# 第三章 多媒体数据编码标准

## 静态图像编码标准JPEG

## 运动图像编码标准MPEG

## 视听通信编码解码标准H.26X

## AVS标准

## 声音压缩标准

# 3.1 JPEG标准

# 3.1.1 JPEG标准主要内容

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

## 该标准为保证通用性,包含以下两种方式：

### 空间方式 可逆编码

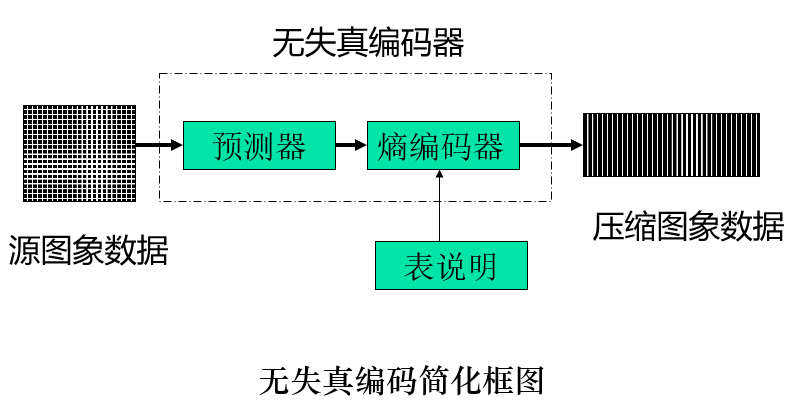
### 空间方式对于基本系统和扩展系统来说,被称为独立功能。

### DCT方式 非可逆编码, 包含基本系统(必须保证的功能)和扩展系统(扩充功能)

## 基本系统是实现DCT编码与解码所需的最小功能集, 大多数的应用系统只需要应用此标准, 就能满足要求（baseline）。

## 扩展系统是为了满足更广领域的应用要求而设置的。

# 1.基于DPCM的无失真编码（空域无失真编码）



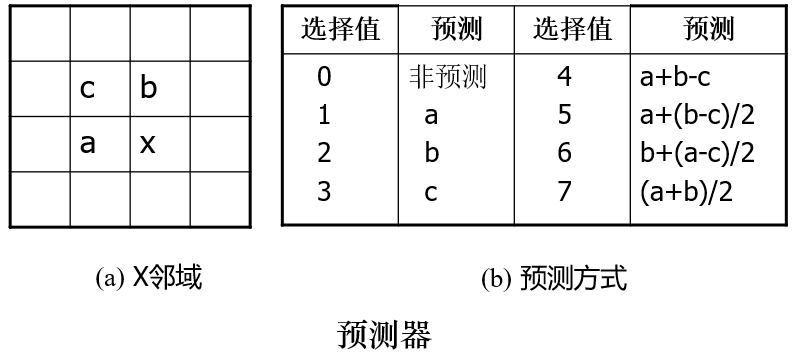
## 这里的预测器采用了基于DPCM（差分脉冲编码调制）的无失真编码技术，优点是硬件易实现，重建图象质量好。 缺点是压缩比太低, 大约为2:1。

## 熵编码器采用huffman编码或算术编码，同样也是无失真编码，在进行编码或解码时需要有Huffman码表或算术编码表。

## 经过这样的无失真编码后，总体压缩比可以达到2:1到5:1.得到的压缩图像可以无损恢复原图像 。

# DPCM无失真编码

工作原理是对X，不直接存储X，而是求X的预测值X’,将X-X’进行无失真熵编码，然后存储。对X’的求法见图给出的预测方式（共7种预测算子）。



# 2. 基于DCT的有失真压缩编码

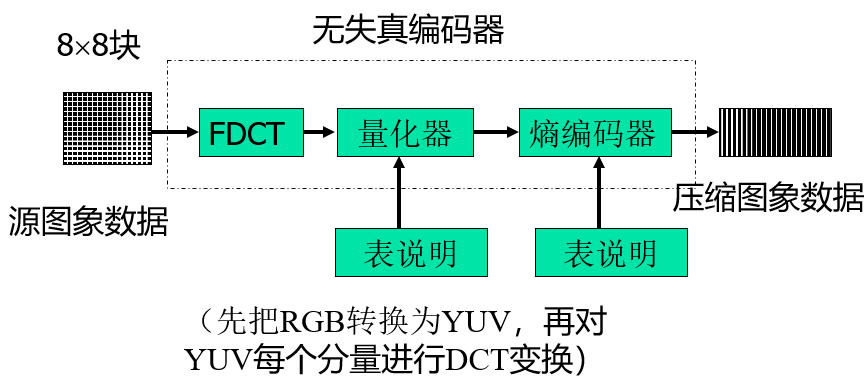
## 离散余弦变换

## 量化处理

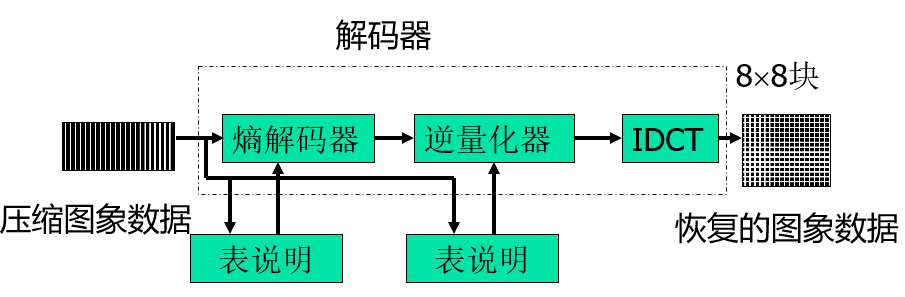
## DC系数的编码和AC系数的行程编码

## 熵编码

# 基于DCT编码过程解码过程



解码过程：



# 离散余弦变换

## (1)首先把原始图象顺序分割成8×8子块;

## (2)进行重采样，采样精度为P位(二进制), 把[0, 2P-1]（0-255）范围的无符号数变换成[-2P-1,2P-1]（-128-127）范围的有符号数, 作为离散余弦正变换(FDCT)的输入（减小绝对值波动）;

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

## 这里用的8×8 FDCT的数学定义为：

## *F(u, v)*=(1/4)C(u) C(v)[∑x=07∑y=07f(x,y)·

## *cos*((2x+1)uπ/16)·*cos*((2y+1)vπ/16)]

## 8×8 IDCT的数学定义为：

## *f(x, y)=* (1/4)[∑u=07∑v=07 C(u)C(v)F(u,v)·

## *cos*((2x+1)uπ/16)·*cos*((2y+1)vπ/16)]

## 其中:C(u), C(v)= 1/√2当u, v=0

## C(u), C(v)=1其它

## 下面的编码针对FDCT输出的64个基信号的幅值(F(0,0),…,F(7,7)称作DCT系数)来进行

# 量化处理

## 量化是一个“多到一”的过程,是造成失真的原因。

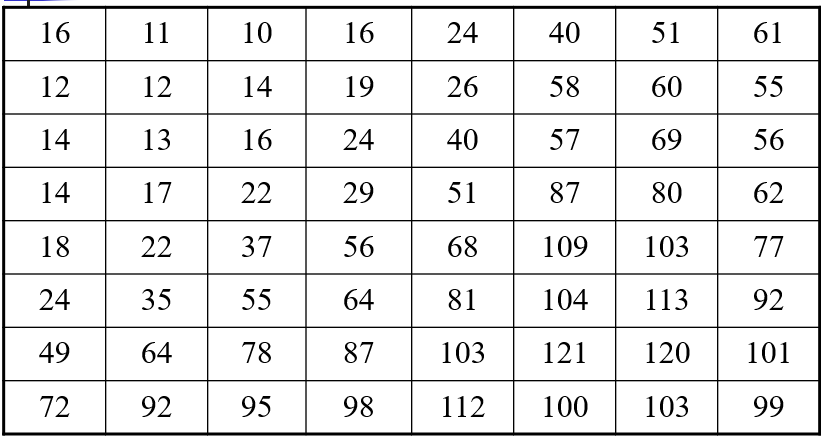
## 关键是找最小量化失真的量化器, JPEG采用线性均匀量化器,定义为对64个DCT系数除以量化步长, 然后四舍五入取整:

