

多媒体技术

多媒体数据压缩技术及标准化



河南大学计算机与信息科学学院

刘扬

2017/11/28 上午11时21分16秒

- 1、数据压缩的基本原理
- 2、常用数据压缩方法
- 3、音频的压缩编码方法
- 4、图像和视频的压缩编码方法
- 5、DCT与JPEG标准
- 6、DWT与JPEG2000标准
- 7、视频压缩标准MPEG概述
- 8、 MPEG音频
- 9、 MPEG视频
- 10、MPEG系统层编码
- 11、AVS与AVC标准

1、数据压缩的基本原理和方法

■ 压缩的必要性

- 多媒体数据可以采用模拟和数字方式两种, 模拟方式信号容易受衰减和噪音干扰, 可控制性差, 复制和传输容易失真。采用数字信号方式主要困难是数据量大, 多媒体数字数据存储量大、快速、宽频带、实时性的环境, 模拟数据转换为数字数据时, 音频、视频的数据量很大, 如果不进行处理, 计算机系统几乎无法对它进行存取和交换。
- 例如, 彩色电视信号PAL制式 (25P/S)、NTSC制式 (30P/S), 中等分辨率 640×480 的24位真彩色数字视频7.37Mbit/帧, PCM采样频率44.1KHz精度为16Bit的2声道音频171Kbit/帧, 如果不压缩1分钟将产生1G数据; 一张 640×480 分辨率24位真彩色图像需900KB. 普通以太网数据传输率为10Mb/s

1分钟数字音频信号需要的存储空间

数字音频格式	频带 (HZ)	带宽 (KHZ)	取样率 (KHZ)	量化 位数	存储容量 (MB)
电话	200~3400	3.2	8	8	0.48
会议电视伴音	50~7000	7	16	14	1.68
CD-DA	20~20000	20	44.1	16	5.292×2
DAT	20~20000	20	48	16	5.76×2
数字音频广播	20~20000	20	48	16	5.76×6

1分钟数字视频信号需要的存储空间

数字电视 格 式	空间分辨率 ×时间	取样率 (MHZ)	量化位数	存储容量 (MB)
公用中间 格式 (CIF)	352×288 $\times 30$	亮度 3; 4:1:1	亮度、色差 共 12	270
CCIR 601 号 建议	PAL 720×480 $\times 30$	亮度 13.5 4:2:2	亮度、色差共 16	1620
	NTSC $720 \times$ 576×25			1620
HDTV 亮度信号	$1280 \times 720 \times$ 60	60	8	3600

■ 视频、图像、声音有很大的压缩潜力

- 信息论认为：若信源编码的熵大于信源的实际熵，该信源中一定存在冗余度。
- 原始信源的数据存在着很多冗余度：空间冗余、时间冗余、视觉冗余、听觉冗余等。

■ 数据压缩的好处

- 时间域压缩的可以迅速传输媒体信源；频率域压缩的可以并行开通更多业务；空间域压缩的可以降低存储费用；能量域压缩的可以降低发射功率。
- 节省图像或视频的存储容量，增加访问速度，使数字视频能在PC机上实现，需要进行视频和图像的压缩。数据压缩的实质就是编码，数据压缩技术实现的衡量标准



- 1. 压缩比：压缩性能常常用压缩比定义（输入数据和输出数据比）
 - 例： 512×480 , 24bit/pixel (bpp)
 - 输出15000byte 输入 = 737280byte
 - 压缩比 = $737280 / 15000 = 49$
- 2. 质量
 - 压缩方法：无损压缩（质量不变）和有损压缩
 - 有损压缩：失真情况很难量化，只能对测试的对象进行估计。

- -客观评价数据损失指标：均方差(MSE)、信噪比(SNR)、峰值信噪比(PSNR)，但必须在主观观察了实际数据信息以后。

$$-MSE = \frac{1}{N} \sum_i (X(i) - N(i))^2$$

$$-PSNR = 10lg \frac{MAX(X)^2}{MSE}$$

■ 3. 压缩解压速度

- 在许多应用中，压缩和解压可能不同时用，在不同的位置不同的系统中。所以，压缩、解压速度分别估计。
- 静态图像中，压缩速度没有解压速度严格；动态图像中，压缩、解压速度都有要求，因为需实时地从摄像机或VCR中抓取动态视频。

■ 4. 硬软件系统

- 有些压缩解压工作可用软件实现。设计系统时必须充分考虑：

算法复杂 — 压缩解压过程长

算法简单 — 压缩效果差

- 目前有些特殊硬件可用于加速压缩/解压。硬接线系统速度快，但各种选择在初始设计时已确定，一般不能更改。因此在设计硬接线压缩/解压系统时必须先将算法标准化。

■ 1. 数据冗余的类型

- 空间冗余：在同一幅图像中，规则物体和规则背景的表面物理特性具有相关性，这些相关性的光成像结果在数字化图像中就表现为数据冗余。如一幅图像中同一种颜色不止一个像素点，若相邻的像素点的值相同，像素点间（水平、垂直）冗余；当图像的一部分包含占主要地位的垂直的源对象时，相邻线间存在冗余（空间冗余）
- 时间冗余：时间冗余反映在图像序列中就是相邻帧图像之间有较大的相关性，一帧图像中的某物体或场景可以由其它帧图像中的物体或场景重构出来。音频的前后样值之间也同样有时间冗余。如若图像稳定或只有轻微的改变，运动序列帧间存在冗余
- 视觉冗余：人眼对于图像场的注意是非均匀的，人眼并不能察觉图像场的所有变化。事实上人类视觉的一般分辨能力为 2^6 灰度等级，而一般图像的量化采用的是 2^8 灰度等级，即存在着视觉冗余。
- 听觉冗余：人耳对不同频率的声音的敏感性是不同的，并不能察觉所有频率的变化，对某些频率不必特别关注，因此存在听觉冗余

- 信息熵冗余：信源编码时，当分配给第 i 个码元类的比特数 $b(y_i) = -\lg(p_i)$ ，才能使编码后单位数据量等于其信源熵，即达到其压缩极限。但实际中各码元类的先验概率很难预知，比特分配不能达到最佳。实际单位数据量 $d > H(S)$ ，即存在信息冗余熵。

$$H(X) = - \sum_x P(x) \log_2 [P(x)]$$



1948年香农提出 “信息熵” 概念，解决了对信息的量化度量问题

- 其它冗余：包括结构冗余、知识冗余等。

■ 老三论(SCI论)：

- 系统论 (Systems theory)：美籍奥地利生物学家贝塔朗菲 (Bertalanffy). 把对象当作一个整体或系统来研究其整体与局部、动态和等级关系，并用数学模型去描述和确定系统的结构和行为。
- 控制论 (Cybernetics)：美国数学家维纳 (Wiener). 研究系统的状态、功能、行为方式及变动趋势，控制系统稳定性的规律。
- 信息论 (Information theory)：美国数学家香农 (Shannon). 用概率论和数理统计方法，从量的方面来研究系统的信息如何获取、加工、处理、传输和控制。

■ 新三论(DSC论)：

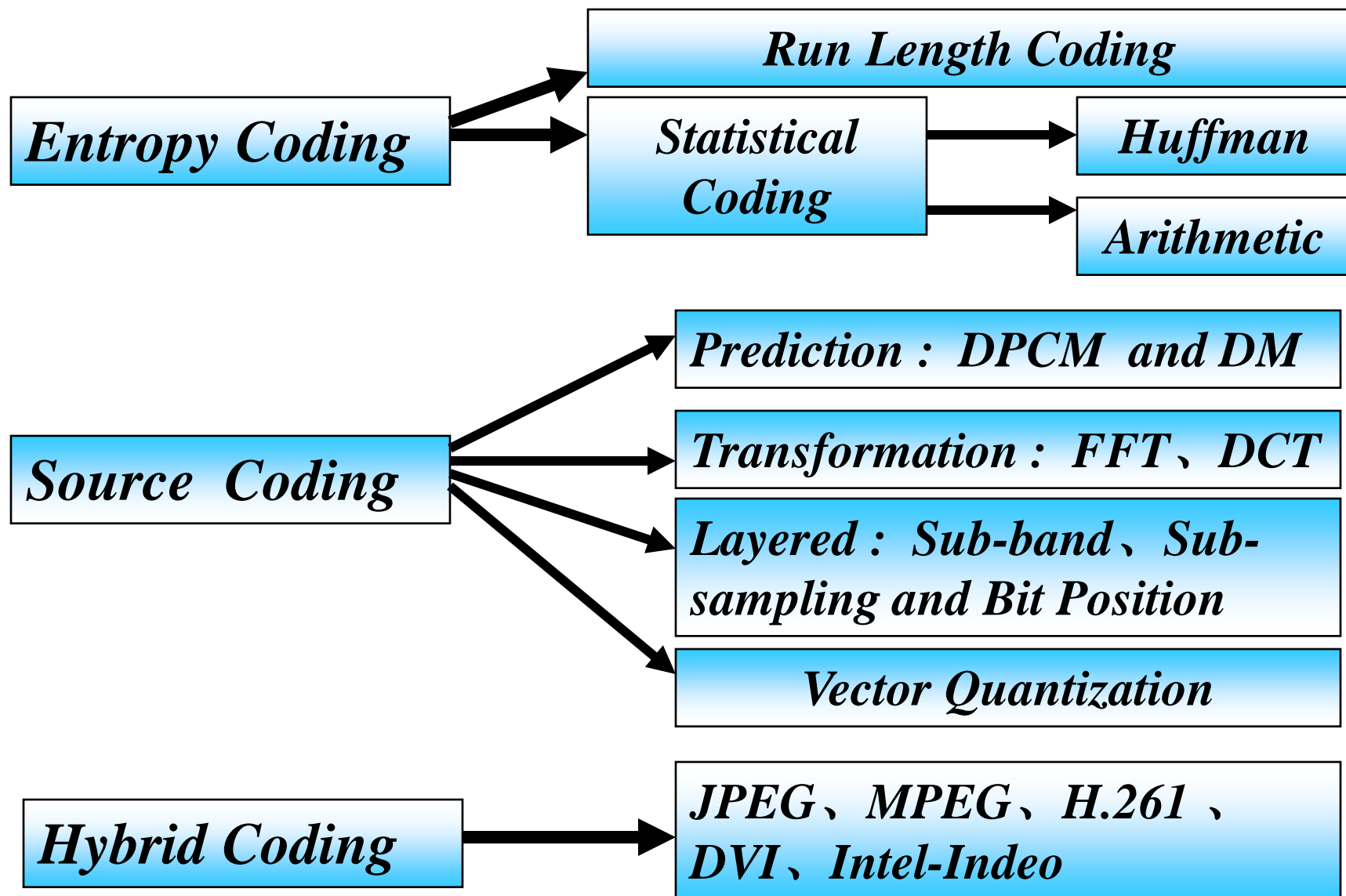
- 耗散结构论：(Dissipative structure theory)：在远离平衡的条件下，才有可能向着有秩序、有组织、多功能的方向进化
- 协同论：(Synergy theory)：大系统结构的小系统相互作用、制约、平衡
- 突变论：(Catastrophe theory)：研究系统突变式质变的奇点

■ 2. 数据压缩技术分类

- 根据解码后数据与原始数据是否完全一致可以分为两大类：一类是熵编码、冗余压缩法，也称无损压缩法、无失真压缩法；二是熵压缩法，也称有损压缩法、有失真压缩法。
- 从“熵”损失角度分为无损压缩和有损压缩两种：
 - 无失真压缩，又称熵编码、无损压缩，它是一种可逆处理，信源的熵不受损失的编码。压缩的仅仅是冗余度信息，解压缩时可以完全再现。由于不会失真，多用于文本、数据的压缩，但也有例外，非线性编辑系统为了保证视频质量，有些高档系统采用的是无失真压缩方法
 - 有失真压缩，又称熵压缩法、有损压缩。是一种不可逆处理，压缩编码过程中损失部分熵，因此解压缩时无法完全再现（熵是从统计特性观点得出的信源平均信息的量度，借用热力学概念称为熵）大多数图像、声音、动态视频等数据的压缩是采用有失真压缩。
- 按照压缩技术分预测编码、变换编码、统计编码和混合编码等。预测编码包括帧内预测（利用帧内像素之间的相关性）和帧间预测（利用帧间顺序的时空相关性）。变换编码是将原始图像分割成多个子像块，对每个子像块进行某种形式正交变换，生成变换域（频域）中的系数阵列，经过滤波、量化、编码、传输，在解压缩时进行逆变换。正交变换常见有快速傅立叶变换（FFT）、离散余弦变换（DCT）、小波变换（WT）等。

- 按照编码方式分为分组编码(Block Codes)和非分组编码(Nonblock Codes):分组编码将连续的非重叠块处理成编码块数据组, 分组长度可以定长码和非定长码方式进行. 非分组编码实质是一个具有记忆和延迟的非线性滤波器, 编码输出与上下文相关, 被处理成连续相互重叠的数据组
- 从信息语义角度分为“熵(平均信息量)编码”和“源编码”两种:
 - 熵(平均信息量)编码(Entropy Coding) 是一种泛指那些不考虑被压缩信息的性质的编码和压缩技术。它是基于平均信息量的技术把所有的数据当作比特序列, 而不根据压缩信息的类型优化压缩。也就是说, 平均信息量编码忽略被压缩信息的语义内容。
 - 熵编码分为: 重复序列消除编码(含: 消零、行程编码)、统计编码等。
 - 源编码(Source Coding) 的冗余压缩取决于初始信号的类型、前后的相关性、信号的语义内容等。源编码比严格的平均信息量编码的压缩率更高。当然压缩的程度主要取决于数据的语义内容, 比起平均信息量编码, 它的压缩比更大。
 - 源编码主要分为: 预测编码、变换编码、向量量化等。

数据冗余的类型与压缩方法分类



- 1. 统计编码：方法是识别一个给定的流中出现频率最高的比特或字节模式，并用比原始比特更少的比特数来对其编码。即就是，频率越低的模式，其编码的位数越多，频率越高的模式编码位数越少。若码流中所有模式出现的概率相等，则平均信息量最大，信源就没有冗余。
- (1) 行程编码(Run Length Coding)
 - 是最简单、最古老的压缩技术之一，主要技术是检测重复的比特或字符序列，并用它们的出现次数取而代之。该方法有两大模式：一是消零(消空白)，二是行(游)程(run length)编码。
 - 消零(或消空白)法：将数字中连续的“0”或文本中连续的空白用一个标识符(或特殊字符)后跟数字N(连续“0”的个数)来代替。

如数字序列：742300000000000000000055 编码为：7423Z1855

- 行程编码法：任何重复的字符序列可被一个短格式取代。该算法适合于任何重复的字符。一组 n 个连续的字符 c 将被 c 和一个特殊的字符取代。当然，若给定字符仅重复两次就不要用此方法。任何重复4次或4次以上的字符由“该字符 + 记号(M) + 重复次数”代替。

例如序列：Name: CR 编码为：Name: . M10 CR

■ (2) LZW编码

- LZW算法在压缩文本和程序数据的压缩技术中唱主角。它的压缩率高，压缩处理所化费的时间比其它方式要少。
 - LZW编码时，首先将原始的数据分成多个条纹，每个条纹都单独进行压缩。
 - LZW算法基于一个转换表或字串表，它将输入字符映象到编码中，使用可变长代码，最大代码长度为12位。
- LZW算法中的字串表对于每个条纹都不同，并且不必保留给解压缩程序，因为解压缩过程中能自动建立完全相同的字串表。实际上，它是通过查找冗余字符串并将此字符串用较短的符号标记替代的压缩技术。

■ (3) 哈夫曼编码

- 1952年Huffman提出了对统计独立信源能达到最小平均码长的编码方法，也即最佳码。最佳性可从理论上证明。这种码具有即时性和唯一可译性。该编码是常见的一种统计编码。对给定的数据流，计算其每个字节的出现频率。根据频率表，运用哈夫曼算法可确定分配各字符的最小位数，然后给出一个最优的编码。代码字存入代码表中。
- 基本原理是对出现频率较高的符号分配短码字；对出现频率较低的符号分配较长的码字，使整体平均编码缩短。编码方法是构造一个哈夫曼树。由于各个字符在不同场合出现的频率不同，可以将哈夫曼编码分为动态和静态两种方式。霍夫曼码没有错误保护功能的可变长度码，编码方案不唯一，自含同步码。
- Huffman编码的优缺点
 - 优点：当信源符号概率是2的负幂次方时，Huffman 编码法编码效率达到100%。一般情况下，它的编码效率要比其它编码方法的效率高，是最佳变长码。
 - 缺点：Huffman 码依赖于信源的统计特性，必须先统计得到信源的概率特性才能编码，这就限制了实际的应用。通常可在经验基础上预先提供Huffman码表，此时性能有所下降。

■ (4) 算术编码

- 算术编码把一个信源集合表示为实数线上的0到1之间的一个区间。这个集合中的每个元素都要用来缩短这个区间。信源集合的元素越多，所得到的区间就越小，当区间变小时，就需要一些更多的数位来表示这个区间，这就是区间作为代码的原理。算术编码首先假设一个信源的概率模型，然后用这些概率来缩小表示信源集的区间。

新子区间的起始位置

= 前子区间的起始位置 + 当前符号的区间左端 × 前子区间长度

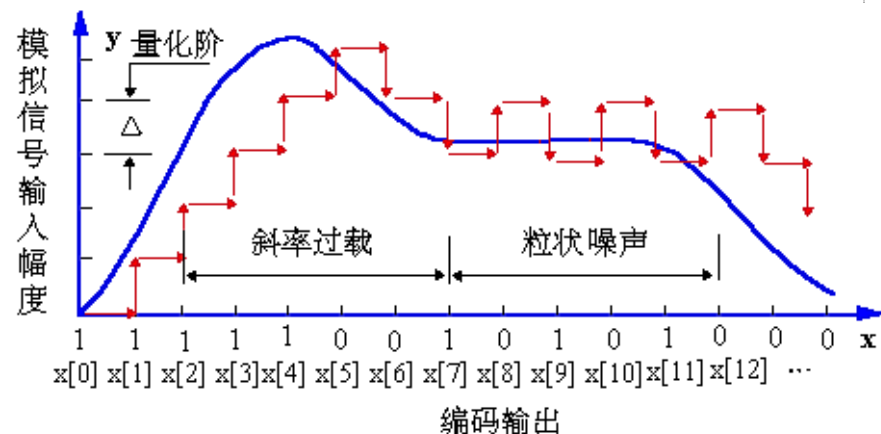
新子区间的长度

= 前子区间的长度 × 当前符号的概率（等价于范围长度）

- 最后得到的子区间的长度决定了表示该区域内的某一个数所需的位数。

■ 2. 预测编码(Prediction Coding)

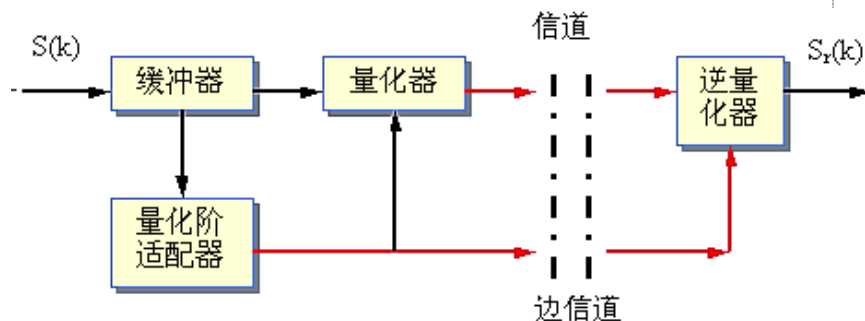
- 预测编码是指利用前面的一个或多个信号对下一个信号进行预测，然后对实际值和预测值的差进行编码。DPCM与ADPCM是两种典型的预测编码。
- (1) PCM(Pulse Code Modulation), 原始的模拟信号经过时间采样, 然后对每一样值进行量化, 作为数字信号传输。PCM编码的一种变形是增量调制, 也称 Δ 调制(delta modulation, DM), 它是一种预测编码技术, PCM是对每个采样信号的整个幅度进行量化编码, 因此它具有对任意波形进行编码的能力; DM是对实际的采样信号与预测的采样信号之差的极性进行编码, 将极性变成“0”和“1”这两种可能的取值之一。如果实际的采样信号与预测的采样信号之差的极性为“正”, 则用“1”表示; 相反则用“0”表示, 或者相反。由于DM编码只须用1位对话音信号进行编码, 所以DM编码系统又称为“1位系统”。



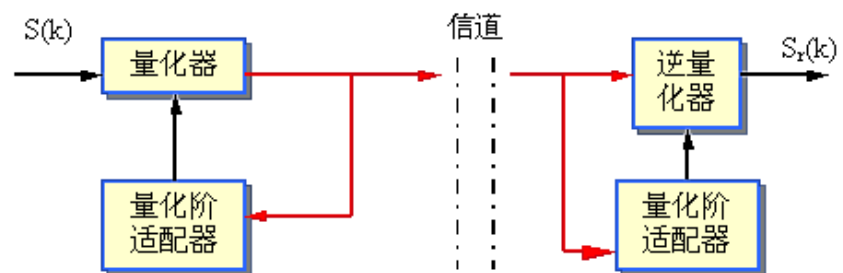
(2) 自适应脉冲编码调制 (APCM)

- 为了使增量调制器的量化阶 Δ 能自适应，也就是根据输入信号斜率的变化自动调整量化阶 Δ 的大小，以使斜率过载和粒状噪声都减小到最小，在检测到斜率过载时开始增大量化阶 Δ ，而在输入信号的斜率减小时降低量化阶 Δ 。这就是自适应增量调制器 (ADM)

- APCM (adaptive pulse code modulation) 是根据输入信号幅度大小来改变量化阶大小的一种波形编码技术。这种自适应可以是瞬时自适应，即量化阶的大小每隔几个样本就改变，也可以是音节自适应，即量化阶的大小在较长时间周期里发生变化。



(a) 前向自适应

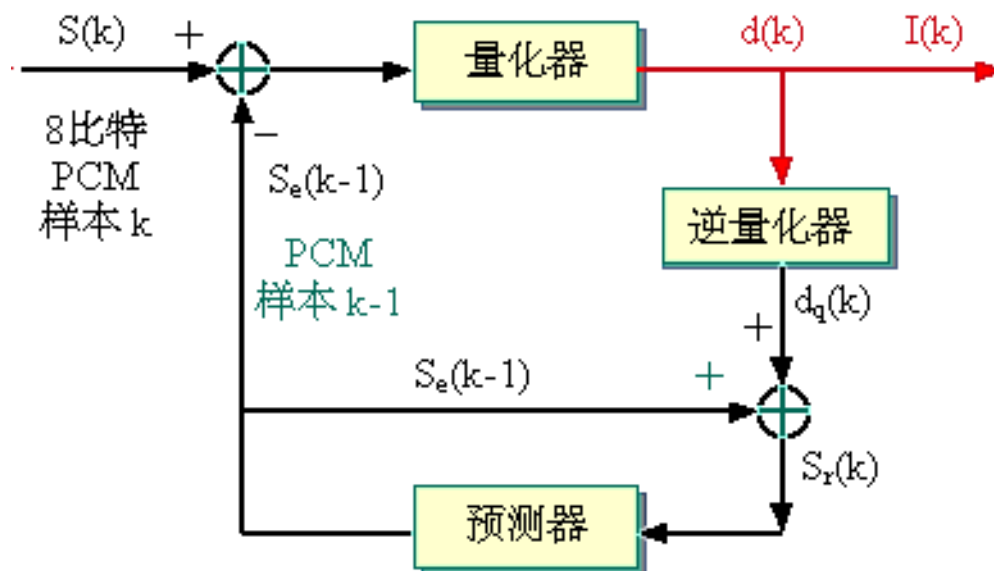


(b) 后向自适应

(3) 差分脉冲编码调制 (DPCM)

- 差分脉冲编码调制
DPCM(differential pulse code modulation)是利用样本与样本之间存在的信息冗余度来进行编码的一种数据压缩技术。差分脉冲编码调制的思想是，DPCM不对每一样值都进行量化，而是根据过去的样本去估算(estimate)下一个样本信号的幅度大小，这个值称为预测值，然后对实际信号值与预测值之差进行量化编码，从而就减少了表示每个样本信号的位数。DPCM是基本的编码方法之一，在大量的压缩算法中被采用，比如JPEG的DC分量就是采用DPCM编码的。

- 例如：设DPCM系统预测器的预测值为前一个样值，假设输入信号已经量化，差值不再进行量化。差值的范围比输入样值的范围有所减小，可以用较少的位数进行编码。



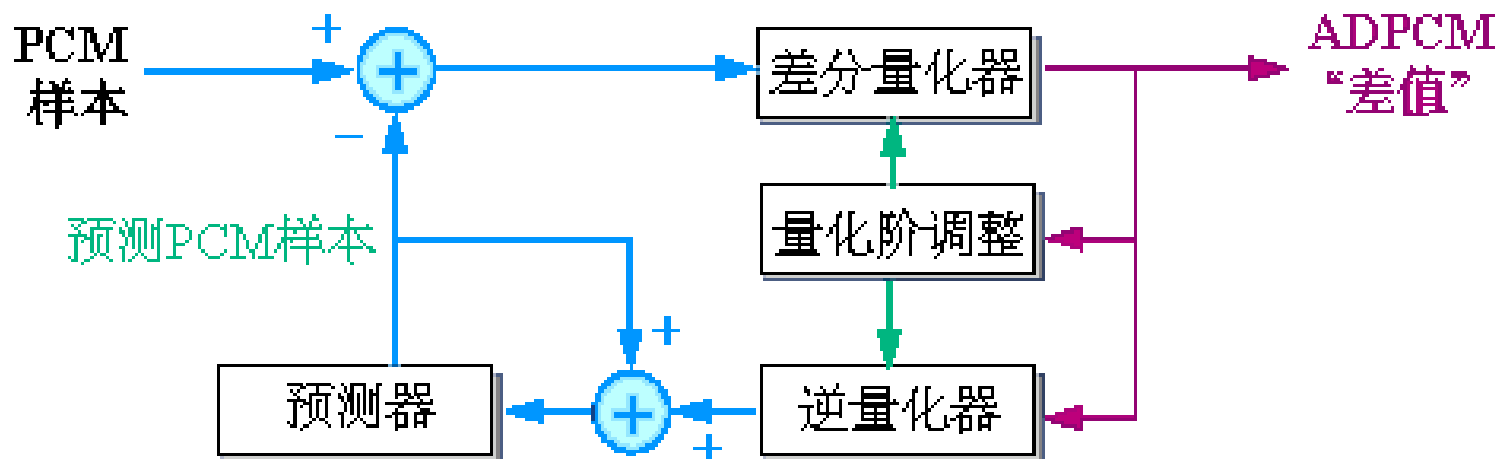
输入值:{0 1 2 1 1 2 3 3 4 4 ...}

预测值:{0 0 1 2 1 1 2 3 3 4 ...}

差分值:{0 1 1 -1 0 1 1 0 1 0 ...}

(4) 自适应差分脉冲编码调制 (ADPCM)

