

## 4.2 MPEG-2 基本码流结构

基本码流是指包含视频、音频或数据的连续码流。基本码流的结构和内容是根据各种数据的编码格式而定的。

### 4.2.1 MPEG-2 视频基本码流

为了便于对图像序列的随机访问和编辑，MPEG-2 对视频基本数据流规定了分层的结构，和详细的语法和语义。如图 4-3 所示，从顶层开始依次为：第一层是图像序列层，它包含序列头、若干个图像组层数据以及序列终止符。下面的第二层是图像组，它由定义的一组或多组帧内编码帧（I 帧）以及/或非帧内编码帧（P 帧以及/或 B 帧）图像组成。每组包括组头和图像层数据。第三层是图像层本身，由图像头和宏块条层数据组成。第四层是宏块条层，它由条层头和若干个连续的宏块组成。第五层是宏块层，由宏块头加块层数据组成，图像以亮度数据矩阵为基准，分为  $16\times 16$  像素的宏块，作为运动补偿的基本单元宏块分为亮度宏块和色度宏块，色度宏块大小与抽样格式有关。第六层是块层由图像数据和块结束符组成，是进行离散余弦变换的单元。

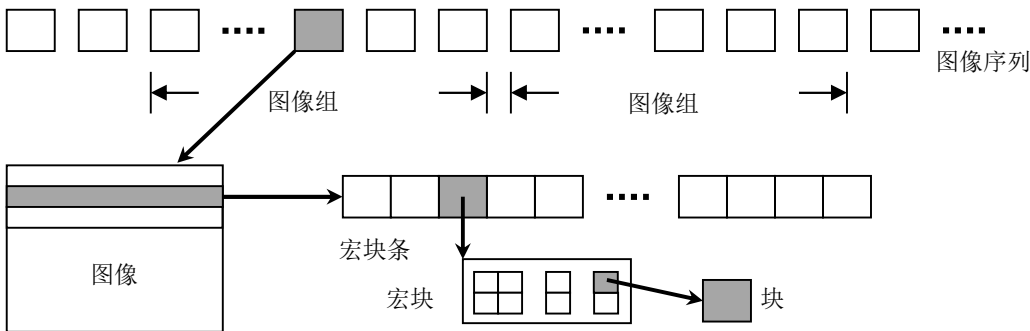


图 4-3 MPEG-2 视频数据流结构分层

在这六层数据结构相应的序列头、图像组头、图像头、宏块条头、宏块头等头部包含了对 MPEG-2 码流进行解码所需的起始码、定时、以及其他相关的参数和信息，如图 4-4 所示。

块是 DCT 变换运算的基本单位，代表亮度或色度的  $8\times 8$  矩阵，在基本视频码流中，首先发送直流系数，然后是其它频率系数，最后以 EOB 结束。块组合成宏块，MPEG-2 编码算法的基础是混合的 DCT/DPCM 编码方法与宏块结构、运动补偿的结合，宏块是运动估计和运动补偿的基本单元，宏块分为亮度宏块和两个色度宏块，色度宏块的大小与抽样格式有关，抽样格式为 4:2:0，色度宏块大小为  $8\times 8$ ，4:4:4 格式的色度块大小为  $16\times 16$ ，4:2:2 格式的色度块大小为  $8\times 16$ ，如图 8-1-3a、b、c 所示。