## *FQ(u, v)=Integer Round*[F(u, v)/Q(u, v)]

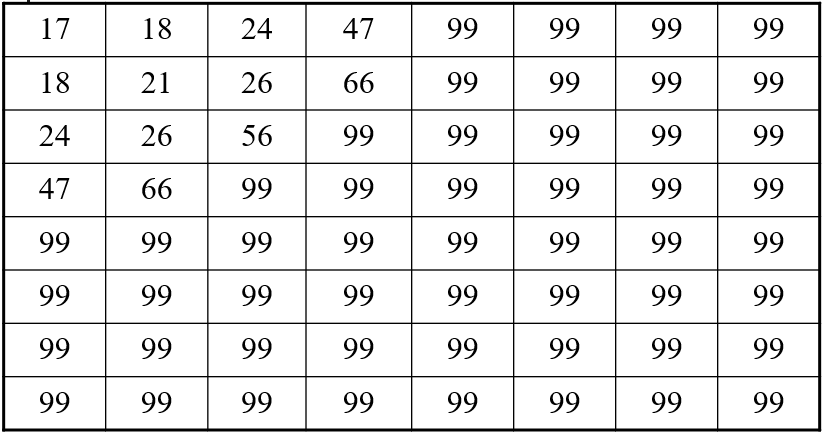
## *Q(u,v)*是量化器步长,它是量化表的元素。量化表元素随DCT系数的位置和彩色分量不同有不同的值,量化表尺寸为8×8与64个变换系数一一对应。

## 这个量化表应由用户规定(JPEG给出参考值-见表2.2,2.3), 并作为编码器的一个输入。

# 亮度量化表



# 色度量化表

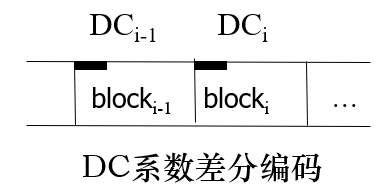


## 量化的作用: 在一定主观保真度图象质量前提下,丢掉那些对视觉影响不大的信息,通过量化可调节数据压缩比。

# DC系数的编码

## 64个变换系数经量化后, 坐标u=v=0的F(0, 0)称DC系数(直流分量), 它即64个空域图象采样值的平均值。

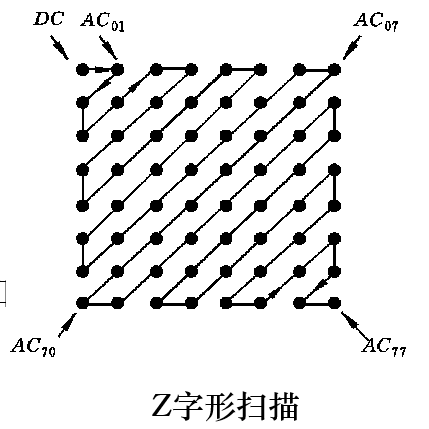
## 相邻8×8块之间DC系数有强相关性。JPEG对量化后的DC系数采用DPCM 编码, 即对DIFF= DCi-DCi-1编码。



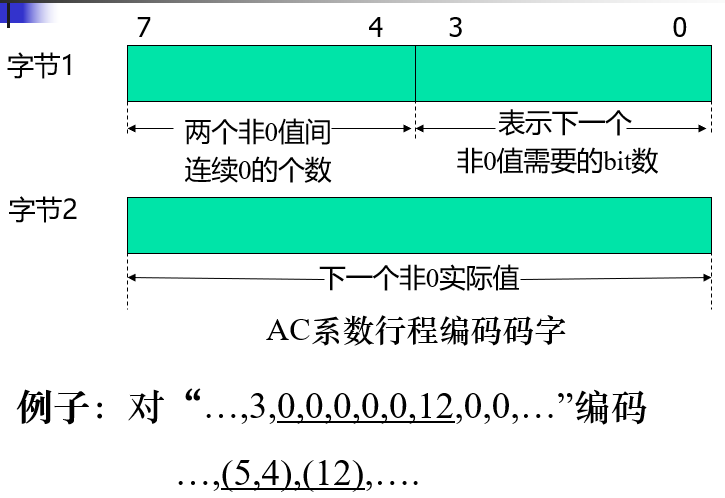
# AC系数的行程编码

## 量化后，AC系数出现较多的零（高频细节），其余63个交流系数(AC)采用行程编码。（zigzag）

## 从左上方AC0,1开始沿对角线方向“Z”字形扫描直到AC7,7扫描结束, 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,如图所示:



# 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:

### 把DC码和AC行程码转换为中间符号序列

### 给这些符号赋以**变长码字**

# AC系数熵编码的中间格式

## 需要进行熵编码的中间符号格式由两个符号组成:

### 符号1: (行程,尺寸)

### 符号2: (幅值)

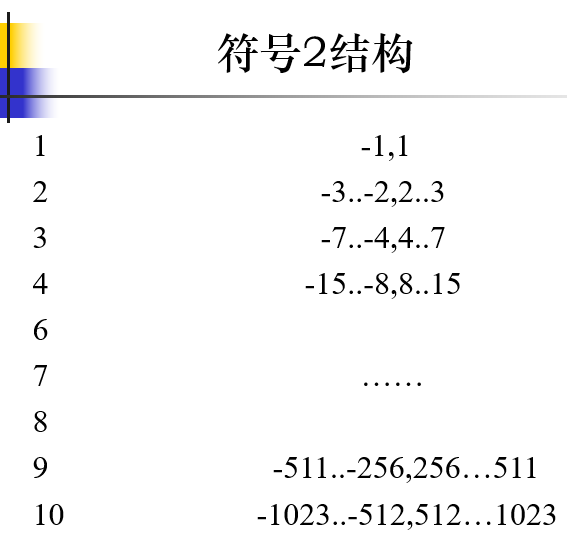
## 第一个信息参数“行程”表示前后两个非0的AC系数之间连续0的个数。

## 第二个信息参数“尺寸”是后一个非0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)（前一个符号1）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。

## “幅值”用以表示非0的AC系数的值, 范围为［-210, 210-1］(最长10bit), 结构形式如表2-4所示。



# DC系数的熵编码

## 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)

## “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［-211,211-1**］**。可在表2.4中多加一级, 幅值尺寸以1到11比特表示。

## 将63个AC系数表示成为符号1和符号2序列,其中连续0的长度超过15时,有多个符号1; 块结束(EOB)时仅有一个符号1(0,0)。

## “…4,0,0,0,0,0,0,0,……0,3,0”. 4,3之间有31个0.

## (15,0),(15,0),(1,2),(3)

## 可变长度熵编码就是对上述序列进行变长编码。

## 对DC系数、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

## JPEG中为了更进一步节约空间，并不直接保存数据的具体数值，而是将数据按照位数分为16组，保存在表里面。这也就是所谓的变长整数编码VLI。即，第0组中保存的编码位数为0，其编码所代表的数字为0；第1组中保存的编码位数为1，编码所代表的数字为-1或者1......，如下面的表格所示，这里，暂且称其为VLI编码表：