- ADPCM (adaptive difference pulse code modulation) 综合了APCM的自适应特性和DPCM系统的差分特性，是一种性能比较好的波形编码。它的核心想法是：
 - 利用自适应的思想改变量化阶的大小，即使用小的量化阶 (step-size) 去编码小的差值，使用大的量化阶去编码大的差值
 - 使用过去的样本值估算下一个输入样本的预测值，使实际样本值和预测值之间的差值总是最小。



■ 自适应差分脉码调制 (ADPCM)

- 具体方法是：预测参数仍采用固定的；但此时有多组预测参数可供选择。这些预测参数根据常见的信源特征求得。编码时具体采用哪组预测参数根据信源的特征来自适应的确定。
- 为了自适应的选择最佳参数，通常将信源数据分区间编码，编码时自动地选择一组预测参数，使该区间实际值与预测值的均方误差最小。随着编码区间的不同，预测参数自适应的变化，以达到准最佳预测。

- 例如，Microsoft 的 ADPCM采用二预测参数，提供7组预测系数，如右表所示。编码时，根据选定的准则(如最小均方误差准则)，每个编码区间自动地选取一组最佳的参数。

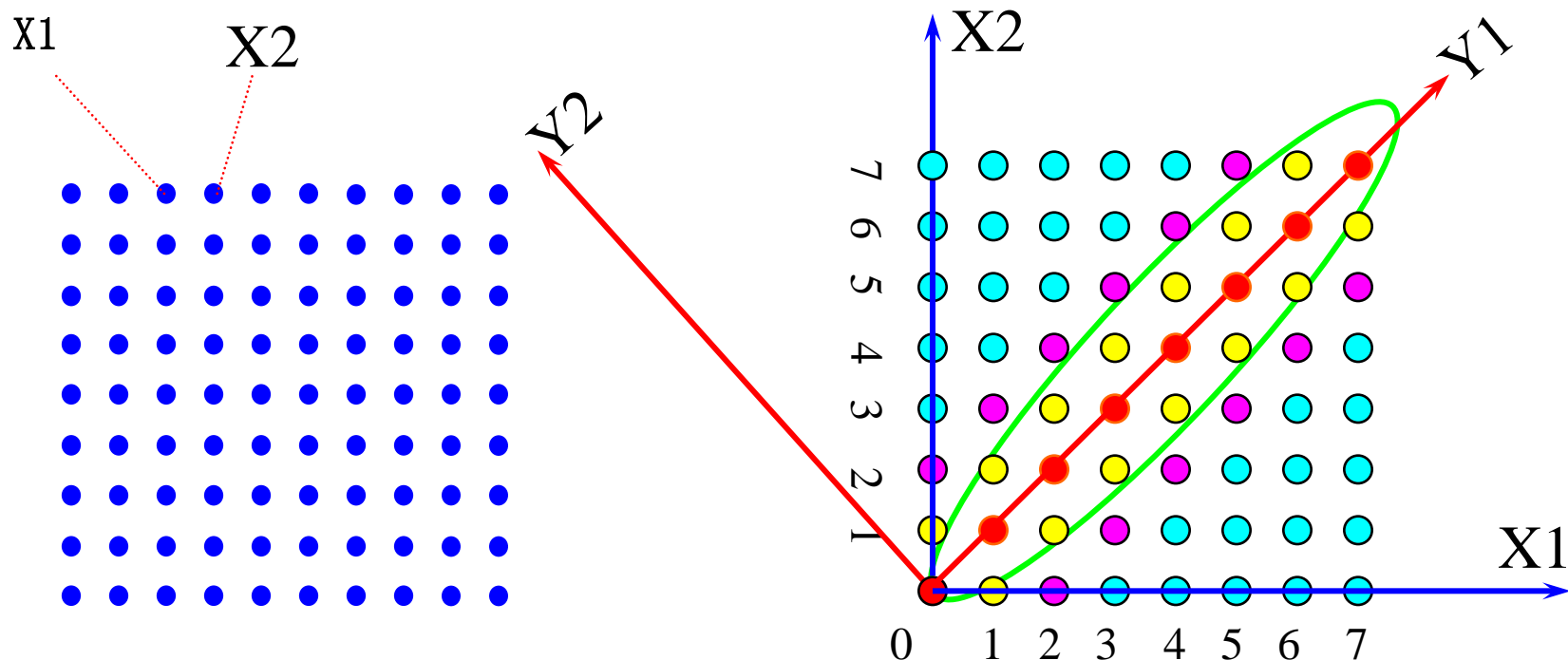
- 自适应量化：根据信号分布不均匀的特点，系统具有随输入信号的变化而改变量化区间大小，以保持输入给量化器的信号基本均匀的能力，这种能力称为自适应量化。

系数集	系数1	系数2
0	256	0
1	512	-256
2	0	0
3	192	64
4	240	0
5	460	-208
6	392	-232

■ 3. 变换编码(Transformation Coding)

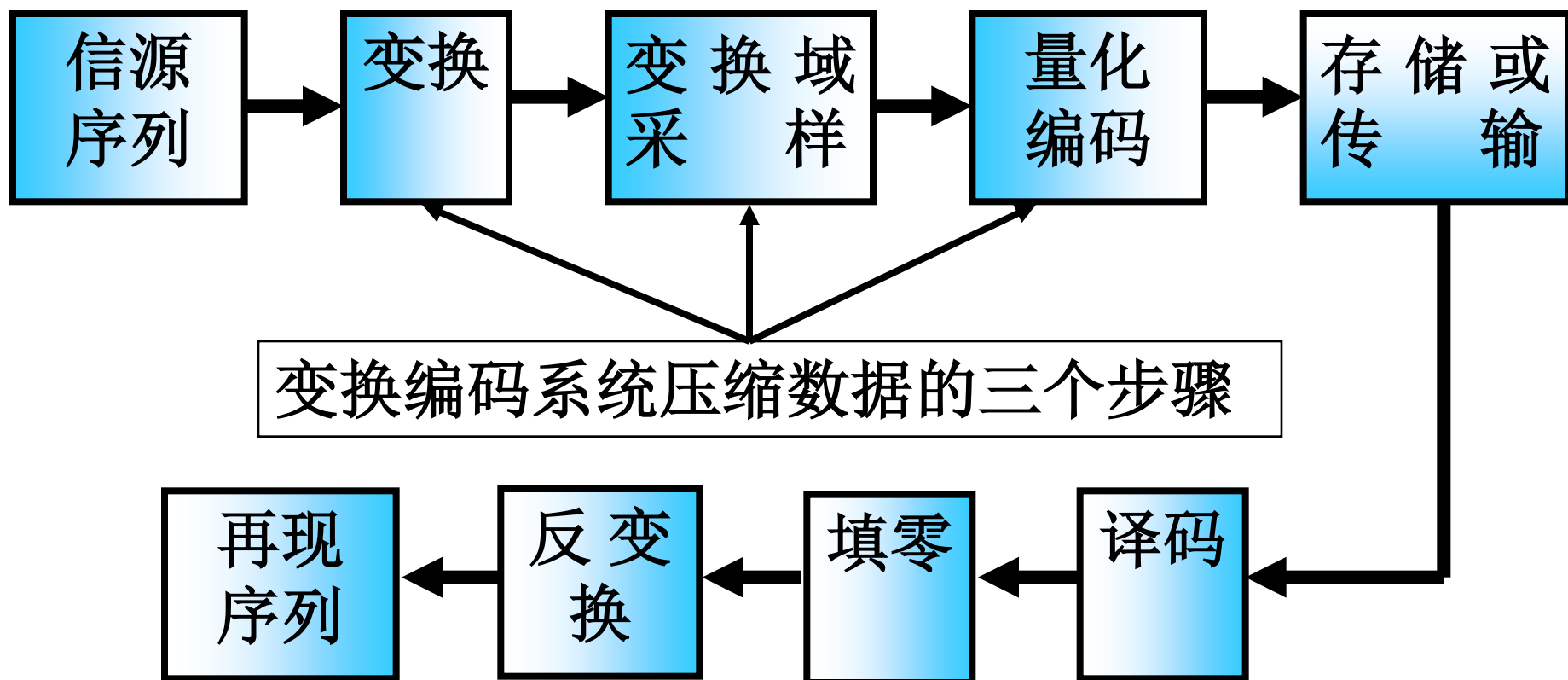
- 在变换编码时，初始数据要从初始空间或时间域进行数学变换，变换为一个更适于压缩的抽象域。该过程是可逆的；即使用反变换可恢复原始数据。如将时域信号变换到频域，因为声音、图像大部分信号都是低频信号，在频域中信号的能量较集中，再进行采样、编码就可以压缩数据。
- 变换本身是可逆的，因而其也是一种无损技术。然而，为了取得更满意的结果，某些重要系数的编码位数比其他的要多，某些系数干脆就被忽略了。这样，该过程就成为有损的了。

- 举例 设灰度图像中的每个像素的亮度值为 $0 \sim 7$ 共八种。相邻两像素的亮度取值共64种不同情况，大部集中在绿线区域内。在新坐标系中， Y_2 与 Y_1 的相关性大大减少。



- 设:原始媒体数据为 X , 变换矩阵为 T , 均为方阵。则: $Y = T \times X$, Y 就是变换后的图像。若变换矩阵 T 满足: $T' \times T = I$, $T^{-1} \times T = I$, 则变换 T 称为正交变换。由 Y 还原出原始媒体 X 的方法是: 因为 $Y = T \times X$, 所以 $X = T' \times Y$
- 各种正交变换
 - 数学家们已经构造了多种数学变换。除了傅里叶变换外, 还有余弦、Hadamard、Haar、Karhunen Loeve变换。最实用最常用的数学变换是离散余弦变换(DCT)。
 - Karhunen-Loeve 变换 (K-L变换)
 - DFT (离散富氏变换)
 - Walsh-Hadamard 变换
 - DCT (离散余弦变换)
 - 仅次于K-L变换的次最佳变换
 - 只有实数运算
 - 有快速算法

■ 典型的变换编码系统框图：



■ (1) 最佳变换 (K-L变换)

- 数据压缩主要是去除信源的相关性。若考虑到信号存在于无限区间上，而变换区域又是有限的，那么表征相关性的统计特性就是协方差矩阵。
- 当协方差矩阵中除对角线上元素之外的各元素都为零时，就等效于相关性为零。所以，为了有效地进行数据压缩，常常希望变换后的协方差矩阵为一对角矩阵，同时也希望主对角线上各元素随 i, j 的增加很快衰减。因此，变换编码的关键在于：在已知 \mathbf{X} 的条件下，根据它的协方差矩阵去寻找一种正交变换 \mathbf{T} ，使变换后的协方差矩阵满足或接近为一对角矩阵。
- 当经过正交变换后的协方差矩阵为一对角矩阵，且具有最小均方误差时，该变换称最佳变换，也称Karhunen-Loeve变换。可以证明，以矢量信号的协方差矩阵的归一化正交特征向量所构成的正交矩阵，对该矢量信号所作的正交变换能使变换后的协方差矩阵达到对角矩阵。

■ (2) 离散余弦变换 (DCT变换)

- 如果变换后的协方差矩阵接近对角矩阵，该类变换称准最佳变换，典型的有DCT、DFT、WHT、HrT等。其中，最常用的变换是离散余弦变换DCT。
- DCT是从DFT引出的。DFT可以得到近似于最佳变换的性能，但DFT的运算次数太多，且需要复数运算。DCT从DFT中取实部，并可用快速余弦变换算法，因此大大加快了运算。同时其压缩性能十分逼近最佳变换的压缩性能。所以，DCT在图像压缩中得到了广泛的应用。

■ 4. 分析合成 编码

- 通过对原始数据的分析，将其分解为一系列更适合于表示的基元或者从中提取出更有本质意义的参数，编码仅对这些基本单元或者特征参数进行，而解码时则借助于一定的规则或者模型，按照一定的算法将这些基元或者参数再合成原始数据的一个逼近。

■ 矢量量化

- 量化编码按照一次量化的码元个数，可分为标量量化和矢量量化两种。对数字化后的数据或PCM数据（样本值）一个一个地进行量化，称为标量量化。而将这些数据分组，每组K维矢量，再以矢量为单元逐个进行量化，称其为矢量量化。矢量量化是标量量化的多维扩展。
- 标量量化中可在随机变量X出现概率比较高的间隔内，选择较小的判决间隔，而在其他区域内选择较大的间隔，这样可以以较小的量化均方误差进行量化。
- 矢量量化基于语义编码，其基本思想是采用非线性量化器，即对空间频率及能量分布较大的系数分配较多比特数；反之分配较少的比特数，从而达到压缩的目的。

■ 小波变换编码

- 小波变换是一个线性变换，能够将一个信号分解成对空间和时间、频率的独立贡献，同时又不失原信号所包含的信息。经过小波变换后的图像能量很集中，便于对不同的分量作不同的处理，达到较高的压缩比。

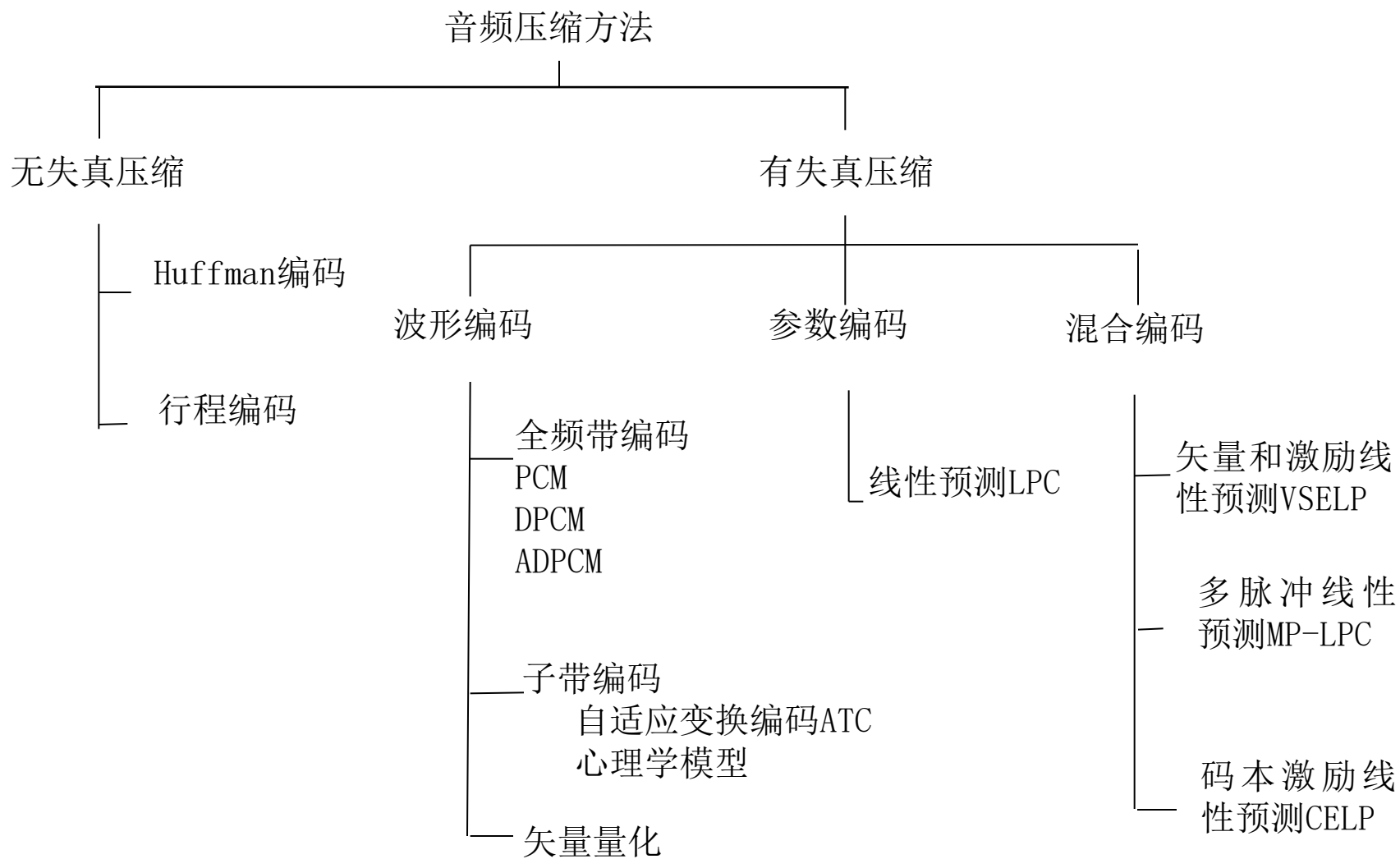
■ 分形编码

- 分形编码是一种模型编码，它利用模型的方法，对需要传输的图像进行参数估测。分形的方法是把一幅数字图像，通过一些图像处理技术，如颜色分割、边缘检测、频谱分析、纹理变化分析等等，将原始图像分成一些子图像。子图像可以是简单的物体，也可以是一些复杂的景物。然后在分形集中查找这样的子图像。分形集实际上并不是存储所有可能的子图像，而是存储许多迭代函数，通过迭代函数的反复迭代，恢复出原来的子图像。表示这样的迭代函数一般只需几个数据即可，这就达到了很高的压缩比。

■ 子带编码：

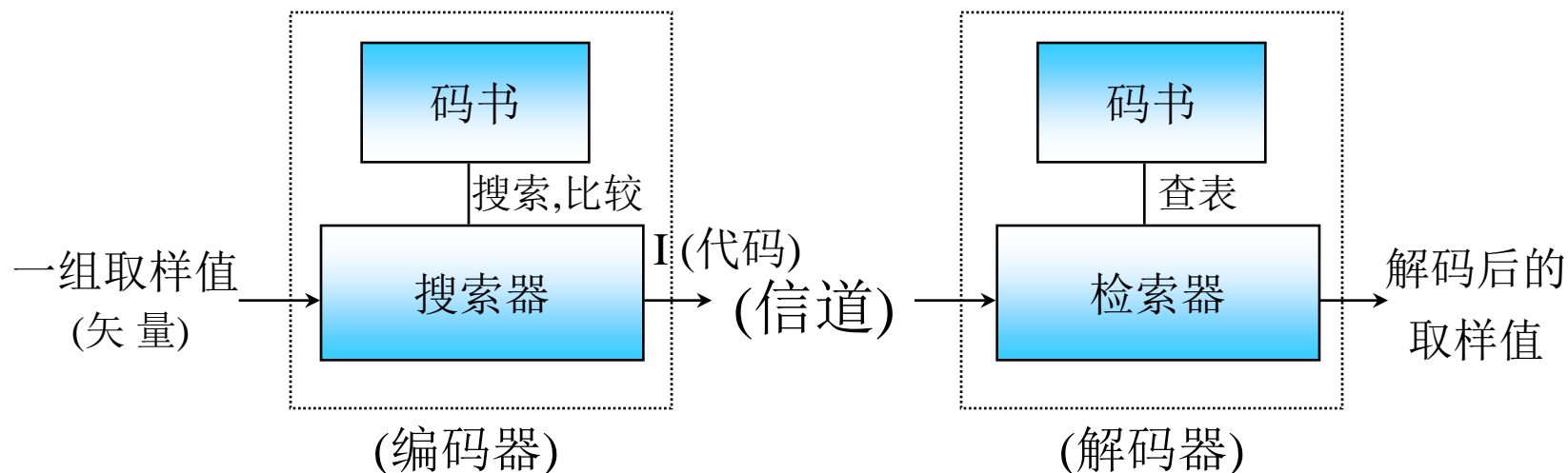
- 利用带通滤波器组把信号频带分割成若干子频带，然后分别处理。通过等效于单边带调幅的调制过程，将各子带搬移到零频率附近以得到低通表示后，再以奈奎斯特速率对各子带输出取样，并对取样值进行通常的数字编码。恢复时，将各子带信号解码并重新调制回其原始位置，再将所有子带输出相加就可得到接近于原始信号的恢复波形。它的复杂度与变换编码差不多，但客观质量高、主观效果好。

3、音频的压缩编码



■ 原理:

- 把一组取样值与预先设计好的若干模板作比较, 选一个最接近的模板, 以它的编号作为其代码.



■ 性能参数:

- 矢量维数 k
- 码书空间大小 N
(每个样本的平均信息量 $r = \log_2 N / k$)
- 搜索时间 T
- 量化失真

■ 特点:

- 压缩效率高
- 解码简单, 编码计算比较复杂
- 有失真
- 关键在于码书的设计 (码书的生成, 码书的组织)

■ 码书的生成

- 初始码书的生成
- 码书的优化（训练法）
 - LBG算法1（概率分布已知）
 - LBG算法2（概率分布未知）
 - SGR算法
- 码书的适应性修改

■ 码书如何组织，使搜索时间得到优化。

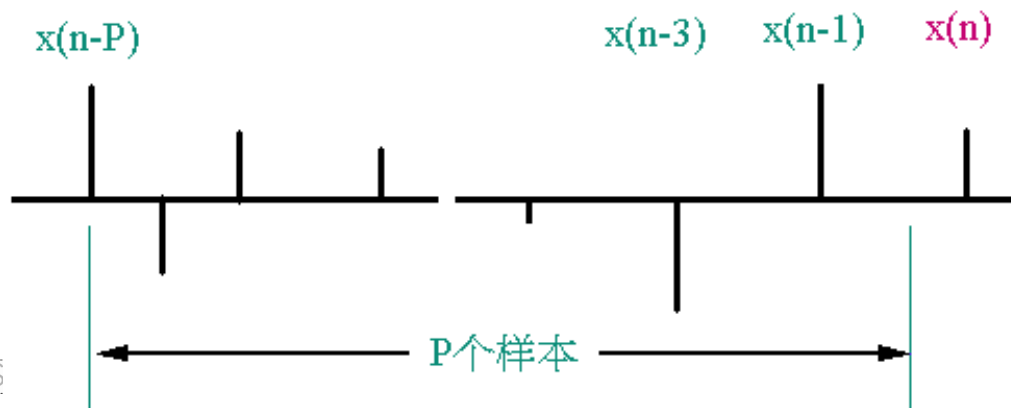
- 线性预测编码(linear predictive coding, LPC)从原理上讲, LPC是通过分析话音波形来产生声道激励和转移函数的参数, 对声音波形的编码实际就转化为对这些参数的编码, 这就使声音的数据量大大减少。在接收端使用LPC分析得到的参数, 通过话音合成器重构话音。合成器实际上是一个离散的随时间变化的时变线性滤波器, 它代表人的话音生成系统模型。时变线性滤波器既当作预测器使用, 又当作合成器使用。分析话音波形时, 主要是当作预测器使用, 合成话音时当作话音生成模型使用。随着话音波形的变化, 周期性地使模型的参数和激励条件适合新的要求。
- 线性预测器是使用过去的P个样本值来预测现时刻的采样值。预测值可以用过去P个样本值的线性组合来表示:

$$x_{pre}(n) = -[a_1x(n-1) + a_2x(n-2) + \dots + a_px(n-p)] = -\sum_{i=1}^p a_i x(n-i)$$

- 残差误差(residual error)即线性预测误差

$$e(n) = x(n) - a_{pre}x(n) = -\sum_{i=0}^p a_i x(n-i)$$

- 在给定的时间范围里，使的平方和即为最小，这样可使预测得到的样本值更精确。通过求解偏微分方程，可找到系数的值。如果把发音器官等效成滤波器，这些系数值就可以理解成滤波器的系数。这些参数不再是声音波形本身的价值，而是发音器官的激励参数。在接收端重构的话音也不再具体复现真实话音的波形，而是合成的声音。



- CCITT和ISO先后提出了一系列有关音频的编码建议。分别应用于窄带语音信号、宽带语音信号、图像伴音信号、数字移动通信GSM信号等等。

- GSM编译码器

—除了ADPCM算法已经得到普遍应用之外，还有一种使用较普遍的波形声音压缩算法叫做GSM算法。GSM是Global System for Mobile communications的缩写，可译成全球数字移动通信系统。GSM算法是1992年柏林技术大学(Technical University Of Berlin)根据GSM协议开发的，这个协议是欧洲最流行的数字蜂窝电话通信协议。

—GSM的输入是帧(frame)数据，一帧(20毫秒)由采样频率为8 kHz的带符号的160个样本组成，每个样本为13位或者16位的线性PCM(linear PCM)码。GSM编码器可把一帧(160×16位)的数据压缩成260位的GSM帧，压缩后的数据率为1625字节，相当于13 kb/s。由于260位不是8位的整数倍，因此编码器输出的GSM帧为264位的线性PCM码。采样频率为8 kHz、每个样本为16位的未压缩的话音数据率为128 kb/s，使用GSM压缩后的数据率为：

— $(264 \text{ 位} \times 8000 \text{ 样本/秒}) / 160 \text{ 样本} = 13.2 \text{ 千位/秒}$

—GSM的压缩比： $128 : 13.2 = 9.7$ ，近似于10:1。

■ ITU—TS建议的语音压缩的标准

- 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。
- G. 721: ITU建议的 G. 721将64Kbps的比特流转换为32Kbps的流, 它是基于 ADPCM技术。每个数值差分用4位编码, 其采样率为8kHz。
- G. 723: G. 723是一种以24Kbps运行的基于 ADPCM 的有损耗压缩标准。其音质不如非压缩的 G. 711PCM 标准以及基于 SB—ADPCM 的 G. 722 标准。G. 723. 1和G. 723. 2用于H. 324标准。
- G. 728: 它的 比特率为16Kbps, 带宽限于3. 4kHz。其音质比 G. 711或 G. 722差得多。它基于一种称为低延迟代码激励线性预测(LD—CELP)的向量量化技术。
- CELP是一种常用的语音压缩技术。它用于美国联邦标准1016, 可将语音压缩至4. 8Kbps。美国联邦标准1015使用 CELP的一个简本, 称为线性预测编码(LPC)。 LPC—10E标准可以运行于2. 4Kbps。采用了一种向量量化方法。声音听起来有点象机器在说话, 但4. 8Kbps与电话差不多。

