



# 多媒体技术

# 回顾

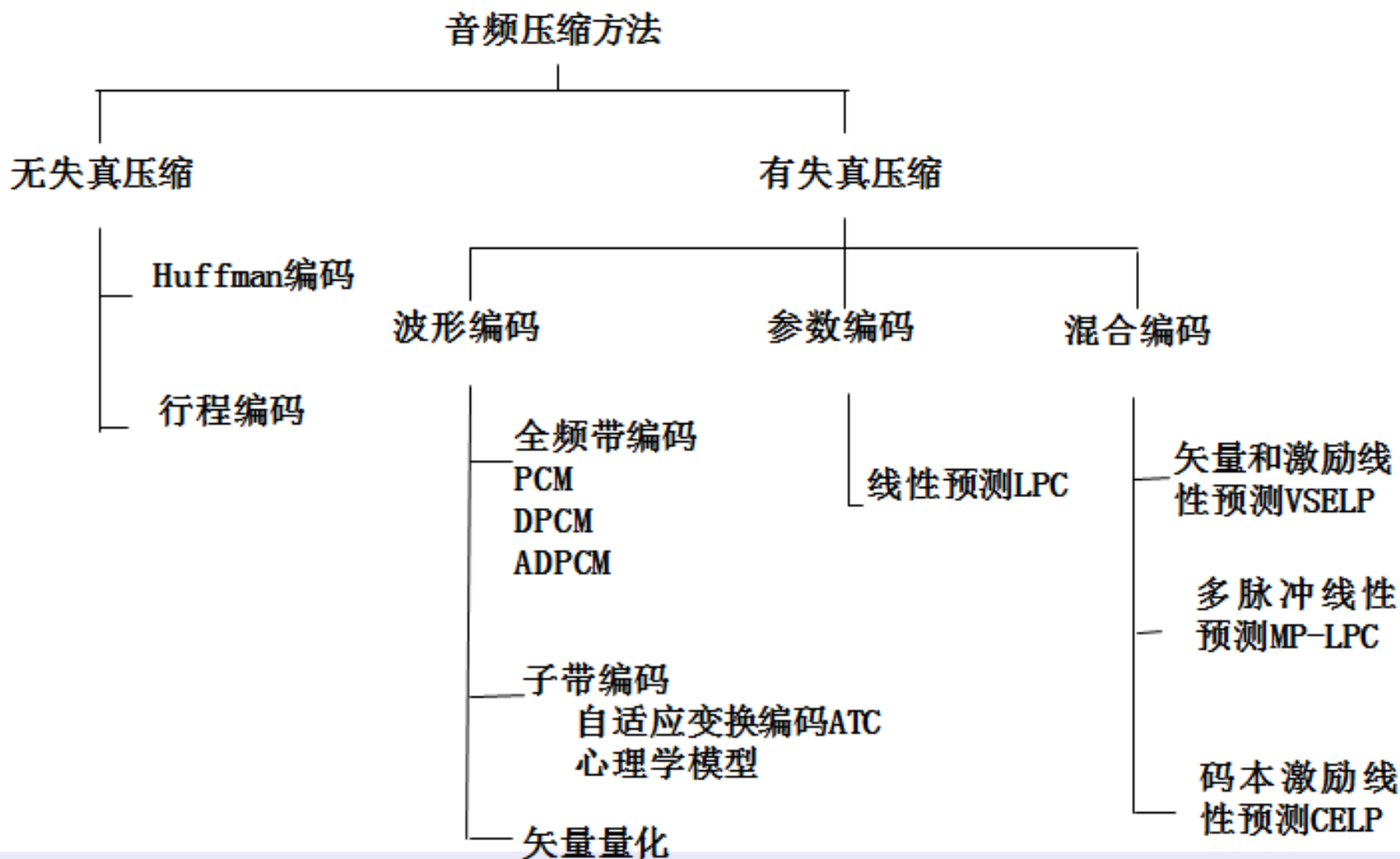
---

- 小波变换
  - 哈尔小波变换
  - **EZW**压缩



## 3.3 音频压缩标准

# 3.3.1 音频压缩编码的基本方法



# 3.3.1 音频压缩编码的基本方法

---

- 波形编码

- 利用**抽样**和**量化**过程来表示音频信号的波形，使编码后的音频信号与原始信号的波形尽可能匹配；
- 在**较高码率**的条件下获得高质量的音频信号；
- 适合**高保真语音**和**音乐**信号；
- 压缩率不大。

# 3.3.1 音频压缩编码的基本方法

- 参数编码

- 从话音波形信号中提取**话音生成模型**的**参数**，使用这些参数通过话音生成模型重构出话音。
- 压缩率很大，计算量大，保真度不高，适合**语音信号**的编码；
- 保密性能好，可用在**军事**上。

- 混合编码

- 介于波形编码和参数编码之间
- 在较低的码率上得到较高的音质。
- **CELP、MPLPC**

## 3.3.2 电话质量的语音压缩标准

---

- 频率范围
  - 300Hz-3.4kHz
  - 用标准的PCM，采样频率为8kHz，量化位数为8bit时对应的速率是64kbit/s。
- ITU压缩标准
  - G.711
    - 采用PCM编码，采样速率为8kHz，量化位数为8bit，对应的比特流速率为64kbit/s；
    - 使用非线性量化技术；

## 3.3.2 电话质量的语音压缩标准

- ITU压缩标准

- G.721

- 将64kbps的比特流转换为32kbps的流，基于ADPCM技术；
    - 每个数值差分用4位编码，其采样率为8kHz。

- G.723

- 一种以24kbit/s运行的基于ADPCM的有损压缩标准。其音质不如非压缩的 G.711PCM 标准；
    - 适用于低速率多媒体服务中语音或音频信号的压缩算法。

- G.728

- 比特率为16kbit/s，带宽限于3.4kHz；
    - 其音质比G.711差得多，与G.721相当。



# 3.3.3 调幅广播质量的音频压缩标准

---

- 频率范围
  - 50Hz-7kHz, 称 “7kHz音频信号”
  - G.722标准
    - 基于子带 **ADPCM** 技术 (**SB-ADPCM**) , 它是将现有的带宽分成两个独立的子带信道分别采用差分脉码调制算法;
    - 主要是保持 **64kbit/s** 的数据率。

## 3.3.4 高保真立体声音频压缩标准

---

- 频率范围
  - 50Hz-20kHz
  - MPEG音频
- MPEG声音编码分为：第1层、第2层、第3层。随着层数的增加算法的复杂度也增大。
- 所有3层都分级兼容。层3工作的译码器也可对层2或层1的码流进行译码。

# 3.3.4 高保真立体声音频压缩标准

- 频率范围

- 第1层—MP1(MPEG Audio Layer 1)

- 仅利用频域掩蔽特性，典型的压缩比为1:4，相应的数据率为384 kbps
    - 算法复杂度最低

- 第2层—MP2(MPEG Audio Layer 2)

- 利用频域掩蔽特性和时间掩蔽特性，典型的压缩比为1:6 ~ 1:8，数据率为256 ~ 192 kbps
    - 算法复杂度中等

- 第3层—MP3(MPEG Audio Layer 3)

