

# 第 8 章数字视音频压缩编码

中国传媒大学  
姜秀华

## ◆8.1 数字视频压缩编码概论

### 8.1. 数字视频压缩编码概论

常见的数字视频的原始有效码率

数字视频格式	色度格式	像素格式	量化比特数	无压缩有效码率
HDTV	4:2:2	1920×1080/50I	10	1037Mbps
HDTV	4:2:0	1920×1080/50I	8	622 Mbps
SDTV	4:2:2	720×576/50I	10	207Mbps
SDTV	4:2:0	720×576/50I	8	124Mbps
CIF	4:2:0	352×288/50I	8	30 Mbps

#### 8.1.1 数字视频压缩机理

- (1) 数据中的冗余
- ① 空间冗余---图像内部相邻像素之间存在较强的相关性所造成的冗余。
- ② 时间冗余---视频图像序列中的不同帧之间的相关性所造成的冗余。
- ③ 结构冗余---是指图像中存在很强的纹理结构或自相似性，称之为结构冗余。
- ④ 信息熵冗余---也称编码冗余，如果图像中平均每个像素使用的比特数大于该图像的信息熵，称为信息熵冗余。

- ⑤视觉冗余
- 亮度辨别阈值---人眼刚刚能察觉的亮度变化值称为亮度辨别阈值。
- 视觉对于不同图像内容的空间分辨力不同，对于亮度变化越剧烈如突变的轮廓和边缘处误差容易被掩盖，而在亮度缓慢变化的平坦区误差容易被察觉。
- 人眼对亮度比色度敏感---人眼对图像中的亮度细节的分辨能力比色度高5~7倍，对色度误差的分辨能力也比亮度低。

#### 8.1.2 数据压缩编码方法分类与性能指标

- 一、压缩编码分类
- 根据解码后的数据与压缩前原始数据是否相同，压缩编码可分为两大类：
- (1) 无损压缩编码方式：仅仅去除图像数据中的冗余信息，压缩比较低，能够准确无误地恢复原始数据。
- (2) 有损压缩编码方式：不仅去除图像数据中的冗余信息，而且也会损失一些图像细节，压缩比较高，不能完全地恢复原始图像。
- 在数字电视的信源压缩编码中，由于要求的压缩比高，普遍采用的是有损压缩编码方法。

- 在数字电视中，通常采用混合编码方案, 包括基于运动估计和补偿的DPCM、正交变换编码、熵编码等。

- 二、压缩编码性能指标
- (1) 图象质量评价
- 通常采用主观评价和客观评价方法评价经过压缩编解码后的视频图像质量. 主观评价是观察者对被测视频质量直接打分。客观评价是通过一定的算法模型对视频图像进行评价。
- (2) 压缩编解码的复杂度
- 其中包括算法的运算量、存储量、编码器和解码器的复杂度等等。如果压缩编码的复杂度过高，难以实时处理。
- (3) 压缩解缩后带来的延时量
- 如果延时过大，会造成明显的可察觉到的声音和画面不同步现象，需要声音通道上加入相应的延时器加以调整。

## 8.2 预测编码

- 预测编码广泛地应用于视频图象和声音信号的压缩编码中。
- 预测编码是基于图像数据具有的空间和时间冗余的特性，用相邻的已知像素(或图像块)来预测当前像素(或图像块)的值，然后再对预测误差进行量化和编码。减少数据在时间和空间上的冗余。

- 分类
  - 线性预测
  - 非线性预测（不讨论）
- 线性预测包括：
  - 差分脉冲编码调制（DPCM, Differential Pulse Code Modulation ）
  - 自适应的差分脉冲编码调制（ADPCM, Adaptive Differential Pulse Code ）（不讨论）

### 8.2.1 预测编码基本原理

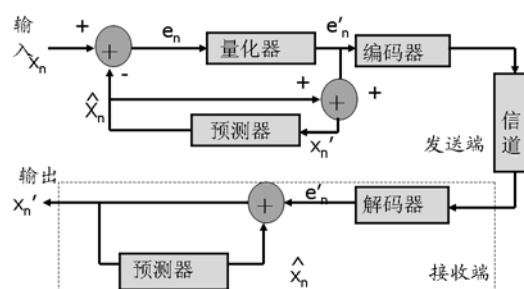
- 预测编码的基本原理
  - 相邻样本之间存在较强的相关性。
  - 差值信号的方差更小，具有更小的动态范围。
  - 编码和传输的并不是采样值本身，而是这个采样值的预测值与其实际值之间的差值。

### 8.2.1 预测编码基本原理

- 差分脉冲编码调制（DPCM, Differential Pulse Code Modulation），在预测编码时，不直接传送图像样值本身，而是对被编码的样值与参考样值间的差值 $e_n$ 进行编码、传送。
- 设 $x_n$ 为被编码图像信号，则对 $x_n$ 的预测值为：  

$$\hat{x}_n = a_1 x_{n-1} + a_2 x_{n-2} + \dots$$
- 其中： $a_1, a_2, \dots$ 为预测系数
- $x_{n-1}, x_{n-2}, \dots$ 为参考样值
- 这是一种线性预测
- 预测误差 $e_n = x_n - \hat{x}_n$

### ◆DPCM系统原理框图



### 8.2.2 帧内预测方法

- 帧内预测是基于同一帧内相邻像素存在很强的相关性。压缩图像的空间冗余。几种常用的线性预测方案

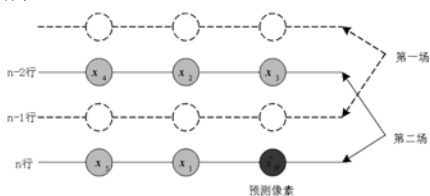
#### ● 前值预测



#### ● 一维预测



#### ● 二维预测

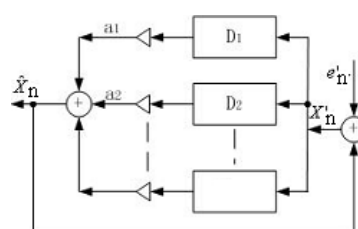


### ■预测器的设计

预测公式为：  

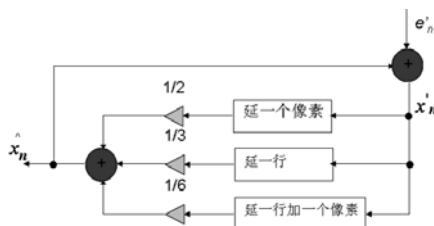
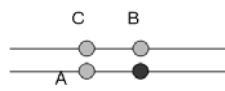
$$\hat{x}_n = a_1 x'_{n-1} + a_2 x'_{n-2} + \dots$$

图中的 $D_1, D_2, \dots$ 为延迟器



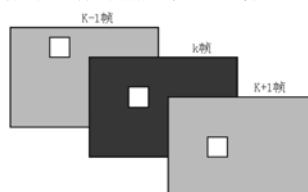
- 例子：设计一个线性预测器，其预测公式为：

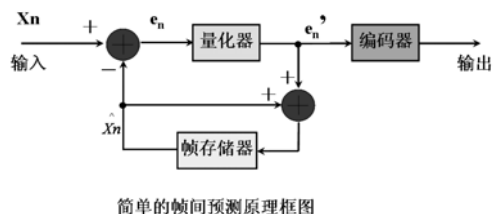
$$\hat{x} = \frac{1}{2}A + \frac{1}{3}B + \frac{1}{6}C$$



### 8.2.3 帧间预测

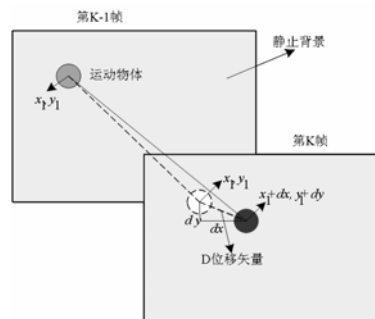
- 帧间预测是利用图像序列在时间上的相关性来压缩图像序列的时间冗余。
- 帧间预测又称为三维预测，是用前一帧重建的像素来预测当前帧像素。这种预测器需要使用大容量的帧存储器存储前一帧的图像。