■ 调幅广播质量的音频压缩标准

- 调幅广播质量：50Hz – 7kHz，称“7kHz音频信号”。
- G. 722：G. 722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。G. 722压缩信号的带宽范围为50Hz到7kHz，而 G. 711 仅限于3.4kHz。其比特率为48、56、64Kbps，在标准模式下，采样速率是16KHz，幅度深度为14比特。

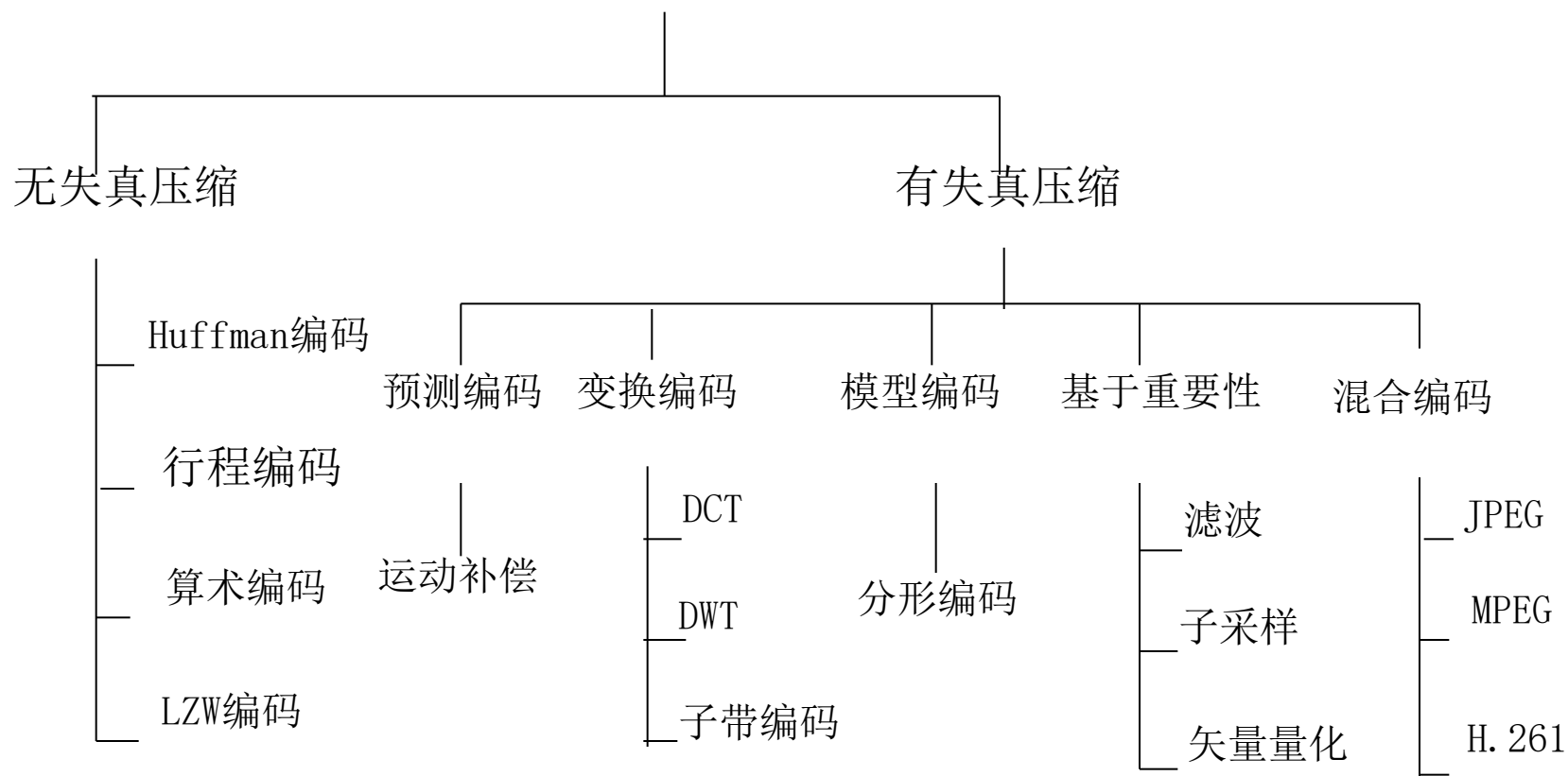
■ 高保真立体声音频压缩标准

- 高质量的声音信号频率范围：50Hz – 20kHz
- 目前国际上比较成熟的高质量声音压缩标准为MPEG音频。MPEG – 1的音频信号在ISO 11172 – 3文档中的描述。MPEG音频不是单个一种压缩算法，而是3种音频编码和压缩方案的一个系列。MPEG 声音编码分为：层 – 1、层 – 2、层 – 3。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。最复杂的译码器（即在层 – 3工作的译码器）也可对层 – 2或层 – 1的码流进行译码

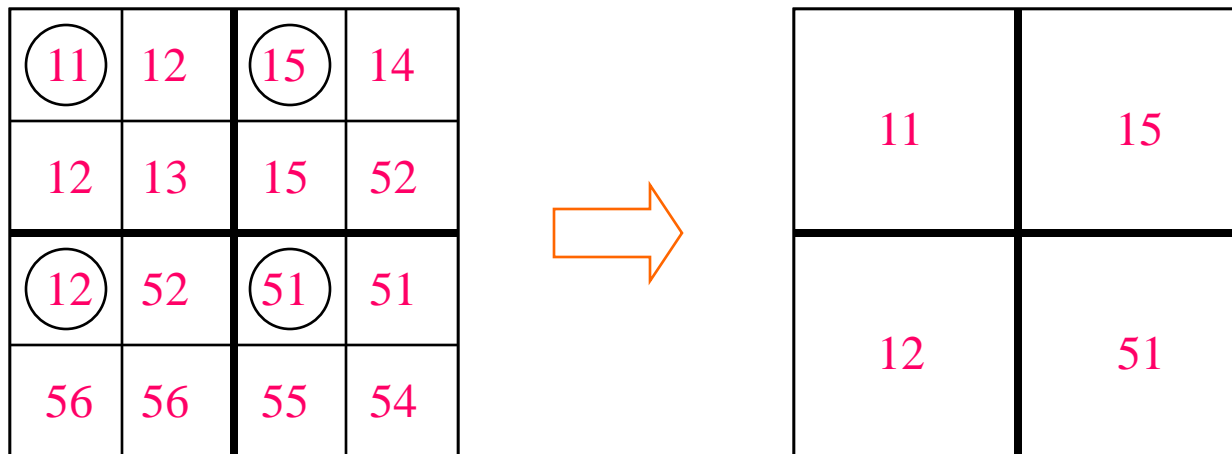
- 所有3层都运用同一原理：变换编码和子带编码。
 - 频谱被分为32个子带。
 - 应用快速傅里叶变换来表示高频域中的信号。
 - 应用心理声学模式来变换信号以估计刚能引起注意的噪音级。
- 层 1、2 和 3主要在最后一个阶段——即量化阶段的方式上有所区别，但不是唯一的差别。

4、图像和视频的压缩编码

图像和视频压缩方法



- 子采样(Subsampling): 编码思想一组像素用一个像素来表示, 达到降低图像分辨率的目的。



■ Run-Length Coding (RLC)

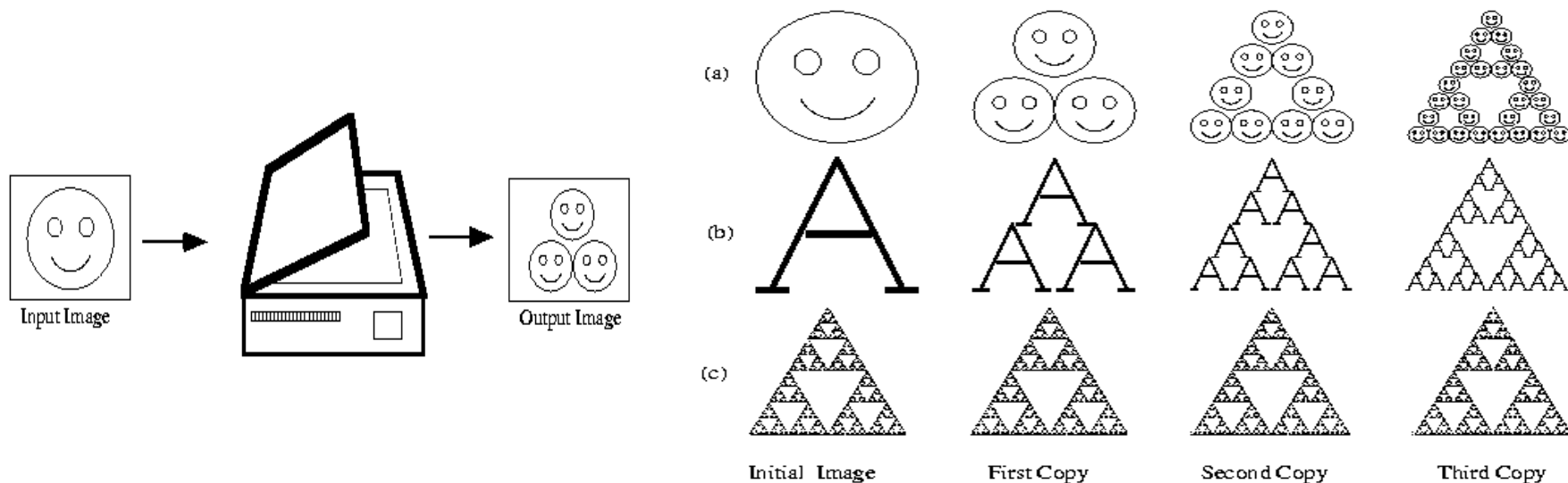
- 例1 (文本) 压缩前: ABCCCCCCDEF~~GGGG~~ 压缩后: ABC!8DEFG!4
- 例2 (黑白图像)

把每一行像素分成: 白段, 黑段, 白段, 黑段, 白段, 黑段, ...

然后, 每一段像素使用其长度值来表示: 计数1, 计数2, 计数3, 计数4, 计数5, 计数6, ...

分形编码 (fractal coding)

- 所谓图像的分形编码,是指利用图像本身的复杂性中所包含的自相似性进行压缩编码的方法,它是分形几何在图像压缩中的应用。
- 优点: 很大的压缩比、极高的分辨率、scalable
- 分形图形原理
 - 基于分形几何,使用少量参数及生成规则来描述有大量细节的自然景物,它们在不同尺度 (scale) 上均显露出统计上的自相似性。



- 任何形状的初始图像均转换成相同的最终图像,这个图像就是这台复印机的吸引子 (attractor). 只有初始图像在其拷贝上的位置和方向才能决定最终图像的形状。

■ 图像分形编码的过程

- 分形图像编码的过程是依据拼贴定理, 通过给定的图像, 寻找一组收缩映射, 使其组成的迭代函数系统的吸引子逼近给定图像, 然后记录下相应参数。解码过程是由相应参数确定迭代函数系统, 并根据迭代函数系统定理, 经过迭代生成图像。

■ 自适应块状分形编码方法

- 首先将图像分割成若干不重叠的值域块range和可以重叠的定义域块domain,
- 接着对每个range寻找某个domain, 使定义域块domain经过某个指定的变换映射到range达到规定的最小误差, 记录下确定range和domain的参数及变换 W_i , 得到一个迭代函数系统,
- 最后对这些参数进行编码。包括: 对图像的分割、搜索最佳匹配、最后记录相关的系数等3个步骤。

- (1) 把图像划分成互不重叠、任意大小、任意形状的range分区，所有range分区拼起来应恰为原图。
- (2) 为每个range分区再确定一个domain分区，domain分区必须大于相应的range分区，domain分区之间可以重叠，所有domain分区拼起来无须恰好覆盖原图全图。domain分区在三维仿射变换后应尽可能与range分区的图像接近。
- (3) 生成FIF(Fractal Image Format)文件。每个range分区都可通过仿射变换(n维空间函数图像经过旋转、平移、拉伸等多种变换的综合过程，仿射变换能使图像构成更好的组形)重建。FIF文件的开头规定如何划分range分区，接着依次记载各range分区对应的仿射变换系数。
- (4) 对FIF文件编码。由于每个range分区对应的仿射变换系数只有为数不多的几个参数(如给出方程 $y = ax+b$ 中的“a、b”值，就可画出一个任意分辨率的直线图形)，所以分形编码可获得很高的压缩比。

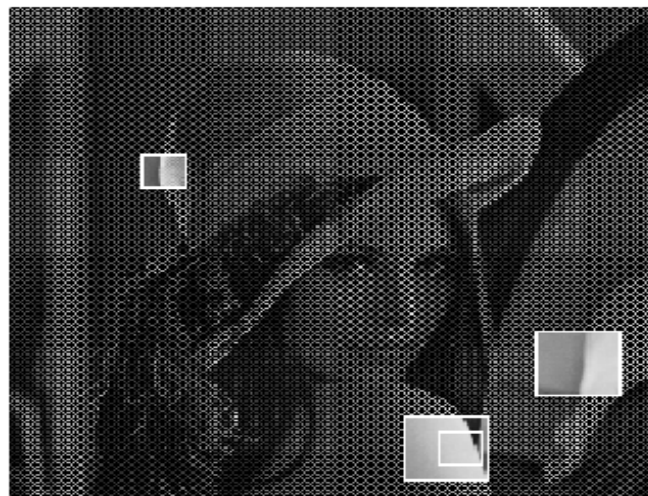
■ 人工干预的交互式分形图像编码方法

- 针对给定图像的形状, 采用边缘检测、频谱分析、纹理分析等传统的图像处理技术进行图像分割, 要求被分开的每部分都有比较直观的自相似特征。
- 然后寻找迭代函数系统, 确定各个变换系统。再由图像中灰度分布求得各个变换的伴随概率。
- 解码过程是采用随机迭代法来生成近似图像。



School of Computer Science and Information Engineering, Henan University, LIU Yang

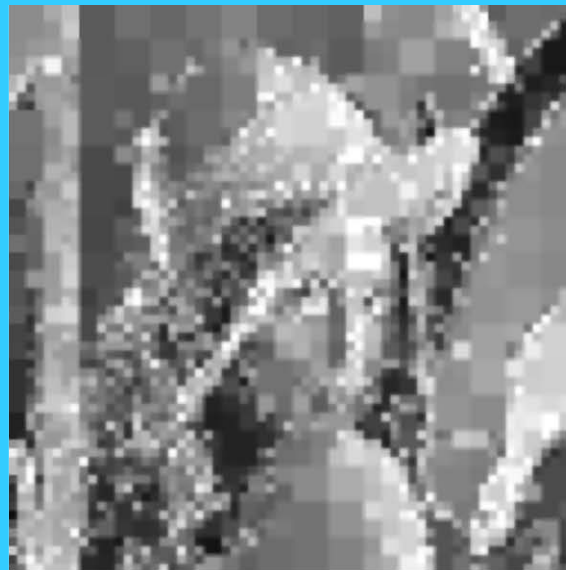
Lena图像



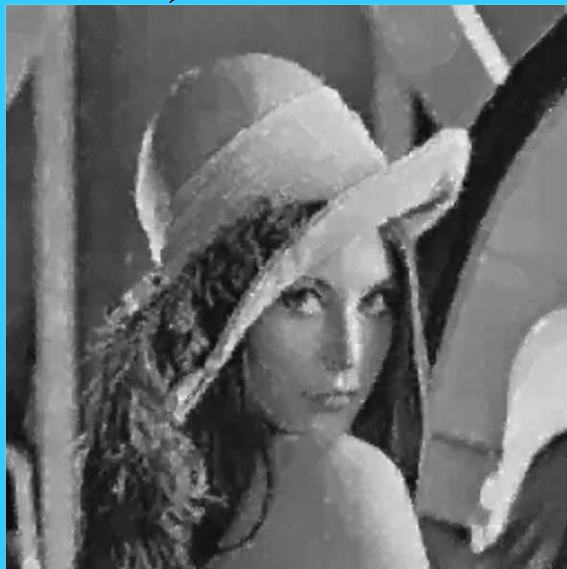
Lena图像中的自相似性



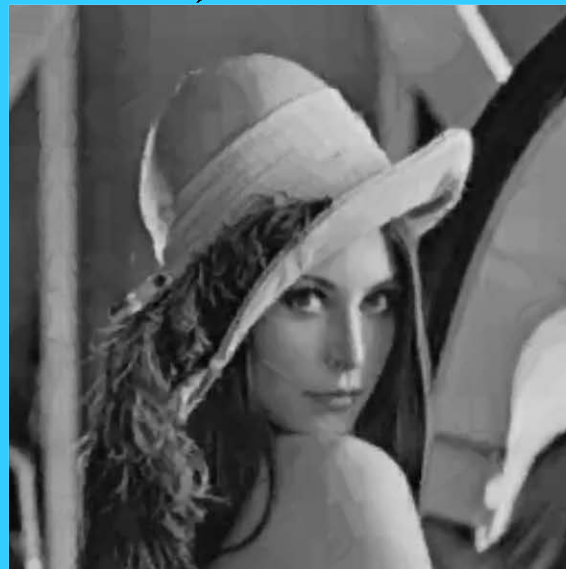
a) 初试化图像



b) 一次迭代后图像



c) 三次迭代后图像



d) 十次迭代后图像

■ JPEG标准简介

- 连续色调静止图像的数字压缩和编码标准JPEG (Joint Photo-graphic Experts Group), 直译为**联合图像专家组**, 1992年7月2日通过标准的第一部分。
- JPEG标准的组成
 1. ISO/IEC 10918-1:1994 | ITU-T T. 81--需求和指导方针
 2. ISO/IEC 10918-2:1995 | ITU-T T. 83--测试
 3. ISO/IEC 10918-3:1997 | ITU-T T. 84--扩展
 4. ISO/IEC 10918-4:1999 | ITU-T T. 86--参数、轮廓、标签、颜色空间、压缩类型和注册机构(REGAUT)等

■ JPEG标准划分为：

● 4种编码模式：

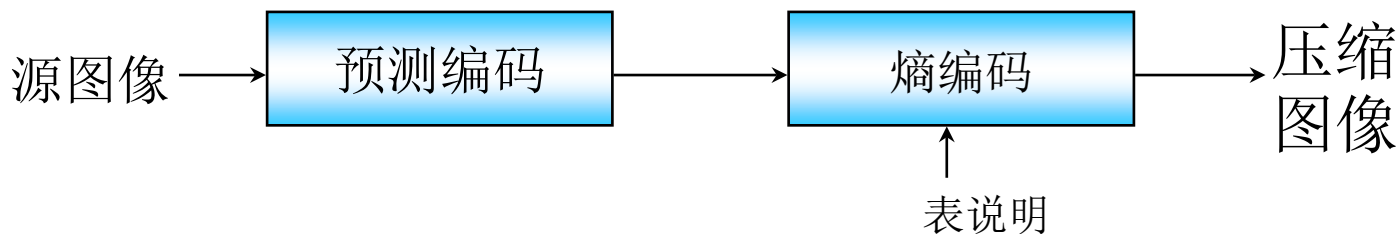
- **无失真编码**模式 (lossless encoding mode)
- 基于离散余弦变换 (Discrete Cosine Transform, **DCT**) 的 **顺序编码**模式 (DCT based sequential encoding mode)
- 基于DCT的**渐进编码**模式 (DCT based progressive encoding mode)
- 基于DCT的**多分辨率编码**模式 (DCT based hierarchical encoding mode)

其中后三种模式采用基于DCT的**有损编码**算法

5.1 无失真编码模式

■ JPEG的无损预测编码算法

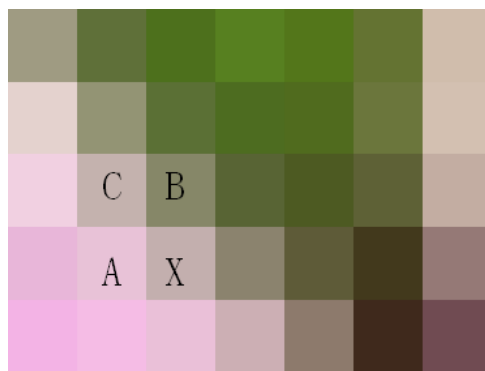
- 原理



- 无损压缩不使用DCT方法，而是采用一个简单的一维预测器 ($i=1, 2, 3$) 或二维预测器 ($i=4, 5, 6, 7$)。预测器可以采用不同的预测方法，不同的预测方法(如**三邻域预测法**)将决定有哪些相邻的像素将被用于预测下一个像素 X 。再用预测值 X' 减实际值 X 进行无失真熵编码。

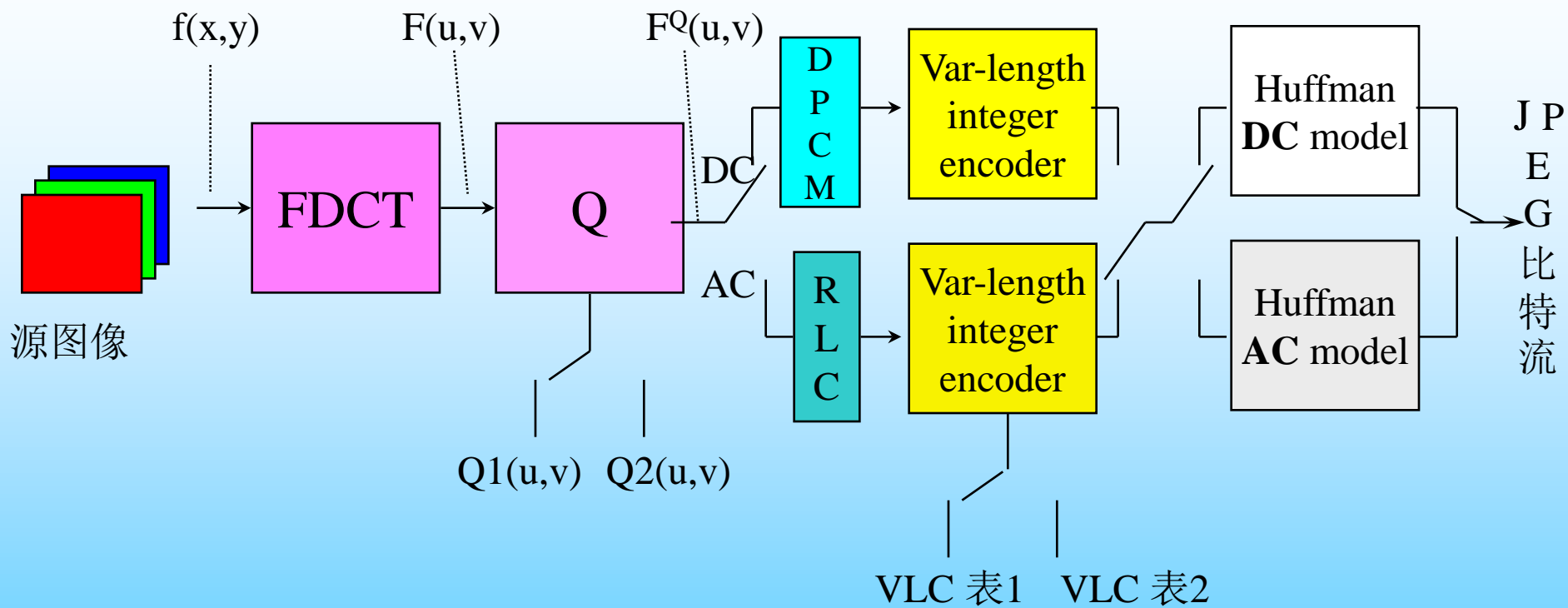
■ 效果

- 无失真
- 压缩比为 $2^{\sim}3$ 倍
- 8种预测公式适用于不同类型图像



$X'(0)=X$	$X'(4)=A+B-C$
$X'(1)=A$	$X'(5)=A+((B-C)/2)$
$X'(2)=B$	$X'(6)=B+((A-C)/2)$
$X'(3)=C$	$X'(7)=(A+B)/2$

5.2 基于DCT的顺序编码模式



1. 预处理
2. DCT变换
3. 量化
4. DPCM/RLC
5. VLC
6. Huffman

- 把每帧图像分量均划分为若干 8×8 的数据子块, 变换前需对像素值进行电平偏移, 即

$$[0, 2^{P-1}] \rightarrow [-2^{P-1}, 2^{P-1}-1] \quad (\text{采样精度 } P=8 \text{ or } 12)$$

- 变换后, 得到的 $F(u, v)$ 其动态范围是

$$[-2^{P+2}-1, 2^{P+2}-1]$$

- 原始的图像数据块经过DCT变换后, 将每个数据块的数据从空间域变换到频率域, 输出64个DCT变换系数。

思考: 为什么要进行电平偏移?