## 对应DC系数中间码的符号1为存储DPCM差值对应需要多少位存储（也就是上表中的组号），符号2为对应编码。比如DC系数的DPCM差值为6，则对应的DC编码为（3）（6）

## JPEG提供2套Huffman码表: 亮度和色度。每套又有DC表和AC表各1个。共有4个表。

## 表定义（亮度DC系数码表）

## 16B说明码字长度：

## X’00 01 05 01 01 01 01 01 01 00 00 00 00 00 00 00’

## *第i个(1-16)元素值表示长度为i的Huffman码个数*。

## 紧跟一组值说明亮度表分类：

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

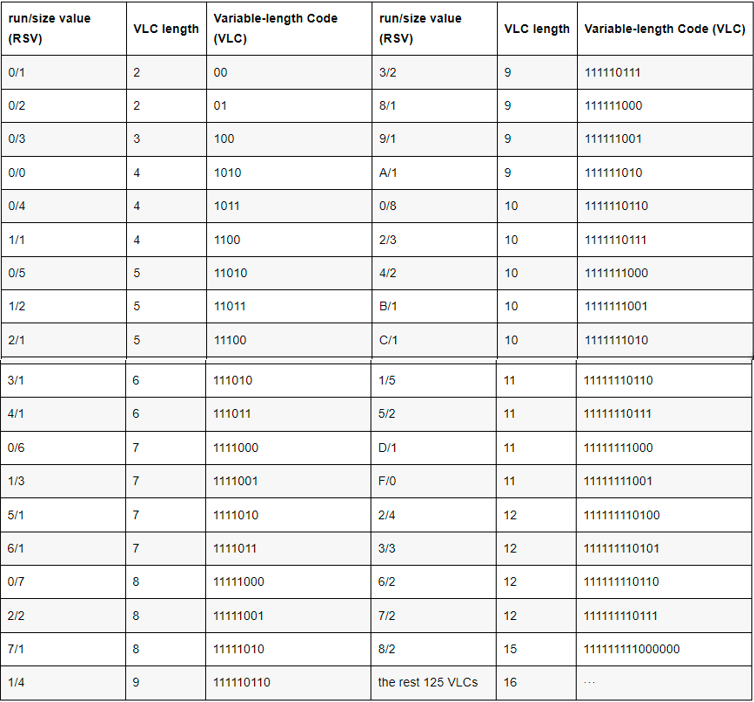
# 亮度DC系数表



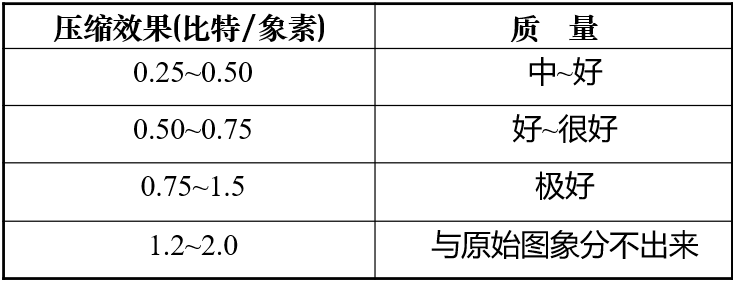
# 色度DC系数表



# 亮度AC系数表



# JPEG压缩效果评价



# 3. 基于DCT的累进操作方式编码

## 顺序方式: 每个图象分量的编码一次扫描完成的;

## 累进方式: 图象分量编码要经过多次扫描才完成。

## 累进方式

## 第一次扫描只进行一次粗糙图象的扫描压缩, 以相对于总的传输时间快得多的时间传输粗糙图象, 并重建一帧质量较低的可识别图象; 在随后的扫描中再对图象作较细的压缩, 这时只传递增加的信息,可重建一幅质量提高一些的图象。这样不断累进, 直到满意的图象为止。

## 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。

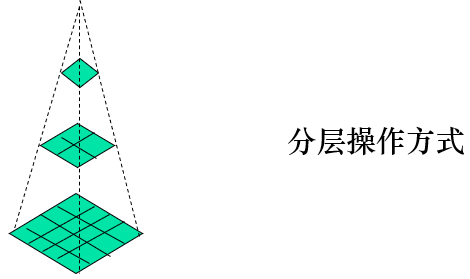
## 两种累进方式：

### 频谱选择法 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, 随后对其它频带编码、传送, 直到全部系数传送完毕为止。

### 按位逼近法 沿着DCT量化系数有效位方向分段累进编码。如第1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。

# 4. 基于DCT的分层操作方式

## 分层方式是对一幅原始图象的空间分辨率,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。



# 分层操作方式的过程

## (1)把原始图象空间分辨率降低。

## (2)对已降低分辨率的图象采用基于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)对低分辨率的图象解码,重建图象,使用插值滤波器,对它插值,恢复图象的水平和垂直分辨率。

## (4)把分辨率已升高的图象作为原始图象的预测值,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)直到图象达到完整的分辨率编码。

# 3.1.2 JPEG2000简介

## 基于Internet的多媒体应用给图像编码提出了新的要求. 2000年12月公布JPEG 2000标准(ISO 15444), 其目标是在高压缩率情况下保证图像传输质量。

## JPEG中采用DCT变换考察整个时域过程的频域特征或整个频域过程的时域特征。JPEG2000采用以小波变换为主的多分辨率编码方式。

## JPEG2000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。与JPEG相比,可修复约30％的速率失真特性。JPEG和JPEG2000在压缩率相同时, JPEG2000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JPEG2000中;

## (3)渐进传输。JPEG2000可实现以空间清晰度和信噪比为首的各种可调节性,从而实现渐进传输,即具有“渐现”特性.

## (4)感兴趣区域压缩。JPEG2000 支持所谓的“感兴趣区域”。

# 3.2 运动图像压缩标准MPEG

# 3.2.1 MPEG标准简介

## MPEG标准是面向运动图象压缩的一个系列标准。

## 最初MPEG专家组工作项目是3个,即在1.5Mbps, 10Mbps,40Mbps传输速率下图象编码, 分别命名为MPEG-1,MPEG-2, MPEG-3。MPEG-3后被取消.

## 为了满足不同的应用要求, MPEG又将陆续增加其它一些标准MPEG-4,MPEG-7,MPEG-21。

## MPEG算法编码过程和解码过程是一种非镜象对称算法(不对称), 解码过程要比编码过程相对简单些。

## MPEG-1和MPEG-2只规定了解码的方案, 重点将解码算法标准化。因而用硬件实现MPEG算法时, 人们首先实现MPEG的解码器,如C-Cube公司CL450解码器系列。

## 随着MPC性能提高,软件解压功能得到支持。

# 3.2.2 MPEG-1系统

## “用于数字存储媒体运动图象及其伴音速率为1.5Mbps的压缩编码”简称MPEG-1, 作为ISO/IEC 11172号建议于1992年通过。

## 主要用于在CD-ROM存储运动视频图像, 它针对标准分辨率(NTSC制为352×240; PAL制为352×288)的图像进行压缩, 每秒30帧画面, 具备CD音质。

## 它还用于数字电话网络上的视频传输, 如非对称数字用户线路(ADSL)、视频点播、教育网络等。

## 使用MPEG-1的压缩算法, 可将一部120分钟长的电影压缩到1.2GB左右。因此, 它被广泛地应用于VCD制作。

# MPEG-1分为5个部分

## (1)MPEG系统(11172-1), 定义音频、视频及有关数据的同步;

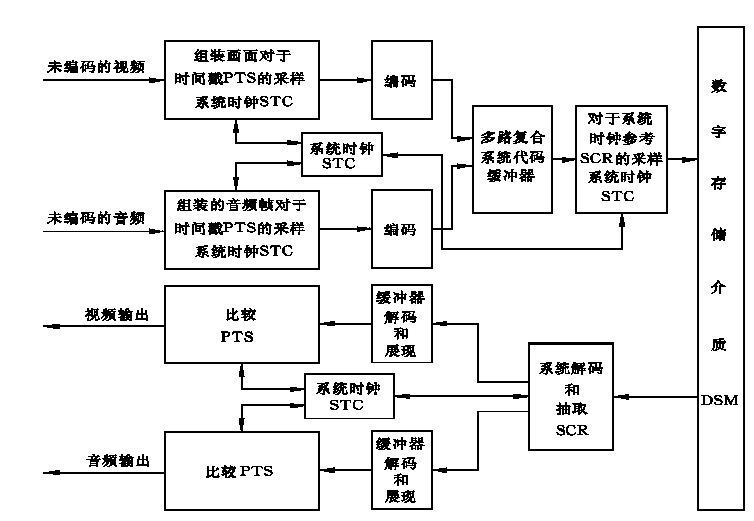
## (2)MPEG视频(11172-2), 定义视频数据的编码和重建图象所需的解码过程,亮度信号分辨率为360×240, 色度信号分辨率为180×120;