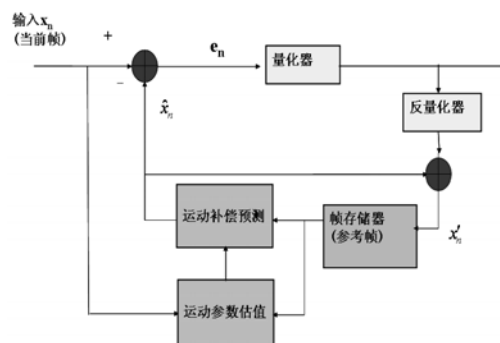
## 8.2.4 运动补偿技术

### ■运动物体的帧间位移



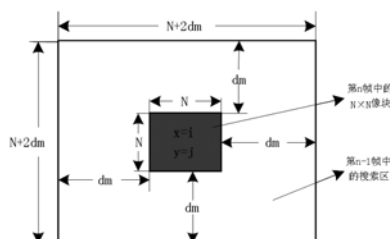
- 如果对运动区域进行预测，首先要估计出物体的运动矢量，这一过程称为运动估计。
- 然后再根据运动矢量找出物体在前一帧的区域位置，用这一区域位置的信号形成当前帧运动物体的预测信号，这叫做运动补偿。
- 运动补偿编码包括以下几个部分：
  - (1) 位移估计：对每一个运动物体进行运动估计，求出运动矢量；
  - (2) 运动补偿：利用运动矢量建立前后帧内同一物体的空间位置对应关系；
  - (3) 预测编码：对经过运动补偿后的运动物体的预测帧差信号、运动矢量进行编码。

在运动补偿技术中主要有两个过程：  
一为运动估计（Motion Estimation, ME），  
二为运动补偿（Motion Compensation, MC）。

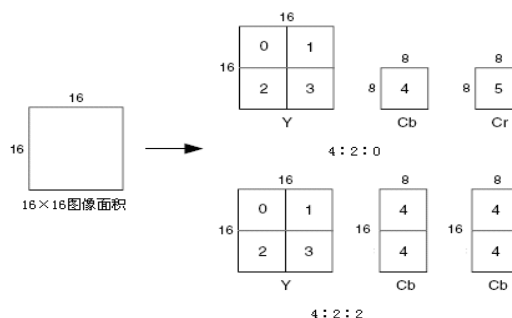


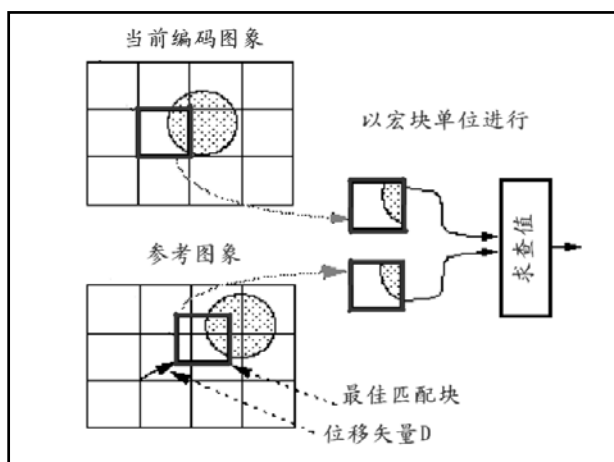
### ■1、运动估计

■如何搜索电视图像中的运动部分是运动补偿预测的关键  
块匹配法是最常用的一种方法。



- 通常一个宏块由一个  $16 \times 16$  像素的亮度阵列和同区域内的 Cb、Cr 色差阵列共同组成，不同的色度格式包括有不同的宏块结构，其中的数字表示块的编码顺序。





### ■ 最佳匹配块

最佳匹配块的判决依据（判据）：

#### 1. 最小均方误差(MSE)

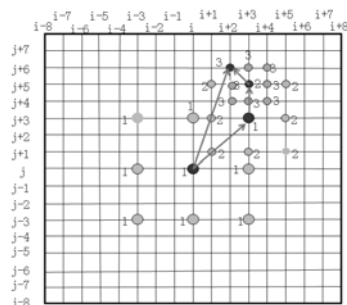
$$MSE(i, j) = \frac{1}{NN} \sum_{x=1}^N \sum_{y=1}^N [f_n(x, y) - f_{n-1}(x+i, y+j)]^2$$

#### 2. 平均绝对帧差 (MAD)

$$MAD(i, j) = \frac{1}{NN} \sum_{x=1}^N \sum_{y=1}^N |f_n(x, y) - f_{n-1}(x+i, y+j)|$$

### ■ 块匹配法的快速搜索

- 三步法搜索



- 运动预测的位移可以是整像素、1/2甚至1/4像素为基本单位。

- 运动矢量位移的精度越高，则帧间预测后的残差就越小，压缩比也越高。



## 运动补偿

- 重建的参考帧存放在帧存储器中，当前帧数据块和参考帧数据块经过运动估值后得到参考帧的匹配块和运动矢量，将当前帧被编码宏块与最佳匹配宏块求差值，得到残差块。因此只需要传输运动矢量和残差块数据，就可以恢复出当前帧的预测图象
- 经过运动补偿后再求残差值，残差数据值将会很小，通过量化可以使传输的数据量大大降低。

## 运动补偿效果实例



前一帧

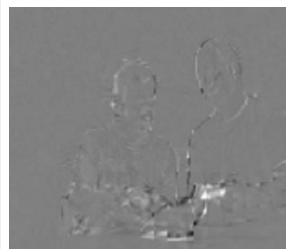
后一帧



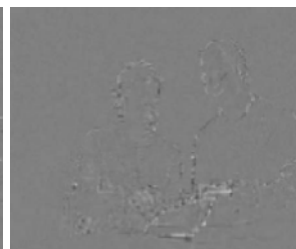
前后帧之间残差 (Residual)  
(no motion compensation)



前后帧之间残差 (Residual)  
(16 × 16 block size运动补偿)



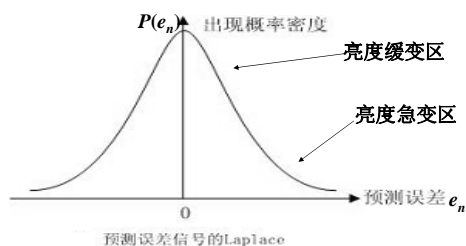
前后帧之间残差 (Residual)  
(8 × 8 block size运动补偿)



前后帧之间残差 (Residual)  
(4 × 4 block size运动补偿)

## 8.25 量化器设计

- 在预测编码中，要对预测误差值 $e_n$ 进行量化，
- 预测误差值 $e_n$ 的概率分布呈拉普拉斯分布，集中在零值附近一个较窄的范围内，随着预测误差绝对值的增大，概率迅速下降。

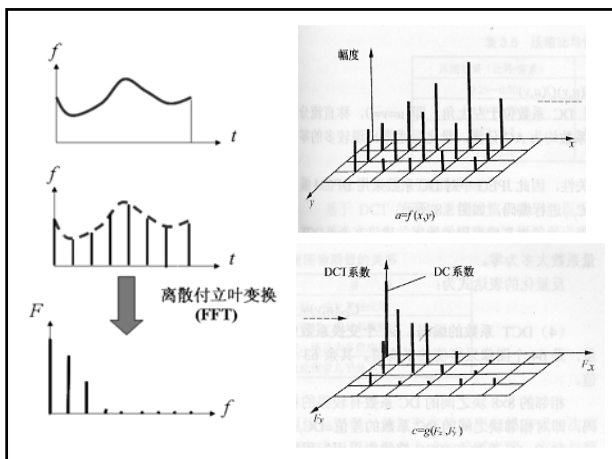


- 预测误差值比较小的部分，对应于图像的平坦区（低频分量），由于人眼对图像的平坦区误差的分辨能力比较高，因此需要细量化；
- 预测误差值较大的部分，对应于图像突变区（高频分量），由于人眼对图像的突变区误差的分辨能力比较低，因此可以粗量化；
- 因此在预测编码的量化器设计中，利用预测误差值的概率分布特性和人眼的视觉特性，对图像的平坦缓变区（对应低频分量）的部分进行细量化，对于图像突变区（对应高频分量）的部分进行粗量化，以尽量减少编码比特数，将量化误差控制到主观视觉可以忍受的范围之内。
- 量化误差是预测编码产生编码失真的主要原因

## 8.3 变换编码

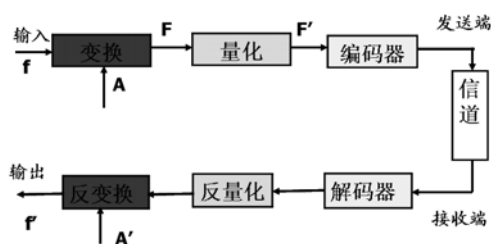
### 变换的物理意义

- 预测编码消除相关性的能力有限，变换编码是一种更高效的压缩编码。
- 变换编码的思想：将原始数据从时间域或者空间域变换到另一个更适合于压缩的域，通常为频域。
- 一般来说，图像数据在空间上具有较强的相关性，变换到变换域的变换系数的可以实现去相关和能量集中。变换系数的能量往往被集中在少数样值上，再通过量化舍弃一些较小的系数，实现数据压缩的目的。