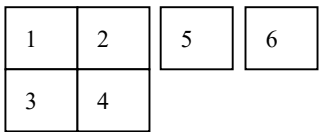


图 4-5a 4:2:0 宏块结构

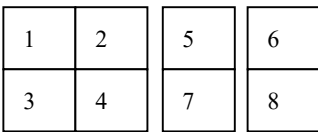


图 4-5b 4:2:2 宏块结构

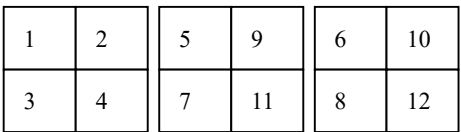


图 4-5c 4:4:4 宏块结构

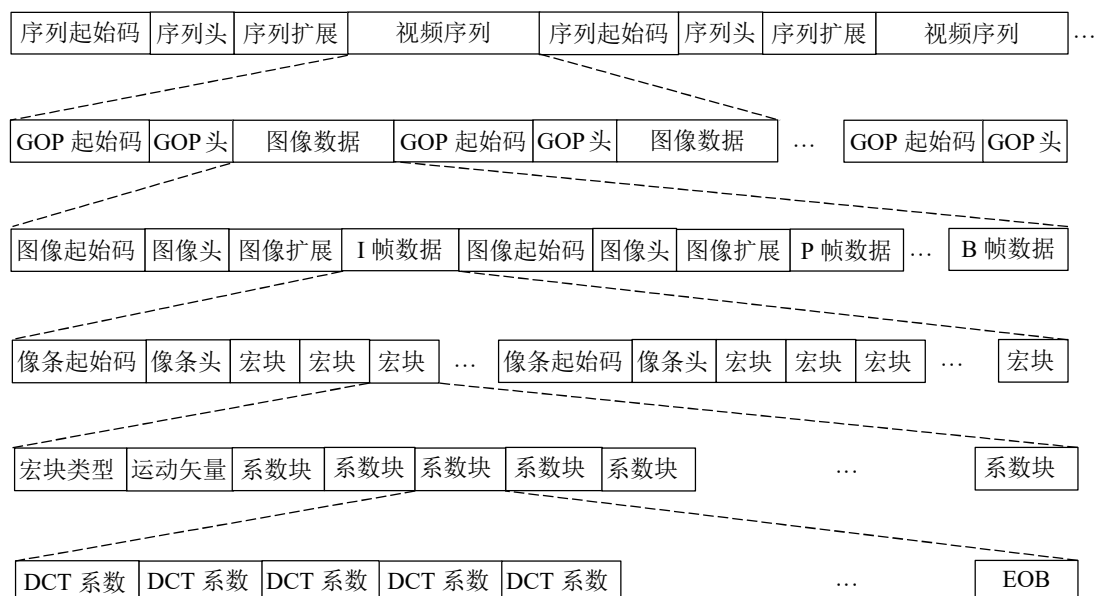


图 4-4 MPEG-2 视频基本码流结构

每个宏块在包头信息中含有二维运动矢量，在 B 帧图像中，运动矢量还可以是前向和后向的，运动补偿可以以场或帧为基础，并在包头信息中标明，系数重新量化使用的量化尺度也已经标明。宏块组合成宏块条，在 MPEG 中，宏块条可以从任何地方开始，并且大小任意，宏块条是可变长度同步和差值编码的基本单位。宏块条的第一个运动矢量必须以完整的方式发送，而剩下的矢量则以差值方式发送。在 I 帧图像中，宏块条的第一个直流系数以完整的方式发送，其余的直流系数则以差值方式发送。宏块条组合起来构成有效的图像帧，图像的包头信息表明了图像是 I、P 或 B 编码，还含有时间标记，确保图像在正确的时间显示，同时还有一个总的运动矢量，而单个宏块的运动矢量则用与总的运动矢量的差值来表示。图像组成一个由 I 帧开始的 GOP（图像组），GOP 是帧间编码的基本单位，GOP 可以是开放的，也可以是封闭的。在封闭的 GOP 中，最后一个 B 帧图像不要求用下一个 GOP 的 I 帧图像来解码，这样数据流可以在 GOP 的末端剪断。若干个 GOP 可以组成一个视频序列，视频序列由序列起始码开始，然后是序列头，最后以序列结束码终止。在序列中还可以设置额外的序列头信息，这样，可以使解码从序列的某段开始。序列头指定了图像的垂直和水平尺寸、宽高比、色度取样格式、帧频、图像的类和级、帧内编码和帧间编码使用的量化矩阵等。序列头是解码器开始正确操作的进入点，没有序列头信息，解码器无法解码数据流。这样，从 DCT 数据块到序列的分层结构形成了视频基本码流。

#### 4.2.2 MPEG-1 音频基本码流

MPEG-1 音频基本码流采用“数据流帧封装”结构，这种数据封装结构实际上是一个多路复用器，它将数字音频取样值、格式标记以及附加的辅助数据按照 ISO/IEC 13818-3 音频标准封装成特定的比特数据流。

##### 1. MPEG-1 层的音频基本码流

MPEG-1 层 1 的音频基本码流的数据帧结构如图 4-6 所示。

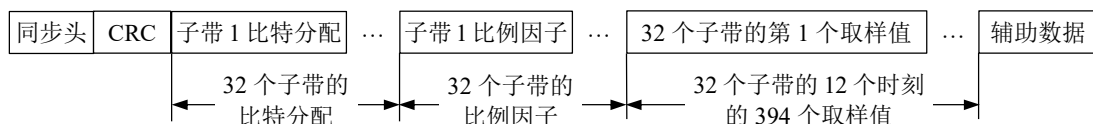


图 4-6 MPEG-1 层 1 的音频基本码流的数据帧格式

在 MPEG-1 层 1 中，将 8ms 长度的音频时间段内的取样值封装在一个数据帧结构数据包中。由于 8ms 长度的音频是划分为 32 个等带宽的子带进行编码，如果取样频率是 48KHz 的话，每个子带的取样频率为  $48\text{KHz}/32=1.5\text{KHz}$ ，那么，每个子带取样得到  $8\text{ms}\times 1.5\text{KHz}=12$  个取样值，32 个子带共有  $32\times 12=394$  个取样值。这 394 个取样值是按照取样的时间次序封装在数据帧中，首先是第一个取样时刻的 32 个子带取样值，然后是第 2 个取样时刻的 32 个子带取样值，直到第 12 个取样时刻的 32 个子带取样值，每个取样时刻的 32 个子带是按照频率从低到高的次序进行排列。

