提出的一种最佳符号编码方法。它完全依据字符出现的概率来构造平均长度最短的异字头码字, 实现霍夫曼编码的顺序是:

- 将待编码的信源符号按其出现概率的大小由小到大依次排列;
- 将概率最小的两个分支进行组合相加, 并将概率较大的分支放在上面, 反复进行这一步直到概率达到 1.0 为止;

- 在组合分支的上边标上“1”，下边标上“0”；
- 画出每个信源符号从其概率开始到概率为 1.0 结束的路径，并记下在该路径上先后出现的“1”或“0”，得到由“1”和“0”组成的一个数码序列；
- 最后将上述数码序列从右到左倒过来排列就可得到霍夫曼编码。

例如，对一个由 7 个符号组成的信源，有如图 1.6 所示的编码过程。

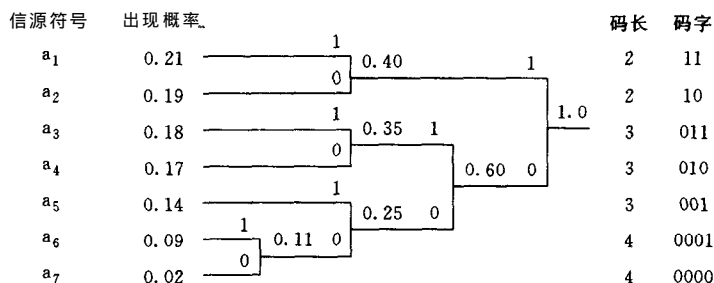


图 1.6 霍夫曼编码过程示意图

(2) 帧间压缩技术

为了使观众感觉不到图像画面的跳动和闪烁，在彩色电视机中采用了每秒传送 25 或 30 幅画面（一幅画面又称一帧图像）的扫描刷新方法。在连续的许多画面中，相邻的两帧图像具有很大的相关性，即同一空间位置处的亮度和色差信号相差不多，存在较多的冗余信息。因此，在记录和保存“动态图像”时（本质上讲是许多静止图片在时间轴上的快速出现），不必把每帧图像的全部信息都拷贝下来，可以只完整记录前一幅画面以及后续的若干帧与前一帧的差别，达到数据压缩的目的。

所谓帧间压缩技术是指：舍弃时间上先后出现的各帧画面的冗余信息，在前一幅画面的基础上，用后一帧与前一帧的差别推算后一帧图像，使图像数据得到压缩。由于帧间压缩针对的主要对象是不同时间的信息冗余现象，故称之为时域冗余压缩技术。

帧间压缩技术是建立在下述三种帧的基础上的，即：

① I 帧

I 帧 (Intracoded Frame) 是场景更换后的第一帧画面，是产生后续其他画面的基础。I 帧编码采用全帧编码传输，解码器仅用 I 帧的数据即可重构一幅完整的图像，要传输的信息量大小由画面本身决定，即 I 帧为解码器随机访问图像信息流提供了一个定位点。I 帧的数据压缩率比较小，在视频数据流中占有比较大的信息量。

② P 帧

P 帧 (Forward Predictive Frame) 是前向预测帧，反映了当前图像画面与前一帧图像的差值信息，可以把它看成是运动图像的变化部分。在参考帧的基础上利用 P 帧可预测出当前帧的完整图像，参考帧可以是前述的 I 帧，也可以是下面将要介绍的 B 帧。同时当前的 P 帧又可作为后续 P 帧或 B 帧的参考帧，因此 P 帧的错误具有扩散性。一旦出现错误，则其错误要传递到下一个 I 帧为止。

③ B 帧

B 帧 (Bidirectionally Predicted Interpolative Coded Frame) 是双向预测内插编码帧，反映了当前图像与前一帧和后一帧的差值信息。B 帧包含的信息最少，不能作为其他帧的参考帧，当然也就不会引起解码错误的扩散。

2. 视频压缩编码器

为了使图像数据得到大量的压缩，MPEG 的视频编码方式没有采用每帧图像依次完全传送的方法，而是采用一种专门设计的图像交互编码方法。

为了能对图像进行及时编辑，MPEG 标准规定每 6 帧或每 5 帧图像为一组，并称之为帧组 GOP(Group of Picture),NTSC 制式和 PAL 制式的帧组排列示意图见图 1.7。

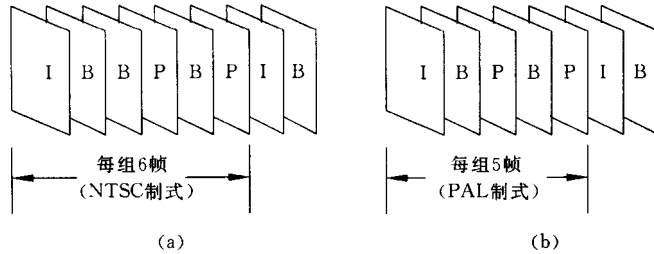


图 1.7 帧组的排列图

由图 1.7 可见,NTSC 制式(帧频为 30Hz)的一个帧组有 6 帧,PAL 制式(帧频为 25Hz)的一个帧组有 5 帧,因此两种制式对应的每个帧组的时间刚好都是 0.2s 保证每 0.2s 有一个图像进出编辑点,并且在每个帧组中安排一个 I 帧,I 帧的前面有一个 B 帧,确保每一个帧组具有独立形成图像的能力。

显然,在一个帧组中数据量多的 I 帧只有一个,其余均为数据量少的 P 帧或 B 帧,大大减少了数据传输的信息量,达到了视频信号压缩的目的。

MPEG-1 视频编码器就是要按照 MPEG-1 视频编码标准将输入的帧组转换为串行输出的码流,在经过其他处理之后刻录到光盘上。MPEG-1 视频编码器主要由帧重排、I 帧编码、P 帧编码、B 帧编码和图像复用编码器等组成,其基本构成框图如图 1.8 所示。

(1) 帧重排

顾名思义,帧重排是指:在编码前应将上述帧组中的 5 个或 6 个帧的顺序重新排列,目的是为了便于解码器对 P 帧和 B 帧的处理。以图 1.7(b)为例,帧组的排列次序是 IBPBP,若不进行帧重排,则先读得的是 I 帧,后读出的是 B 帧,再其次才是 P 帧,...。由于在获得 B 帧时,尚未得到 P 帧,故此时的 B 帧是不可解码的,因 B 帧的解码必须借助于其前后两帧信息。帧重排的原则是将 P 帧排在 B 帧之间,重排后的顺序应是 IPBPB。

同理,对于 NTSC 制式,帧重排前的顺序是 IBBPBP,重排后的顺序应为 IPBBPB。

(2) I 帧编码

当输入 I 帧时,开关 S_1 和 S_2 置于上方, S_3 置于左边。在对 I 帧编码时,以宏块为单位按照图像分割的步骤依次进行,直到 I 帧编码结束为止。I 帧编码的具体顺序是:

- 先对帧内的各个宏块进行离散余弦变换,将 8×8 的像素块变成频率系数;
- 再利用自适应量化器对频率系数进行量化,量化后的结果分成两路输出,即
- 一路送图像复用编码器,先进行“之”形扫描,再进行无损编码,并与其他辅助信息一起经缓冲器输出视频码流;
- 另一路经 S_3 、反量化器和离散余弦逆变换 DCT^{-1} ,还原成变换前的 I 帧图像数据,存入 I 帧存储器,供后续的 P 帧和 B 帧编码使用。

(3) P 帧编码

当输入 P 帧时,开关 S_1 和 S_2 置于下方, S_3 置于左边。P 帧一路送到加法器,另一路送至运