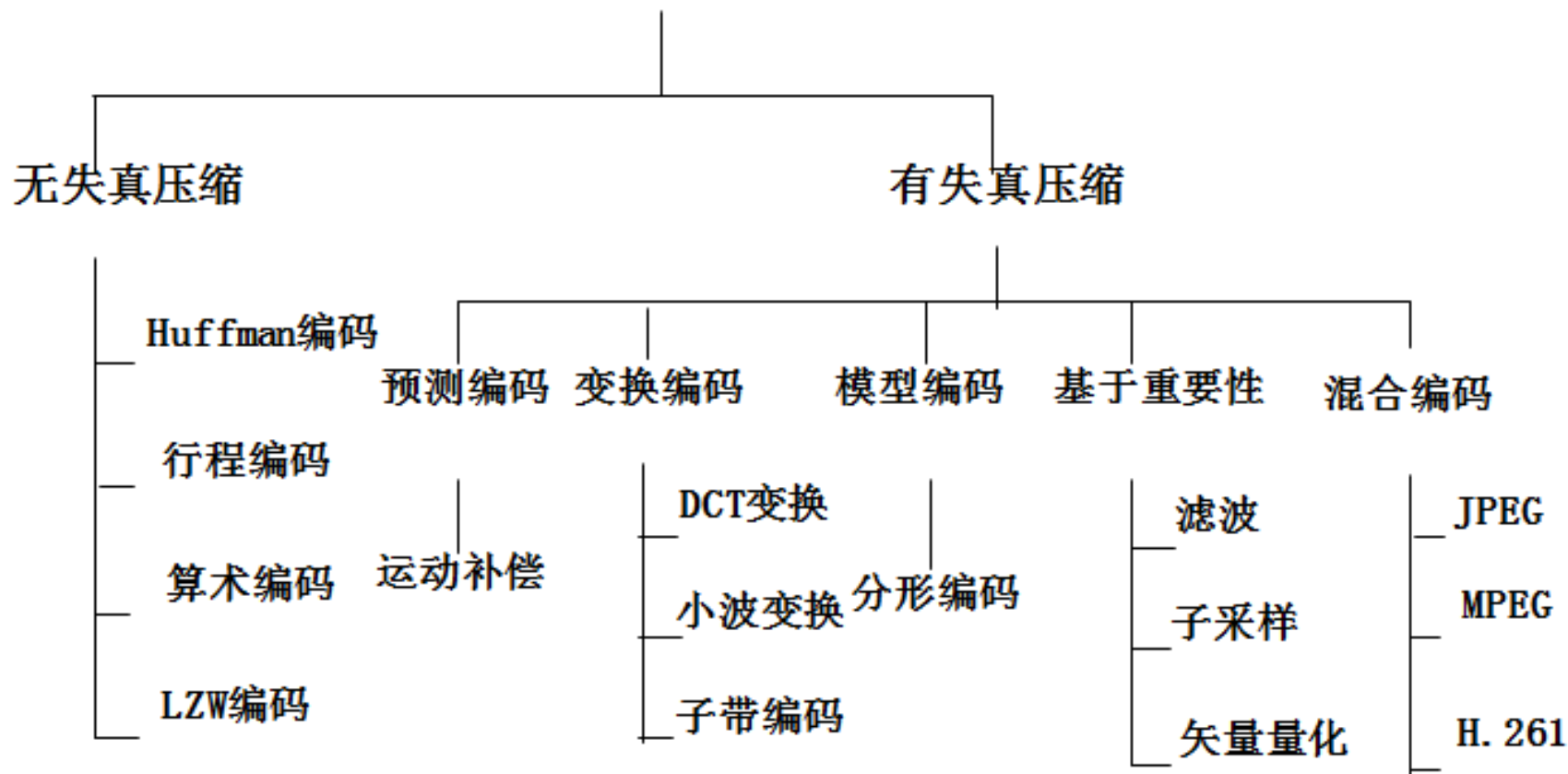
- 利用频域掩蔽特性、时间掩蔽特性和临界频带特性，典型的压缩比为1:10 ~ 1:12，相应的数据率为128 ~ 112 kbps，声音质量接近CD-DA
    - 算法复杂度最高



## 3.4 图像和视频压缩标准

# 3.4.1 图像和视频压缩编码的基本方法

图像和视频压缩方法

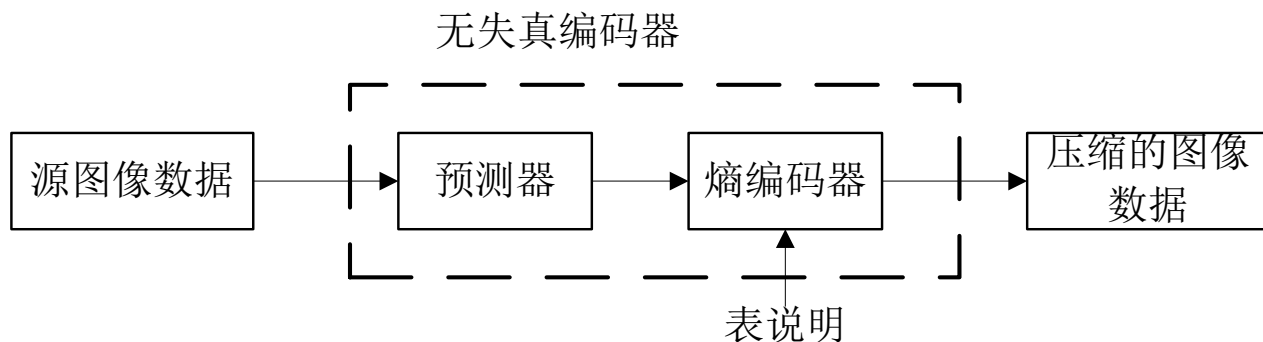


## 3.4.2 静止图像压缩标准

- **JPEG(Joint Photographic Experts Group)**
  - 是一个由 **ISO**和**IEC**两个组织机构联合组成的一个专家组，负责制定**静态的数字图像**数据压缩编码标准，因此又称为**JPEG**标准。
  - **JPEG**是一个适用范围很广的静态图像数据压缩标准，既可用于**灰度图像**又可用于**彩色图像**。
- **JPEG**专家组开发了两种基本的压缩算法
  - 一种是采用以**离散余弦变换**为基础的有损压缩算法；
  - 另一种是采用以**预测技术**为基础的无损压缩算法。

## 3.4.2 静止图像压缩标准

- JPEG的无损预测编码算法



	c	b		
	a	X		

选择值	预测	选择值	预测
0	非预测	4	$a+b-c$
1	a	5	$a+((b-c)/2)$
2	b	6	$b+((a-c)/2)$
3	c	7	$(a+b)/2$