## (3)MPEG音频(11172-3), 定义音频数据的编码和解码;

## (4)一致性测试(11172-4);

## (5) 软件模拟(11172-5)。

# MPEG-1解码器原型



## MPEG-1编解码器原型:

## 多路复合而成的码流假设以介质特定格式存储在数字存储介质(DSM)或网络上, 标准不规定介质特定格式。

## 系统解码器从输入多路复合流中抽取定时信息,并对输入流进行分流处理, 输出两个基本流分别给视频和音频解码器。

## 视频和音频解码器分别解码输出视频和声音信号。

## 系统、视频、音频和介质4个解码器之间用定时信息进行同步。

## 多路复合流构造为2层: 系统层和压缩层。系统解码输入的是系统层; 而视频、音频解码器输入的是压缩层。

## 系统解码器执行两类操作:

### 一类是作用在整个多路复合流上的操作,称为复合流操作;

### 另一类是作用在单个基本流上的操作,称为特定流操作。

## 系统层分为两个子层:

### 一个子层称为包(pack),是复合流操作对象;

### 另一个子层称为组(packet),它用于特定流操作。

## MPEG音频标准的特点:

## 音频信号采样率可以是32KHz,44.1KHz或48KHz。

## 压缩后的比特流可以按以下4种模式之一支持单声道或双声道:

### 提供给单音频通道的单声道模式；

### 提供给两个独立的单音频通道的双-单声道模式；

### 提供给立体声通道的立体声模式；

### 联合立体声模式,利用立体声通道之间的关联或通道之间相位差的无关性,或者对两者同时利用。

## MPEG音频标准提供3个独立的压缩层次,用户可在复杂性和压缩质量之间权衡选择。

### 层1最简单,使用比特率384Kbps,主要用于DCC;

### 层2的复杂度中等,使用比特率192Kbps左右, 主要应用于数字广播的音频编码、CD-ROM上的音频信号以及CD-I和VCD。

### 层3最为复杂,使用比特率64Kbps,尤其适用于ISDN上的音频传输,有损压缩但音质保持逼真效果。

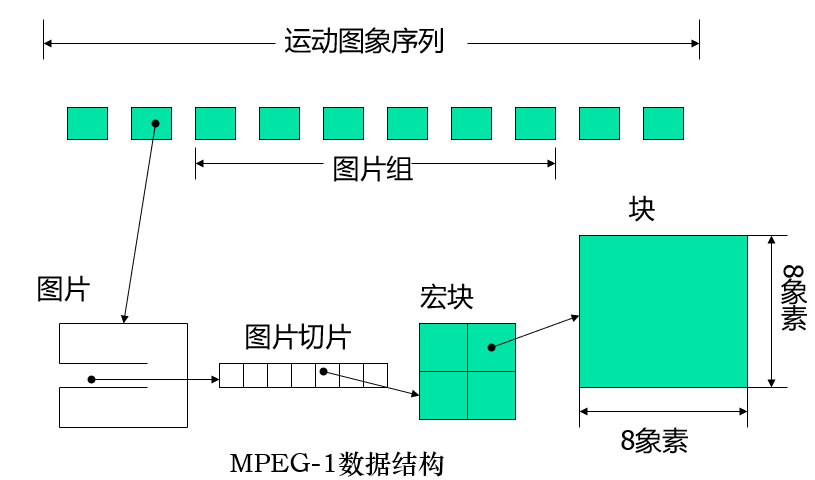
### MP3音乐 是利用 MPEG Audio Layer 3 的技术, 声音采用 1:10 甚至 1:12 的压缩率

## 压缩后的比特流具有预定义的比特率之一。MPEG音频标准也支持用户使用预定义的比特率之外的比特率。

## 编码后的比特流支持循环冗余校验(CRC)。

## MPEG音频标准还支持在比特流中载带附加信息。

# 3.2.3 MPEG视频数据流的结构



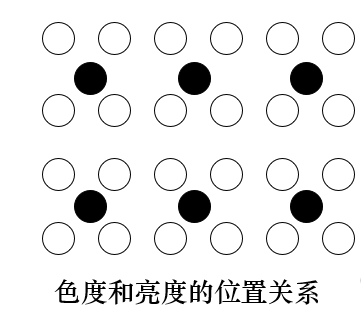
## 运动序列

## 图象组

## 图象信号分3个部分:

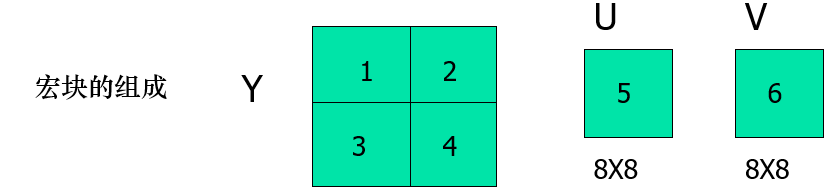
## 一个亮度信号Y和两个色度信号U、V。

亮度信号Y由偶数个行和偶数个列组成, 色度信号U、V分别取Y信号在水平、垂直方向的1/2。如图所示, 黑点代表色度U、V位置,亮度Y位置用白圈表示。



## 块: 一个块由一个8×8的亮度信息或色度信息组成。

## 宏块一个宏块由一个16×16的亮度信息和两个8×8色度信息构成，如图所示。



## 图象切片 由一个或多个连续的宏块构成。

# 3.2.4 MPEG-1视频编码技术

## 主要问题:一方面无法达到很高的压缩比，另一方面用单一的静止帧内编码方法能最好地满足随机存取的要求。

## 解决方法：对这两个方面做了折衷考虑。即为了减少时间上冗余性的基于块的运动补偿技术和基于DCT变换的减少空间上冗余性的ADCT技术

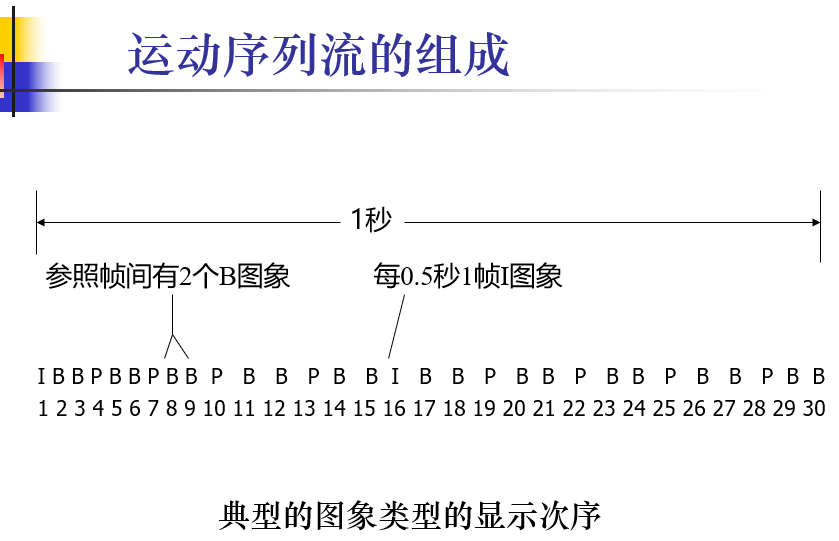
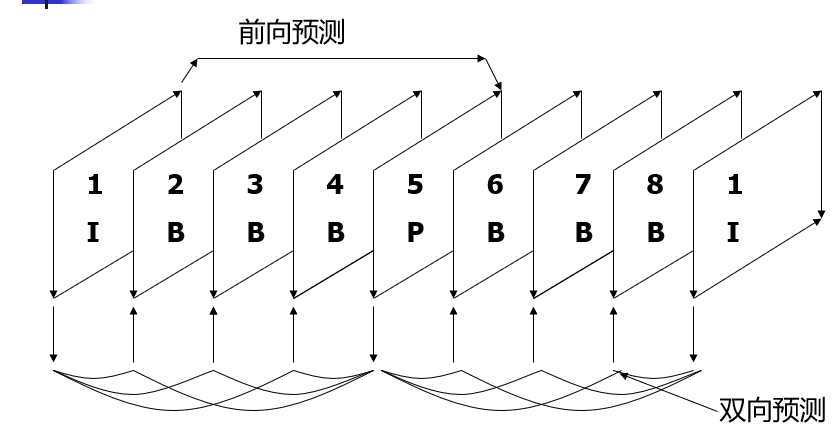
## 在MPEG中将图象分为3种类型：