第2步：DCT变换

■ 8X8像素DCT变换

- 设 $f(x, y) = \begin{bmatrix} f_{0,0} & \dots & f_{0,7} \\ & & \\ & & \\ & & \\ & & \\ & & \\ f_{7,0} & \dots & f_{7,7} \end{bmatrix}$ 是64个像素的子图像块, 则

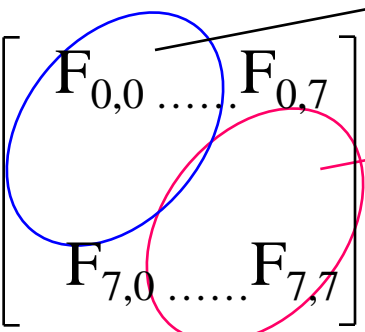
$$\text{FDCT} \quad F(u, v) = \frac{1}{4} C(u) C(v) \left[\sum_{i=0}^7 \sum_{j=0}^7 f(i, j) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} \right]$$

$$\text{IDCT} \quad f(i, j) = \frac{1}{4} C(u) C(v) \left[\sum_{u=0}^7 \sum_{v=0}^7 F(u, v) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} \right]$$

- 当 $u=0, v=0$ 时, $c(u)=c(v)=1/\sqrt{2}$, 否则 $c(u)=c(v)=1$
- MATLAB: $F=\text{dct2}(f)$; $f=\text{idct2}(F)$;

DCT频域图像F(u, v)的解释

设 $F(u,v) = \begin{bmatrix} F_{0,0} & \dots & F_{0,7} \\ \vdots & & \vdots \\ F_{7,0} & \dots & F_{7,7} \end{bmatrix}$



低频成分

高频成分

- $F_{u,v}$ 称为DCT系数，其中 $F_{0,0}$ 是直流系数，其它为交流系数。 $F_{0,0}$ 与原始图像的平均亮度成正比。 $(u+v)$ 较小的系数代表低频成分， $(u+v)$ 较大的系数代表高频成分。
- 多数情况下高频成分数值很小，甚至有许多“0”。

第3步：DCT系数的量化

$$V_l(u,v)=\begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \quad V_{cbcr}(u,v)=\begin{bmatrix} 17 & 18 & 24 & 47 & 99 & 99 & 99 & 99 \\ 18 & 21 & 26 & 66 & 99 & 99 & 99 & 99 \\ 24 & 26 & 56 & 99 & 99 & 99 & 99 & 99 \\ 47 & 66 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \end{bmatrix}$$

亮度量化表 色度量化表

- JPEG量化采用根据视觉心理加权函数计算得到的线性均匀量化器(源于1984年Lohscheller的CCIR-601实验)
- 量化处理:

$$F^Q(u, v) = \text{Integer Round}(F(u, v) / Q(u, v))$$

$$Q(u, v) = (Q/50) * V(u, v)$$

Q越小\CR越低, 图像质量越好

DCT量化处理示例

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

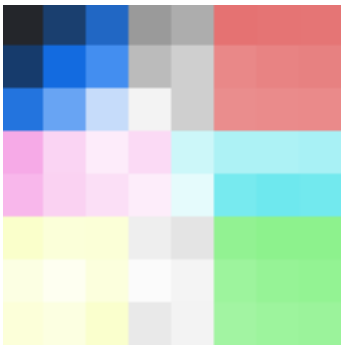
源图像数据



DCT

1259.6	-1.0	-12.1	-5.2	2.1	-1.7	-2.7	1.3
-22.6	-17.5	-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

DCT系数



量化

180	0	-1	0	0	0	0	0
-3	-2	-1	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

量化DCT系数 (Q=20)

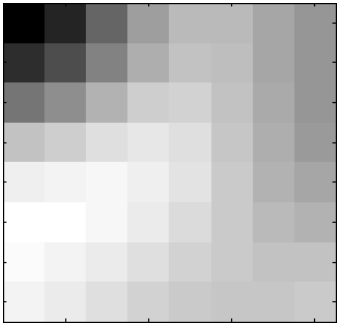
1260	0	-10	0	0	0	0	0
-21	-16	-10	0	0	0	0	0
-10	-10	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

反量化DCT系数

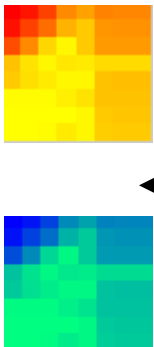
IDCT

142	145	150	155	157	157	156	154
146	149	153	156	158	158	156	154
152	154	157	159	159	158	156	154
158	159	160	161	160	158	156	155
162	162	162	162	160	159	157	156
163	163	162	161	160	159	157	157
163	162	161	160	159	158	158	158
162	161	160	159	159	158	158	159

IDCT图像数据



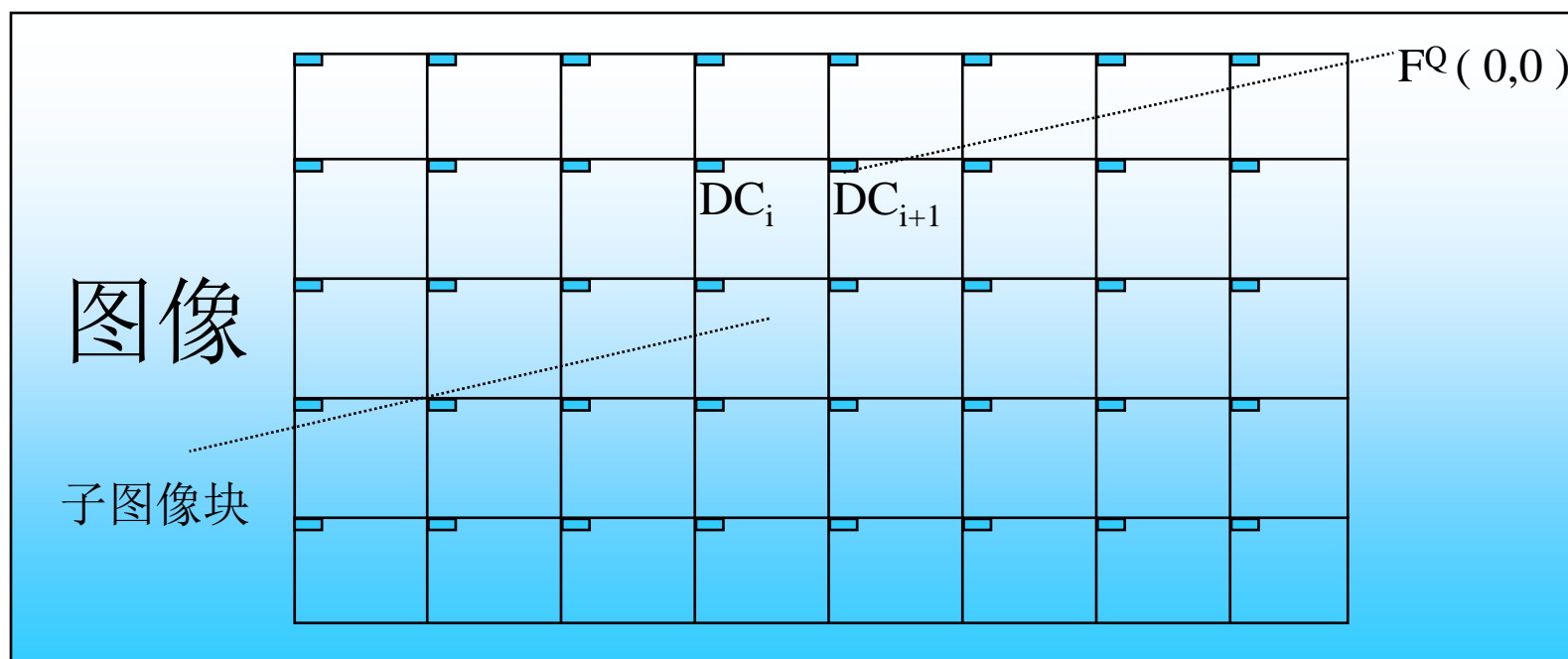
RGB/YCrCb



第4步：DCT系数编码(DC)

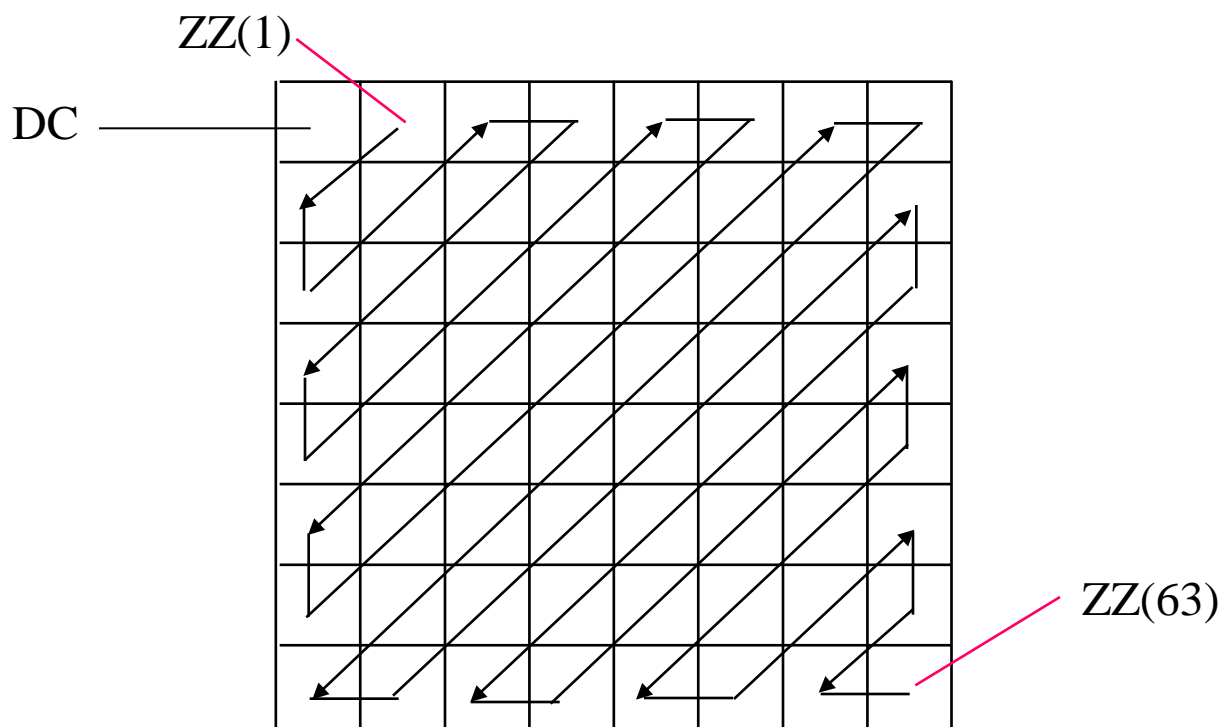
- JPEG中对DC系数采用DPCM编码，63个AC系数在JPEG算法中采用RLC编码。
 - A) 直流系数 DC_i 的处理： $F^Q(0,0)$ 是子图像块的直流系数，一幅图像的所有直流系数使用DPCM编码。

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



第4步：DCT系数编码(AC)

- B) 交流系数AC的处理: 将 $F^Q(u, v)$ 按“Z”字形排成一个一维数组 $ZZ(1..63)$ 。



思考：为什么Z形扫描？

第5步：VLC编码(熵编码的中间格式表示)

- AC系数的符号对:符号1(RunLength, Size)为即连零游程长度(行程)和非零值的位数(尺寸);符号2(Amplitude)为幅度值,即非零值的变长整数编码(VLI)。
- DC系数的符号对:符号1(Size)为非零值的位数(尺寸);符号2为幅度值(Amplitude)。



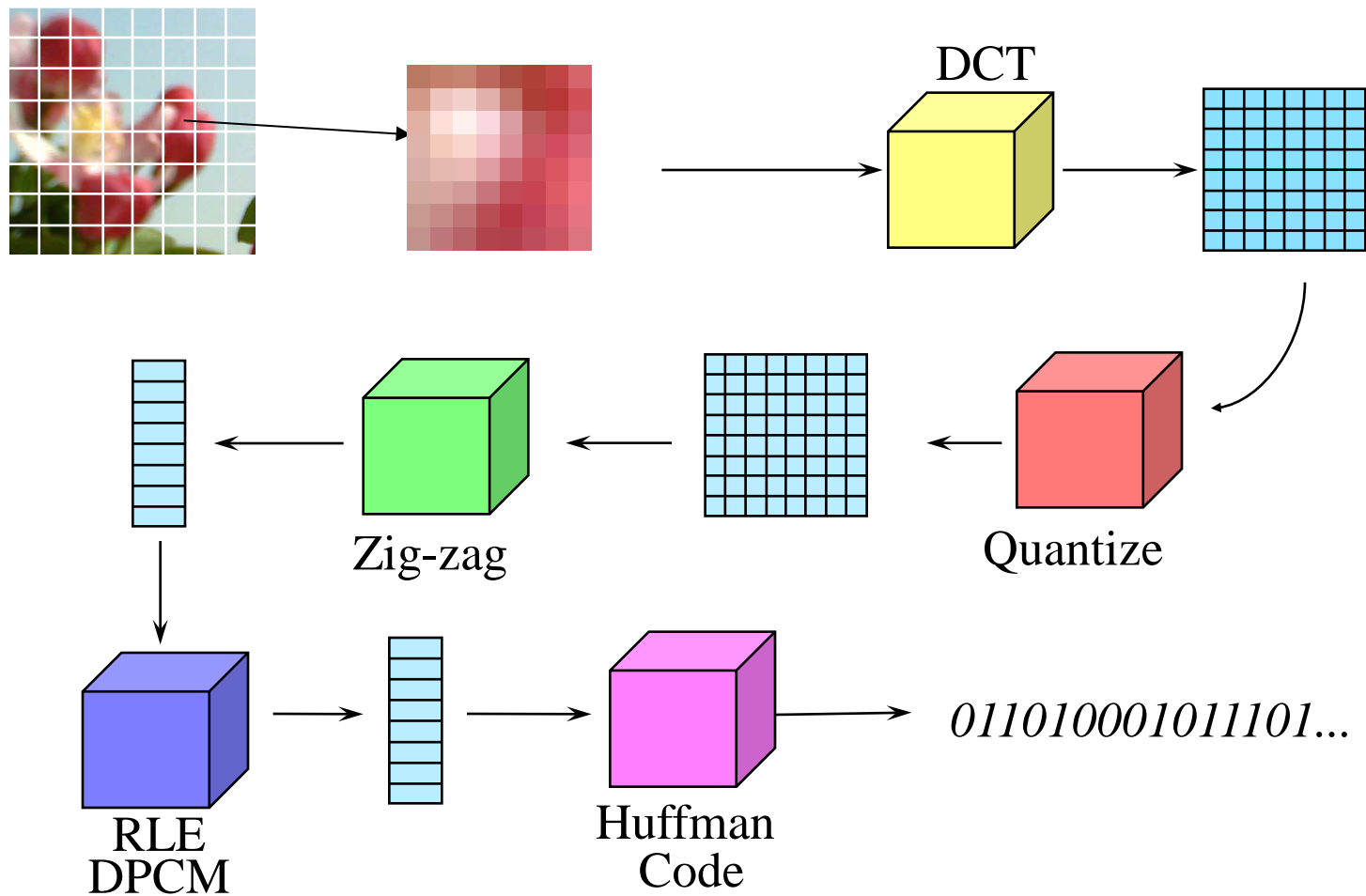
第6步:熵编码

- 经过DPCM编码的直流项和经过行程编码的交流项再进行Huffman编码或自适应二进制算术编码。
- 直流系数符号1的Huffman码表

	亮度分量		色度分量	
SSSS	码长	代码	码长	代码
0	2	00	2	00
1	3	010	2	01
2	3	011	2	10
3	3	100	3	110
4	3	101	4	1110
5	3	110	5	11110
6	4	1110	6	111110
7	5	11110	7	1111110
8	6	111110	8	11111110
9	7	1111110	9	111111110
10	8	11111110	10	1111111110
11	9	111111110	11	11111111110

- 交流系数符号1的Huffman码表（略）

JPEG压缩编码的整体过程图解



问题：JPEG信息熵损失在哪一步？

顺序编码压缩效果



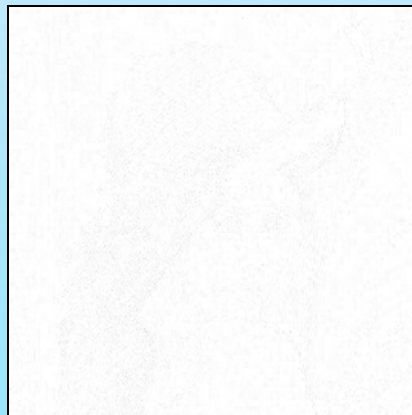
Uncompressed
(262 KB)



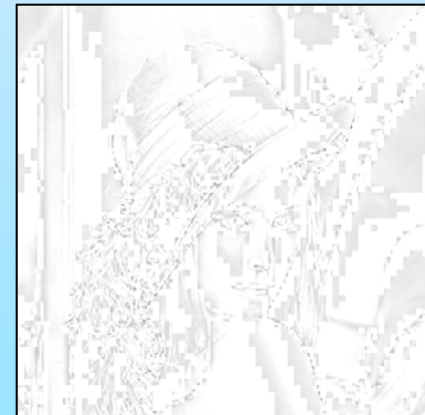
Compressed (50)
(22 KB, 12:1)



Compressed (1)
(6 KB, 43:1)



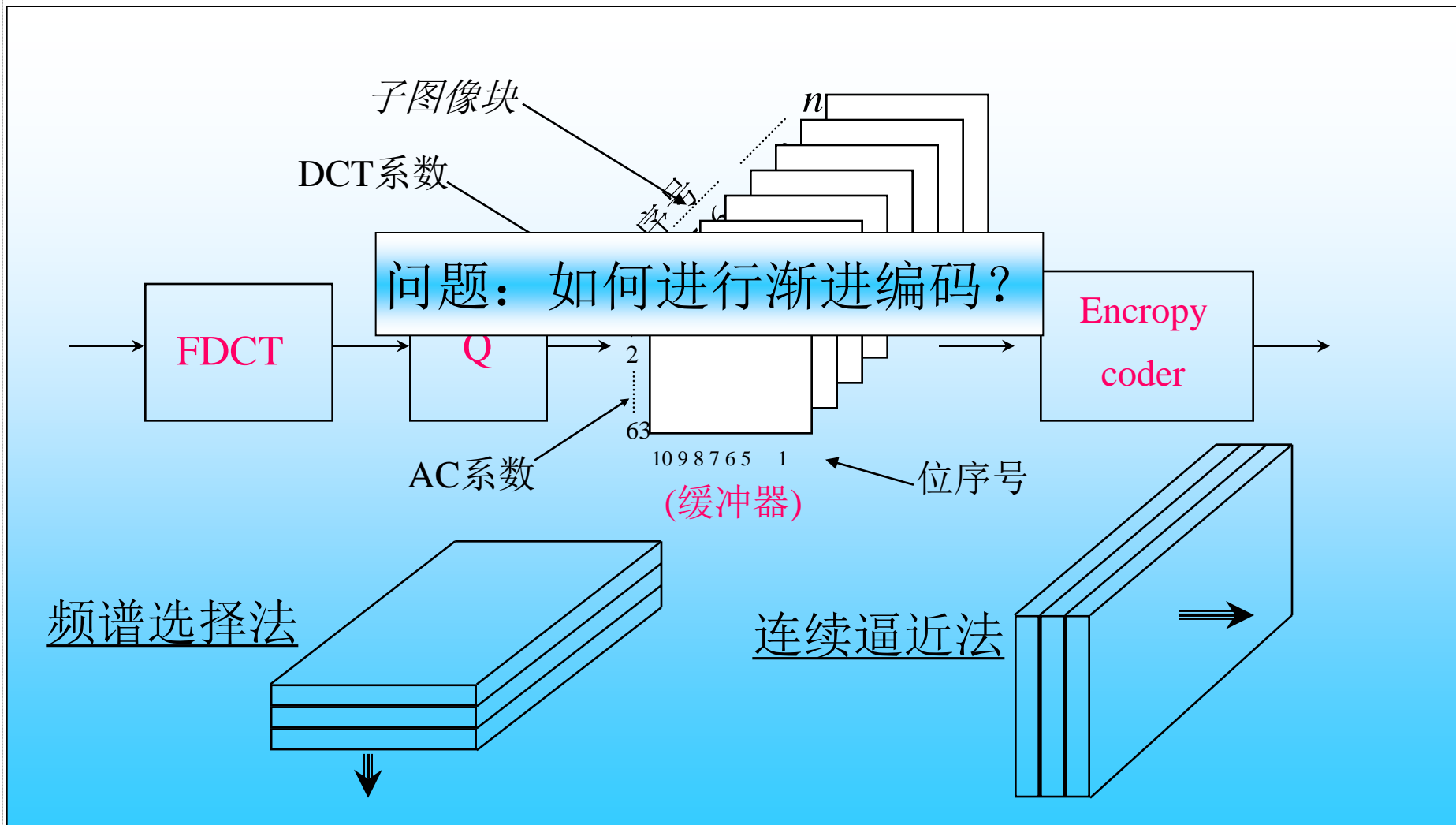
```
imwrite(p,'jpg0.jpg','quality',50);
```



```
imwrite(p,'jpg0.jpg','quality',0);
```

5.3 基于DCT的渐进编码模式

- 目的:重建的图像在屏幕上逐步浮现,改善HCI效果



顺序显示和渐进显示示例

顺序显示和渐进显示示例

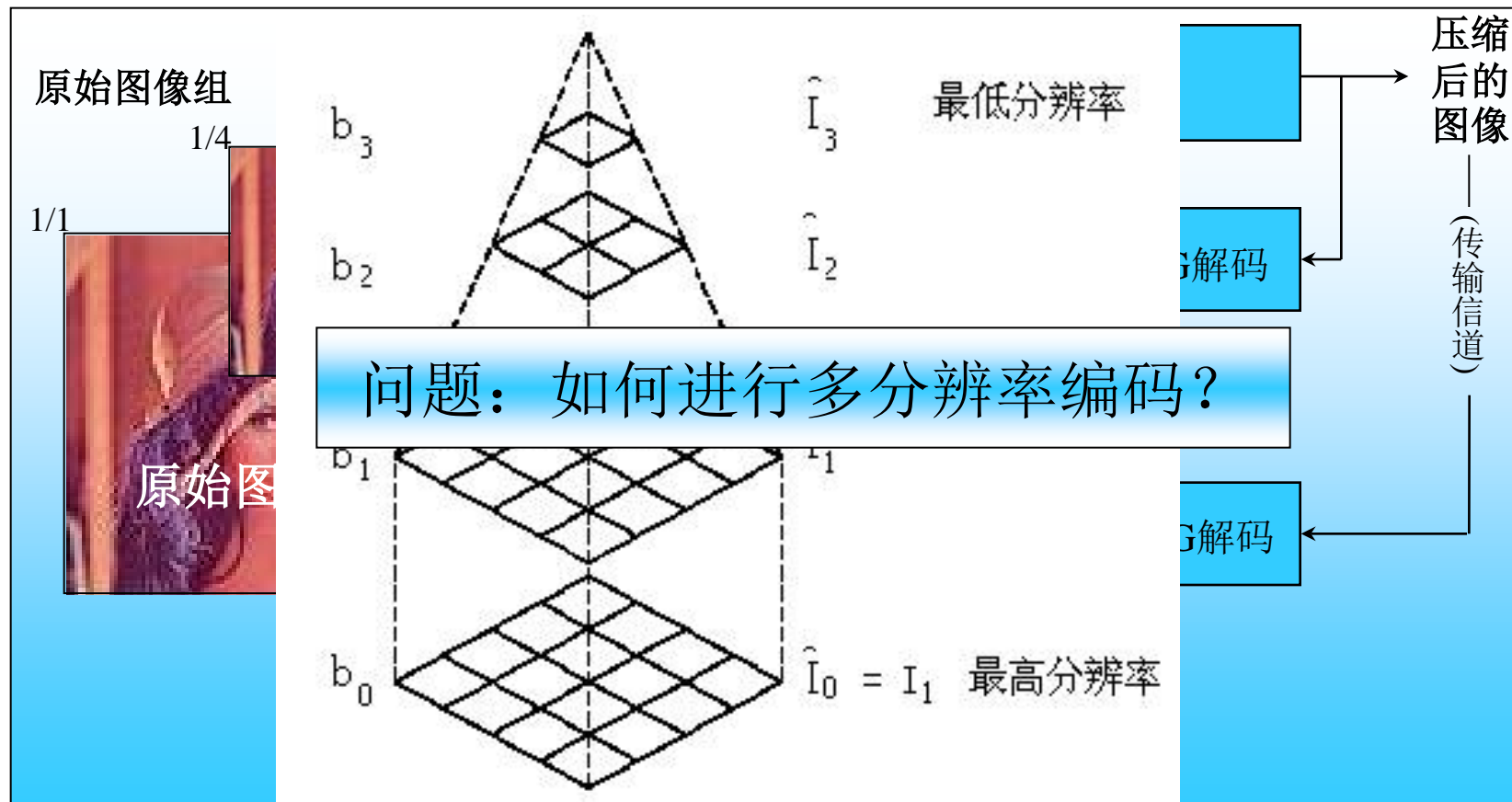


5.4 基于DCT的多分辨率编码模式



5.4 基于DCT的多分辨率编码模式

- 多分辨率编码又称金字塔 (pyramid) 编码或锥形编码。其目的是支持高分辨率图像在任意分辨率 (scalable) 设备上输出。



■ 金字塔分解的数据量计算

第一层数据量 m_0 , 第二层数据量 $m_0 \times 0.25$, ..., 第 N 层数据量 $m_0 \times 0.25^{N-1}$

总计 $M = m_0 \times (0.25^0 + 0.25^1 + \dots + 0.25^{N-1})$

$$= m_0 \times (1 - 0.25^N) \times 4/3$$

$$S_n = \begin{cases} na_1 & (q = 1) \\ \frac{a_1(1 - q^n)}{1 - q} & (q \neq 1) \end{cases}$$

■ 地球高分辨率图像数据量计算

高分辨率影像即单位像素到达分米到米的遥感图像。对于AMD 2GHz, 1.37G单机, 全球米级图像 ($5.11e+14$), 每个像素处理一次, 需要2.5年; 全球分米级图像 ($5.11e+16$), 每个像素处理一次, 需要249年! (人脑神经元数目 $1.0e+11 \sim 1.0e+12$, 每个神经元胞体和树突可与其它神经元建立约1000联系, 每个神经元轴突与其它神经元可与其它神经元建立约1000联系)

■ DCT变换的不足:

- 基于Fourier变换的DCT能分析非平稳信号的频率, 却无法知道频率出现时间段。
- 原始图像的分块DCT变换和量化, 高压压缩比重建图像在水平和垂直方向会出现晕圈和幻影, 产生“方块”效应。
- 不分块或分块很大, 图像块中像素能量集中到少量系数的效果将变得不明显, 不利于对数据进行压缩, 同时还使得计算复杂度增加。

■ 小波变换(Discrete Wavelet Transform, DWT)发展历史

- 1822年Fourier发表“热传导解析理论”论文, 提出FFT。
- 1910年Haar提出最简单的Haar小波。
- 1980年Morlet提出平移伸缩小波公式, 用于地质勘探。
- 1985年Meyer提出“正交小波基”, 形成小波研究的高潮。
- 1988年Mallat提出的多分辨率分析(Multi-resolution Analysis, MRA)理论, 统一了语音识别中的**镜向滤波**, 图像处理中的**金字塔分解**方法, 地震分析中**短时波形处理**等不相关的领域。