## 3.4.2 静止图像压缩标准

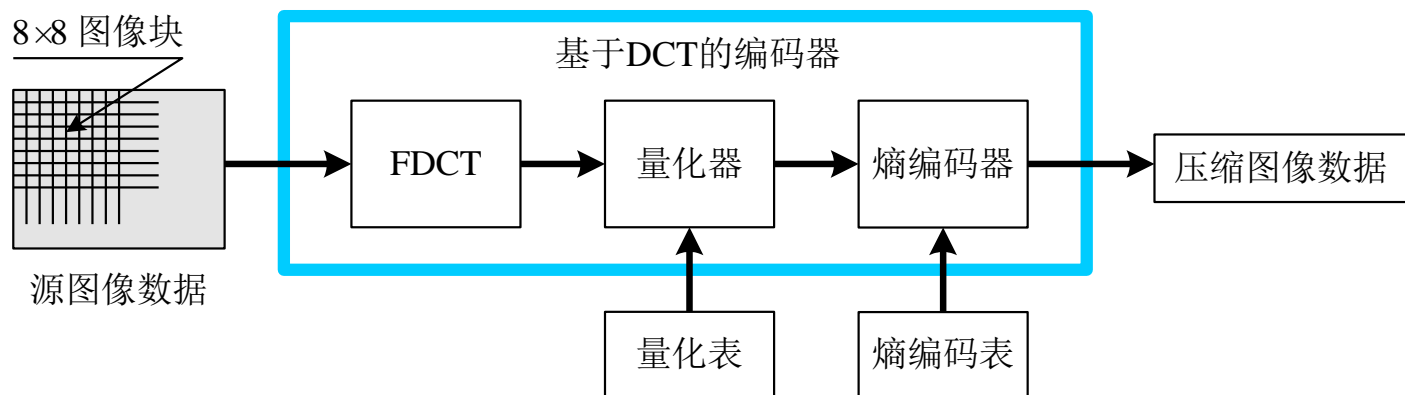
- **JPEG**的有损预测编码算法

- 利用了人的视觉特性，使用**量化**和**无损压缩编码**相结合来去掉视觉的冗余信息和数据本身的冗余信息。压缩编码大致分成三个步骤：

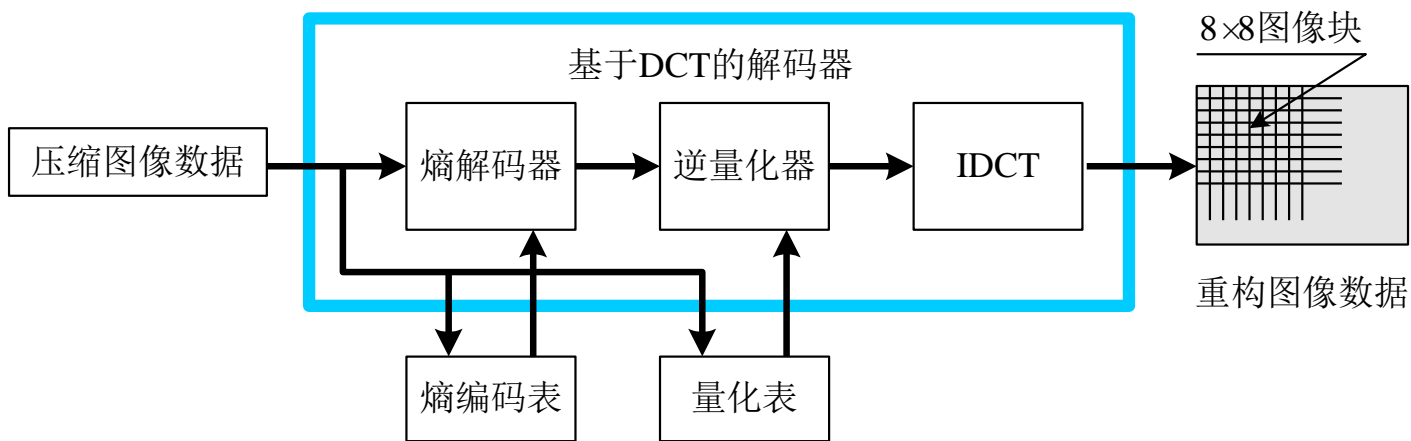
- 使用正向离散余弦变换**FDCT**(forward discrete cosine transform)把**空间域**表示的图变换**频率域**表示的图；
    - 使用加权函数对**DCT系数**进行**量化**，这个加权函数对于人的视觉系统是最佳的；
    - 使用**霍夫曼**可变字长编码器对量化系数进行编码。



## 3.4.2 静止图像压缩标准



(a) JPEG压缩算法框图



(b) JPEG解压缩算法框图

## 3.4.2 静止图像压缩标准

### • 1、块准备

- 将一帧图像分成 **$8 \times 8$** 的数据块。
- 假设彩色图像由Y、U和V三种分量组成，且色度分解为4: 1: 1
- 对于 **$640 \times 480$** 的图像，**亮度分量Y**是一个 **$640 \times 480$** 的数值矩阵；**色差分量**是一个 **$320 \times 240$** 的数值矩阵
- 为了满足DCT过程的要求，块准备必须划分出**4800**个亮度块和两份**1200**个色差块。
- 将原始数据从**无符号整数**变成**有符号整数**。
  - 若采样精度为P位，采样数据范围在 $[0, 2^P-1]$ 内，则变成在范围 $[-2^{P-1}, 2^{P-1}-1]$ 内。

## 3.4.2 静止图像压缩标准

- 2、离散余弦变换DCT

- 对每个单独的彩色图像分量，把整个分量图像分成  $8 \times 8$  的图像块
- DCT变换使用下式计算

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

- 逆变换

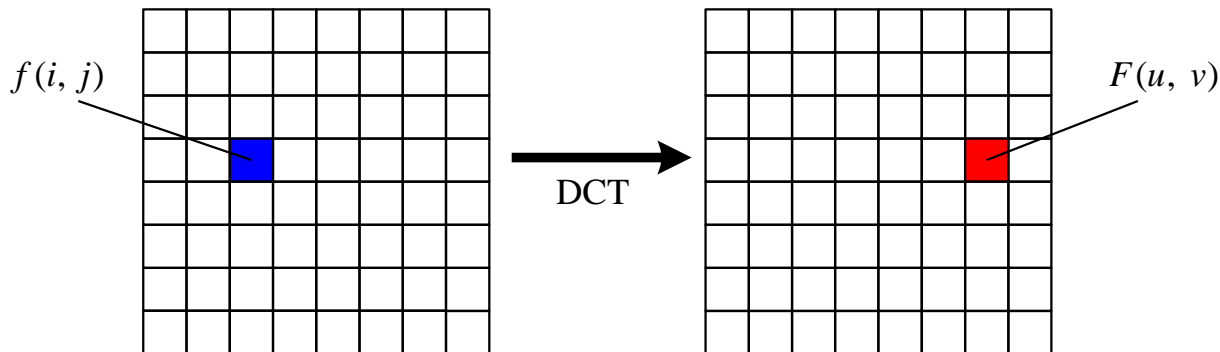
$$f(x, y) = \frac{1}{4} \left[ \sum_{u=0}^7 \sum_{v=0}^7 E(u) E(v) C(u, v) \cdot \cos \frac{(2x+1)u\pi}{16} \cdot \cos \frac{(2y+1)v\pi}{16} \right]$$

$$\begin{cases} E(u) = E(v) = 1/\sqrt{2} & \text{当 } u, v = 0 \\ E(u) = E(v) = 1 & \text{其他} \end{cases}$$

## 3.4.2 静止图像压缩标准

- 2、离散余弦变换DCT

- $f(i, j)$ 经过DCT变换后，输出是64个基信号的幅值，称为DCT系数。
- $C(0,0)$ 的值最大，称为DC（直流系数）；其他称为AC（交流系数）。



## 3.4.2 静止图像压缩标准

### • 3、量化

- 是对经过DCT变换后的频率系数进行量化。
- 量化的目的是减小非“0”系数的幅度以及增加“0”值系数的数目。量化是图像质量下降的最主要原因。
- JPEG算法使用均匀量化器进行量化。

$$C_Q(u, v) = Integer \left( Round \left( \frac{C(u, v)}{Q(u, v)} \right) \right)$$

- $Q(u, v)$ 是量化器步长，是量化表的元素，量化表8\*8大小。
- 因为人眼对亮度信号比对色差信号更敏感，因此使用了两种量化表。

## 3.4.2 静止图像压缩标准

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

色差量化表

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

## 3.4.2 静止图像压缩标准

- 4、DCT系数的编码

- 直流系数

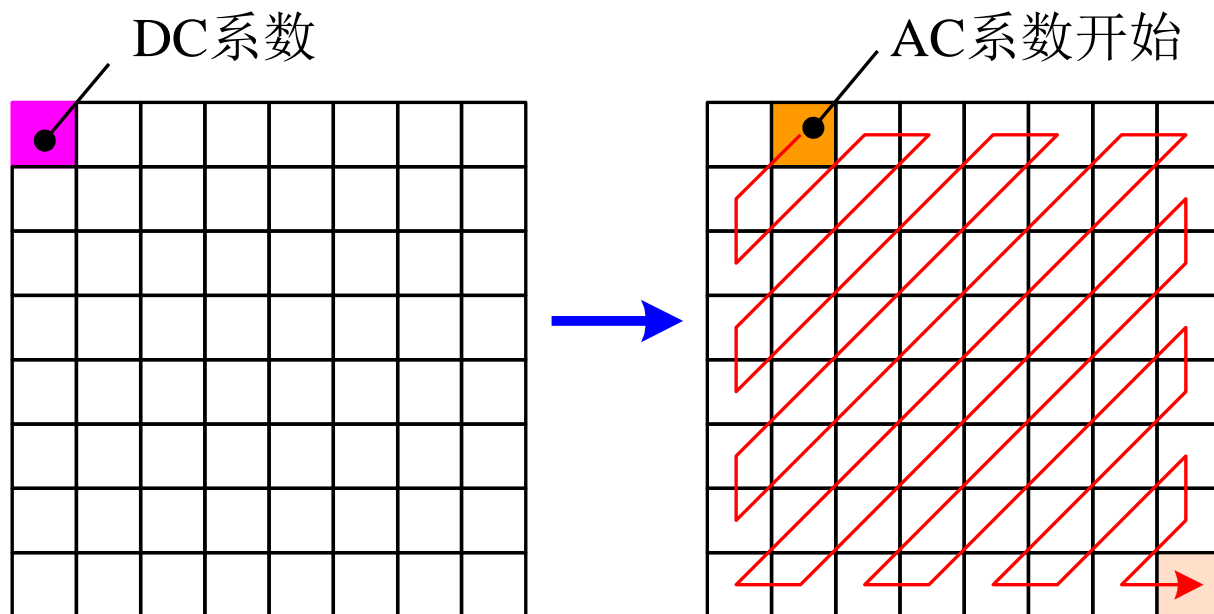
- 相邻的 $8 \times 8$ 块之间的DC系数有强相关性，JPEG对量化后的DC系数采用DPCM编码，即对 $\text{DIFF} = \text{DC}_i - \text{DC}_{i-1}$ 编码。

- 交流系数

- 对于剩下的63个交流系数采用行程编码（RLE）。从左上方 $\text{AC}_{01}$ 开始沿对角线方向“Z”字形扫描直到 $\text{AC}_{77}$ 扫描结束，这样可增加行程中连续0的个数。