### I图象 利用图象自身的相关性压缩, 提供压缩数据流中的随机存取的点.

### P图象 用最近的前一个I图象(或P图象)预测编码得到(前向预测)。

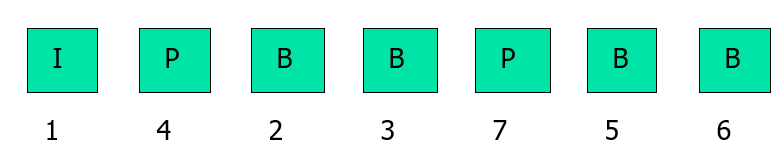
### B图象 B图象在预测时, 既可使用了前一个图象作参照, 也可使用下一个图象做参照或同时使用前后两个图象作为参照图象(双向预测)。

# 帧间预测



# 传输顺序

## MPEG编码器需对上述图象重新排序, 以便解码器高效工作, 因为参照图象必须先于B图象恢复之前恢复。上述1～7帧图象重排后图象组次序为：



# 运动补偿技术

## 运动补偿技术: 用于消除P图象和B图象在时间上的冗余性提高压缩效率。

## MPEG中,运动补偿技术工作在宏块一级。

## B图象宏块有4种类型

## 帧内宏块, 简称I块；

## 前向预测宏块, 简称F块；

## 后向预测宏块, 简称B块；

## 平均宏块, 简称A块。

## 对于P图象, 其宏块只有I块和F块两种。

## 无论B图象和P图象,I块处理技术都与I图象中采用技术一致即ADCT技术。

## 对于F块、B块和A块,MPEG都采用基于块的运动补偿技术。

### F块预测时其参照为前一个I图象或P图象

### B块预测时其参照为后一个I图象或P图象

### 对于A块预测其参照为前后两个I图象或P图象

# 基于块的运动补偿技术

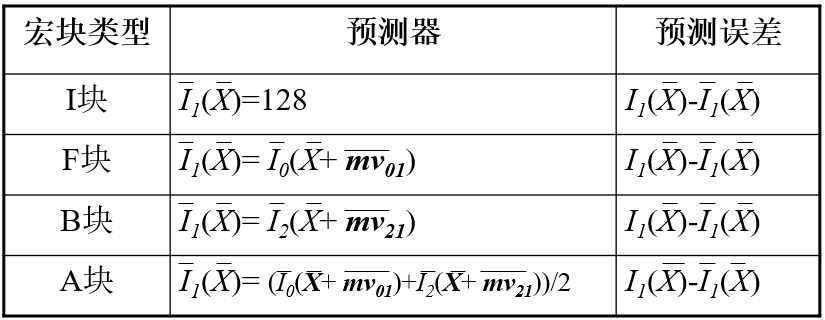
## 基于块的运动补偿技术,就是在其参照帧中寻找符合一定条件,当前被预测块的最佳匹配块。

## 找到匹配块后,有两种处理方法：

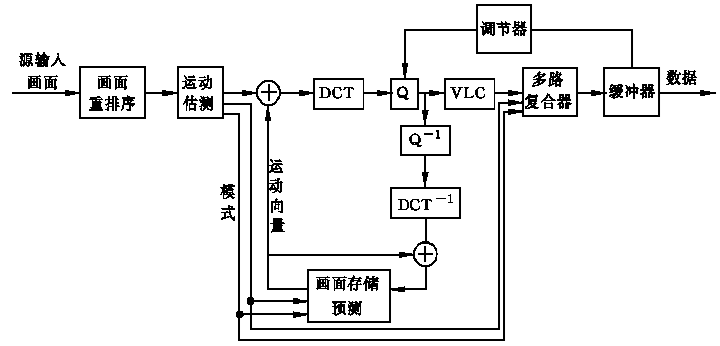
### 一是在恢复被预测块时,用匹配块代替；

### 二是对预测的误差采用ADCT技术编码,在恢复被预测块时,用匹配块加上预测误差。

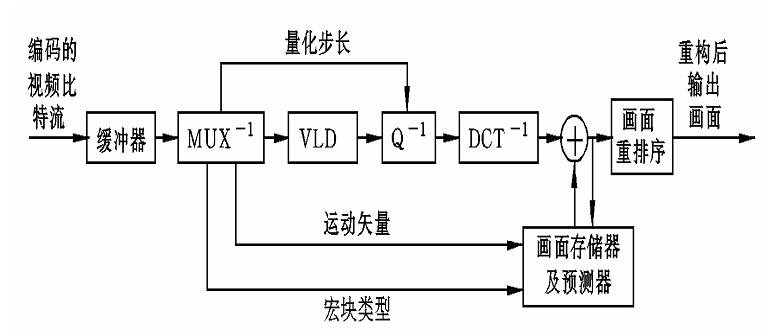
# 宏块的预测方式



# MPEG-1视频系统



## 简化的视频编码框图



## 基本的视频解码器框图

# 3.2.5 MPEG-2标准

## MPEG-2(ISO/IEC 13818)标准制定于1994年。它利用网络3～100Mbps的数据传输率来支持具有更高分辨率图象的压缩和更高的图象质量。

## MPEG-2可支持交迭图象序列,支持可调节性编码,多种运动估计方式,提供一个较广的范围改变压缩比,以适应不同画面质量、存储容量和带宽的要求。

## 它与MPEG-1兼容基础上实现了低码率和多声道扩展：MPEG-2可以将一部120分钟长的电影压缩到4～8GB(DVD质量),其音频编码可提供左右中及两个环绕声道、一个加重低音声道和多达7个伴音声道。

## 除了作为DVD指定标准外，MPEG-2还可用于为广播、有线电视网、电缆网络等提供广播级数字视频.

## MPEG-2 Video 定义了不同的功能档次(Profiles)

## 每个档次又分为几个等级(Levels),一个等级为N的解码器能够对最高为该等级的数码流解码。

# 5个档次

## • 简单型（Simple）

## • 基本型（Main）

## • 信噪比可调型（SNR Scalable）

## • 空间可调型（Spatial Scalable）

## • 增强型（High）

# 4个等级

## • 低级（Low） 352×288×30, 它面向VCR并与MPEG-1兼容；

## • 基本级（Main） 720×460×30或720×576×25，它面向视频广播信号；

## • 高1440级（High-1440） 1440×1080×30或1440×1152×25，它面向HDTV；

## • 高级（High） 1920×1080×30或1920×1152×25，它面向HDTV。

# 11种规范

## 高级的基本型 MP@HL

## 高级的增强型 HP@HL

## 高-1440级的基本型 MP@H1440

## 高-1440级的空间可调型 SSP@H1440

## 高-1440级的的增强型 HP@H1440

## 基本级的简单型 SP@ML

## 基本级基本型 MP@ML

## 基本级的信噪比可调型 SNP@ML

## 基本级的增强型 HP@ML

## 低级的基本型 MP@LL

## 低级的信噪比可调型 SNP@LL

# MPEG-2音频

## 基本特性之一是向后与MPEG-1音频兼容。

## 可以是5.1也可以是7.1通道的环绕立体声。

### 5.1也称为“3/2-立体声加LFE”,其含义是播音现场前面可有3个喇叭通道(左、中、右),后面可有2个环绕声喇叭通道。LFE是低频音效加强通道。

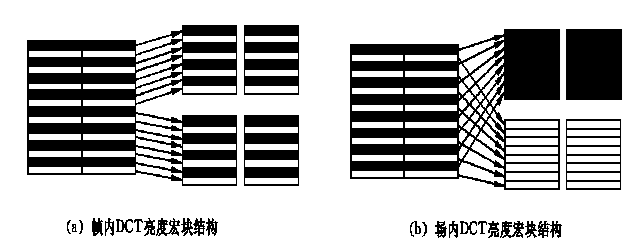
### 7.1通道环绕立体声与5.1类似,它另有中左、中右2个喇叭通道。

# MPEG-2编码方法

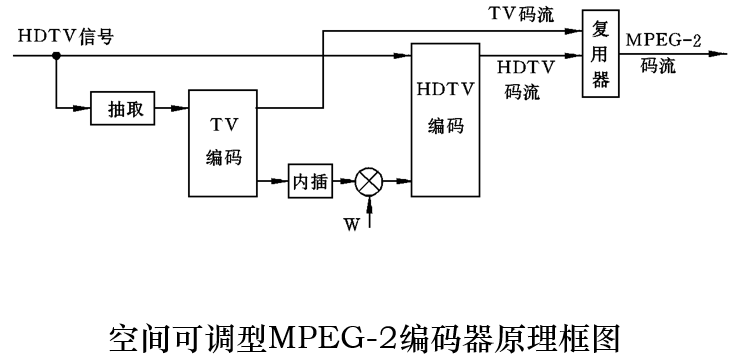
## MPEG-2的编码方法和MPEG-1区别主要是在隔行扫描制式下, DCT变换是在场内还是在帧内进行由用户自行选择, 亦可自适应选择。