当在某一个分辨率检测不到的现象, 在另一个分辨率却很容易观察处理, 如:



不识庐山真面目,只缘身在此山中



6.1 离散小波变换(DWT)

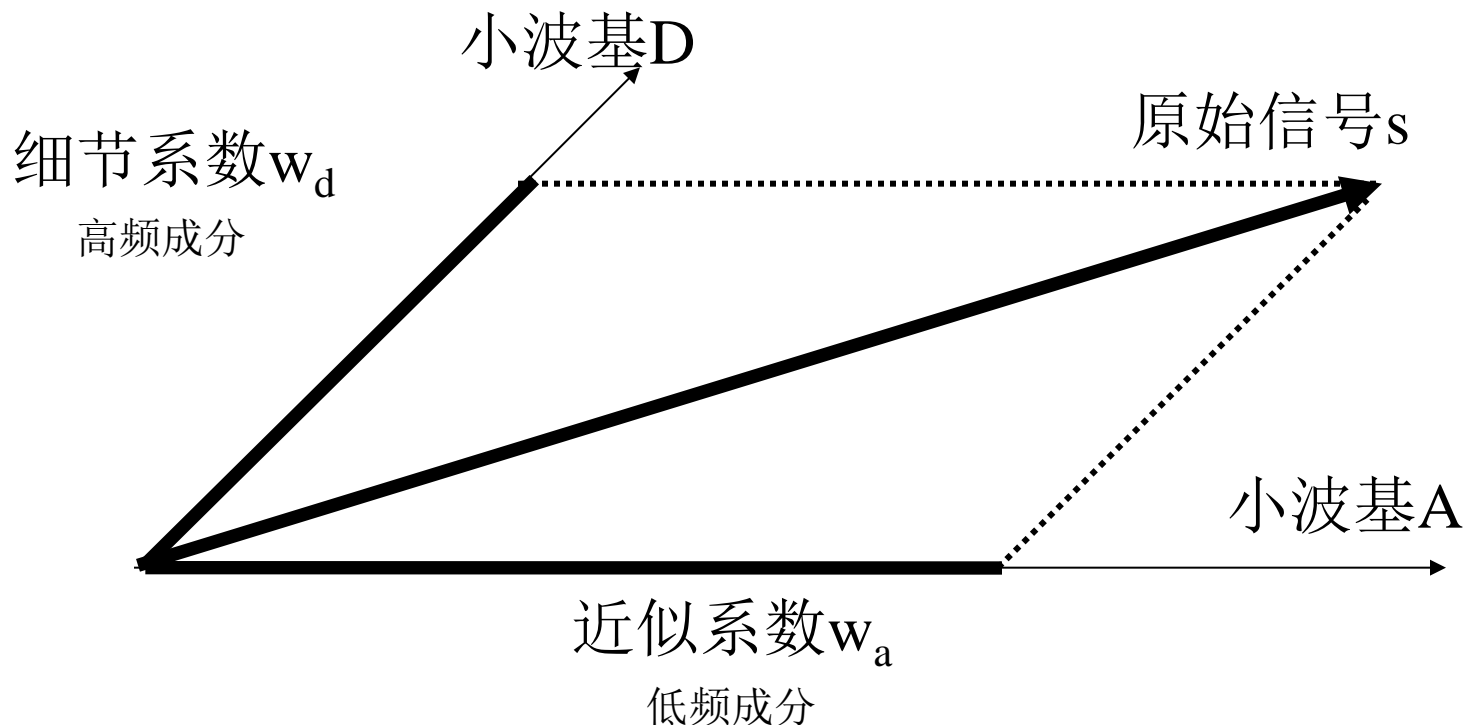
- 设信号 s 有 M 个样本, $n=1\dots M$
则信号 s 的 J 级小波变换 $w_J = [w_{aJ}, w_{dJ}, \dots, w_{d1}]$
其中:

$$\begin{aligned} DWT \quad w_{aJ} &= \langle s(n) \cdot A_J(n) \rangle \\ w_{dj} &= \langle s(n) \cdot D_j(n) \rangle \quad j = J, \dots, 1 \end{aligned}$$

$$IDWT \quad s(n) = a_J(n) + \sum_{j=1}^J d_j(n) = w_{aJ} A_J(n) + \sum_{j=1}^J w_{dj} D_j(n)$$

$A_J(n), D_j(n)$ 为小波基函数

6.2 小波分解和小波基



正变换：原始信号在小波基上，获得“小波系数”分量

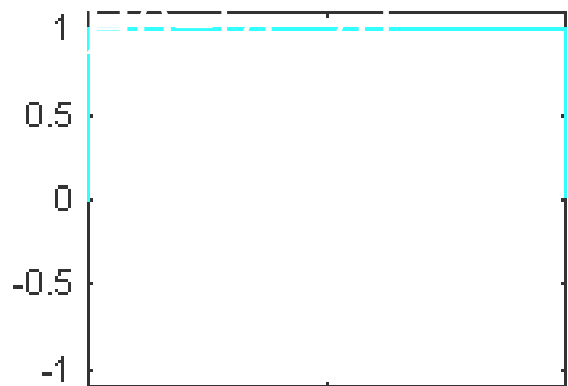
反变换：所有“小波分解”合成原始信号

例如：小波分解 $a = \text{小波系数 } w_a \times \text{小波基A}$

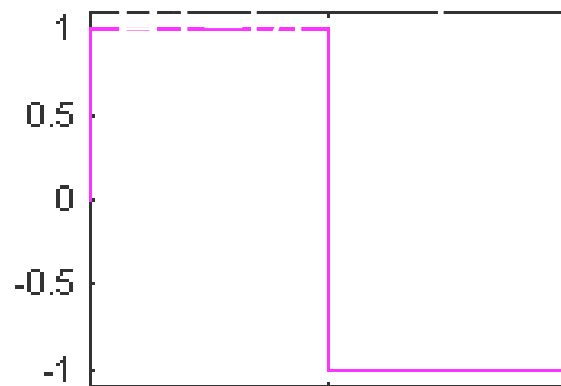
Haar小波基函数

“近似”基函数

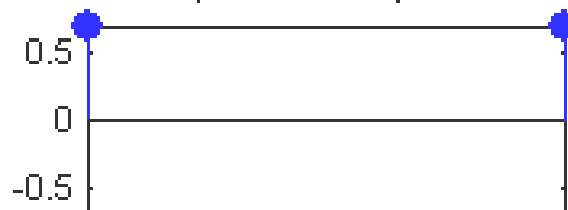
“细节”基函数



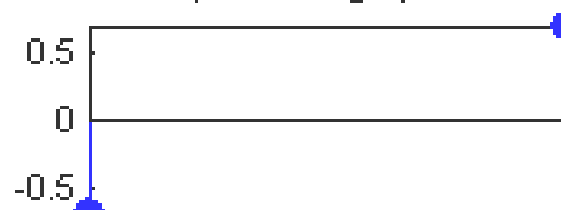
Decomposition low-pass filter



Decomposition high-pass filter



Reconstruction low-pass filter



Reconstruction high-pass filter

Wavelet

Refinement

Display

Information on

“正变换” 低频
和

高频 “滤波系数”

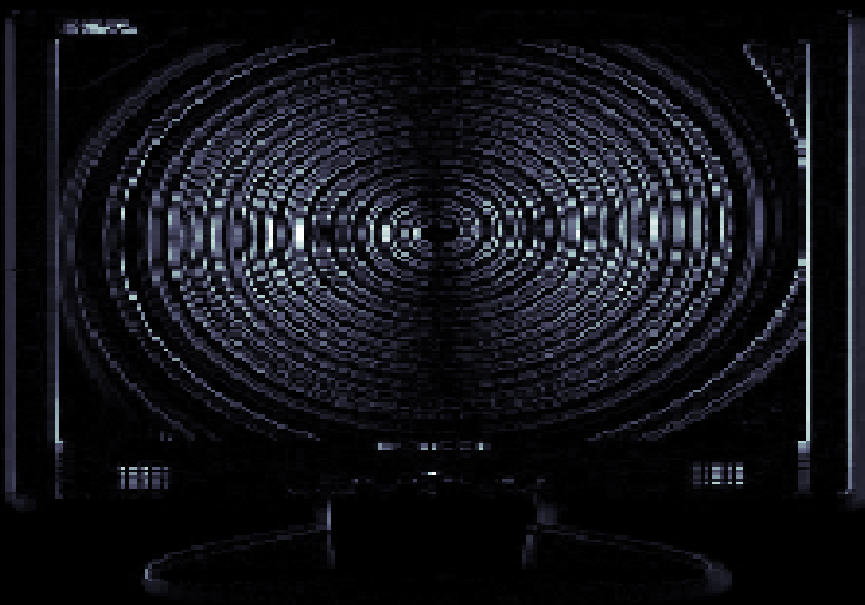
“反变换” 低频
和

高频 “滤波系数”

6.3 图像的小波分解

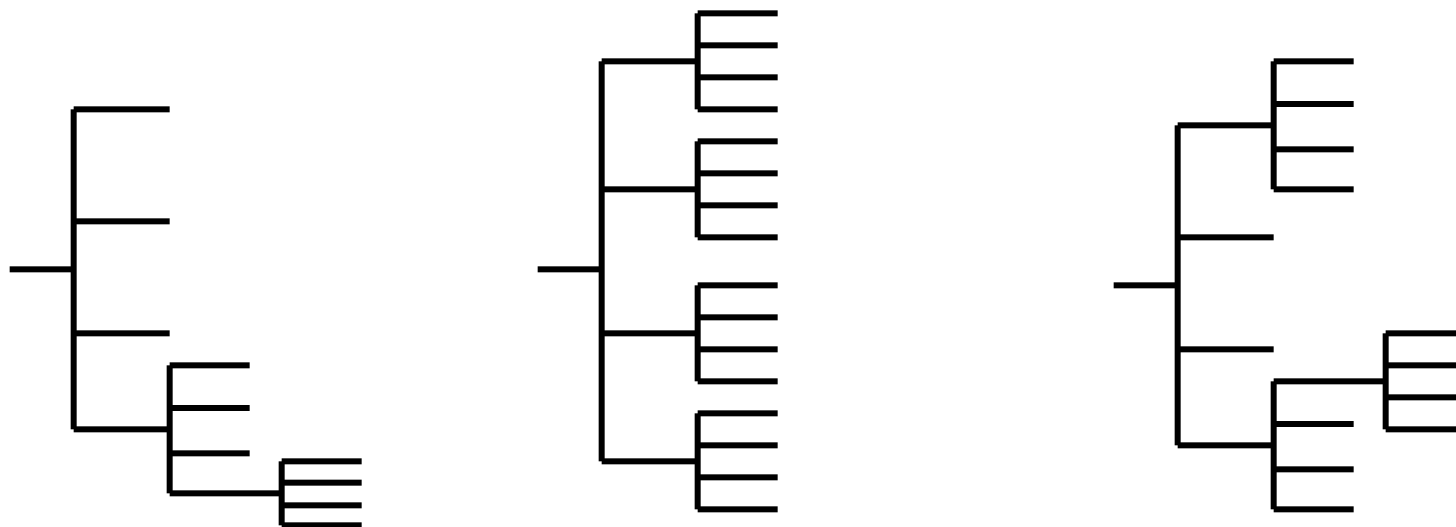
问题：HL与LH子图有什么不同？

MfT



图像小波分解的常用方式

- a. 金字塔式分解
- b. 完全平衡的分解
- c. 小波包 (Wavelet Packet) 分解





小波系数分级树形表示法

Original Image



Synthesized Image



Approximation coef. at level 3



Image Selection : Decomposition at level 3

 L_1 第 1 级 L_1 分辨率

Decomposition at level : 3

View mode : Tree

Full Size

1

2

 L_2 第 2 级 L_2 分辨率

Operations on selected image

Visualize

Full Size

Reconstruct

 L_3 第 3 级 L_3 分辨率

Colormap

pink

Nb. Colors

Brightness

Approximations

Horizontal Details

Diagonal Details

Vertical Details



Wavelet 2.0

File

Options

小波系数分级方块表示法

Original Image



Approximation coef. at level 3



近似图像

Synthesized Image



dwt

idwt

Image Selection

第3级 L_3

第2级 L_2 细节

第1级 L_1

水平细节

第1级 L_1
垂直细节

第1级 L_1

斜线细节

Decomposition at level 3

Image

cameraman

Wavelet

haar

Level

3

Analyze

Statistics

Compress

Histograms

De-noise

Level: 3

Full Size

1

2

3

4

on selected image :

Visualize

Full Size

Reconstruct

map

pink

Colors

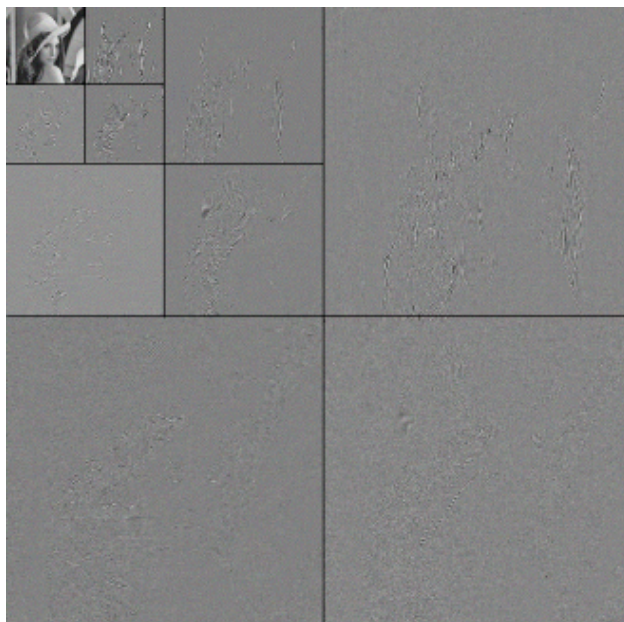
25

Brightness

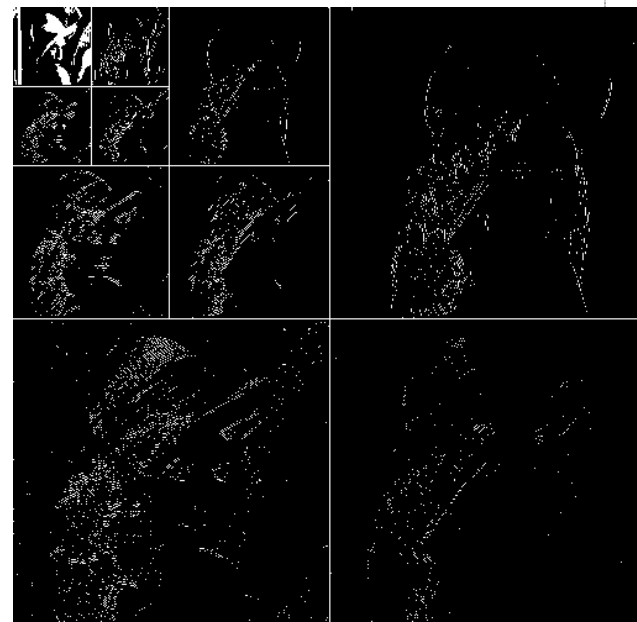
原始图像与三级小波分解后图像的比较



a. 原图 Lena 512*512



b. 分解3层



c. 分解图进行二值化后的图形显示

$[CA, CH, CV, CD] = \text{dwt2}(X, \text{Lo_D}, \text{Hi_D})$

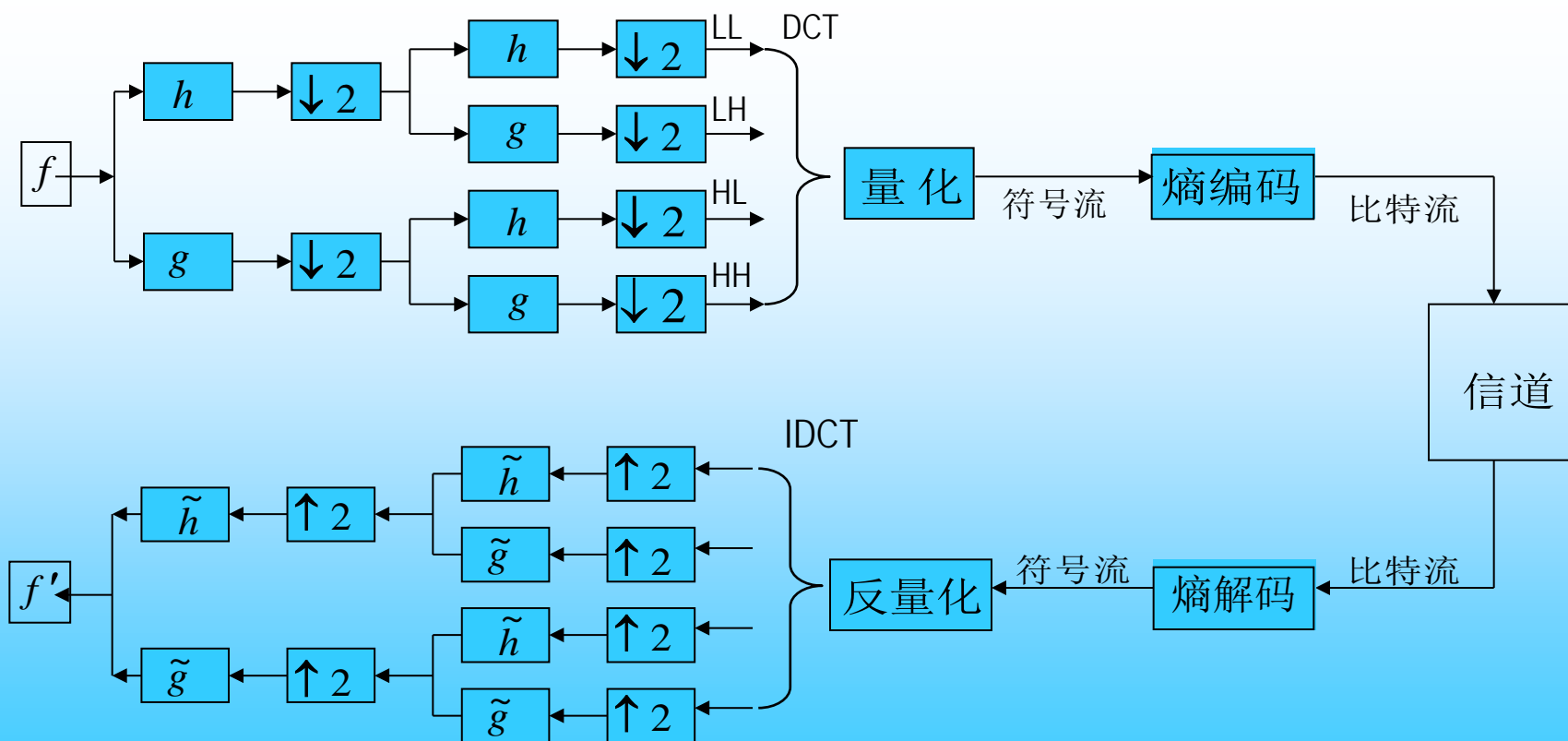
$[CA, CH, CV, CD] = \text{dwt2}(X, \text{'wname'})$

$X = \text{idwt2}(CA, CH, CV, CD, \text{Lo_R}, \text{Hi_R})$

$X = \text{idwt2}(CA, CH, CV, CD, \text{'wname'})$

6.4 小波变换用于图像压缩

- DWT后的子图统计特性有改善，可消除行列之间的相关关系。根据人类的视觉生理和心理特点分别作不同策略的量化和编码处理。



基于DWT的图像编码框架

小波压缩试验结果



BMP 509X375



J2K CR=60

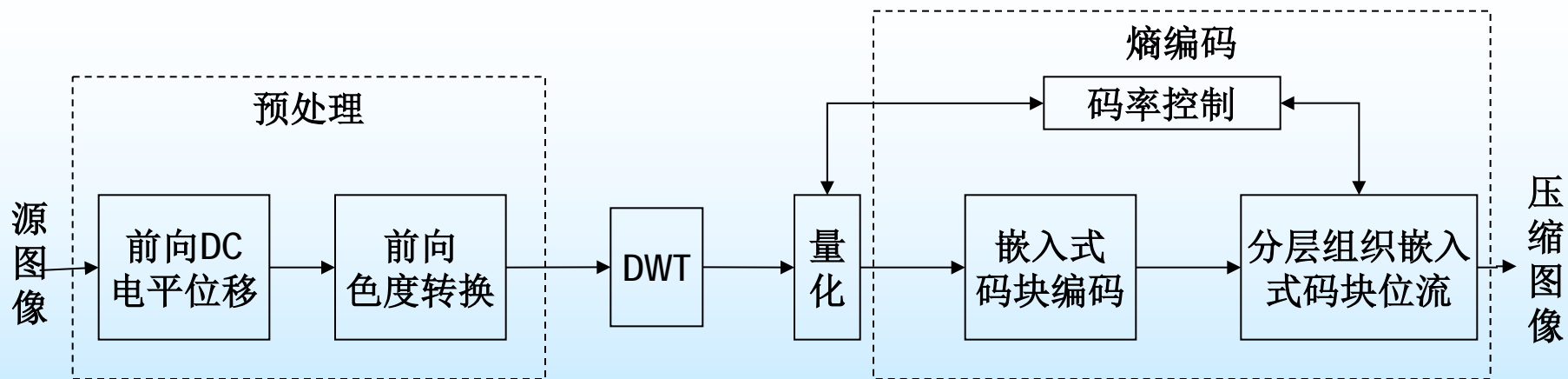


J2K CR=240

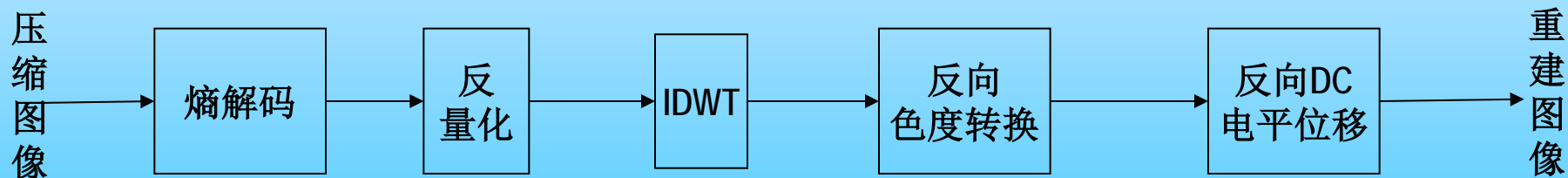


J2K CR=480

6.5 JPEG2000简介



JPEG2000编码器



JPEG2000解码器

JPEG与JPEG 2000的性能比较



标准	JPEG	JPEG 2000
编码技术	DCT 知觉量化 Zigzag扫描 Huffman编码 算术编码	DWT 最佳截断嵌入码块编码(EBCOT) 感兴趣区域(ROI)编码 空间可扩展编码 质量可扩展编码 面向对象编码 位图形状编码 容错编码,固定码率 (TCQ)
压缩比	2~30	2~50
算法效率	30:1以上急剧下降	100:1以上急剧衰减
速率失真特性	J2K比JPEG提高30%	

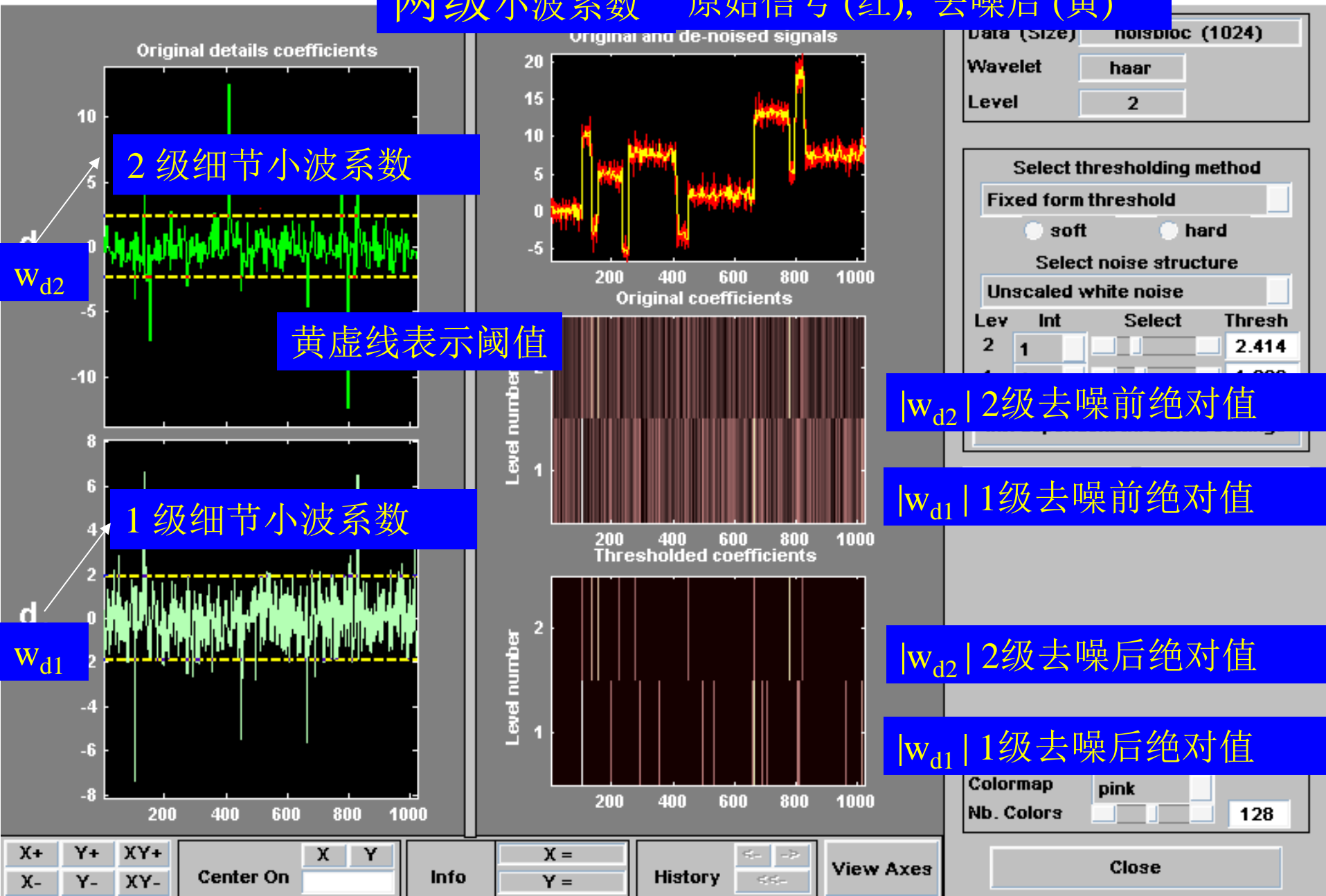


JPEG

JPEG2000

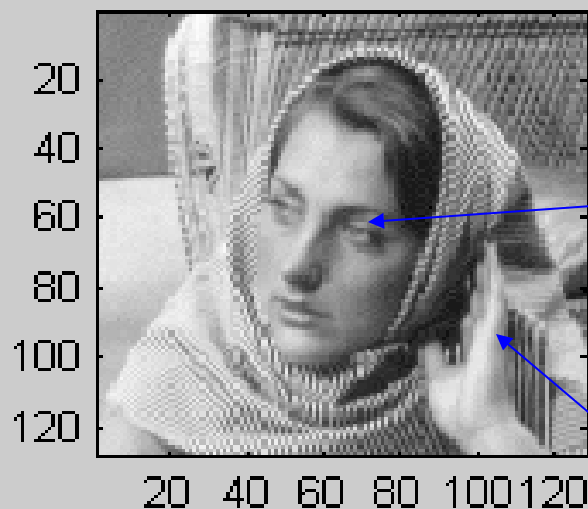
6.6 小波的其他应用—滤噪

两级小波系数 原始信号 (红), 去噪后 (黄)