## 3.4.2 静止图像压缩标准

- 4、DCT系数的编码



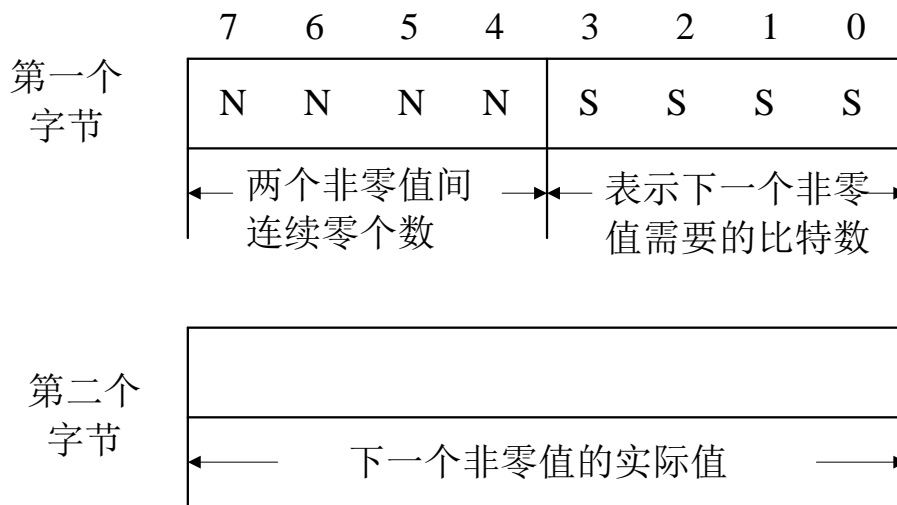


## 3.4.2 静止图像压缩标准

### • 4、DCT系数的编码

#### – 两个字节

- **DC系数**: 第1个字节的高4位是0;  
低4位是差值的幅值编码需要的**比特数**;  
第2个字节表示**DC差值的幅值**。
- **AC系数**:



## 3.4.2 静止图像压缩标准

---

- 5、熵编码

- 为了进一步压缩数据，需要对**DC码**和**AC行程编码**的码字再作基于统计特性的熵编码
  - Huffman编码
  - 自适应二进制算术编码

## 3.4.2 静止图像压缩标准

---

- 5、熵编码

- Huffman变长码表必须作为JPEG编码器的输入
- JPEG解码器能够同时存储最多4套不同的熵编码表
  - 亮度DC系数表
  - 色度DC系数表
  - 亮度AC系数表
  - 色度AC系数表

## 3.4.2 静止图像压缩标准

DC 系数差值幅度范围	分类	亮度编码
0	0	00
-1, 1	1	010
-3, -2, 2, 3	2	011
-7,...,-4,4,...,7	3	100
-15,...,-8,8,...,15	4	101
-31,...,-16,16,...,31	5	110
-63,...,-32,32,...,63	6	1110
	7	11110
-255,...,-128,128,...,255	8	111110
	9	1111110
	10	11111110
-2047,...,-1024;1024...2047	11	111111110

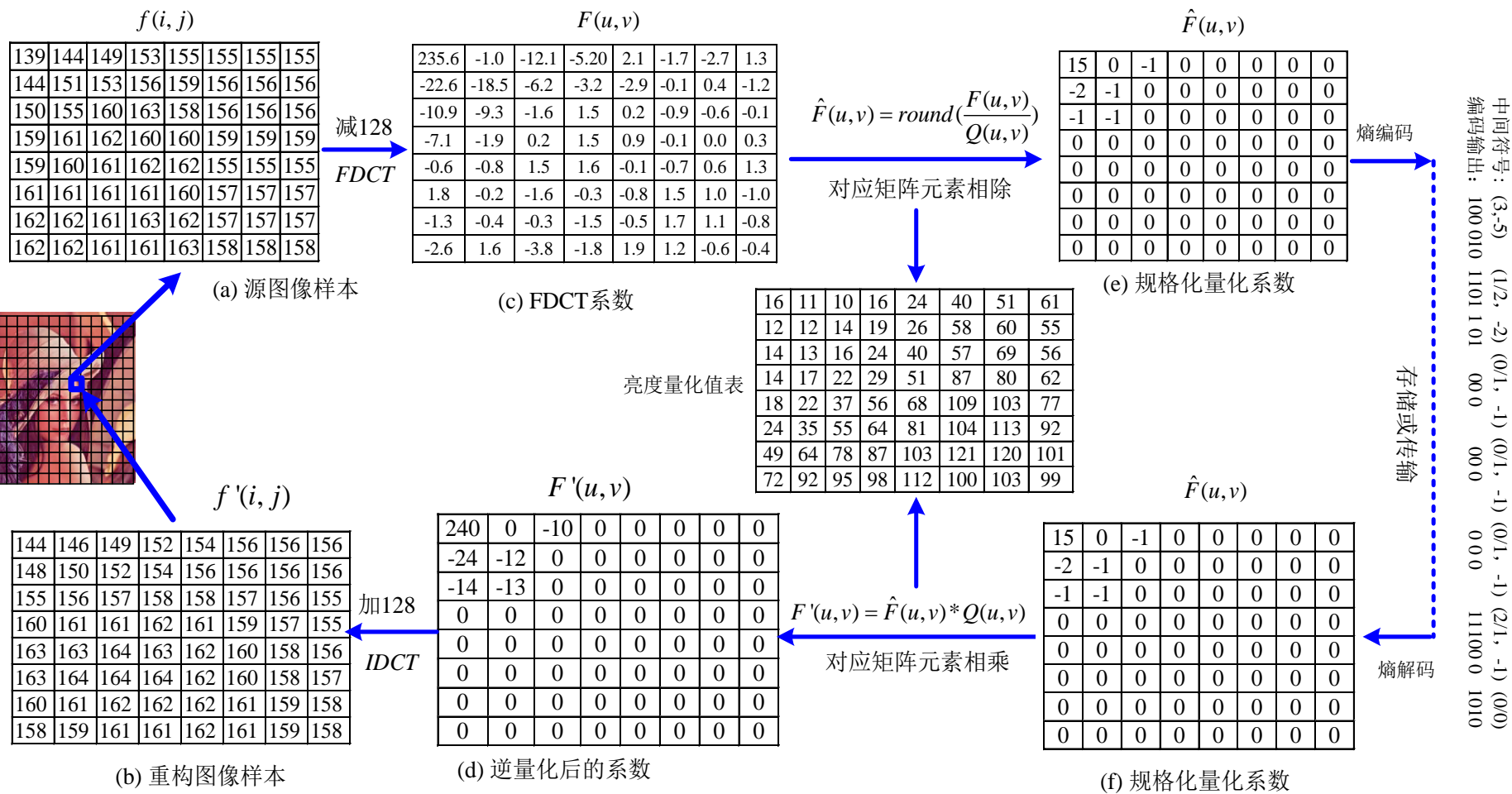
## 3.4.2 静止图像压缩标准

- 5、熵编码

- 亮度分量的前5个图像块的DC系数是150、155、149、152、144

- 使用DPCM编码时产生的输出时150、5、-6、3、-8;
    - 中间符号: (8, 150)、(3, 5)、(3, -6)、(2, 3)、(4, -8);
    - 符号编码: (111110、10010110)、(100、101)、(100、001)、(011、11)、(101、0111)。
      - 5: 3位2进制表示;
      - -6: -6-1的补码的后3位。

# 3.4.2 静止图像压缩标准



## 3.4.2 静止图像压缩标准

---

- **JPEG2000**

- 内容:

- **JPEG2000**图像编码系统（核心部分）
    - 应用扩展（在核心上扩展更多特性）
    - **运动JPEG2000**
    - 兼容性（即包容性与继承性）
    - 参考软件（目前主要为**JAVA**与**C**程序）
    - 复合图像文件格式（如传真式的服务等）

- 核心:

- **离散小波变换**算法

## 3.4.2 静止图像压缩标准

---

- **JPEG2000**

- 特点:

- 高压缩率: 比JPEG高20%-40%左右
    - 无损压缩: 集成预测法无损压缩
    - 渐进传输: 先传输轮廓, 再逐步传输其他数据
    - 感兴趣区域压缩: 指定区域的压缩质量, 或解压要求
    - 色彩模式
    - 图像处理简单: 图像有多个分辨率基准