## 一般情况下, 对细节多、运动部分少的图象在帧内进行DCT, 而细节少、运动分量多的图象在场内进行DCT。

## MPEG-2采用可调型和非可调型两种编码结构。还可以使用一个基本层加上多个增强型的多层编码结构,这由用户按质量和压缩比要求选择使用.



# MPEG-2亮度宏块结构



# 3.2.6 MPEG-4标准

## MPEG-4“甚低速率视听编码”1998年11月公布, 它针对低速率视音频编码,更加注重系统交互性和灵活性。

## MPEG-4引入了AV对象(AVO), 使得更多的交互操作成为可能：“AV对象”可以是一个孤立的人,也可以是这个人的语音或一段背景音乐等。

## MPEG-4对AV对象的操作主要有：

## 采用AV对象来表示听觉、视觉或者视听组合内容；

## 组合已有AV对象来生成复合的AV对象，并生成AV场景；

## 对AV对象的数据灵活地多路合成与同步，以便选择合适的网络来传输这些AV对象数据；

## 允许接收端用户在AV场景中对AV对象进行交互操作等。

## MPEG-4**标准主要构成部分**：

## (1)**传输多媒体集成框架(DMIF)**。主要用于解决交互网络中、广播环境下以及光盘应用中多媒体应用操作问题。它是MPEG-4制订会话协议,用来管理多媒体数据流。通过传输多路合成比特信息来建立客户端和服务器端的连接与传输。

## (2)**场景描述**。场景声音视频对象间的关系的描述体现在两个层次：BIFS描述场景中对象的空间时间安排,观察者可以有与这些对象交互的可能性；在较低的层次上,对象描述子定义针对每个对象的基本流的关系，并提供诸如访问基本流需要的URL地址、译码器的特性、知识产权等其它信息。

# 

## MPEG-4具备与Web3D X3D和W3C SMIL的互操作性。XMT格式可在SMIL播放器、VRML 和 MPEG-4 播放器间互换。

## (3)**音频编码。**MPEG-4不仅支持自然声音,而且支持合成声音。MPEG-4的音频部分将音频的合成编码和自然声音的编码相结合,并支持音频的对象特征。支持MIDI和TTS.

## (4)**视频编码**。MPEG-4也支持对自然和合成的视觉对象的编码。合成的视觉对象包括2D、3D动画和人面部表情动画等。

## (5)**缓冲区管理和实时解码**。MPEG-4定义了一个系统解码模式(SDM), 该解码模式描述了一种理想的处理比特流句法语义的解码装置,它要求特殊的缓冲区和实时模式。通过有效地管理,可以更好地利用有限的缓冲区空间。

## 与MPEG-1和2相比,MPEG-4更适于交互AV服务以及远程监控,其设计目标使它具有更广适应性和可扩展性：传输速率可在4.8-64kbps之间,分辨率为176×144, 可以利用很窄带宽通过帧重建技术压缩和传输数据。

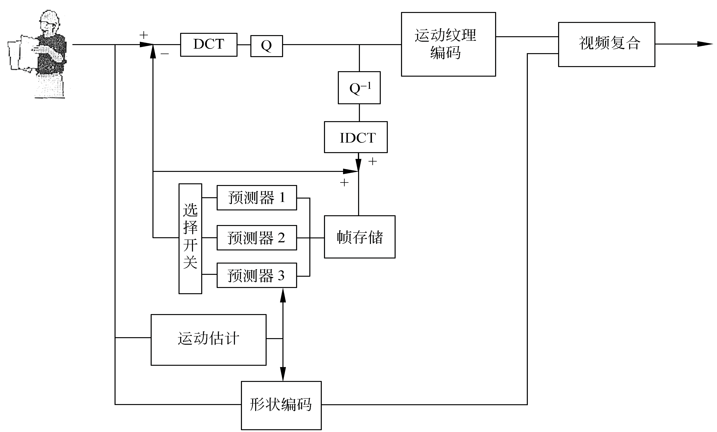
## 它将应用在数字电视、动态图像、互联网、实时多媒体监控、移动多媒体通信、Internet/Intranet上的视频流与可视游戏、DVD上的交互多媒体等方面。

## MPEG-4用MPEG-4压缩算法的ASF可以将120分钟的电影压缩为300MB左右的视频流；采用MPEG-4压缩算法的DIVX编码技术可以将120分钟的电影压缩600MB左右，也可以将一部 DVD影片压缩到 2 张 CD-ROM上.

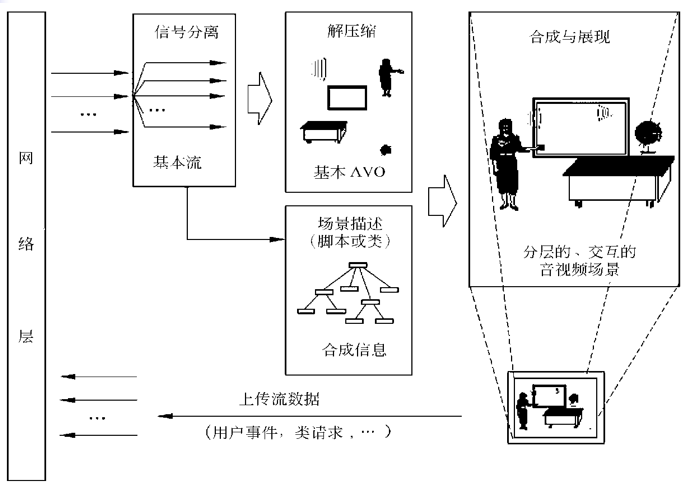
## MPEG-4属于一种高比率有损压缩算法,其图像质量始终无法和DVD的MPEG-2相比, 毕竟DVD的存储容量较大。

## 要想保证高速运动的图像画面不失真, 必须有足够的码率,目前MPEG-4的码率虽然可以调到和DVD差不多,但总体效果还有不小的差距。因此, 对图像质量要求较高的专业视频领域暂时还不能采用。

# MPEG-4视频编码器的算法方框图



# MPEG-4终端的构成(接收端)



## 背景全景图+视频对象(VO)=合成图象

# 3.3 视听通信编码解码标准H.26X

# 3.3.1 H.261

## ITU推荐H.261方案标题“64Kbps视声服务用视象编码方式”, 又称为**P×64Kbps视频编码标准。**

## P取值范围为1-30。P=1或2时,仅能支持QCIF(176×144)分辨率格式, 每秒帧数较低的可视电话;当P≥6时,则可支持图象分辨率格式为CIF(352×288)的电视会议。

## P×64Kbps压缩算法采用基于DCT的变换编码和带有运动预测的DPCM预测编码的混合方法。

## P×64Kbps标准的压缩算法与MPEG-1标准有许多共同之处, 只是传输速率P×64Kbps覆盖较宽的信道频带, 而MPEG-1是基于较窄的频带上传输。

# 3.3.2 视频层次数据结构

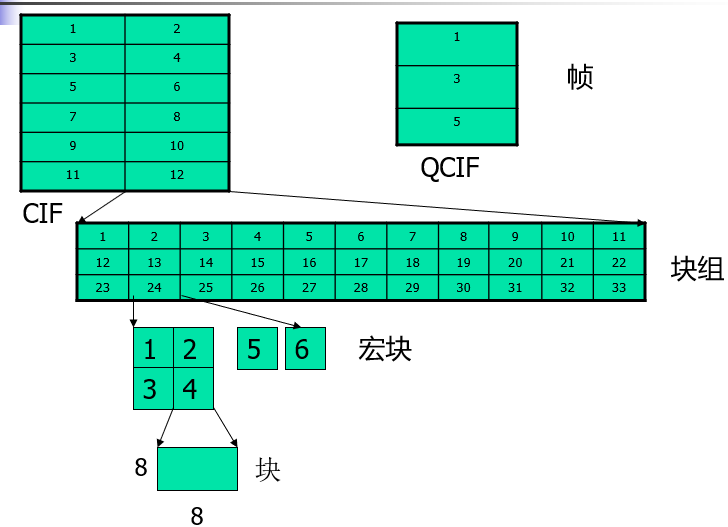
## P×64Kbps标准采用层次块的视频数据结构形式, 使高压缩视频编码算法得以实现。