- 变换的关键在于去除相关性，即使得变换后的图像数据互相之间都不相关，互相独立
- 常用的图像正交变换有离散傅立叶变换（DFT）、离散余弦变换（Discrete Cosine Transform，DCT）及沃尔什变换（WH）等。
- 由于DCT变换具有较好的去除相关的能力，实现快速算法比较容易，因此，在数字视频压缩中得到了广泛的应用。

### ◆变换编码原理框图



- 变换本身并不会产生压缩，实现压缩的是变换后的量化编码。

### 8.3.1 二维离散余弦变换(DCT)

- 二维  $N \times N$  图像块的DCT变换

- 正变换 DCT

- 图像样值  $f(x, y)$   $x, y = 0, 1, \dots, N-1$

变换系数  $F(u, v) = C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N}$   
 $u, v = 0, 1, \dots, N-1$ , 分别表示图像在变换域中的水平空间频率和垂直空间频率

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{N}} & \text{当 } u, v = 0 \\ \sqrt{\frac{2}{N}} & \text{当 } u, v = 1, 2, \dots, N-1 \end{cases}$$

- 反变换 IDCT

$$f(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)F(u, v) \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N}$$

- 二维DCT的矩阵表示式

- 正变换DCT:  $F(u, v) = [A][f][A^T]$

- 正交变换矩阵[A]

- 转值矩阵[A]<sup>T</sup>

- 根据矩阵的正交性，有[A]<sup>T</sup>=[A]<sup>-1</sup>

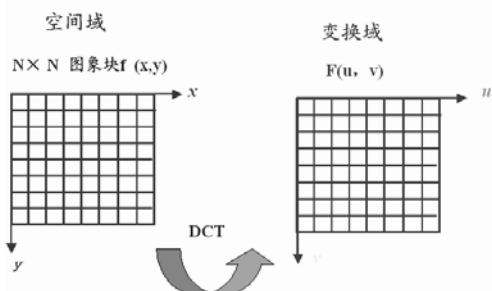
- 反变换IDCT:

$$f(i, j) = [A^T]F[A]^{-1} = [A^T]F[A]$$

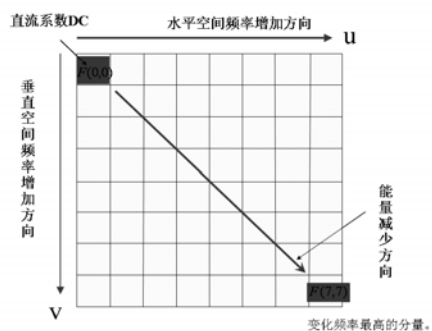
- 二维8×8 DCT变换 矩阵

$$A = \frac{1}{2} \begin{bmatrix} \frac{\sqrt{2}}{8} & \frac{\sqrt{2}}{8} & \frac{\sqrt{2}}{8} & \frac{\sqrt{2}}{8} & \frac{\sqrt{2}}{8} & \frac{\sqrt{2}}{8} & \frac{\sqrt{2}}{8} & \frac{\sqrt{2}}{8} \\ \cos \frac{\pi}{16} & \cos \frac{3\pi}{16} & \cos \frac{5\pi}{16} & \cos \frac{7\pi}{16} & -\cos \frac{7\pi}{16} & -\cos \frac{5\pi}{16} & -\cos \frac{3\pi}{16} & -\cos \frac{\pi}{16} \\ \cos \frac{2\pi}{16} & \cos \frac{6\pi}{16} & -\cos \frac{6\pi}{16} & -\cos \frac{2\pi}{16} & -\cos \frac{2\pi}{16} & -\cos \frac{6\pi}{16} & \cos \frac{6\pi}{16} & \cos \frac{2\pi}{16} \\ \cos \frac{3\pi}{16} & -\cos \frac{7\pi}{16} & -\cos \frac{\pi}{16} & -\cos \frac{4\pi}{16} & \cos \frac{4\pi}{16} & \cos \frac{\pi}{16} & \cos \frac{7\pi}{16} & -\cos \frac{3\pi}{16} \\ \cos \frac{4\pi}{16} & -\cos \frac{4\pi}{16} & -\cos \frac{4\pi}{16} & \cos \frac{4\pi}{16} & \cos \frac{4\pi}{16} & -\cos \frac{4\pi}{16} & -\cos \frac{4\pi}{16} & \cos \frac{4\pi}{16} \\ \cos \frac{5\pi}{16} & -\cos \frac{5\pi}{16} & \cos \frac{7\pi}{16} & \cos \frac{3\pi}{16} & -\cos \frac{3\pi}{16} & -\cos \frac{7\pi}{16} & \cos \frac{5\pi}{16} & -\cos \frac{5\pi}{16} \\ \cos \frac{6\pi}{16} & -\cos \frac{2\pi}{16} & \cos \frac{2\pi}{16} & -\cos \frac{6\pi}{16} & -\cos \frac{6\pi}{16} & \cos \frac{2\pi}{16} & -\cos \frac{2\pi}{16} & \cos \frac{6\pi}{16} \\ \cos \frac{7\pi}{16} & -\cos \frac{5\pi}{16} & \cos \frac{5\pi}{16} & -\cos \frac{7\pi}{16} & \cos \frac{7\pi}{16} & -\cos \frac{5\pi}{16} & \cos \frac{5\pi}{16} & -\cos \frac{7\pi}{16} \end{bmatrix}$$