### 3.4.3 视频压缩标准

---

- **ITU-T**
  - **H.26x ( H.261、 H.262、 H.263、 H.26L )**
- **ISO/IEC**
  - **MPEG-x ( MPEG-1、 MPEG-2、 MPEG-4 )**
  - **Moving Pictures Experts Group**

# 3.4.3 视频压缩标准

- **MPEG-1**

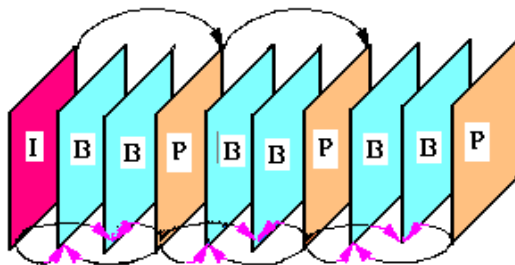
- 目标是以约**1.5Mbps**的速率传输电视质量的视频信号，**亮度**信号的分辨率为 **$360 \times 240$** ，**色度**信号的分辨率为 **$180 \times 120$** ，每秒**30**帧。
- **MPEG-1**标准包括：
  - **MPEG**系统（**ISO/IEC11172-1**）
  - **MPEG**视频（**ISO/IEC11172-2**）
  - **MPEG**音频（**ISO/IEC11172-3**）
  - 测试验证（**ISO/IEC11172-4**）
- 两个基本技术
  - 基于 **$16 \times 16$ 子块**的**运动补偿**，减少帧序列的时域冗余度；
  - 基于**DCT**的**压缩**技术，减少空间冗余度。

# 3.4.3 视频压缩标准

## (1) 时间冗余量的减少

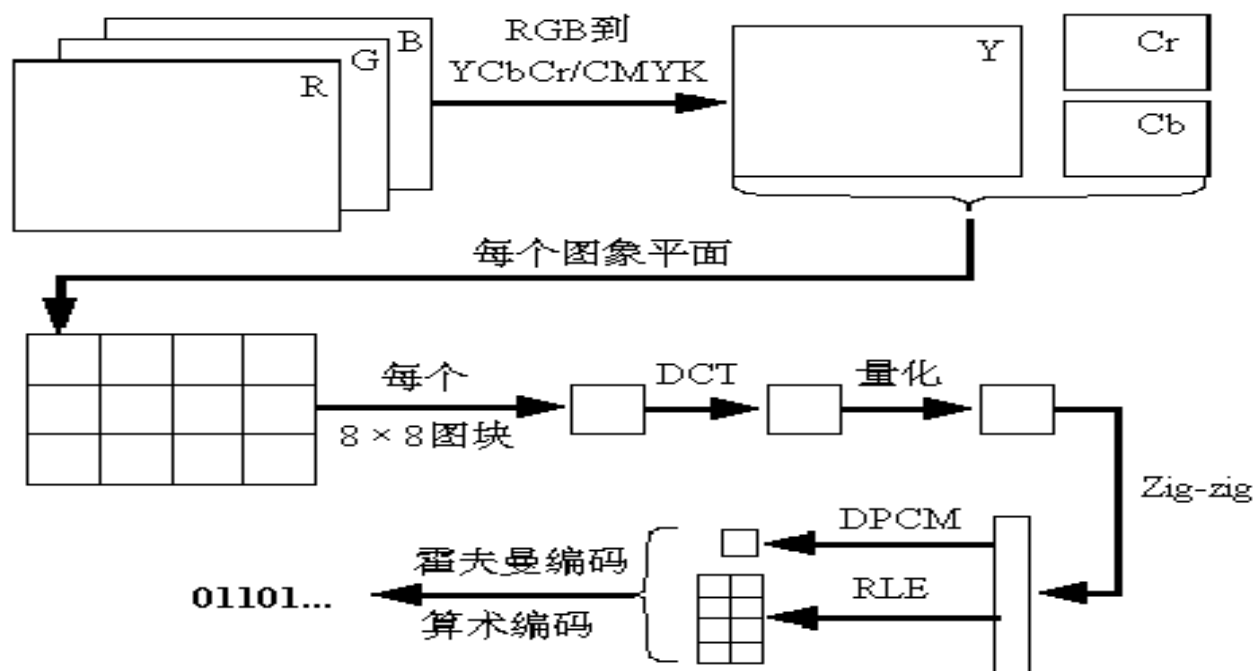
— 为了减少时间冗余量，MPEG将1/30秒时间间隔的帧序列电视图像，以三种类型的图像格式表示

- 内码帧 (I)：是完整的独立编码的图像，是不能由其他帧构造的帧，必须存储或传输。
- 预测帧 (P)：通过对它之前的I帧进行预测，对预测误差作有条件的存储和传输。
- 插补帧 (B)：称为双向帧或插补帧，是根据其前后的I帧或者P帧的信息进行插值编码而获得。该过程有时也称为双向插值。