6.7 小波的其他应用—图像特征抽取

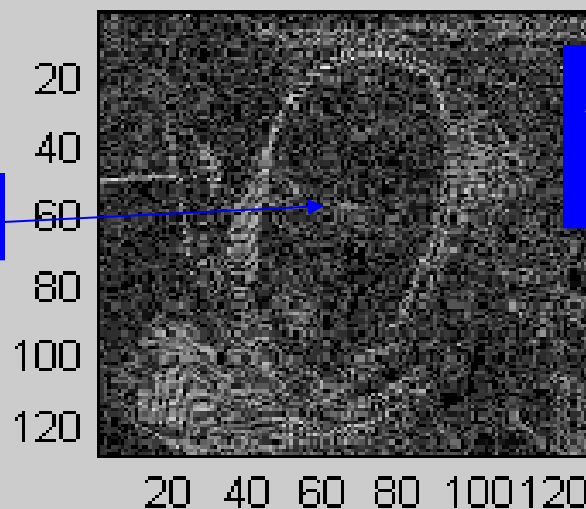
编码后低频系数图像



近似
图像

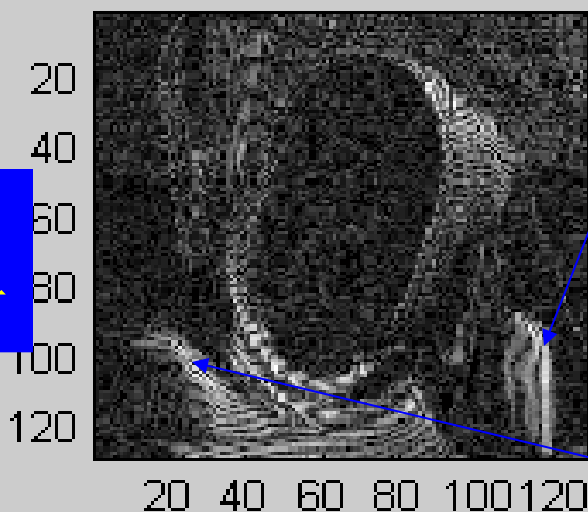
水平细节

编码后水平高频图像



第1级
水平细节

编码后垂直高频图像

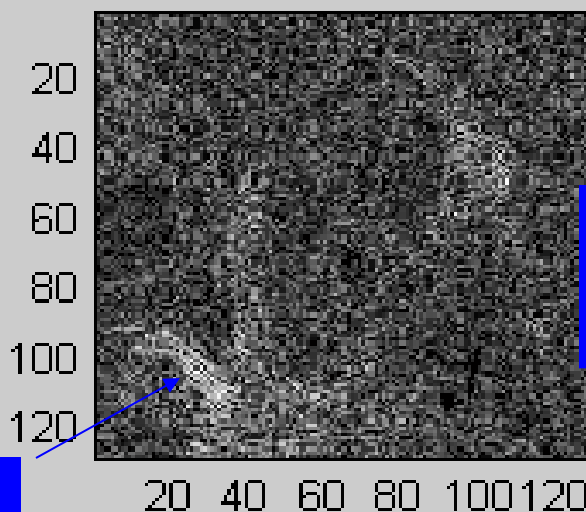


垂直细节

第1级
垂直细节

斜线细节

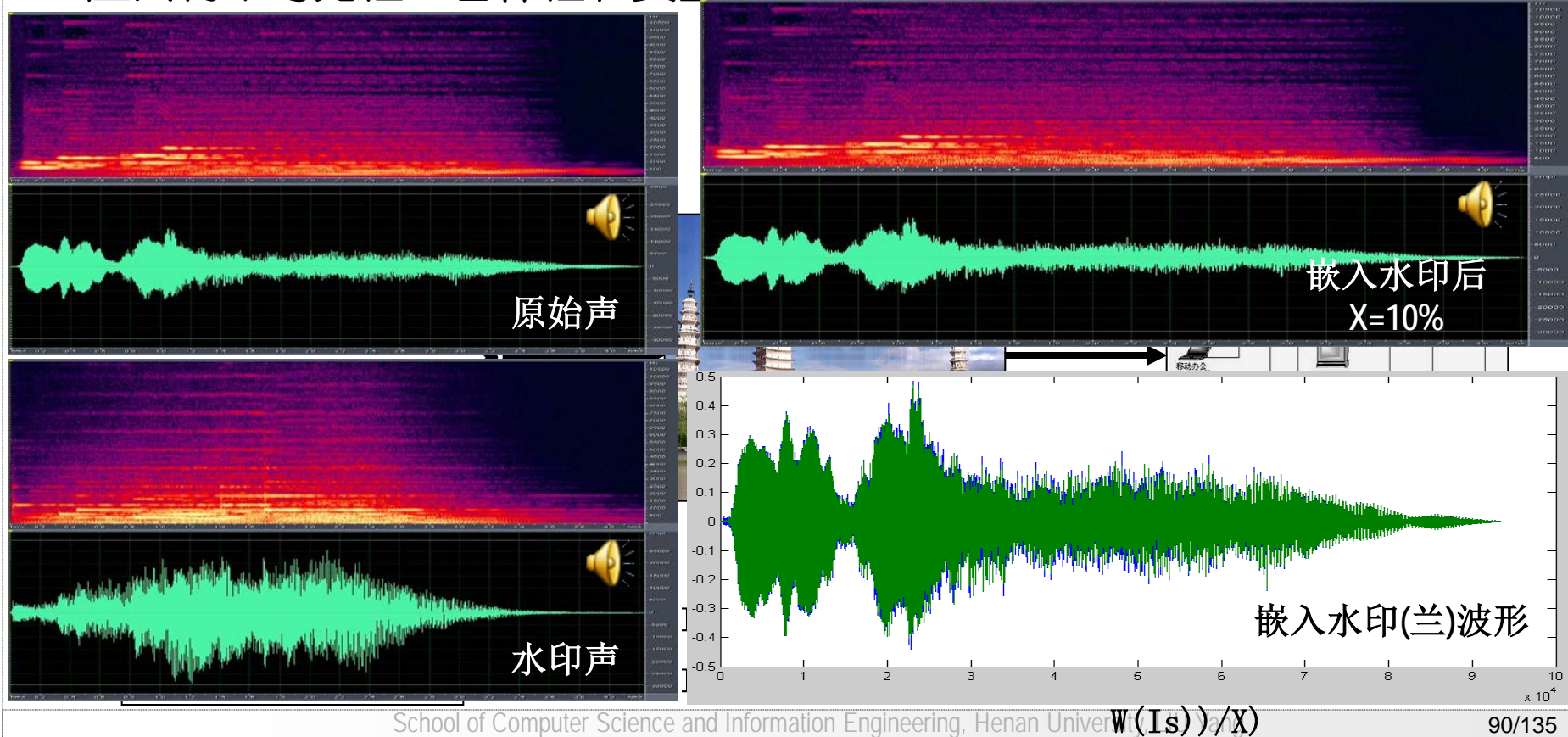
编码后斜线高频图像



第1级
斜线细节

6.8 小波的其他应用—数据隐藏

- 数字水印(Digital Watermark)技术是将与多媒体内容相关或不相关的秘密水印信息嵌入多媒体内容当中,但不影响原内容的使用价值,不易被人的知觉系统觉察或注意到。
- 数字水印可证实数据的所有权归属和保证数据的完整性。嵌入的水印一般应具有不可见性、鲁棒性和安全性。



6.9 图像的DWT优点

- **高压缩比**：图像的DWT压缩率比高，重建图无“方块”效应。
- **速度快**：DWT可一次变换整幅图像，速度比FFT快一个数量级。

$$\text{DWT} = O(N)$$

$$\text{FFT} = O(N \log_2 N)$$

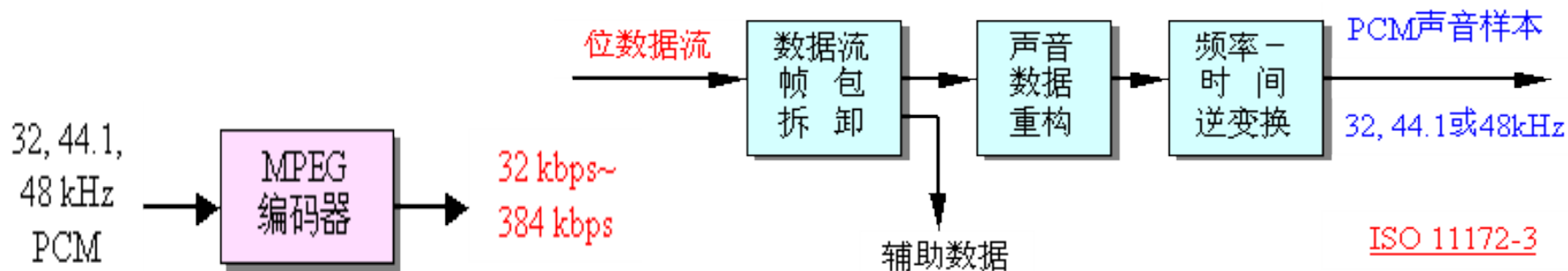
- **渐进传输编码**：可以将图像分层次按小波基展开，根据图像信号的性质及事先给定的图像处理要求确定到底要展开到哪一级为止，从而实现**渐进传输编码**。
- **多分辨率编码**：利用DWT具有平移和缩放的数学显微镜的功能进行**多分辨率编码**，适应于不同分辨率的图像I/O设备和不同传输速率的通信系统。

7 视频压缩MPEG标准概述

- MPEG (Motion Picture Experts Group) 是视频压缩的一个重要标准, MPEG标准面向问题是视频压缩、音频压缩以及多样压缩数据流的复合和同步问题。
- MPEG内容主要包括: MPEG视频组、MPEG音频组和MPEG系统组三部分。
 - MPEG视频组任务是研究视频信号压缩传输
 - MPEG音频组任务是研究音频信号压缩传输
 - MPEG系统组任务解决多道压缩数据流的同步及合成问题

- ISO/IEC 11172 MPEG-1标准 1992, 1.5MbpsAV编码应用于VCD、MP3。
- ISO/IEC 13818 MPEG-2标准 1994, 1.5Mbps~60MbpsAV编码应用于DVD、HDTV。
- ISO/IEC 14496 MPEG-4标准 1999 基于对象的视音频编码标准(多媒体信息内容的访问), 可根据不同的应用需求现场配置解码器。
- ISO/IEC 15938 MPEG-7标准 2001 多媒体内容描述接口(Multimedia Content Description Interface), 确定各种类型的多媒体信息的标准描述方法, 应用于数字图书馆、各种多媒体目录服务、广播媒体的选择及多媒体编辑等领域。
- ISO/IEC 18034 MPEG-21标准, 多媒体框架(Multimedia Framework)。适合于多种应用领域中各种类型用户对各种类型内容的传送和使用, 可通过大范围的网络和设备允许对多媒体资源的透明和增强使用。

- MPEG的音频**压缩编码**不依据波形本身的相关性和模拟人的发音器官的特性，而是利用人的听觉系统的特性来达到压缩声音数据的目的，这种压缩编码称为**感知声音编码** (perceptual audio coding, PAC)。MPEG Audio采用**PAC编码**和**Dolby AC-3编码**。PAC既可压缩声音数据又尽可能保留声音原有质量。
- MPEG Audio**解码器**对位数据流进行解码，恢复被量化的子带样本值以重建声音信号。由于**解码器无需心理声学模型**，只需拆包、重构子带样本和把它们变换回声音信号。

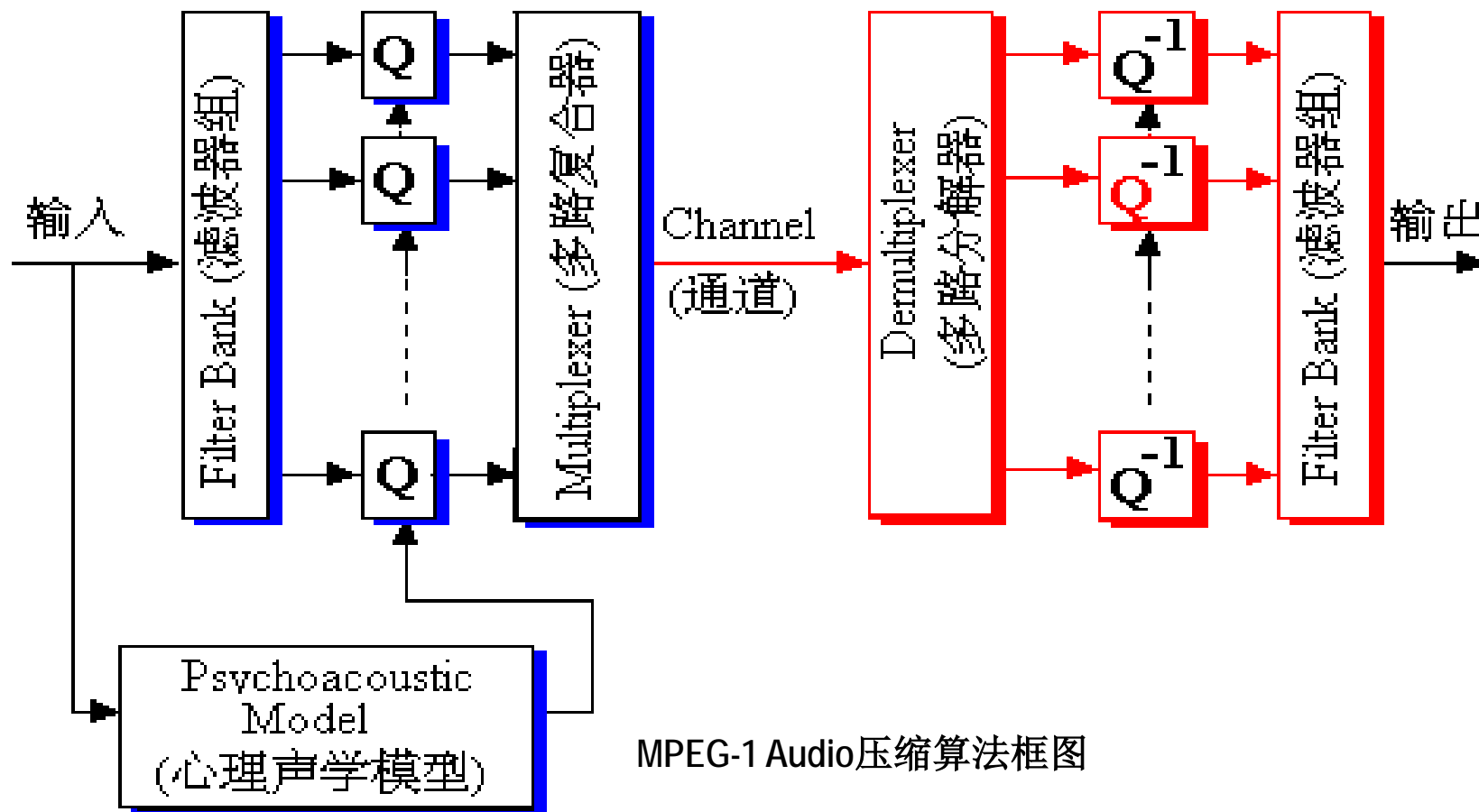


8.1 MPEG-1 Audio

■ MPEG-1 Audio压缩

编码器

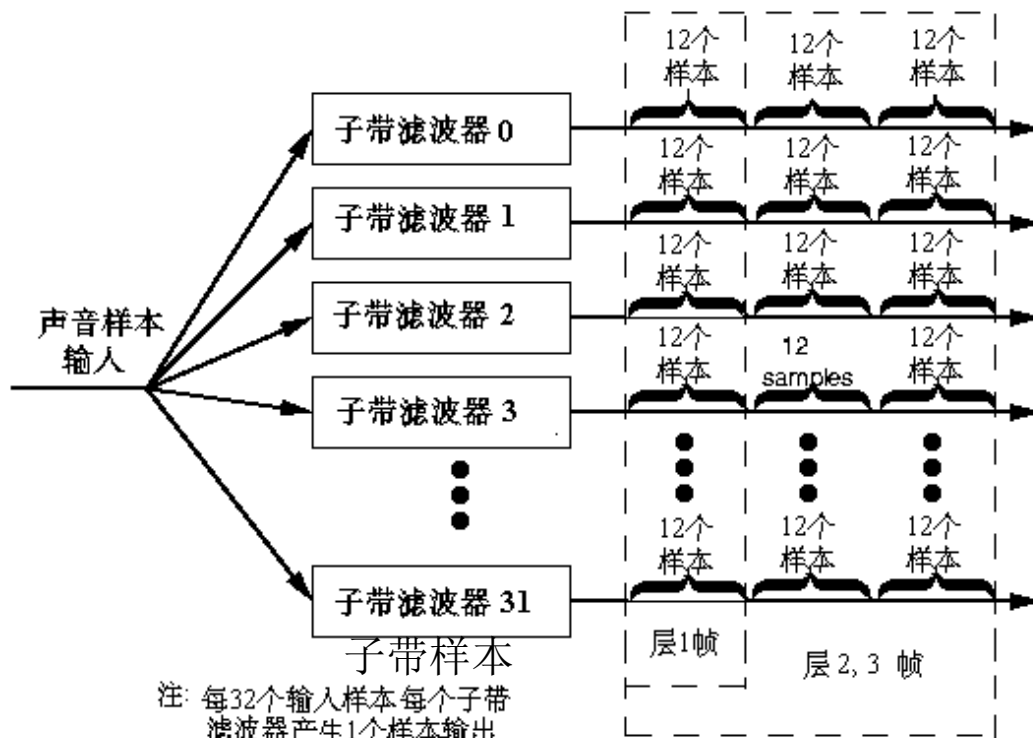
译码器



MPEG-1 Audio压缩算法框图




问题:为什么安非线性划分?

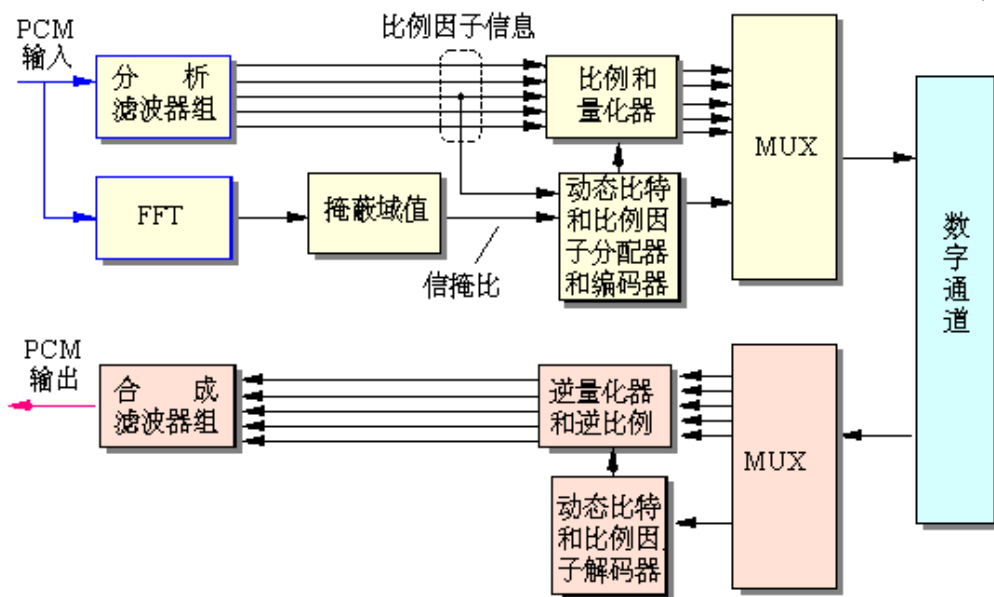
- MPEG的声音数据分成帧(frame)，层1每帧包含384个样本的数据，每帧由32个子带分别输出的12个样本组成。层2和层3每帧为1152个样本。MPEG编码器的输入以12个样本为一组，每组样本经过时间-频率变换之后进行一次比特分配并记录一个比例因子(scale factor)。



■ MPEG-1音频的三层压缩编码

- 1, 2层采用自适应声音掩蔽特性的通用子带综合编码和复合技术 (MUSICAM)
- Layer1: 编码器最为简单, 使用心理学模型**频域掩蔽**特性。编码器的输出数据率为384 kb/s, CR=4:1, 主要用于小型数字盒式磁带 (DCC)。
- Layer2: 的编码器的复杂程度属中等, 使用心理学模型的**频域和时域掩蔽**特性。编码器的输出数据率为256 kb/s~192 kb/s, CR=6:1~8:1, 其应用包括数字广播声音 (DBA)、数字音乐、CD-I和VCD等。

WAV		1614K
MP1		512K
MP2		292K



MP1-2编/解码结构

■ MPEG-1音频的三层压缩编码

- Layer3使用心理学模型的频域和时域掩蔽特性, 考虑立体声数据冗余, 使用Huffman编码, 高质量音乐信号自适应谱感知熵编码(ASPEC)和频域最佳编码(OCF), 并采用时域混叠取消(TDAC)技术的改进离散余弦变换(MDCT)细分子带频域, 消除多相滤波器混叠效应。

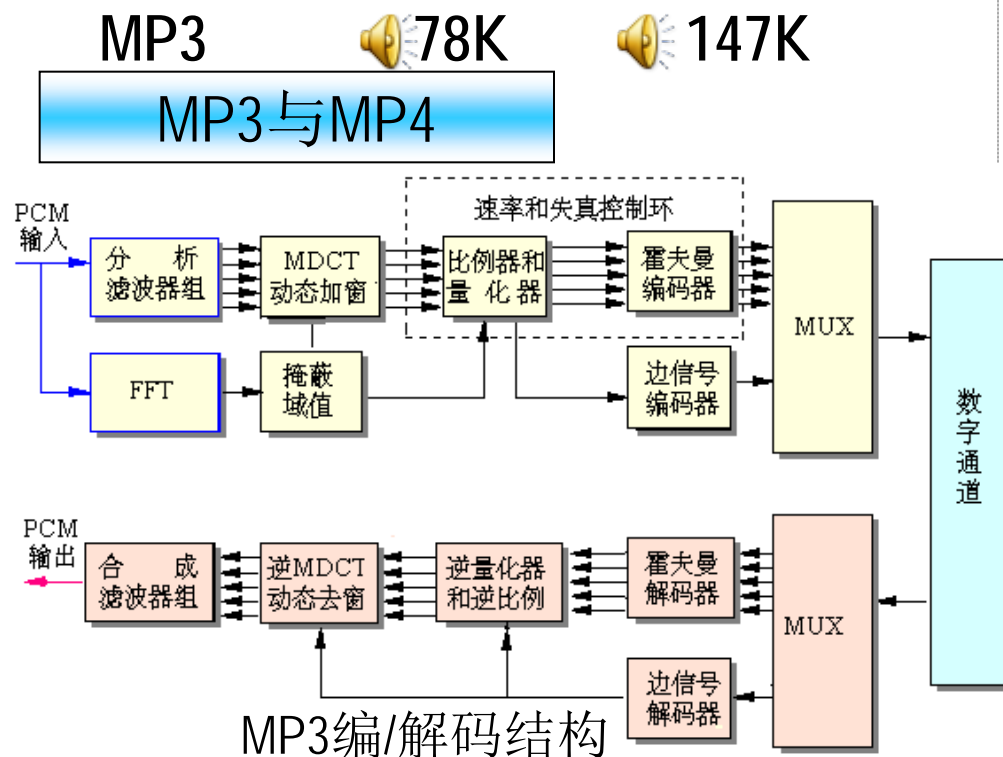
- Layer3编码器的输出数据率为64 kb/s, CR=10:1~12:1, 主要应用于ISDN上的声音传输。

- MP3提供多种CR和速率(固定速率CBR, 可变速率VBR)

$$\text{MDCT} \quad X_{ik} = 2 \sum_{n=0}^{N/2-1} x_{in} \cos\left[\frac{2\pi}{N}(n+n_0)(k+\frac{1}{2})\right], k=0, \dots, \frac{N}{2}-1$$