二维DCT变换示意图

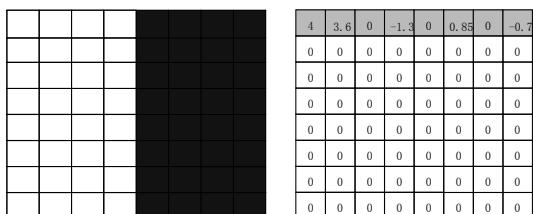


变换系数矩阵 $F(u, v)$

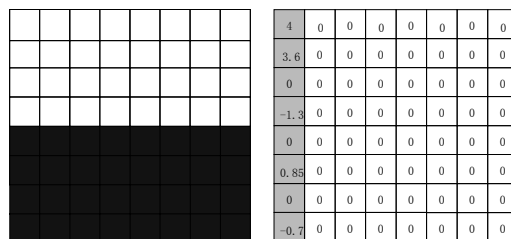


### 8.3.2 DCT变换举例

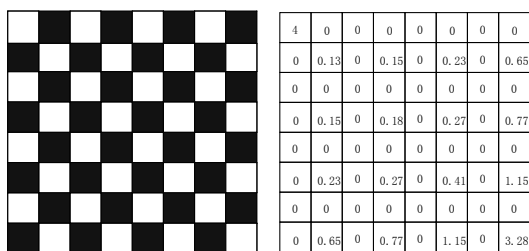
- (a): 仅具有水平空间频率成分



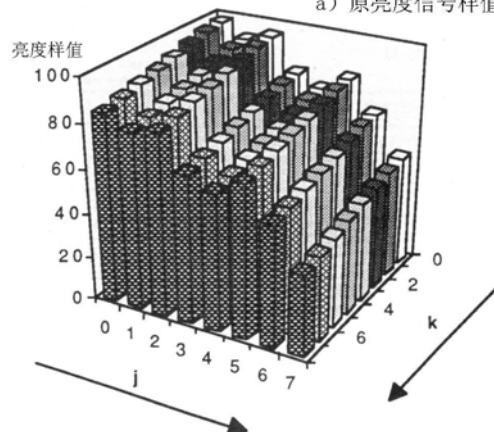
- (b): 仅具有垂直空间频率成分



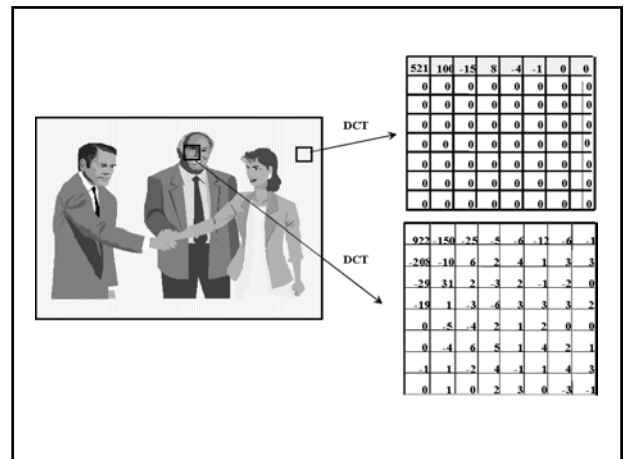
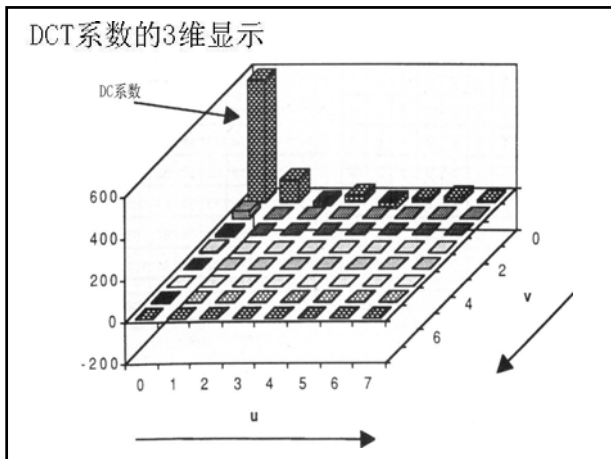
- (c): 既具有水平空间频率成分, 又具有垂直空间频率成分



a) 原亮度信号样值的3维显示







### 8.3.3 量化器

- 变换只是将图像从空间域变换到变换域中，其元素个数并未减少，变换本身并不能压缩码率，只有通过量化器对系数进行量化和编码后才能降低要传输的数据量。
- 由于量化过程是在变换域进行的，因此设计量化器应根据图像在变换域中的分布特性。因为大多数电视信号如背景等部分亮度值变化很少，变换域后的变换系数的大部分能量集中在直流和低频区域，又因为亮度突变区域如轮廓、边缘等高频细节部分，变换后的系数能量一般都很小。为了得到好的编码效果，应该根据系数块中的不同位置采用不同的量化精度。

JPEG压缩算法中给出的量化矩阵 $Q(u,v)$

<div><div>LL</div><div>HL</div></div> <div><div>LH</div><div>HH</div></div>								LL:细量化 LH、HL:中等量化 HH:粗量化							
16	11	10	16	24	40	51	61	17	18	24	47	99	99	99	99
12	12	14	19	26	58	60	55	18	21	26	66	99	99	99	99
14	13	16	24	40	57	69	56	24	26	56	99	99	99	99	99
14	17	22	29	51	87	80	62	47	66	99	99	99	99	99	99
18	22	37	56	68	109	103	77	99	99	99	99	99	99	99	99
24	35	55	64	81	104	113	92	99	99	99	99	99	99	99	99
49	64	78	87	103	121	120	101	99	99	99	99	99	99	99	99
72	92	95	98	112	100	103	99	99	99	99	99	99	99	99	99
亮度量化矩阵								色度量化矩阵							

- 量化矩阵 $Q(u,v)$ 充分利用了人眼对图像的低频分量比对高频分量更敏感的视觉特性，量化因子随空间频率 $u,v$ 的增加而增大，即在低频区进行细量化，在高频区进行粗量化。
- 量化公式为
 
$$[F(u,v)]_Q = [F(u,v) / Q(u,v)]_{\text{取整}}$$
- 其中 $[F(u,v)]_Q$ 是系数 $F(u,v)$ 的量化近似值， $Q(u,v)$ 是量化系数。

- 量化矩阵 $Q(u,v)$ 充分利用了人眼对图像的低频分量比对高频分量更敏感的视觉特性，量化因子随空间频率 $u,v$ 的增加而增大，即在低频区进行细量化，在高频区进行粗量化。
- 量化公式为
 
$$[F(u,v)]_Q = [F(u,v) / Q(u,v)]_{\text{取整}}$$
- 其中 $[F(u,v)]_Q$ 是系数 $F(u,v)$ 的量化近似值， $Q(u,v)$ 是量化系数。

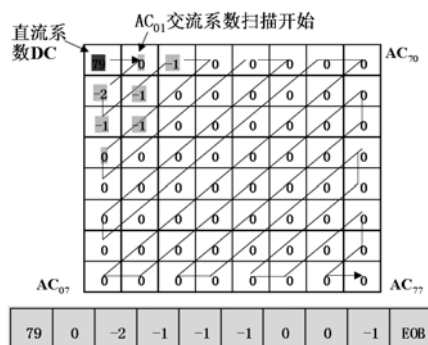
## DCT变换及压缩编码举例

8×8亮度块 $f(x, y)$								8×8DCT系数块 $F(u, v)$							
139	144	149	153	155	155	155	155	1260	-1	-12	-5	2	-2	-1	1
144	151	153	156	159	156	156	156	-23	-17	-6	-3	-3	0	0	-1
150	155	160	163	158	156	156	156	-11	-9	-2	2	0	-1	-1	0
159	161	162	160	160	159	159	159	-7	-2	0	1	1	0	0	0
159	160	161	162	162	155	155	155	-1	-1	1	2	0	-1	1	1
161	161	161	161	160	157	157	157	2	0	2	0	-1	1	1	-1
162	162	161	163	162	157	157	157	-1	0	0	-1	0	2	1	-1
162	162	161	161	163	158	158	158	-3	2	-4	-2	2	1	-1	0
亮度量化矩阵 $Q(u, v)$								量化后的DCT系数 $[F(u, v)]_q$							
16	11	10	16	24	40	51	61	79	0	-1	0	0	0	0	0
12	12	14	19	26	58	60	55	-2	-1	0	0	0	0	0	0
114	13	16	24	40	57	69	56	-1	-1	0	0	0	0	0	0
14	17	22	29	51	87	80	62	0	0	0	0	0	0	0	0
18	22	37	56	68	103	103	77	0	0	0	0	0	0	0	0
24	35	55	64	81	104	113	92	0	0	0	0	0	0	0	0
49	64	78	87	108	121	120	101	0	0	0	0	0	0	0	0
72	92	95	98	112	100	103	99	0	0	0	0	0	0	0	0

- $F(u, v)$ 中高频系数都比较小，经过粗量化以后，高频系数已经大部分为0，只有少数几个低频系数集中在左上角，从而大大压缩了需要传输的数据量。
- 量化不仅压缩了数据量，也给图像带来了失真。在DCT编码中，主要给图像带来如下失真：舍去高频系数而使图像产生模糊，对某些系数采用粗量化而产生颗粒状结构，在压缩倍数较大的情况下，会产生块效应失真。

(a) 解码后重建图像数据								(b) 重建图像与原图像块的差值 $e(x, y)$							
144	146	149	152	154	156	156	156	5	2	0	-1	-1	1	1	1
148	150	152	154	156	156	156	156	4	-1	-1	-2	-3	0	0	0
155	156	157	158	158	157	156	155	5	1	-3	-5	0	1	0	-1
160	161	161	162	161	159	157	155	1	0	-1	2	1	0	-2	-4
163	163	164	163	162	160	158	156	4	3	3	1	0	5	3	1
163	164	164	164	162	160	158	157	2	3	3	3	2	3	1	0
160	161	162	162	162	161	159	158	-2	-1	1	-1	0	4	2	1
158	159	161	161	162	161	159	158	-4	-3	0	0	-1	3	1	0
139	144	149	153	155	155	155	155								
144	151	153	156	159	156	156	156								
150	155	160	163	158	156	156	156								
159	161	162	160	160	159	159	159								
159	160	161	162	162	155	155	155								
161	161	161	161	160	157	157	157								
162	162	161	163	162	157	157	157								
162	162	161	161	163	158	158	158								

### ZigZag(Z字形)扫描输出路径



## 8.4 变字长编码 (VLC)

- 变字长编码 (Variable-length Coding, VLC) 是将输入的特编码的符号映射成一系列可变长的码字，并且根据符号发送概率的不同而分配以不同的码长的码字，对于出现概率大的符号给以短码，对于概率小的符号给以长码，通过大量的编码后，平均码长小于按其他方法编码得到的平均码长，可以使总码率降低，达到去除数据的统计冗余，压缩码率的目的。
- VLC编码方法有很多种，如哈夫曼编码（利用概率分布特性）、游程编码（利用相关性）、算术编码（利用概率分布的编码方法）等。Huffman编码是在数字电视压缩编码中常用的一种效率高、方法简单的编码方法。