## P×64Kbps标准的视频编码定义一个视频数据结构CIF保证解码器对接收到的比特流进行没有二义性的正确解码。

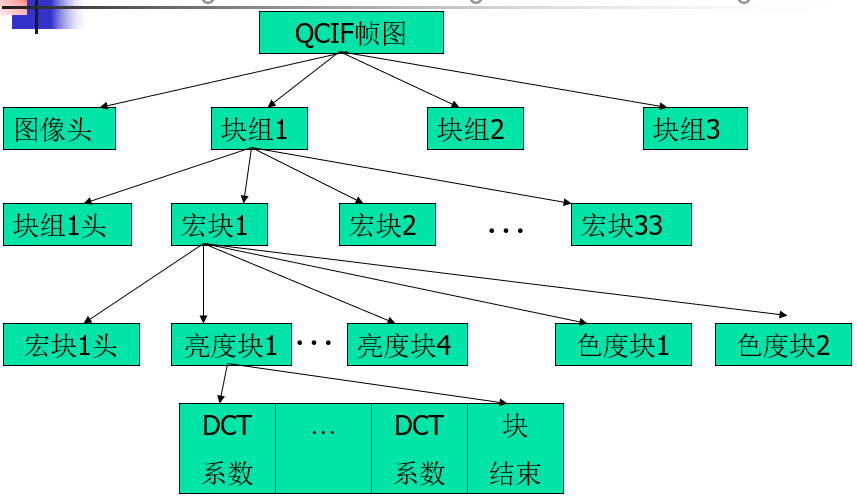
## 利用CIF格式, 可使不同制式的各国电视信号变换为统一的中间格式, 然后输入给编码器, 从而使编码器本身不必意识信号是来自哪种制式的。

## H.261标准适合各种各样实时视觉应用,如位率不同(P不同), 运动效果和图象质量不同,位率提高、画面质量改善。

# 图象数据层次结构



# 视频数据结构图



## 为了适应B-ISDN ATM传输需要, ITU与MPEG联合发布ISO/IEC 13818, 分别称为H.262和MPEG-2标准, 它与H.261和MPEG-1兼容。

## H.263是ITU-T制定的适合于低速视频信号压缩标准。大多数用户最方便的是公用电话线, 以V.34为标准的调制解调器支持在电话线中传输速率可达28.8kbps或33.6kbps, 甚至56kbps.

## 与MPEG-4基于对象编码不同,H.263采用基于帧编码.

## H.263是在H.261基础上扩展形成的, 支持的图象格式包括Sub-QCIF(128×96), QCIF, CIF, 4CIF, 16CIF (1408×1152)等。其中主要采用的改进技术有：

# 

## **(1)半象素精度的运动补偿** 在H.261中,运动矢量的精度为1个象素, H.263运动矢量的估值精度达到半个象素。精度的提高使运动补偿后的帧间误差减少,从而降低了码率。

## **(2)不受限的运动矢量** 当运动跨越图象边界时, 由运动矢量所确定的宏块位置可能有一部分落在边界之外, 此时可以用边界上的象素值表示界外的象素值,从而降低预测误差。

## **(3) 用基于句法的算术编码代替Huffman编码**(可选项) 这是一种效率较高的自适应算术编码。

# 

## **(4) 先进的预测模式(可选项)** 对宏块中的4个亮度块分别进行运动估值获得4个运动矢量。虽然此时传输运动矢量的比特数增加一些,但由于预测误差的大幅度降低,仍然使总码率降低。

## **(5) PB帧模式(可选项)** 虽然使用双向预测B帧可以降低码率, 但却要引入附加的编码延时和解码延时.为降低延时, H.263采用了P帧和B帧作为一个单元来处理的方式, 即将P帧和由该帧与上一个P帧所共同预测的B帧一起进行编码。

# 2.5.4 H264/AVC标准

## ITU-T和ISO/IEC联合开发组共同开发的最新标准，同等质量下压缩效率比提高了2倍以上。

## H.264最大的技术优势体现在4个方面：

## 1）将每个视频帧分离成由像素组成的块，因此视频帧的编码处理的过程可以达到块的级别。

## 2）采用空间冗余方法，对视频帧一些原始块进行空间预测、转换、优化和熵编码。

## 3）对连续帧的不同块采用临时存放的方法，这样只需对连续帧中有改变的部分进行编码。

## 4）采用剩余空间冗余技术,对视频帧里残留块进行编码。对源块和相应预测块不同,再次采用转换、优化和熵编码。

## H.264是保留了以往压缩技术的长处又具有其它压缩技术无法比拟的许多优点。

## 1）低码流：采用H.264技术压缩后的数据量只有MPEG-2的1/8, MPEG-4的1/3。

## 2）高质量的图像：H.264能提供连续、流畅的高质量图像（DVD质量）。

## 3）容错能力强：H.264提供了解决在不稳定网络环境下容易发生的丢包等错误的必要工具。

## 4）网络适应性强：H.264提供了网络适配层 , 使得H.264的文件能容易地在不同网络上传输(例如互联网, CDMA, GPRS, WCDMA, CDMA2000等)。

# 3.4 AVS标准

## AVS标准是《信息技术-先进音视频编码》系列标准的简称，它包括9个部分：系统(第1部分)、视频(第2部分)、音频(第3部分)、数字版权管理(第6部分)技术标准、一致性测试(第4部分)、参考软件(第5部分)、移动视频(第7部分)、系统知识产权IP(第8部分)、文件格式(第9部分)等支撑标准。