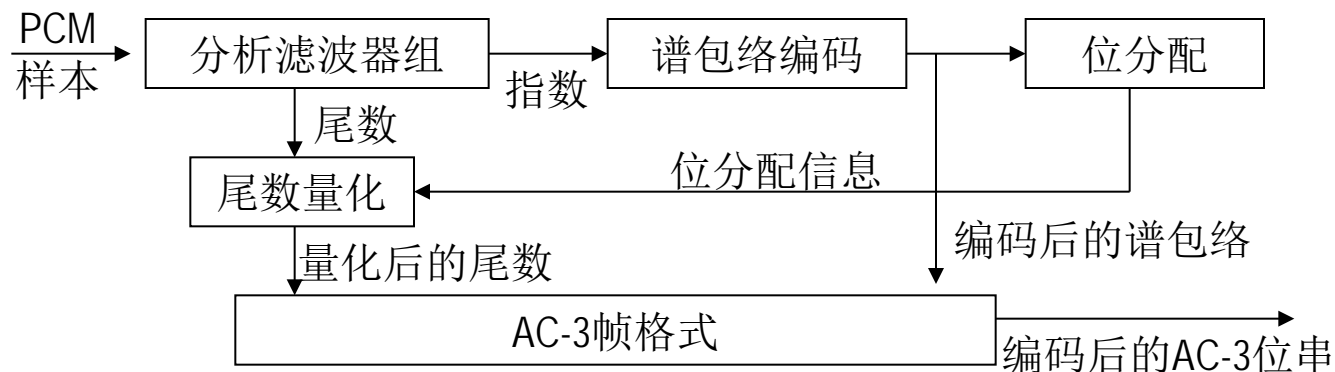
$$\text{IMDCT} \quad x_{in} = \frac{2}{N} \sum_{k=0}^{N/2-1} X_{ik} \cos\left[\frac{2\pi}{N}(n+n_0)(k+\frac{1}{2})\right], k=0, \dots, \frac{N}{2}-1$$

n = 样本号, N = 变换块长度, i = 块号, $n_0 = N/4 + 0.5$



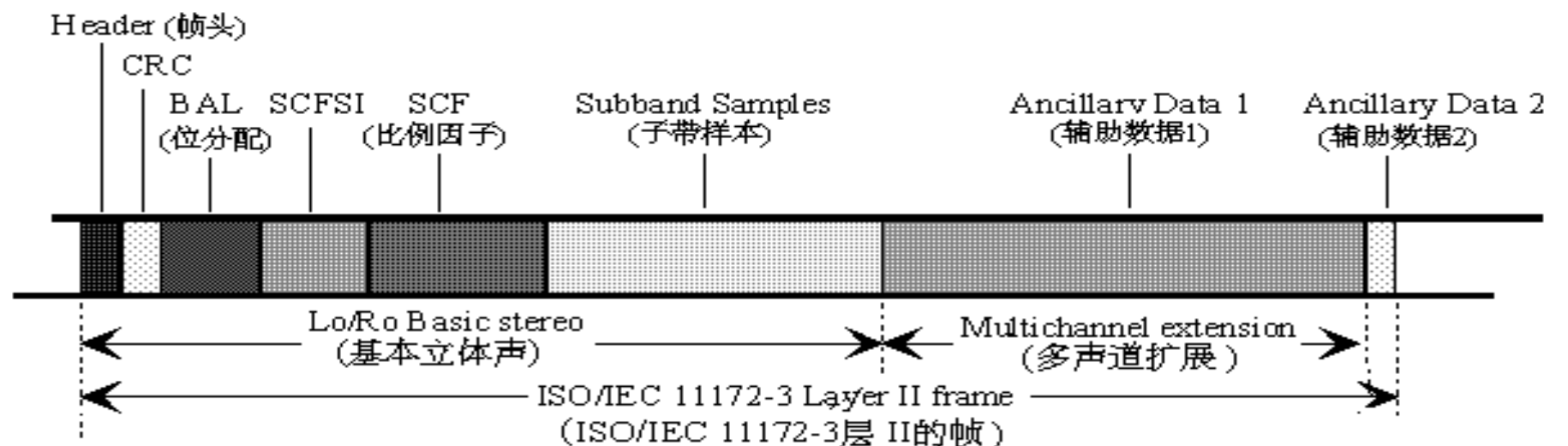
8.2 MPEG-2 Audio

- MPEG-2标准委员会定义MPEG-2多通道(MPEG-2 BC)和MPEG-2 AAC (Advanced Audio Coding)两种声音数据压缩格式, BC和MPEG-1兼容, AAC非后向兼容.
- AAC标准定义了三种配置
 - 基本配置(Main Profile): 音质最好, 算法复杂对计算机性能要求高。
 - 低复杂性配置(Low Complexity Profile): 采用瞬时噪声定形(Temporal Noise Shaping, TNS)滤波器音质中等。
 - 可变采样率配置(Scalable Sampling Rate Profile): 使用增益控制对信号作预处理, 不使用预测模块, TNS滤波器的级数和带宽也都有限制, 提供可变采样频率信号。

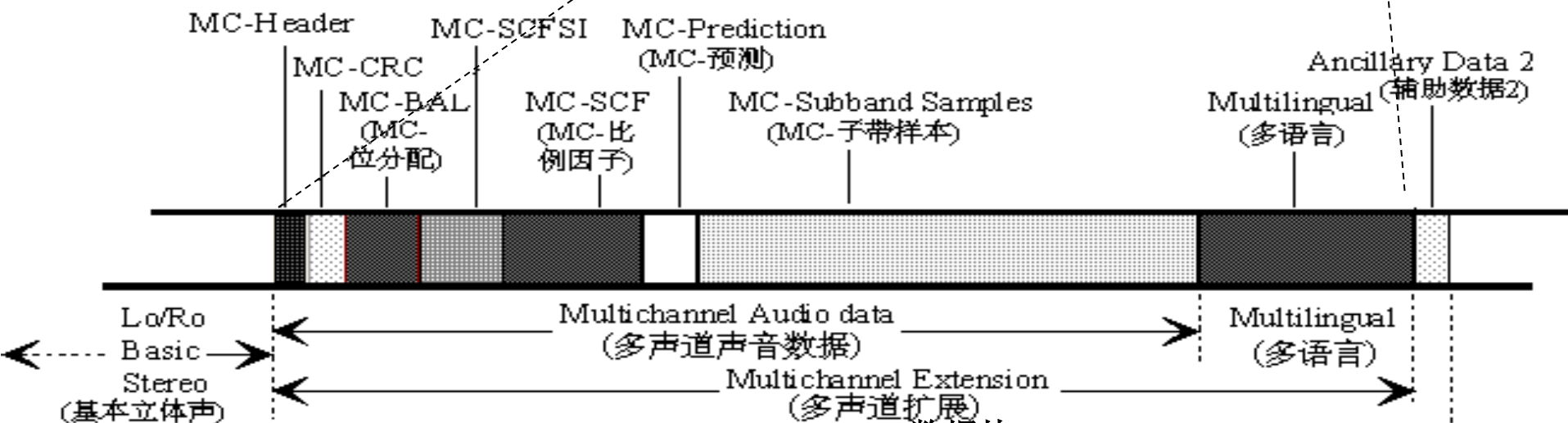


Dolby AC-3压缩编码算法框图

MPEG-2 AAC (advanced audio coding)



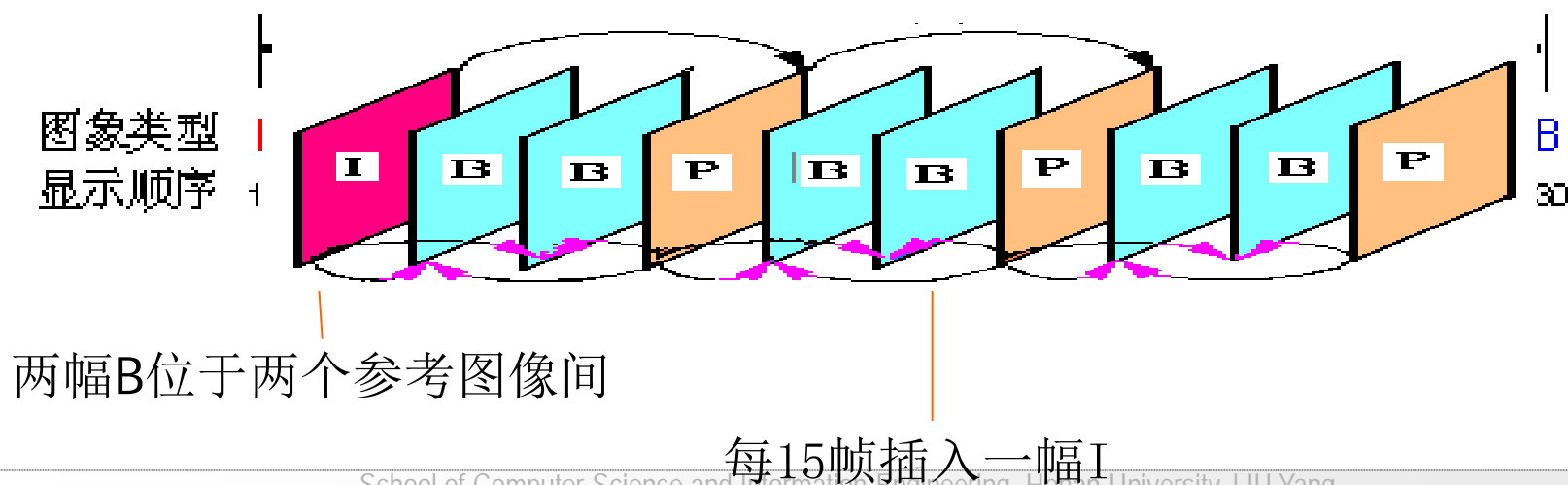
SCFSI: 比例因子选择信息
MC: 多通道
MC-SCFSI: MC比例因子选择信息



MPEG-2 Audio AAC数据块

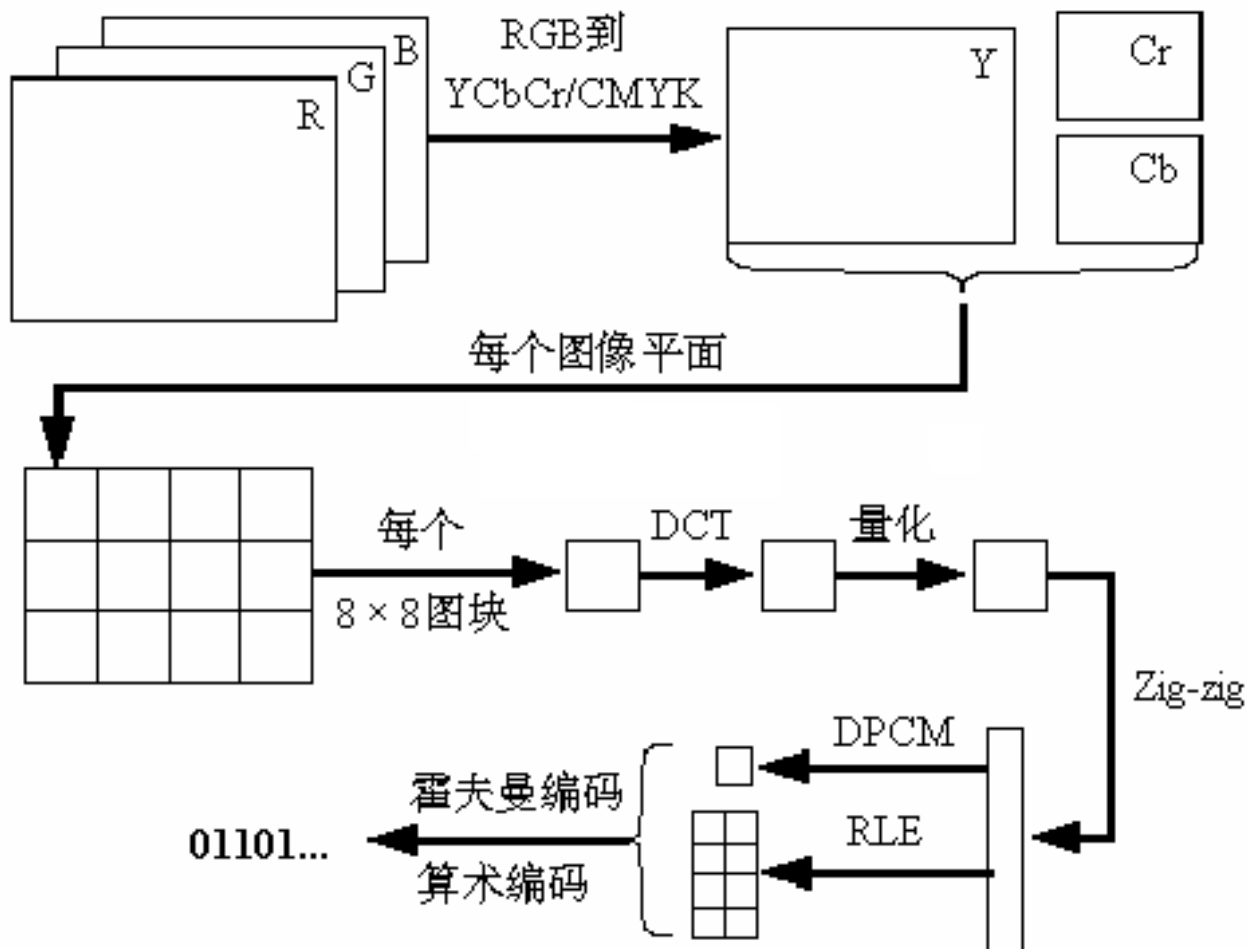
■ MPEG视频压缩的技术基础

- 时间域：块基运动补偿技术缩减画面间时间冗余度
- 空间域：JPEG域基变换压缩技术以缩减空间冗余度
- 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)、预测帧(P)和插补帧(B)。另有第四种类型帧是D帧，它是一种专用帧格式，仅仅用于实现快速查询中。