# 3.4.3 视频压缩标准

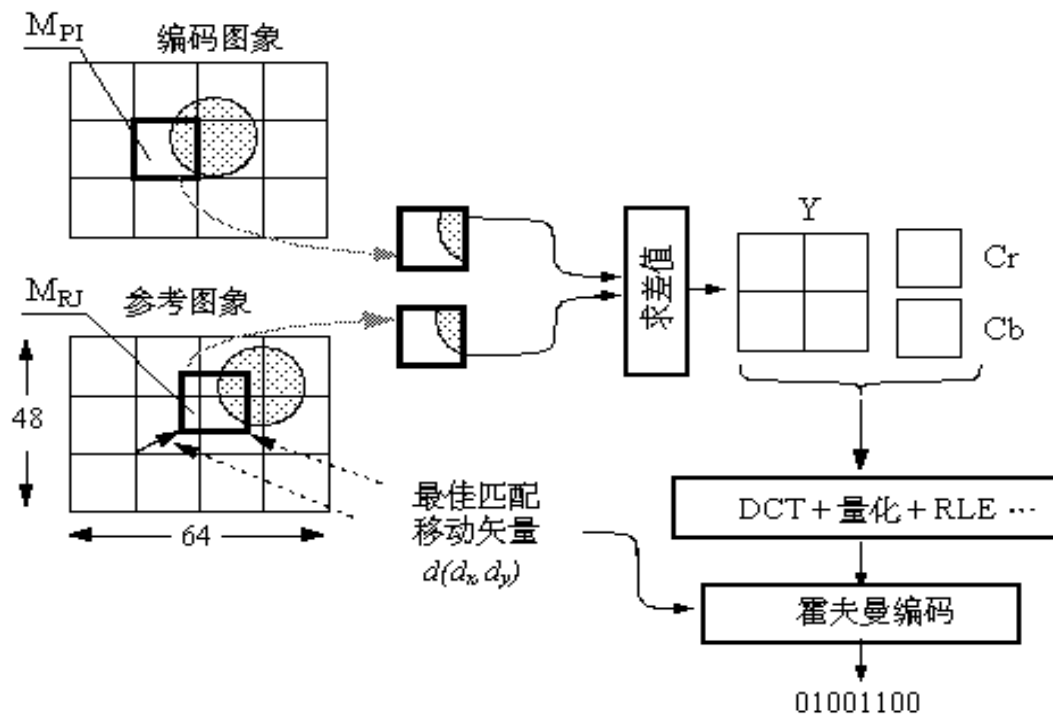
## (1) 时间冗余量的减少



## | 画面压缩编码的流程

# 3.4.3 视频压缩标准

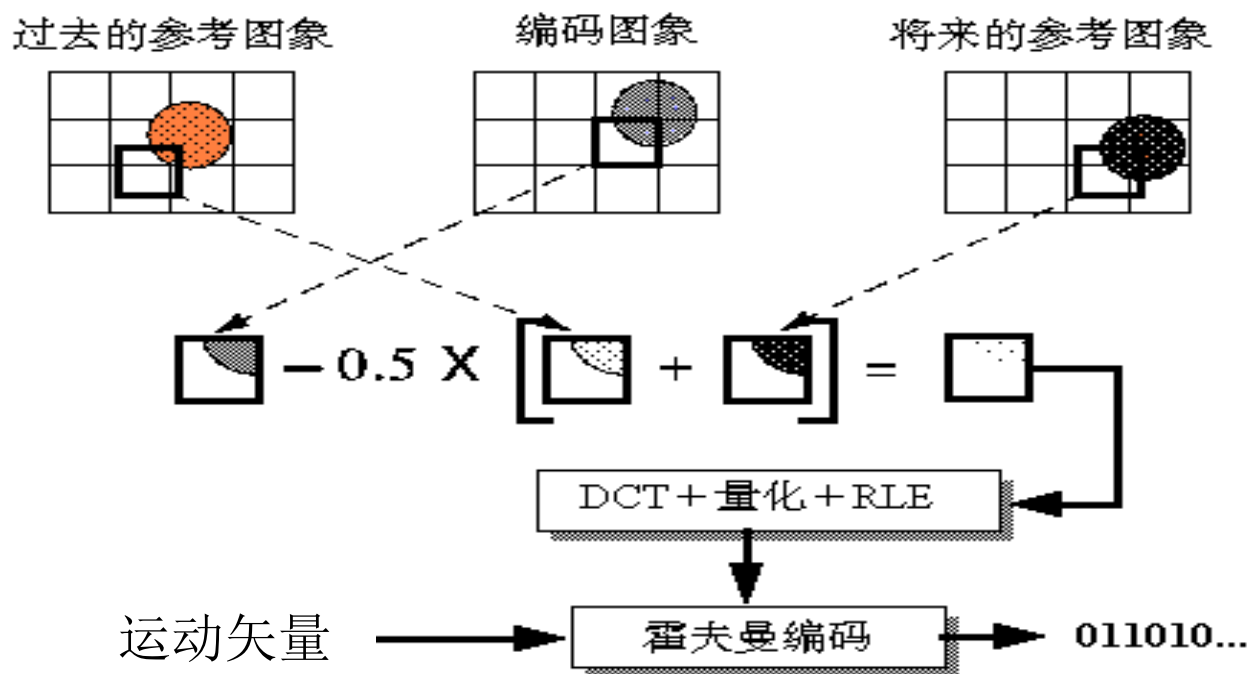
## (1) 时间冗余量的减少



P压缩编码的流程

# 3.4.3 视频压缩标准

## (1) 时间冗余量的减少



B压缩编码的流程

# 3.4.3 视频压缩标准

---

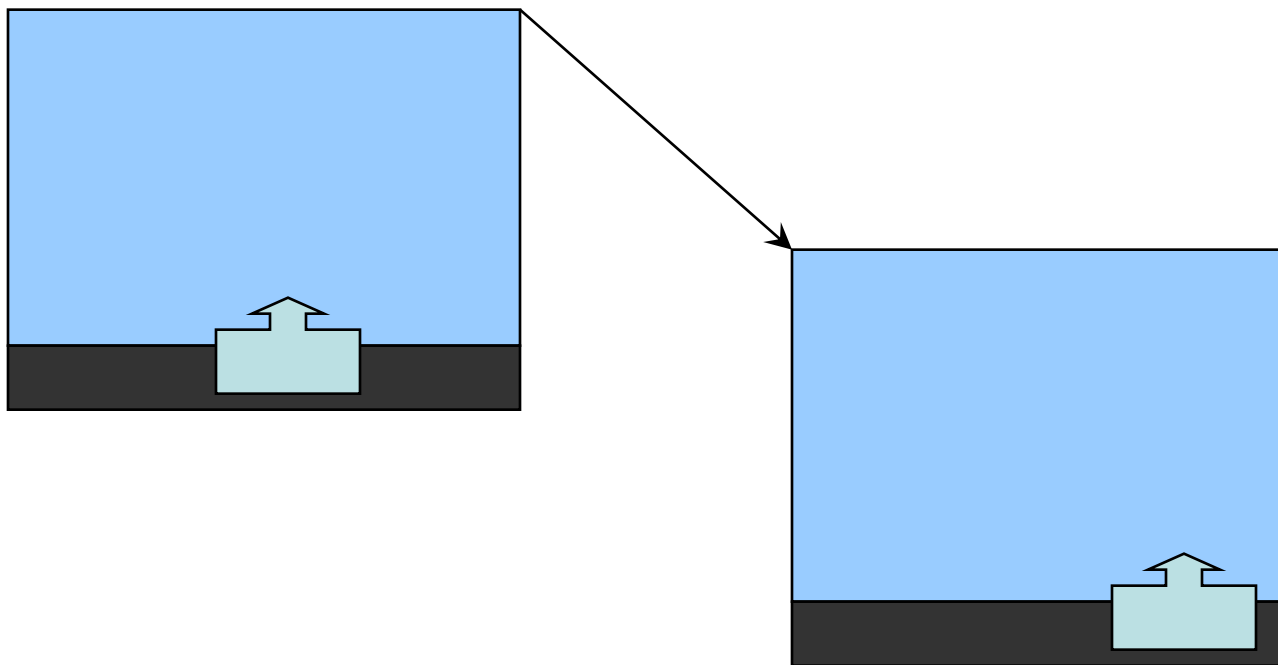
## 运动补偿技术

- 将当前的图像画面看作是前面某时刻图像的位移，位移的幅度和方向在图像画面的各处可有不同。
- 运动补偿技术主要用于消除P图像和B图像在时间上的冗余，以此提高压缩效率。
- 运动补偿在宏块一级进行（16x16）。

## 3.4.3 视频压缩标准

---

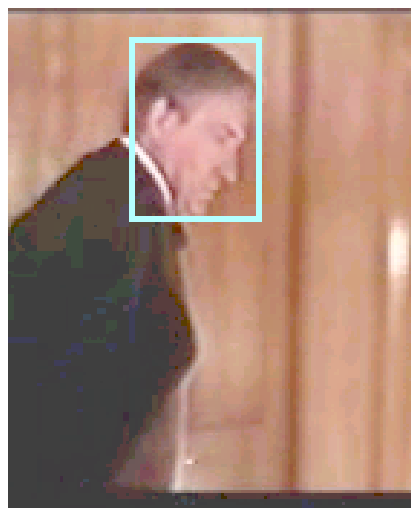
### 运动补偿技术





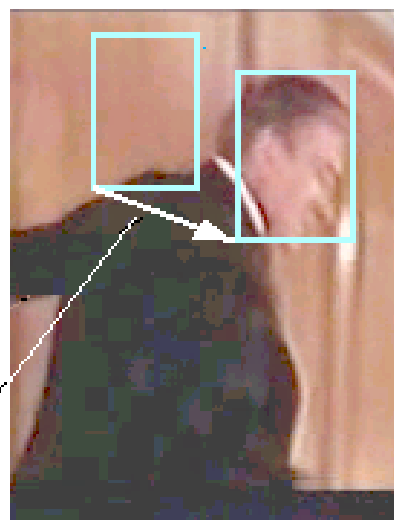
# 3.4.3 视频压缩标准

## 运动补偿技术



时刻1

《泰坦尼克》电影上的镜头



时刻2

移动矢量

## 3.4.3 视频压缩标准

### 运动补偿技术

#### — 块匹配

- 在运动估计时，编码器在参考帧中寻找与**当前宏块匹配**的 $16 \times 16$ 像素区域，使**匹配准则**达到最小的 $16 \times 16$ 区域即为最佳匹配。
- 当前宏块的位置与最佳匹配宏块的**相对位置**为**运动矢量**，当前宏块减去最佳匹配宏块所得到的**残差宏块**将与**运动矢量**一同被编码和传输。