为了有效降低码率，根据量化噪声低于掩蔽阈值的要求，可以给不同的子带信号分配不同的最低量化比特数。同时，每个子带中的 12 个取样值并成一组，求出幅度最大值，对该子带进行归一化，使各子带电平一致，以便进行合适量化。这样，每个子带中的 12 个取样值作为一个处理单元要进行一次比特分配并记录一个比例因子。在 MPEG-1 层 1 的数据帧中，先给出了按频率从高到低排列的 32 个每个长度为 4 比特的比特分配，然后是同样按频率次序排列的 32 个每个长度为 6 比特的比例因子。

在数据帧的前面分别是 32 比特的帧同步和 16 比特的 CRC（循环冗余码）。在 32 比特的帧同步中，12 比特作为帧同步，其余的 20 比特为系统信息，如 MPEG-1 层的标识、错误保护标志、比特率索引、取样频率、通道方式选择、版权标志、原版/复制指示、去加重类型等。在数据帧的最后是附加的辅助数据。

## 2. MPEG-1 层 2 的音频基本码流

为了进一步提高音频的压缩编码效率，MPEG-1 层 2 在层 1 的基础上对压缩算法和数据帧结构进行了改进。在算法技术层面，层 2 采用 1024 点的 FFT 运算，提高了频率的分辨率，可以得到比层 1 更准确的瞬间频谱特性。

不同与层 1 的 8ms 音频长度，层 2 将 24ms 长度的音频时间段内的取样值封装在一个数据帧结构数据包中，如图 4-7 所示，层 2 对一个子带的 3 个 8ms 音频块进行取样，每个块有 12 个取样值，共 36 个取样值，全部 32 个子带共有  $32\times 36=1152$  个取样值。这样，在码流中可以节省出 2 个 32 比特的同步头和 2 个 16 比特 CRC，合计 96 比特。

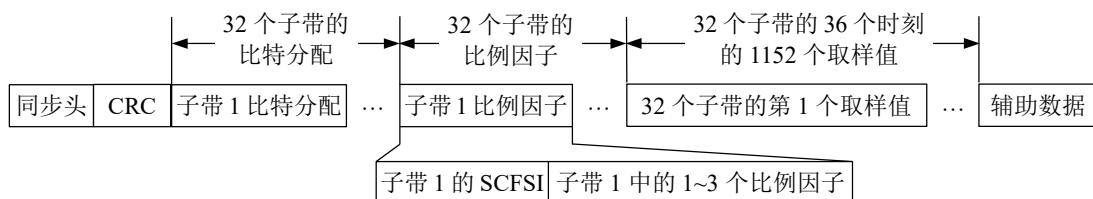


图 4-7 MPEG-1 层 2 的音频基本码流的数据帧格式

层 2 的每个子带的比特分配字段长度随子带频段高低不同而不同。低频段子带用 4 比特

描述，中频段子带用 3 比特描述，高频段子带用 2 比特描述。每 36 个取样值进行一次比特分配，可以节省  $32 \times 4 + 32 \times 4 + 32 \times 1 = 288$  比特。

与层 1 类似，层 2 对每个子带内的 12 个取样值进行一次归一化并传输一次比例因子，因此子带内分成 3 个块（每个块 12 个取样值），每个块使用一个比例因子，共传送 3 个比例因子。为了降低传送比例因子的码率，层 2 采取了进一步的附加编码措施，将一个数据帧内的 3 个连续的比例因子按照不同的组合共同地进行编码和传送。信号变换平稳时，只传送一个或两个较大的比例因子；对变化剧烈的峰值信号，三个比例因子都传输。同时，还传输一个 2 比特比例因子选择信息（SCFSI）描述比例因子的传送情况。SCFSI 编码为 00，表示传送 3 个比例因子；SCFSI 编码为 01，表示传送第 1 个和第 3 个比例因子；SCFSI 编码为 10，表示传送第 1 个比例因子；SCFSI 编码为 11，表示传送第 1 个和第 2 个比例因子。当不需要传送比例因子的子带，也不需要传送比例因子选择信息。采用这种附加编码措施，用于传输比例因子所需的码率平均可压缩约三分之一。