## 8.5 视频压缩标准概述

- 国际标准化组织(International Standardization Organization, ISO)、国际电工委员会(International Electrotechnic Committee, IEC)、国际电信联盟(International Telecommunication Union, ITU)等国际组织,陆续制定了多种视音频压缩编码标准。
- 压缩标准的制定大大推动了数字电视技术及多媒体技术的发展,以下主要介绍在数字电视中得到广泛应用的数字视音频压缩编码标准JPEG、MPEG-2和H2.64。

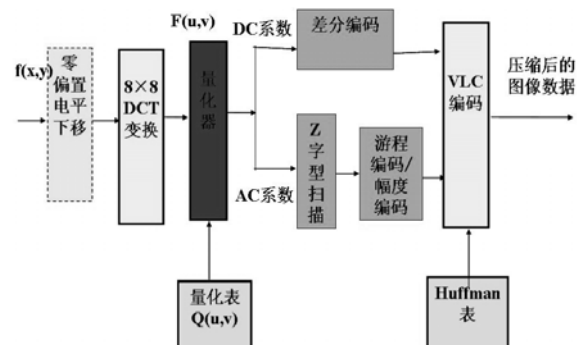
常用名称	制定机构	发布日期	标准编号	标题
H.261	ITU-T	1990	ITU-T H.261	运动图像编码标准
MPEG-1	ISO/IEC	1991	ISO/IEC 11172	运动图像和伴音的编码第2部分视频
JPEG	ISO/IEC	1992	ITU-T T.81 ISO/IEC 10918	静止图像编码
MPEG-2	ISO/IEC ITU-T	1994	ITU-T H.262 ISO/IEC 13818-2	运动图像和伴音的通用编码第2部分视频
H.263	ITU-T	1995	ITU-T H.263	运动图像编码标准
MPEG-4	ISO/IEC	2000	ISO/IEC 14496-2	基于对象的音视频编码第2部分视频
JPEG2000	ISO/IEC	2000	ISO/IEC 15444-3	静止图像编码
H.264/AVC /MPEG-4-10	ITU-T ISO/IEC	2003	ITU-T Rec.H.264 /ISO/IEC 14496-10	高级视频编码
AVS	中国国家 标准	2006	GB/T 20090-2006	先进音视频编码 第2部分

## • 8.6 JPEG压缩编码标准

- **JPEG** (Joint Photo-Graphic Experts Group, 联合图像专家组) 组织,负责制定与静止图像编码的相关标准,所制定的标准有JPEG、JPEG2000等,这些标准在数码相机、数字视频制作、数字电影等领域中得到广泛应用。
- **JPEG**组织由多个国际组织联合制定:国际标准化组织ISO/IEC、国际电信联盟ITU)、国际电工委员会(IEC)也参加合作。
- **JPEG**组织从1986年正式开始制订,该标准于1992年正式通过,称为JPEG标准。

- JPEG包含两种基本压缩方法:
  - 有损压缩:基于DCT的有损压缩,
  - 无损压缩:又称预测压缩方法
- JPEG编码压缩是在本帧内进行,与前后帧无关;
- JPEG也可用于活动图像压缩编码,只不过压缩方法只限于帧内进行,不作运动补偿和帧间差运算。这种用于活动图像的压缩技术称作:M-JPEG(Motion-JPEG),主要用于高质量图像压缩编码方案中,例如电视台节目编辑中。

### 1. JPEG编码系统框图



- **JPEG编码算法主要有以下几个重要步骤:**
- **(1) 通过DCT变换减少数据的空间冗余;**
- **(2) 用与人眼视觉特性相关的量化矩阵量化DCT系数 $F(u,v)$ , 减少视觉冗余;**
- **(3) 对量化后的直流系数DC和交流系数AC分别进行可变长编码 (VLC), 减少编码冗**

### 1. 图像分块与电平偏置

- 首先将图像分割成相互不重叠的矩形块, 每一个像块作为一个独立的单元进行变换和编解码。
- 由于像素的平均亮度较高, 为了提高编码效率, 在DCT变换前, 对像块进行零偏置, 即对于灰度级是256的像素, 通过减去128。

139	144	149	153	155	155	155	155
144	151	153	156	159	156	156	156
150	155	160	163	158	156	156	156
159	161	162	160	160	159	159	159
159	160	161	162	162	155	155	155
161	161	161	161	160	157	157	157
162	162	161	163	162	157	157	157
162	162	161	161	163	158	158	158

11	16	21	25	27	27	27	27
16	23	25	28	31	28	28	28
22	27	32	35	30	28	28	28
31	33	34	32	32	31	31	31
31	32	33	34	34	27	27	27
33	33	33	33	32	29	29	29
34	34	33	35	34	29	29	29
34	34	33	33	35	30	30	30

- **2. 离散余弦变换和DCT系数量化**
- **JPEG标准是将“二维”空间的图像数据用离散余弦变换DCT公式变换到二维频域, 成为二维频率系数。变换后的系数经量化矩阵量化**

亮度量化矩阵

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

色度量化矩阵

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

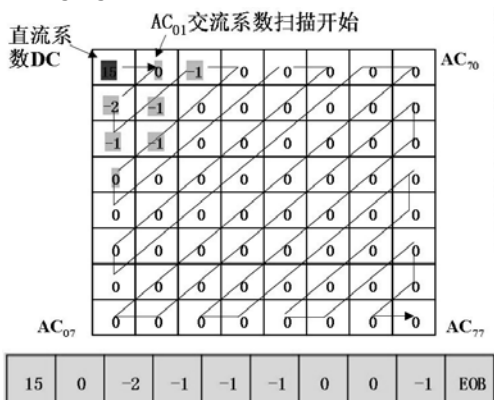
DCT系数 $F(u, v)$

235	-1.0	-12.1	-5.2	2.1	-1.7	-2.7	1.3
-22.6	-18.5	-6.2	-3.2	-2.9	-0.1	0.4	-1.2
-10.9	-9.3	-1.6	1.5	0.2	-0.9	-0.6	-0.1
-7.1	-1.9	0.2	1.5	0.9	-0.1	0.0	0.3
-0.6	-0.8	1.5	1.6	-0.1	-0.7	0.6	1.3
1.8	-0.2	-1.6	-0.3	-0.8	1.5	1.0	-1.0
-1.3	-0.4	-0.3	-1.5	-0.5	1.7	1.1	-0.8
-2.6	1.6	-3.8	-1.8	1.9	1.2	-0.6	-0.4

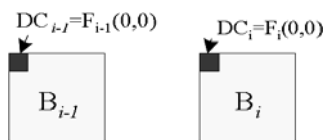
量化后的DCT系数 $[F(u, v)]_q$

15	0	-1	0	0	0	0	0
-2	-1	0	0	0	0	0	0
-1	-1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

### 3. ZigZag(Z字形)扫描



### 4. 对于直流分量DC进行差分脉冲编码



$$\Delta DC_i = DC_i - DC_{i-1}$$

$$\text{设 } \Delta DC_i = 3$$

符号[ (Size位长), (Amplitude幅度) ],

2

3

(查表8-5)

(查表8-4)

011

11

01111

幅度值 (Amplitude) VLC码表

位长 (Size)	幅度 (Amplitude)	幅度 (Amplitude) 码表
0	0	-
1	-1, 1	0, 1
2	-3, -2, 2, 3	00, 01, 10, 11
3	-7, ..., -4, 4, ..., 7	000, ..., 011, 100, ..., 111
4	-15, ..., -8, 8, ..., 15	0000, ..., 0111, 1000, ..., 1111
...	...	...
16	32768	-

8-5 亮度直流系数的Huffman码表

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

## 5. AC系数的游程-幅度编码

- 在AC系数编码中，0值不单独编码，而是将非0系数的幅值和0的游程这两个事件合并成一个二维事件，根据其出现的联合概率设计Huffman码表。
- (1) 将每个AC非0系数和它前面相邻的全部0系数分在一组内：
- {(0, -2); (-1); (-1); (-1); (0, 0, -1); EOB}。
- (2) 将每一组用两个符号表示(Run, Size), (Amplitude)
- 其中“Run”表示该组内0的游程；
- “Amplitude”表示该组内非0系数的“幅度”值；
- “Size”表示该组非0系数的“Amplitude”的编码位长
- 转换为：
- {[(1,2), (-2)]; [(0,1), (-1)]; [(0,1), (-1)]; [(0,1), (-1)]; [(2,1), (-1)]; (0,0)}，其中，(0, 0)表示块结束符EOB。
- 查8-5后：AC系数全部码字为
- {1101101, 000, 000, 000, 111000, 1010}