## 3.4.3 视频压缩标准

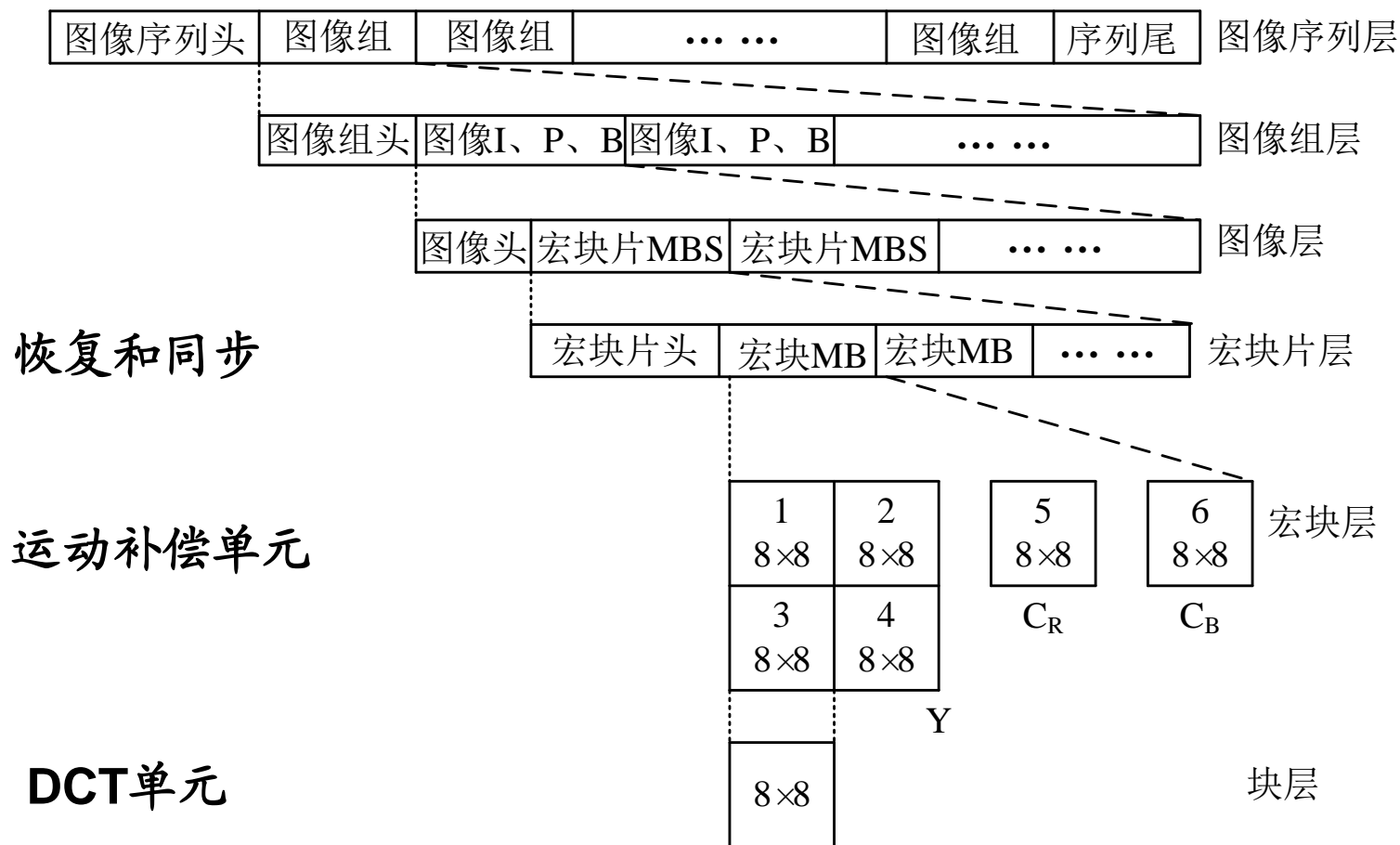
---

### (2) 空间冗余量的减少

- 帧内图和预测图有很高的空间冗余度，采用基于**DCT**的方法。
- 类似于**JPEG**，量化器的设计需作特殊考虑。

# 3.4.3 视频压缩标准

## (3) MPEG的分层结构和位流



# 3.4.3 视频压缩标准

---

## (3) MPEG的分层结构和位流

### — 序列层

- 序列头包括图片的信息，如图像水平大小、垂直大小、像素长宽比、帧率、码率等。

### — 图像组层

- 包含一张或多张图片，其中一张必为I图。GOP头包含如时间码等信息，用来标识从序列开始到现在的时-分-秒-帧。

### — 图像层

- I图、P图和B图

# 3.4.3 视频压缩标准

---

## (3) MPEG的分层结构和位流

### – 宏块片层

- 用来进行**码率控制**以及在位丢失或受损后进行**恢复**和**同步**。
- 包含多个宏块，每个宏块片的长度和位置在头部指定。

### – 宏块层

- 由四个Y块，一个C<sub>b</sub>块，一个C<sub>r</sub>块组成
- 每个块是8×8像素

### – 块层

# 3.4.3 视频压缩标准

---

- **MPEG-2**

- **MPEG系统**

- 将一个或更多的音频、视频或其他的基本数据流合成单个或多个数据流，适应于存储和传送。

- **MPEG视频**

- 位速率范围大约为**2到15Mbps**。
    - 分辨率有低（ $352 \times 288$ ）、中（ $720 \times 480$ ）、次高（ $1440 \times 1080$ ）、高（ $1920 \times 1080$ ）等。
    - 压缩编码方式从简单到复杂不同等级。

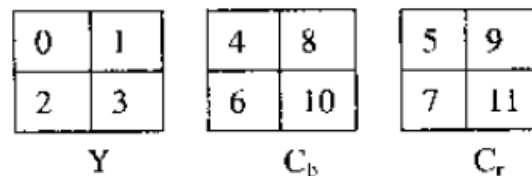
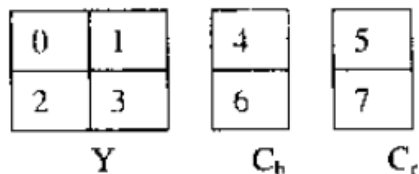
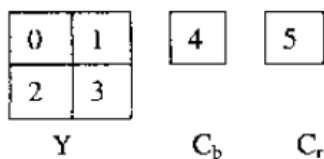
- **音频及其他**

# 3.4.3 视频压缩标准

- **MPEG-2**

- 特色

- 支持隔行扫描视频
    - **MPEG-II** 视频压缩编码的数据结构
      - 宏块结构有三种格式。4:2:0格式，4:2:2格式，4:4:4格式

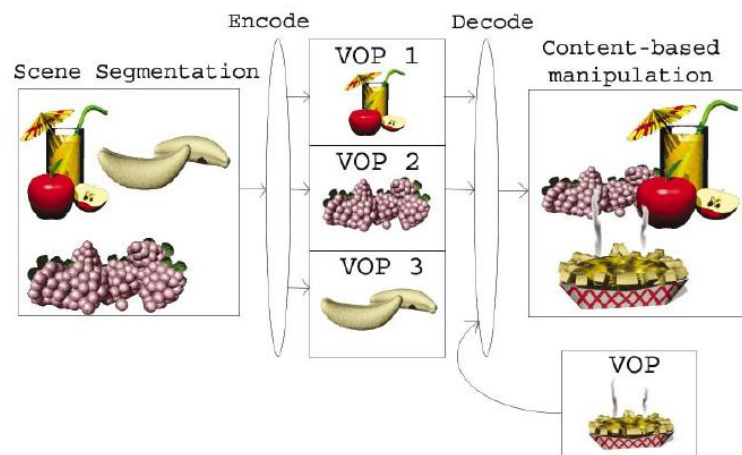




# 3.4.3 视频压缩标准

## • MPEG-4

- 所涉及的应用范畴包括有线、无线、移动通信和因特网等领域。
- MPEG-4采用**基于内容**的压缩编码，把图像**按照内容分块**，如图像的场景、物体被分割成不同的子块。
- 采用**对象**的概念
  - 视频对象Video Object, VO
  - 音频对象Audio Object, AO



# 3.4.3 视频压缩标准

---

- **MPEG-4**

- 基于内容的视频编码过程可由三步完成：

- **VO的形成**。先从原始视频流中分割出VO。
    - **编码**。对各VO分别独立编码，即对不同VO的**运动信息**、**形状信息**和**纹理信息**这三类信息分别编码，分配不同的码字。
    - **复合**。将各个VO的码流**复合**成一个符合**MPEG-4**标准的位流。

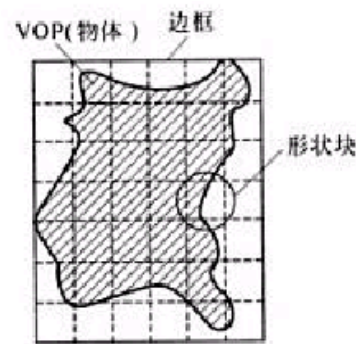
# 3.4.3 视频压缩标准

## • MPEG-4

- 基于VOP的编码：VOP是某一时刻某一帧画面的VO，VOP编码就是对某一时刻该帧画面VO的形状、运动和纹理等三类信息进行编码。

- 形状编码

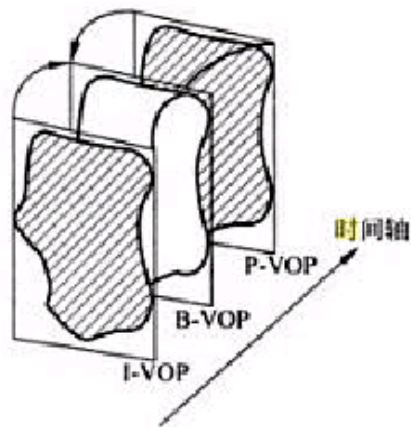
- 位图法。VOP被一个边框框住，边框长、宽均为16的整数倍，同时保证边框最小。
- 位图法实际是一个边框矩阵，对矩阵编码。
- 二值图（0~1）
  - » 0: 非VOP区域（背景）
  - » 1: VOP区域
- 灰度图（0-255）
  - » 0: 非VOP区域
  - » 1~255: VOP区域