## 2006年2月22日，国家标准化管理委员会颁布通知：《信息技术-先进音视频编码》第2部分视频（GB/T 20090.2）于2006年3月1日起开始实施。标准其他部分将继续开展工作, 陆续进入标准报批和审核程序。

## 我国具备自主知识产权的第二代信源编码标准，它是数字音视频产业的共性基础标准。

## AVS标准具备先进性、自主性、开放性。

## AVS是基于我国创新技术和部分公开技术的自主标准，技术方案简洁,芯片实现复杂度低,达到第2代标准最高水平；

## AVS通过简洁的一站式许可政策，解决了AVC专利许可问题死结，制定过程开放、国际化，是开放式制订的国家、国际标准，易于推广；

## AVC仅是一个视频编码标准,而AVS是一套包含系统、视频、音频、媒体版权管理在内的完整标准体系，为数字音视频产业提供更全面的解决方案。

# 2.5 声音压缩标准

# 3.5.1 声音编码

## 声音包括语音和音乐, 是多媒体系统中两类重要数据。

## 声音数据表征是一个一维时变系统,特别对于语音数据,人们已经找到了较合理的声道模型,因此声音数据的压缩要比图象数据的压缩容易。

## 统计表明,语音过程是一个近似的短时平稳随机过程. 短时是指在10～30ms的范围。

## 由于语音信号的这一性质,使得我们有可能将语音信号划分为一帧一帧进行处理,每一帧内的信号近似地满足同一模型—这是本方法假设的基本前提。

## 在实用中,一般一帧的宽度为20ms。

## 语音的基本参数包括

### 基音周期

### 共振峰

### 语音谱

### 声强。

# **语音生成机构的模型**

## 语音生成机构的模型由3部分组成

### 声源

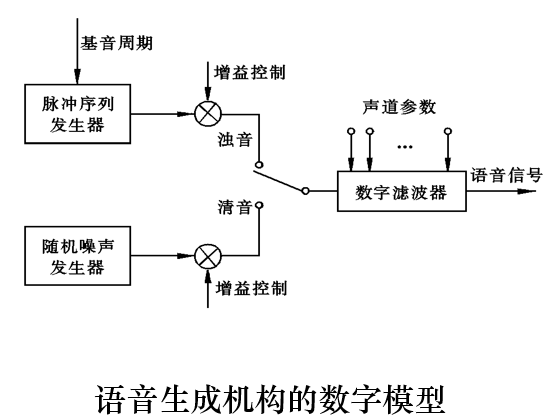
### 共鸣机构

### 放射机构

## 声源共有3类: 元音、摩擦音、爆破音。

## 共鸣机构 也称声道,由鼻腔、口腔与舌头组成。

## 放射机构 由嘴唇和鼻孔组成,是功能是发出声音并传播出去。



# 模型的描述

## 与此语音生成机构模型相对应的声源由基音周期参数描述, 声道由共振峰参数描述, 放射机构则由语音谱和声强描述。

## DPCM,ADPCM等波形预测技术是音乐和实时语音数据压缩技术的主要方法。

## 虽然该方法与基于语音识别的方法和基于参数分析合成的方法相比有压缩能力差的缺点, 但算法简单,容易实现, 以及能够较好地保持原有声音的特点等优点, 因而在语音数据压缩的标准化推荐方案中最先被考虑

## 参数编码的压缩率很大, 但计算量大, 保真度不高, 适合语音信号的编码。

## 混合编码介于波形编码和参数编码之间, 集中了两者优点。

# 2.5.2 ITU语音标准化方案

## 16Kbps ITU语音标准化方案G.728

## 32Kbps ITU语音标准化方案G.721

## 64Kbps ITU语音标准化方案G.722

# 1. 16Kbps语音标准化方案G.728

## 使用领域统一在包括可视电话、数字移动通信、无绳电话、卫星通信、DCME、ISDN等范围内。

## 对于以上所提到的应用范围，约束条件是语音质量在32Kbps ADPCM的同等或以上，且编码延迟时间在5ms以下。

## 1992年,CCITT制定了基于短延时码本激励线性预测编码(LD-CELP)的G.728标准, 这是一种基于AbS原理并考虑了听觉特性的编码方法, 它具有以下特征:

### 以块为单位的后向自适应高次线性预测：

### 后向自适应型增益量化；

### 以向量为单位的激励信号量化。

## G.728用在64Kbps的ISDN线路的可视电话,带宽分配为语音16Kbps,图象48Kbps。语音的多重化传送装置和个人计算机用的编码也是有希望的应用领域。

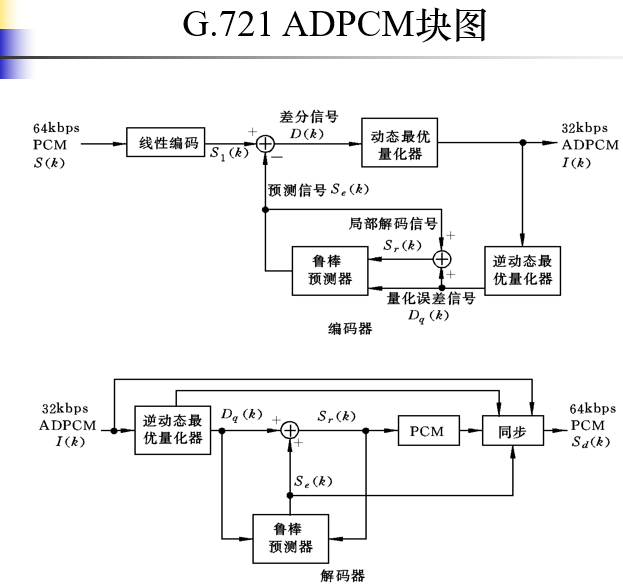
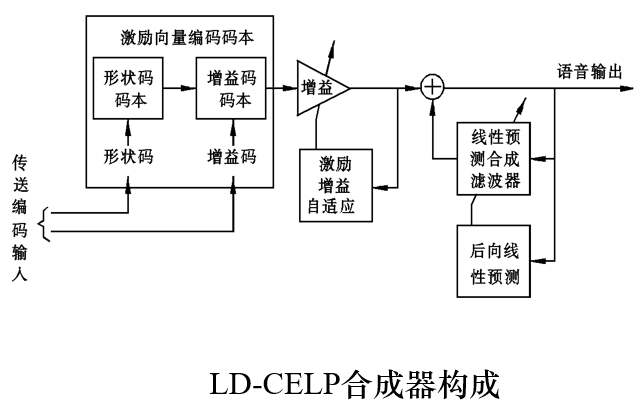
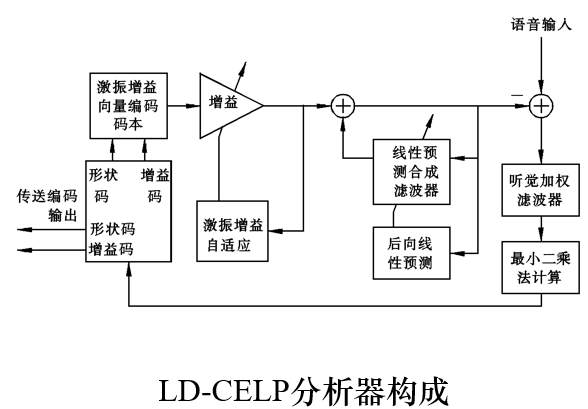
# 2. 32Kbps标准化方案G.721

## 1984年10月CCITT公布了使用ADPCM的标准G.721,速率为32Kbps.目的是最终取代PCM电路传送方式.

## 作为对象的信号包括在电话线中流通的所有的信号, 如语音、个人计算机通信的调制解码信号, 按键电话的信号等等。

## 本方案针对PCM(8KHz采样,8比特样点)规定G.721用PCM的一半速率(8KHz采样,每样点4比特)完成。

## G.721方案采用算法是编码符号延迟为0的且对传送通道的误码率要求不高的ADPCM方式。



## ADPCM方式在算法上的特征

## (1)为了提高预测精度(特别对于性质相差很大的语音信号和调制解调器对的信号),采用了动态对数量化器。

## (2)ADPCM本身采用了按每个采样点进行自适应控制的鲁棒自适应预测器。

## (3)追加了PCM和ADPCM间不论进行多少次转换都不会引起特性降低的同步功能。

## G.721方案的应用 最初是面向卫星通信,长距离通信以及信道价格很高的语音传输。

## 目前的应用领域除了最初的目标外, 还被使用在包括电视会议的语音编码, 为提高线路利用率的多媒体多路复用装置, 数字录音电话的数字记录部件, 以及高质量的语音合成器等等。

# 3. 64Kbps标准化方案G.722

## 1983年开始,CCITT开始了7KHz带宽以语音和音乐为对象的标准化音响编码方案,1988年公布为G.722标准。

## G.722方案编码方法是使用在64Kbps位速率以内工作的SB-ADPCM方法音频编码,它将50Hz到7KHz间频带从4KHz处分割为高频区和低频区,分别利用ADPCM算法编码。

## 算法分为3种基本工作模式,即64Kbps,56Kbps和48Kbps模式。56Kbps和48Kbps两工作模式可分别在总体64Kbps位速率中设置一个8Kbps或14Kbps的数据通道.

## 编码器的构成

## (1) 将输入信号以16KHz的速率采样后,每个样点量化为14比特,然后到进行发送的功能模块。

## (2) 编码速率为64Kbps的SB-ADPCM的编码器。

## 解码器的构成

## (1) 与编码器操作完全相反的(逆)SB-ADPCM解码处理。解码器的速率依工作模式的选择可以在64Kbps, 56Kbps和48Kbps之间变化。

## (2) 生成在16KHz采样速度下与14比特的线性量化精度的数字信号相对应的模拟信号的接收功能模块。

## G.722的应用

## 主要应用对象是电视会议系统, 这是多媒体通信一个子领域。要解决的主要问题是高质量的语音传送。语音通信会议一般涉及3个或3个以上的不同地点, 因此64Kbps音频编码标准必须支持多地点间的会议系统。

# **本章小结**

## 首先介绍媒体数据的表示方法,重点是表示图象数据的彩色空间和文件格式, 并介绍了常用的数据压缩方法。

## 本章重点讨论多媒体系统中几个重要的图象压缩标准, 其中JPEG标准、MPEG标准以及H.26X标准是最重要的得到广泛应用的图象压缩标准。

## 最后简要介绍了多媒体系统的几个重要的声音压缩标准。