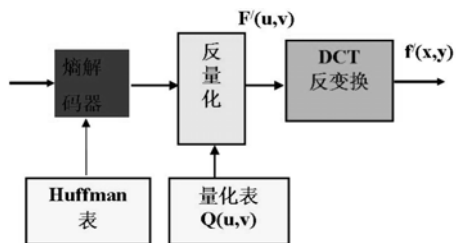
8-6 亮度交流系数的Huffman码表

游程, 位长	Huffman码字	码长	游程, 位长	Huffman码字	码长
(0,0)(EOB)	1010	4	(2,1)	11100	5
(0,1)	00	2	(2,2)	11111001	8
(0,2)	01	2	(2,3)	1111110111	10
(0,3)	100	3	(2,4)	11111110100	12
(0,4)	1011	4	(2,5)	111111110001001	16
(0,5)	11010	5	(2,6)	1111111110001010	16
(0,6)	1111000	7	(2,7)	1111111110001011	16
(0,7)	11111000	8	(2,8)	1111111110001100	16
(0,8)	1111110110	10	(2,9)	1111111110001101	16
(0,9)	1111111110000010	16	(2,A)	1111111110001110	16
(0,A)	11111111110000011	16	(3,1)	111010	6
(1,1)	1100	4	(3,2)	111110111	9
(1,2)	11011	5	(3,3)	11111110101	12
(1,3)	1111001	7	(3,4)	111111110001111	16
(1,4)	111110110	9	(3,5)	1111111110010000	16
(1,5)	11111110110	11	(3,6)	1111111110010001	16
(1,6)	1111111110000100	16	(3,7)	1111111110010010	16
(1,7)	1111111110000101	16	(3,8)	1111111110010011	16
(1,8)	1111111110000110	16	(3,9)	1111111110010100	16
(1,9)	1111111110000111	16	(3,A)	1111111110010101	16
(1,A)	11111111100001000	16	(4,1)	111011	6

- 整个编码过程如下
- 之字形扫描后：
- (15, 0, -2, -1, -1, -1, 0, 0, -1, EOB)
- DC系数差值后
- (3, 0, -2, -1, -1, -1, 0, 0, -1, EOB)
- (Run, Size), (Amplitude):
- {(2),(3); (1,2), (-2); (0,1), (-1); (0,1), (-1); (0,1), (-1); (2,1), (-1); (0,0)}
- 查表后：全部码字为
- {01111, 1101101, 000, 000, 000, 111000, 1010}

- 在本例中，原像块为 $8 \times 8 = 64$ 个像素，压缩前的总比特位数为 $8 \times 8 \times 8 = 512$ 。
- 经过JPEG编码后，输出的总比特位数为31位，压缩比为 $512/31 = 16.5$ 。

## JPEG解码过程



(a) 解码后重建图像数据

144	146	149	152	154	156	156	156
148	150	152	154	156	156	156	156
155	156	157	158	158	157	156	155
160	161	161	162	161	159	157	155
163	163	164	163	162	160	158	156
163	164	164	164	162	160	158	157
160	161	162	162	162	161	159	158
158	159	161	161	162	161	159	158

(b) 重建图像与原图像块的差值 $e(x, y)$

5	2	0	-1	-1	1	1	1
4	-1	-1	-2	-3	0	0	0
5	1	-3	-5	0	1	0	-1
1	0	-1	2	1	0	-2	-4
4	3	3	1	0	5	3	1
2	3	3	3	2	3	1	0
-2	-1	1	-1	0	4	2	1
-4	-3	0	0	-1	3	1	0

## • 8.7 MPEG-2视频压缩编码

### 8.7.1 MPEG-1和MPEG-2概述

- 运动图像专家组（Moving Picture Experts Group, MPEG）于1988年开始工作，其任务是对应用于数字存储媒介、广播电视和通信的运动图像及其相关声音制定一种通用的编码标准。针对不同的应用目的，MPEG专家组制定了MPEG系列标准。

### MPEG-1概述

- ◆ MPEG-1的全称为“Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s”，标准的文件编号为ISO/IEC11172，标准的发布日期是1991年。
- ◆ MPEG-1由多个部分组成，其中主要的部分有三个：
- ◆ 第一部分系统（ISO/IEC11172-1），是关于数字视频、数字音频和辅助数据等多路压缩数据流复用和同步的规定。
- ◆ 第二部分视频（ISO/IEC11172-2），是关于位速率约为1.5 Mbps的视频信号的压缩编码的规定。
- ◆ 第三部分音频（ISO/IEC11172-3），是关于每通道位速率为64kbps、128kbps、192kbps的数字音频信号的压缩编码的规定。

- ◆ MPEG-1处理的视频图像格式为CIF格式，即  $352 \times 240/30P$  和  $352 \times 288/25P$ 。

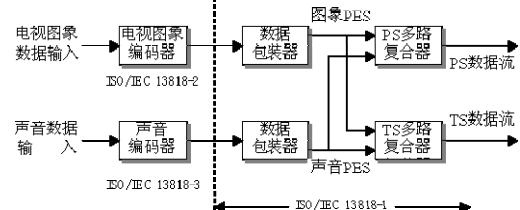
- ◆ MPEG-1只能压缩逐行扫描视频，如果待处理的视频信号是隔行扫描格式，则编码前需要转换为逐行格式后再进行压缩。
- ◆ MPEG-1视频编码方式采用DPCM、DCT和运动补偿编码技术结合的混合编码方式，较大的提高了编码效率。
- ◆ MPEG-1视频主要应用于VCD等领域。

## MPEG-2概述

■ MPEG-2全称为“运动图像及有关声音信息的通用编码；Coding of moving pictures and associated audio”，标准的文件编号为ISO/IEC13818，标准的发布日期是1994年。

● MPEG-2由多个部分组成，其中主要部分有三个；

- 第一系统部分：系统（Systems）部分（ISO/IEC13818-1），是关于多路音频、视频和数据的复用和同步的规定；
- 第二视频部分：图像（Video）部分（ISO/IEC13818-2），主要涉及各种比特率的数字视频压缩编解码的规定；
- 第三音频部分：声音（Audio）部分（ISO/IEC13818-3），扩充了MPEG-1的音频标准，使之成为多通道音频编码系统；



## MPEG-1与MPEG-2的异同

- MPEG-1 和 MPEG-2的视频编码方式是类似的，都是基于块的DCT变换编码、帧间预测和运动补偿编码技术的混合编码。
- MPEG-2支持的图像格式不仅包括了CIF图像格式，还包括了SDTV和HDTV数字图像格式，能够向后兼容MPEG-1。
- MPEG-2中，视频序列可以是隔行扫描的，也可以是逐行扫描的，而在MPEG-1中视频序列只能是逐行的。
- MPEG-1中，图像的色度格式是4:2:0，而在MPEG-2中图像的色度格式还包括4:2:2格式。
- MPEG-1中，像条不必在同一宏块水平行中开始和结束，而在MPEG-2中像条永远在同一宏块水平行内开始和结束。
- 在MPEG-1中，运动矢量搜索的精度是以整像素为单位的，而在MPEG-2中运动矢量搜索的精度是以半像素为单位。

## ●MPEG-2的应用

- MPEG-2已广泛应用于DVD，SDTV 和 HDTV数字电视广播系统中。美国的ATSC、欧洲的DVB、日本的ISDB都将MPEG-2作为信源编码标准。我国制定的数字卫星电视，数字有线电视和数字地面电视标准中，信源编码也采用MPEG-2标准。
- 由于MPEG-1和 MPEG-2的视频压缩编解码原理基本相同，因此以下主要介绍MPEG-2视频压缩编解码原理。

## 8.7.2 MPEG-2的型和级

■ 为了使MPEG-2能适应图像质量从低到高的各种应用领域，MPEG-2标准定义了“型”（Profile，）和“级”（Level）

- “型”（Profile），规定了不同的压缩处理方法，共有5种配置；
- “级”（Level），规定了不同的输入视频格式（Video Format），共有4种等级；