# 3.4.3 视频压缩标准

## • MPEG-4

- 基于VOP的编码: VOP是某一时刻某一帧画面的VO, VOP编码就是对某一时刻该帧画面VO的**形状**、**运动**和**纹理**等三类信息进行编码。
  - 运动估计和运动补偿
    - I-VOP, P-VOP, B-VOP
    - 外加了边框, 分成 $16 \times 16$ 的宏块, 宏块由 $8 \times 8$ 的块组成。
    - 运动补偿可基于**宏块或块**。



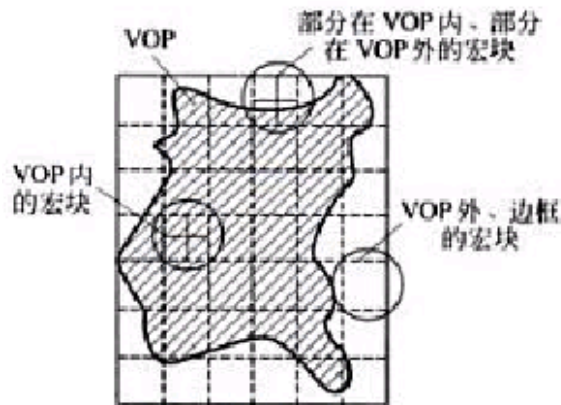
# 3.4.3 视频压缩标准

## • MPEG-4

- 基于VOP的编码：VOP是某一时刻某一帧画面的VO，VOP编码就是对某一时刻该帧画面VO的**形状**、**运动**和**纹理**等三类信息进行编码。

- 纹理编码

- VOP外、边框内的块：不编码；
- VOP内的块：传统的DCT编码；
- 部分在VOP内的块：用“重复填充”方法将该块在VOP外的部分进行**填充**，再用DCT编码；
- 对DCT系数进行量化、Z扫描、游程及霍夫曼编码。



# 3.4.3 视频压缩标准

---

- H.26L

- 是由ITU-T的视频编码专家组(VCEG)及ISO/IEC的活动图像专家组(MPEG)大力发展研究的、适应于低码率传输的新一代压缩视频标准。
- 2003年3月由两个专家组组成的联合视频专家组(JVT)公布了这一压缩视频标准的最终草案。
- 此标准被称为ITU-T的H.264协议或ISO/IEC的MPEG-4的高级视频编码部分。
- 比H.263和MPEG-4节约了50%的码率，而且对网络传输具有更好的支持功能。

## 3.4.3 视频压缩标准

---

- **H.26L**
  - 宏块
    - **CIF:  $352 \times 288$ ,  $22 \times 18 = 396$ 宏块**
    - **QCIF:  $176 \times 144$ ,  $11 \times 9 = 99$ 宏块**
  - **I、P、B帧**

# 3.4.3 视频压缩标准

- H.26L预测编码特点

- 帧内预测编码

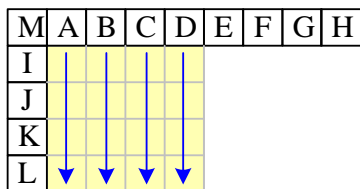
- 亮度块:  $4 \times 4$ , 适用于图像复杂的**细节**部分;  $16 \times 16$ , 适用于图像中变化不大的**平坦**部分。
    - 色度块:  $4 \times 4$ 预测模式。
    - $4 \times 4$ 模式
      - 有13个参考像素 ( $8 + 4 + 1$ ), 支持9种预测方式
    - $16 \times 16$ 模式
      - 垂直预测、水平预测、直流预测和平面预测



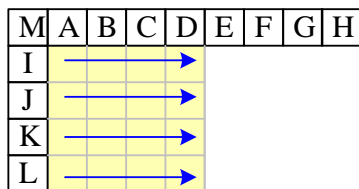
# 3.4.3 视频压缩标准

## • H.26L预测编码特点

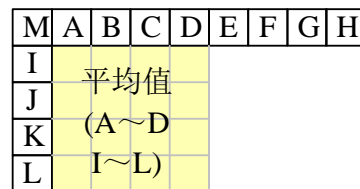
### — 帧内预测编码



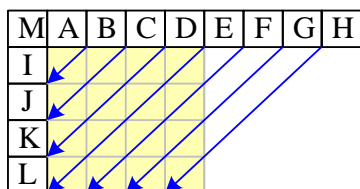
Mode 0 (垂直)



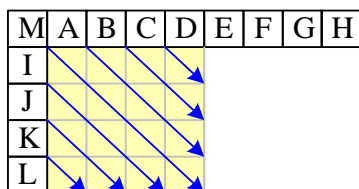
Mode 1 (水平)



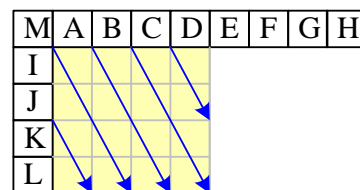
Mode 2 (DC预测)



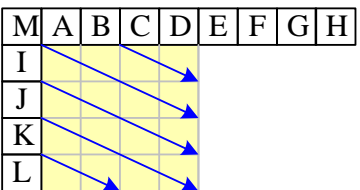
Mode 3 (对角左下45度)



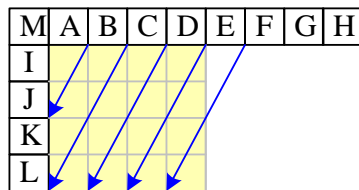
Mode 4 (对角右下45度)



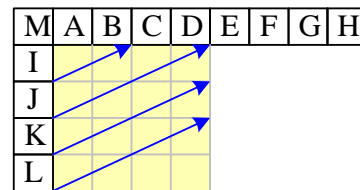
Mode 5 (垂直向上26.6度)



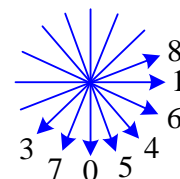
Mode 6 (水平向下26.6度)



Mode 7 (垂直向上26.6度)



Mode 8 (水平向上26.6度)



预测方向

# 3.4.3 视频压缩标准

---

- H.26L预测编码特点

- 帧间预测编码

- 从 $16 \times 16$ 到 $4 \times 4$ 有多种选择
    - 编码
      - 大块: 分割模式占用较小比特, 但运动残差占用比特多
      - 小块: 运动残差比特少, 但表示块分割模式占用较多比特
    - 支持更高精度的运动向量,  $\frac{1}{4}$ 像素精度
    - 支持多参考帧, 最多5个参考帧
    - 去块滤波器

# 3.4.3 视频压缩标准

- H.26L变换编码特点

- 采用**4 × 4的整数变换**作为基本变换编码

- 减小方块效应

- 变换并量化后高频交流系数损失导致细节丢失，4 × 4比8 × 8更小

- 算法简单明了

- 运算结果精度高且不溢出

- 运算速度快

- 占有内存小

**4 × 4变换**

$$A = 13a + 13b + 13c + 13d$$

$$B = 17a + 7b - 7c - 17d$$

$$C = 13a - 13b - 13c + 13d$$

$$D = 7a - 17b + 17c - 7d$$

**4 × 4反变换**

$$a' = 13A + 17B + 13C + 7D$$

$$b' = 13A + 7B - 13C - 17D$$

$$c' = 13A - 7B - 13C + 17D$$

$$d' = 13A - 17B + 13C - 7D$$

## 3.4.3 视频压缩标准

---

- **H.26L熵编码特点**
  - 基于通用变长编码 (**UVLC**)
  - 是基于上下文的自适应二进制算术编码 (**CABAC**)

# 思考题

- **JPEG**编码中，一个亮度子块经过**DCT**、量化等步骤后得到的系数如下图所示，给出对**AC**系数进行**Huffman**编码的中间格式以及编码位流。

提示：

(1) 注意如何得到**AC**系数的中间格式

(2) **JPEG**编码中**Huffman**的亮度**AC**系数表

(3) 注意如何处理连续多个零

12	5	0	0	0	0	0	0
-2	2	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	0	-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