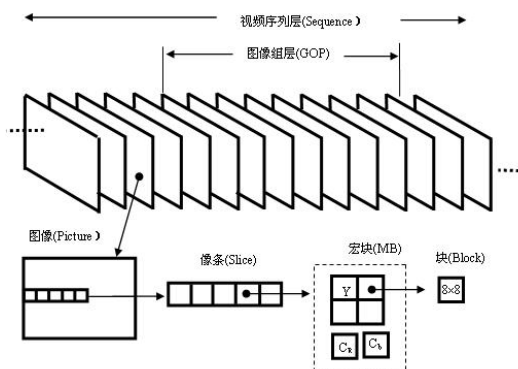
MPEG-2的级和型

型profile 级level	简单型 SP (4:2:0) (IP)	主型 MP (4:2:0) (IBP)	信噪比可级 SNRP (4:2:0) (IBP)	空间可分 级SSP (4:2:0) (IBP)	高型HP (4:2:2) 4:2:0 (IBP)
高级HL 1920×1080×30 1920×1080×25		MP@HL			HP@HL
1440×1080×30 1440×1080×25		MP@H14 40L		SSP@H1 440L	HP@H14 0L
级ML 720×480×30 720×576×25	SP@ML (无B帧)	MP@ML	SNRP@ML		HP@ML
低级LL 352×248×30 352×288×25		MP@LL	SNRP@LL		

- MPEG-2将处理的视频图像格式分成四级：低级（Low Level, LL）为CIF视频图像格式、主级（Main Level, ML）为标清视频图像格式、高1440级（High-1440 Level）H-1440为准高清视频图像格、高级（High Level）为高清视频图像格式。分别对应不同的应用领域。

- 在数字电视传输应用中，主要的应用模式为MP@ML和HP@ML、4: 2: 0格式、长GOP（15或12帧）。
- 但是在要求较高图像质量的演播室制作环境中，MPEG-2也用于压缩4: 2: 2视频，短GOP模式，如GOP为4或全I帧。

### 8.7.3 MPEG-2视频结构



### MPEG-2对视频数据规定了六层结构

- 视频序列层（Video Sequence, VS）
- 图像组层（Group of Pictures, GOP）
- 图像层（Picture）
- 像条层（Slice）
- 宏块层（MacroBlock, MB）
- 像块层（Block），

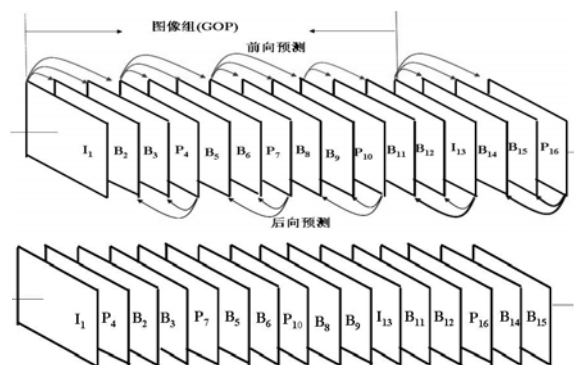
- 视频序列层（Video Sequence, VS）：
- 视频序列是随机选取节目的一个基本单元。
- 图像组层（Group of Pictures, GOP）：
- GOP是将一个图像序列中连续的几个图像组成一个小组，简称为GOP。它是对编码后的视频码流进行编辑存取的基本单元。在MPEG-2编码规范中，对于GOP的长度及GOP内部的I, P, B帧的分布，没有作规定，这些均由设计者根据需要确定。常用的GOP的组成是：IBBPBBPBBPBB，即每12帧中包含1个I帧、3个P帧和8个B帧。

- 图像(Picture):
- 图像是一个独立的显示单元，也是图像编码的基本单元，可分为I, P, B三种编码图像。一个编码图像可以是帧图像或场图像，对于逐行扫描图像来说，图像总是帧结构编码，即所谓的I, P, B帧，对于隔行扫描图像来说，图像可以是场结构编码也可以是帧结构编码。
- 像条(Slice):
- 像条由一系列连续的宏块组成。像条是发生误码后且不可纠正时，数据重新获得同步从而能正常解码的基本单元。



- 宏块(MacroBlock, MB):
- 宏块是运动预测的基本单元。一个宏块由一个  $16 \times 16$  像素的亮度阵列和同区域内的Cb、Cr 色差阵列共同组成, 因此一个宏块是由多个  $8 \times 8$  的块亮度块和色度块组成, 如图8-9所示。
- 像块 ((Block)):
- 一幅图像以像素数据阵列为基准划分成为若干个  $8 \times 8$  像素的阵列, 称为像块, 是DCT变换的基本单元。

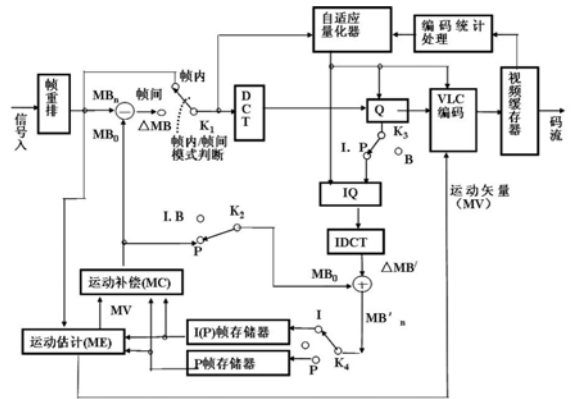
#### 8.7.4 I、B、P帧编码原理



#### ■ MPEG-2标准中, 定义了三种图像类型:

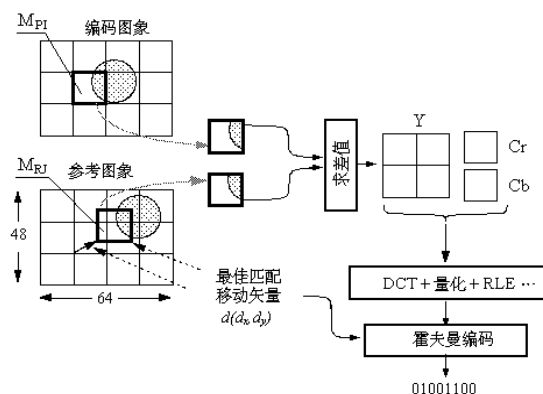
- 帧内编码图像 (Intra coded pictures), 简称I帧, 采用帧内编码, 不参照其他图像。I帧可以作为P和B帧的参考帧, 为了减少误差的传递, I帧的压缩比较低, 一般在  $(2 \sim 5):1$  的范围内。GOP的第一个编码帧为I帧。
- 预测编码帧图像 (Predictive coded pictures), 简称P帧, 参照前一幅I帧或P帧图像做运动补偿编码; P帧压缩比高于I帧, 可达到  $(5 \sim 10):1$ 。P帧可以作为B帧和后面的P帧的参考帧。
- 双向预测编码图像 (Bidirectionally-predictive coded pictures), 简称B帧, 参照前一幅和后一幅I帧或P帧图像做双向运动补偿编码。B帧的预测精度能做到很高, 其压缩比更大, 可达到  $(20 \sim 30):1$ 。

#### 8.7.5 MPEG编码器工作原理



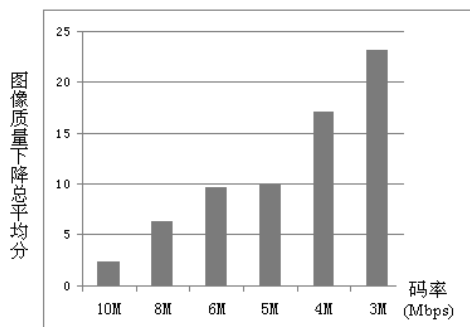
- MPEG-2标准采用了一系列技术以获得高压缩比
- (1) 利用运动补偿预测编码方法减少电视图像信号的时间冗余信息
- (2) 利用离散余弦变换 (DCT), 去掉经过运动补偿预测的图像差信号中的空间冗余信息
- (3) 对DCT系数进行量化, 舍去人眼不敏感、幅度又比较小高频系数;
- (4) 利用可变长编码技术, 对运动向量和DCT变换系数进行编码, 以达到充分利用通道带宽的目的

#### ■ 预测图像P的压缩编码算法框图

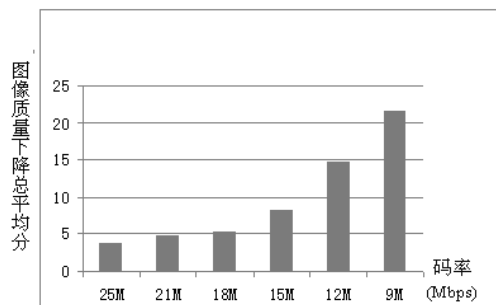




SDTV测试序列在不同的码率下图像质量主观评价分

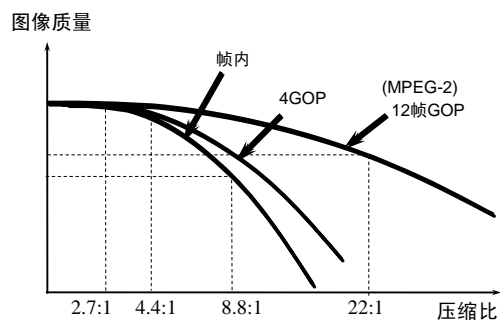


HDTV测试序列在不同码率下图像质量主观分



- 压缩编码后图像质量的下降程度与视频图像内容密切相关。
- 对于快速随机运动的序列、并有丰富的亮度细节成分的序列，帧间相关性非常小，运动估计时匹配程度差，压缩后图像质量下降也最多；

图像质量与压缩比的关系



#### ● 编解码延时问题

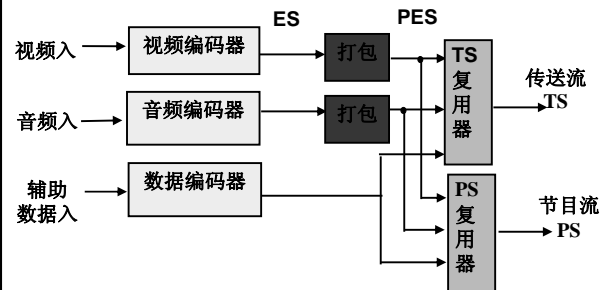
- 由于编码顺序和显示顺序之间的差异,带来编、解码总延时达0.8~1秒时间
- 这种延时对广播性质的图像传输不会带来任何影响,因为用户并不能发现发射端与接收端信号起始时间的差异
- 对于交互性质的图像传输,例如,数字会议电视则带来十分不便。

#### • 8.8 MPEG-2系统复用

### 8.8.1 系统复用

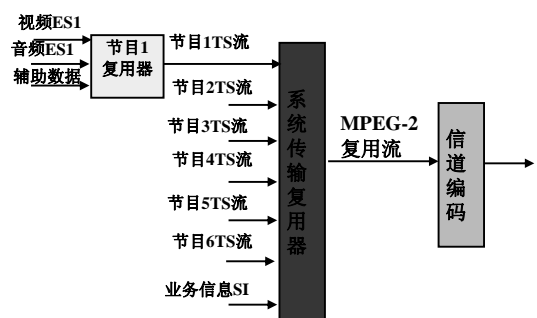
- 复用器是数字电视系统的关键设备之一，它将编码器来的视频和音频数据流，并按照一定的复用规范将其交织复用成符合MPEG-2系统层规范的单一的码流。
- 为了实现视音频的解码同步，在码流中还需插入各种时间标记和系统控制等信息，复用器还提供整个编码系统的启动信号及27MHz的时钟信号。

### 节目复用



- ES流(Elementary Stream):经过视音频编码器压缩编码后的视音频码流，称为基本流。
- 打包基本码流 (Packetised Elementary Streams, PES) : 将ES流按一定的长度分段，构成具有特定结构和长度的一个个单元包，称为打包的基本码流(PES)
- 节目码流 (Program Stream, PS流) : ，包的长度是不固定的，适用于信道较好的存储媒介中环境。
- 传输流 (Transport Stream, TS流) : TS包的长度是固定的，适用于性能差些的信道环境，如广播传输信道。

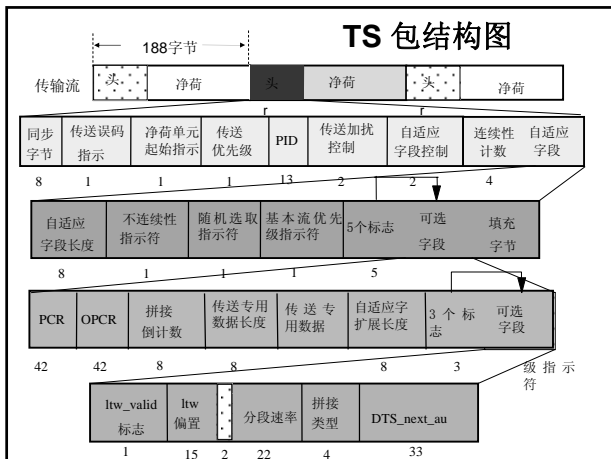
### 传输复用



- 在数字电视广播中，一个频道往往能传输多套电视节目，例如，一个8MHz带宽的有线电视频道，按照我国的数字有线电视传输标准，可以传送38Mbps的码率。
- 一般高质量的数字电视节目压缩后的每套标清视频码率为6Mbps、音频码率为256kbps，则每套节目的码率为6.256Mbps，因此需要将6套节目的TS复用到一个传输流中。

### 8.8.2 TS流分析

- ◆ TS包长共188字节，分为包头和净荷。
- ◆ 净荷部分放置的是PES包，
- ◆ 包头提供关于传输方面的信息，例如，同步、有无差错、有无加扰等标志。
- ◆ 在包头中的包标识符 (PID) 是13比特，用来标识数据包属于哪一个节目流的视频、音频或辅助数据。
- ◆ TS包头的长度是不固定的，分为前32比特 (4个字) 和自适应字段，包头的前32比特是最小包头。



### 8.8.3 节目专用信息 (PSI)

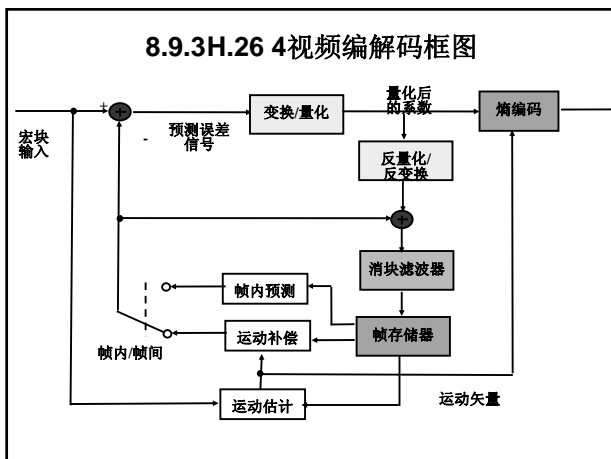
- 在MPEG-2的码流中必须包含向接收机提供选择控制作用的信息，以帮助接收端正确地进行解码。为此MPEG-2系统标准中定义了节目专用信息 (Program specific information, PSI)，它是MPEG码流中的重要组成部分。
- PSI信息主要由以下几种类型的表构成。
- 节目关联表(Program Association Table, PAT)
- 节目映射表(Program Map Table, PMT)
- 条件接收表(Conditional Access Table, CAT)
- 网络信息表(Network Information Table, NIT)

## • 8.9 H.264视频压缩编码

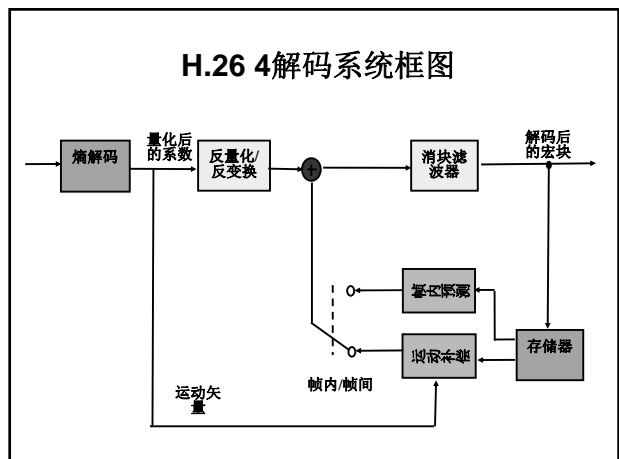
### 8.9.1 H.264概述

- H.264主要特点
- (1) 码率低，在同等图像质量下，采用H.264技术压缩后的数据量约有MPEG-2的1/2。
- (2) 容错能力强，H.264码流具有较强的抗误码特性，可适应丢包率高、干扰严重的信道中的视频传输，如IP和无线网络的应用。
- (3) 网络适应性强，H.264提供了网络适应层，使得H.264的文件可以在不同网络上传输。
- (4) 计算复杂度高，据估计，H.264编码的计算复杂度相当于MPEG-2的2倍~3倍，增加了编码器实现的难度。

### 8.9.3H.264视频编解码框图



### H.264解码系统框图



- H. 264之所以能达到较高的编码效率，与MPEG-2相比主要增加了以下方法，
- (1) 帧内预测 (Intra Prediction)、
- (2) 高精度运动估计和运动补偿技术、
- (3) 整数DCT变换和Hadamard变换编码
- (4) 消块滤波器等。