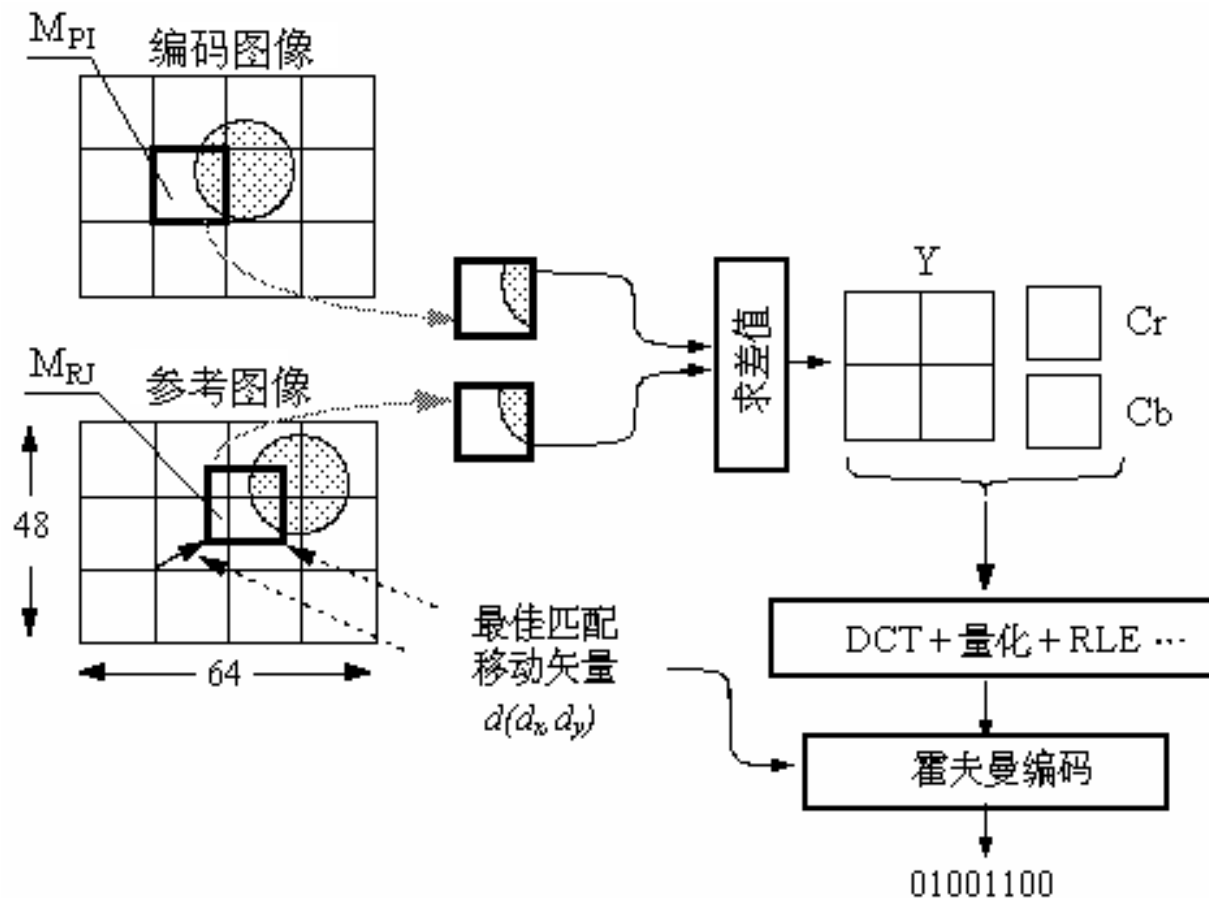
9.1 帧内图像I的压缩编码算法

- 帧内图像I 不参照任何过去的或者将来的其他图像帧，压缩编码采用类似JPEG压缩算法

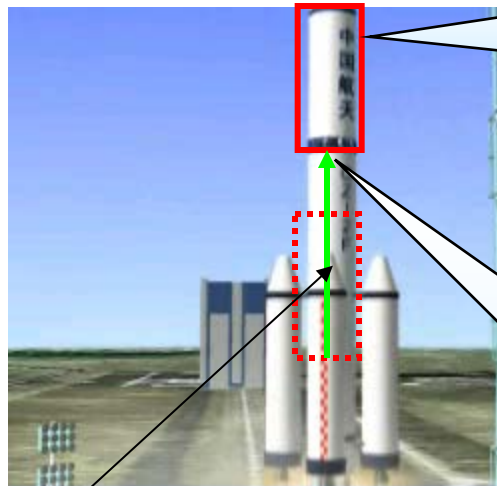


9.2 预测图像P的压缩编码算法

- 预测图像P使用两种类型的参数来表示：一种参数是当前要编码的图像宏块与参考图像的宏块之间的差值，另一种参数是宏块的移动矢量。



运动矢量与帧间预测



由于画面内容有连贯性,因此当前画面的内容可以看作是前面画面内容的位移(运动)。

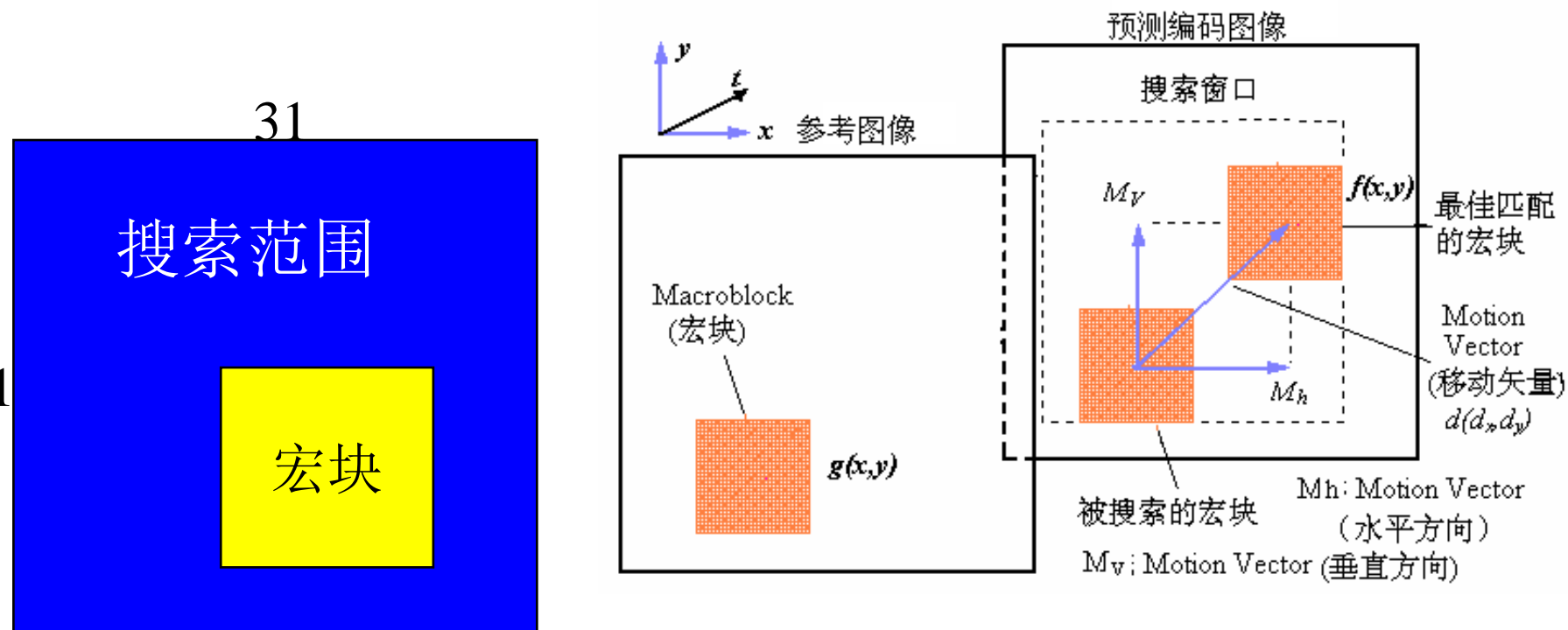
运动矢量



画面各处相应的位移量(幅度、方向)各有不同,因此画面应细分成块(宏块),以宏块为单位,找出两帧画面中相应宏块之间的位移量---运动矢量

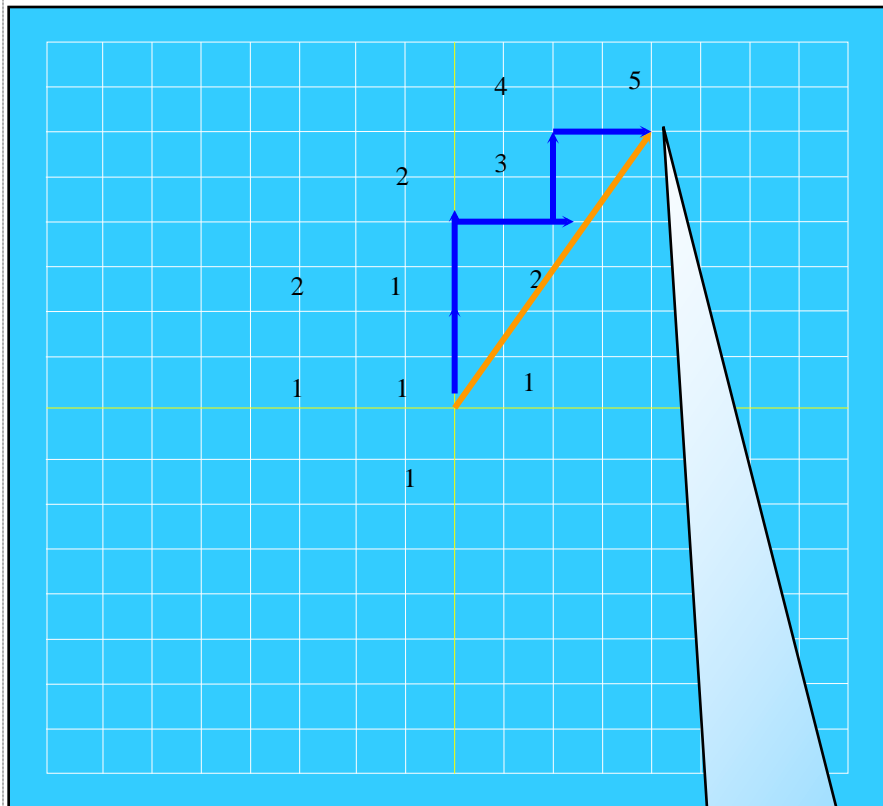
■ 预测图像P的压缩编码算法

- 搜索范围31, 63, 127, ..., 2047, 7种可选



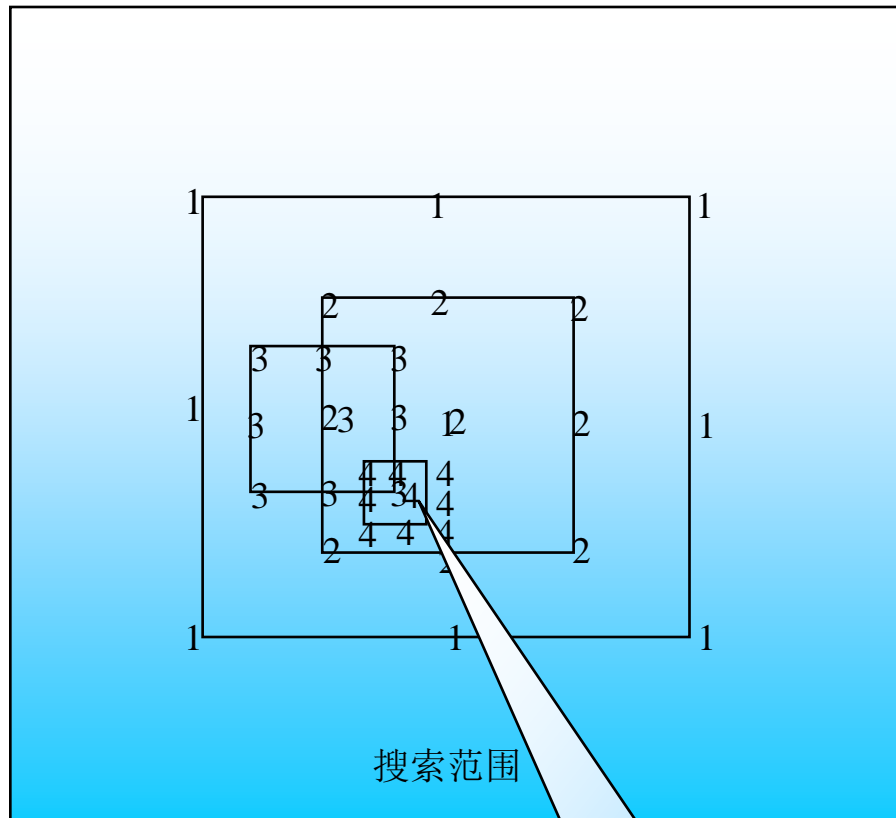
问题：16X16宏块搜索范围有多少种情况？

MPEG-1的运动矢量估计



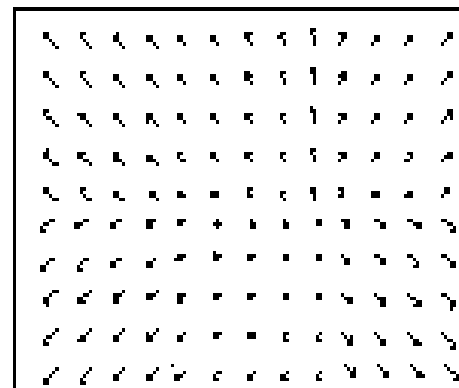
方向搜索法

沿着最小失真方向搜索，每一次检查5个搜索点，搜索步长固定。

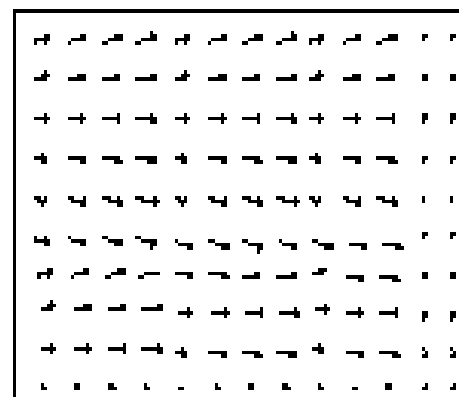


对数搜索法

每一步测试8个搜索点，并逐步缩小搜索步长



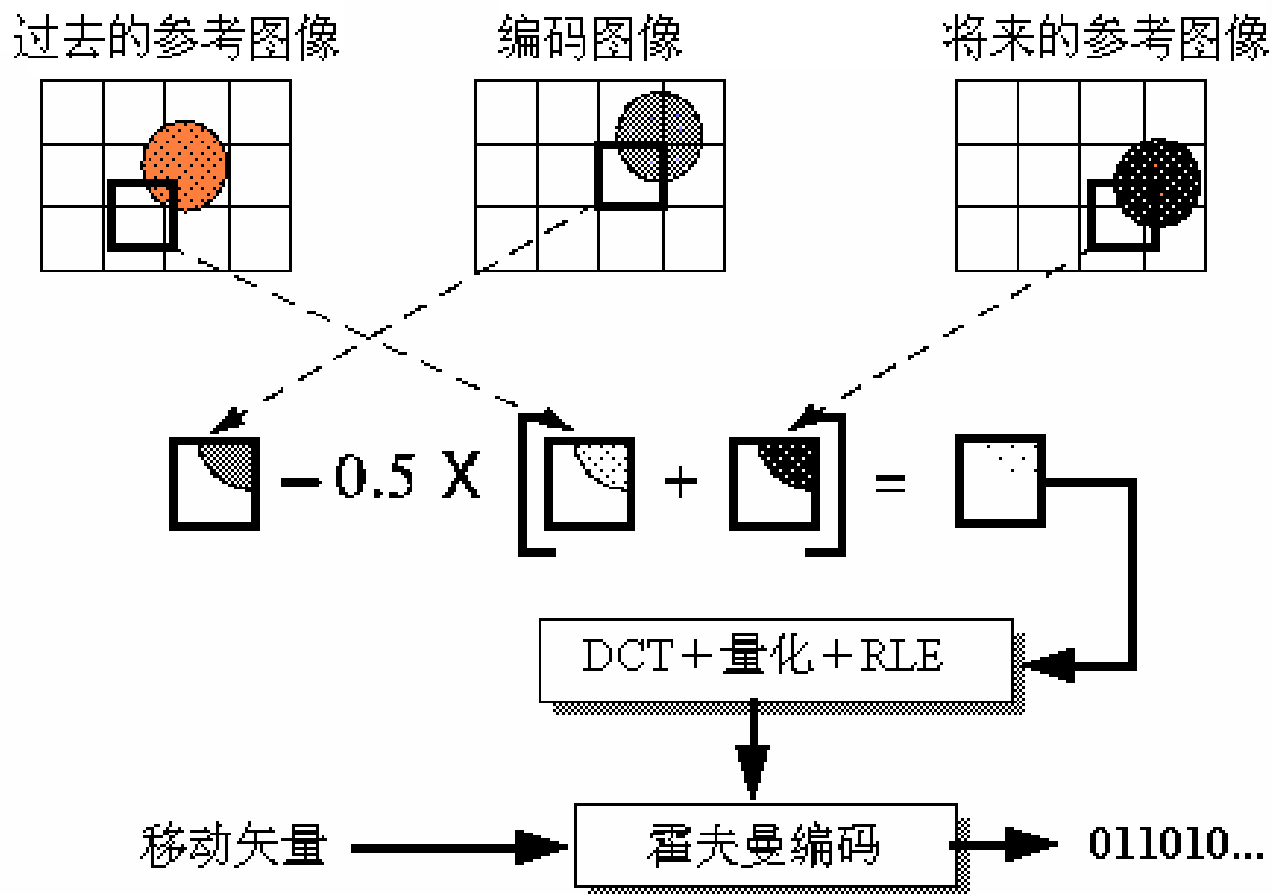
变焦的镜头及其运动场



平移的镜头及其运动场

9.3 插补图像B的压缩编码算法

■ 算法框图



MPEG三种图像的压缩后的典型值(比特)

图像类型	I	P	B	平均数据/ 帧
MPEG-1 CIF 格式 (1.15 Mb/s)	150 000	50 000	20 000	38 000
MPEG-2 601 格式 (4.00 Mb/s)	400 000	200 000	80 000	130 000

MPEG-1
352x240, PAL (25 fps)
Video Data Rate 1200 Kpbs
Audio Data Rate 224 Kb/s
Layer 2, 44.1 KHz, stereo

- MPEG-2视频体系与MPEG-1向下兼容，并满足高清晰度电视(HDTV)等应用领域对多媒体视频、音频通用编码需求。
- MPEG-2分辨率有低、中、次高、高不同档次，压缩方法从简单到复杂有不同级别。
- MPEG-2 Video 特色
 - Profile(配置或框架, 是算法集合)和Level(级别, 是能力的体现如分辨率、解码速度等)
 - 视频压缩编码的数据分层比特流结构
 - 对比特流制定了更高层语法结构的语义规范
- MPEG-2 Video 编码方式
 - 信噪比可变性(Signal-to-Noise Scalability)
 - 空间分辨率可变性(Spatial Scalability)
 - 时间分辨率可变性(Temporal Scalability)

MPEG-2 Video 配置等级和参数级

Profile Level	Simple	Main	SNR scalability	Spatial scalability	High
High		4:2:0 1920×1152×60 80 Mb/s I, P, B			4:2:0, 4:2:2 1920×1152×60 80 Mb/s I, P, B
High -1440		4:2:0 1440×1152×60 60 Mb/s I, P, B		4:2:0 1440×1152×60 60 Mb/s I, P, B	4:2:0, 4:2:2 1440×1152×60 60 Mb/s I, P, B
Main	4:2:0 720×576×30 15 Mb/s I, P	4:2:0 720×576×30 15 Mb/s I, P, B	4:2:0 720×576×30 15 Mb/s I, P, B		4:2:0 720×576×30 20 Mb/s I, P, B
Low		4:2:0 352×288×30 4 Mb/s I, P, B	4:2:0 352×288×30 4 Mb/s I, P, B		MPEG-2演示

- MPEG-4 Video算法的核心是**基于内容(content-based)**的编码和解码功能，也就是对场景中使用分割算法抽取的单独的**媒体对象(Media Object)**进行编解码。
- MPEG-4的MO类型有IO、VO、AO、TO、GO等，每一个MO与周围的对象或背景**独立编码**。
- MPEG-4以一种标准化的方法**重用**音视频对象AVO(Audio Visual Object)来描述和构造**视听场景(Plane)**。
- MPEG-4的场景描述是一棵**树**，叶节点是基本媒体对象，中间节点是**复合媒体对象**。因而内容的创作者可以构造复杂的场景，用户则可选择其中感兴趣的对象进行操纵处理。

由媒体对象组合而成的音视频场景

MPEG参考VRML语言开发了一种称为BIFS的 (Binary Format for Scenes) 语言, 用来对场景进行描述。

场景

(场景的逻辑结构)

人

2D背景

家具

演示

声音

教师

地球仪

讲台

下载的数据/
控制复合流

上载的数据/
控制复合流

视频复合
投影平面

假想的观察者

audiovisual objects

voice

sprite

audiovisual
presentation

2D background

3D objects

场景坐标
系

user events

audio
composer

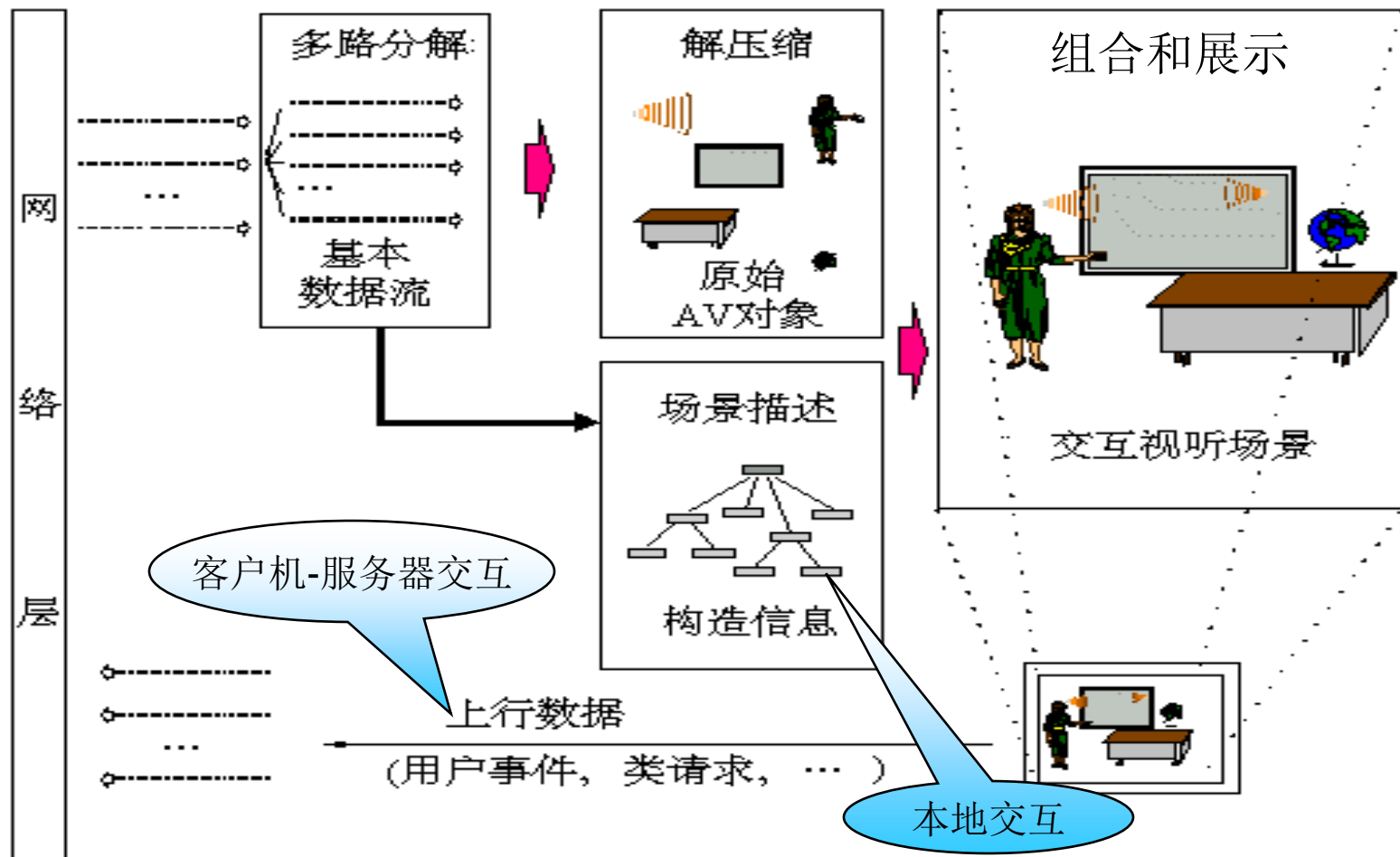
display

speaker

用户输入

思考：是否和矢量动画原理类似？

MPEG-4接收端的组成和交互操作



Generic MPEG-4 Coder



Video
Object
Plane

**Motion
(MV)**

**Texture
(DCT)**

bitstream
(位流)

(与MPEG-1/H.261类似)

MPEG-4 VLBV Core Coder



Video
Object
Plane

Shape

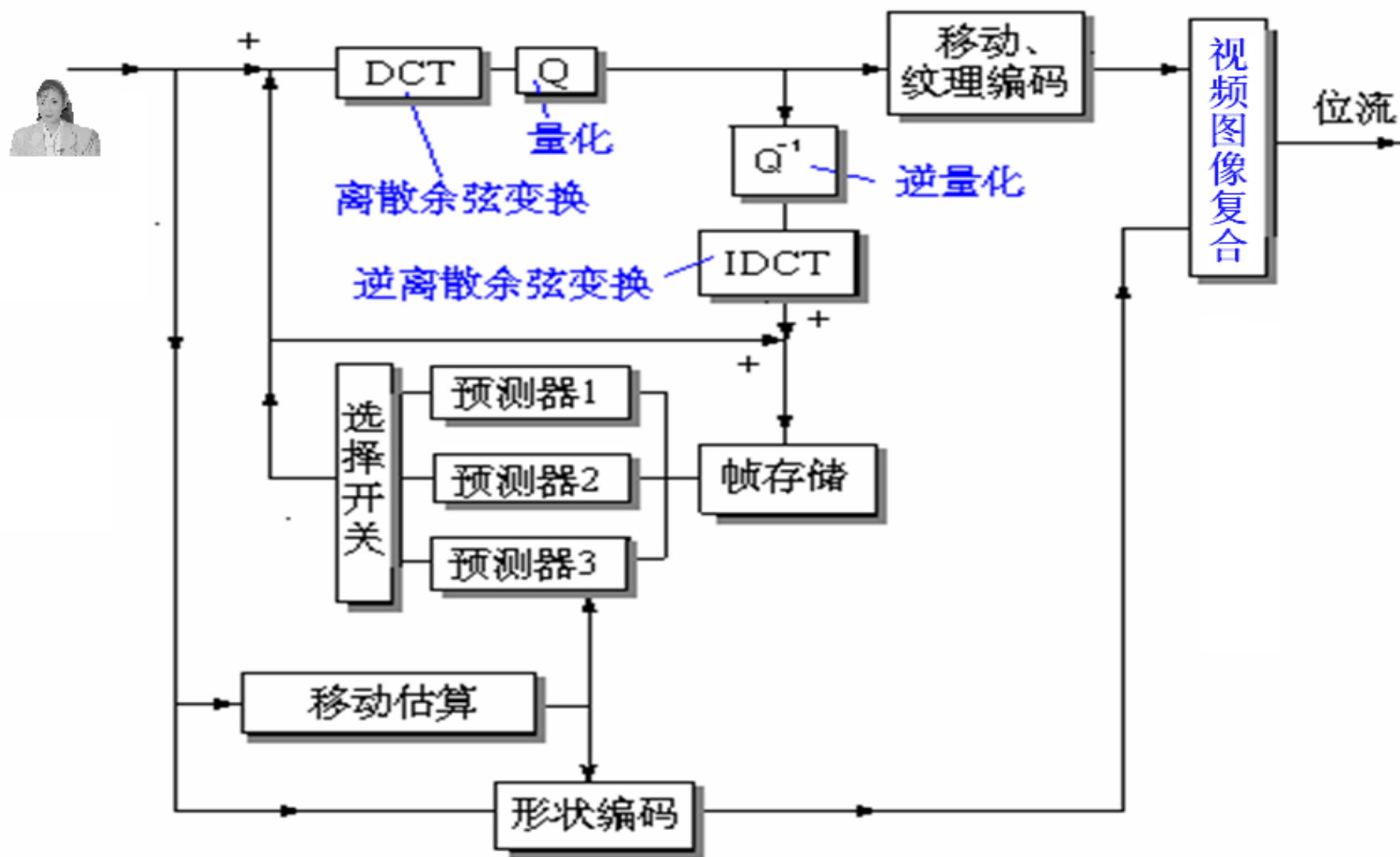
**Motion
(MV)**

**Texture
(DCT)**

bitstream
(位流)

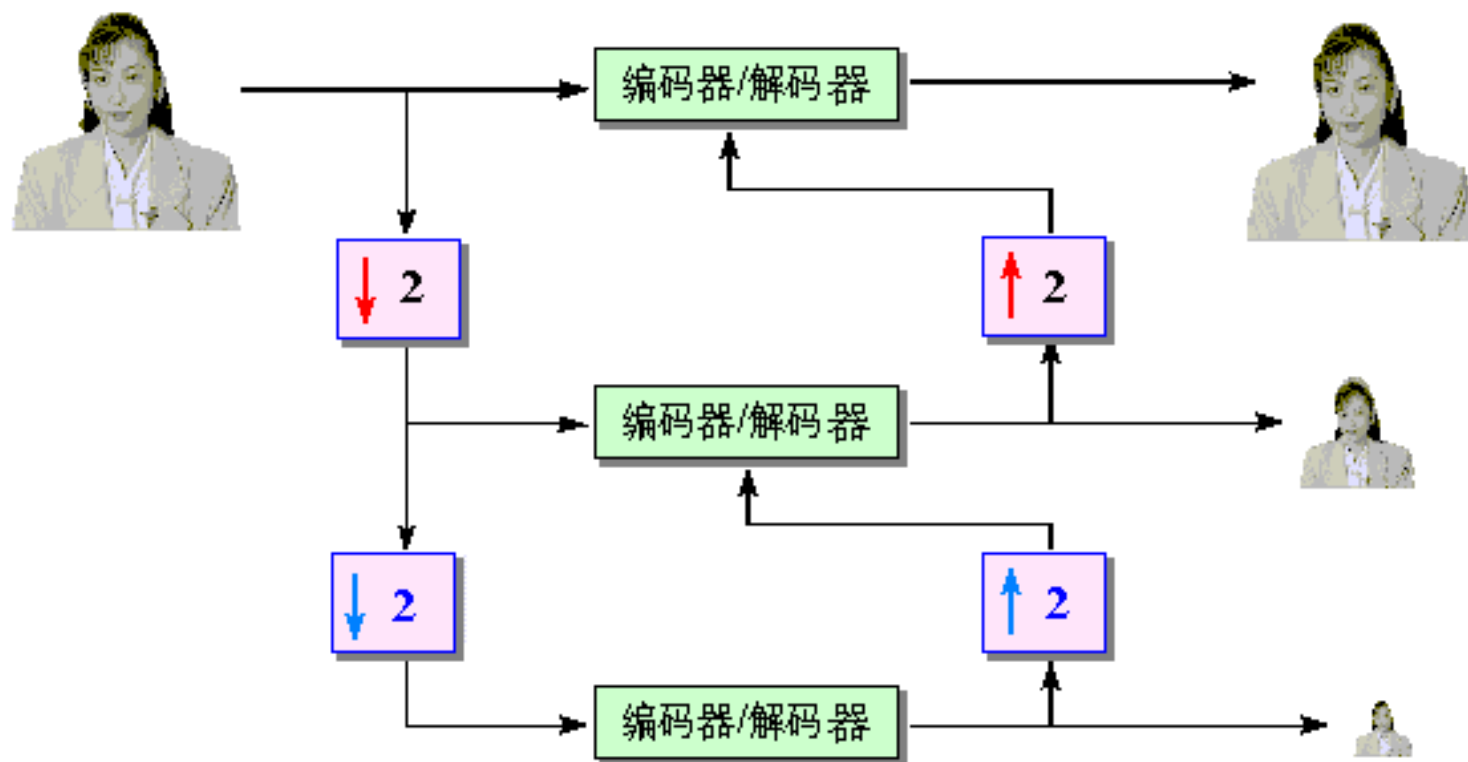
思考：如何对不规则形状图像编码？

MPEG-4 Video 编码方案

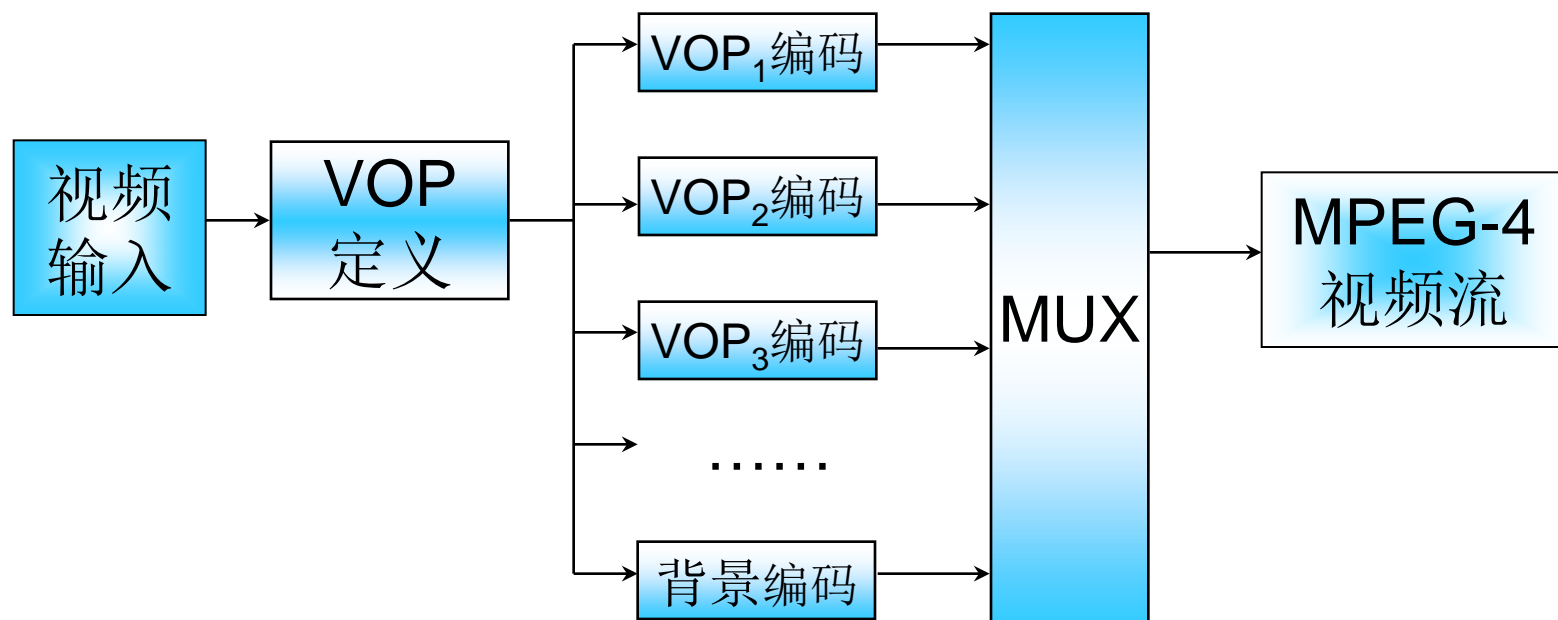


MPEG-4 Video编码器算法框图

MPEG-4 Video VOP分辨率可变编码



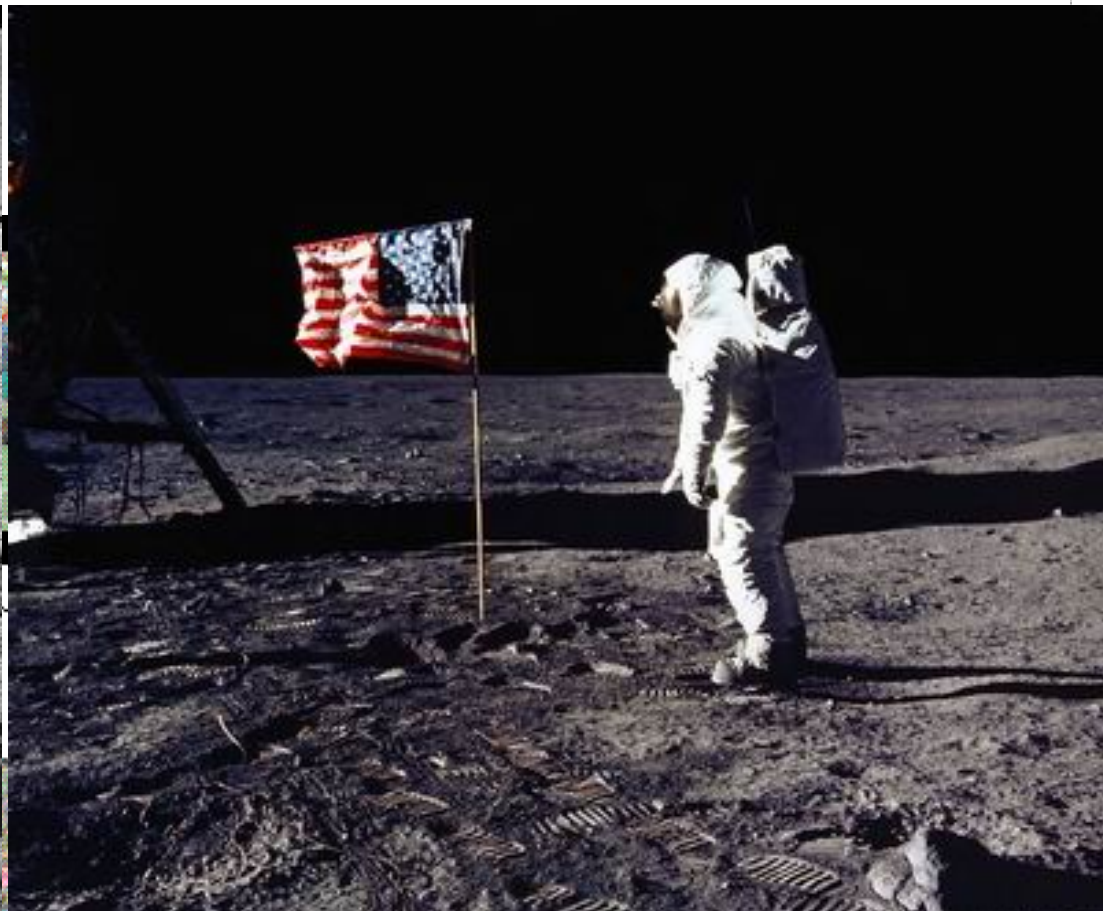
- MPEG-4在比特流中按照大致固定的距离插入再同步标志(Resync Marker)以增强容错性, 使用3层结构流式传送。



MPEG-4举例:任意形状视频对象编码的应用

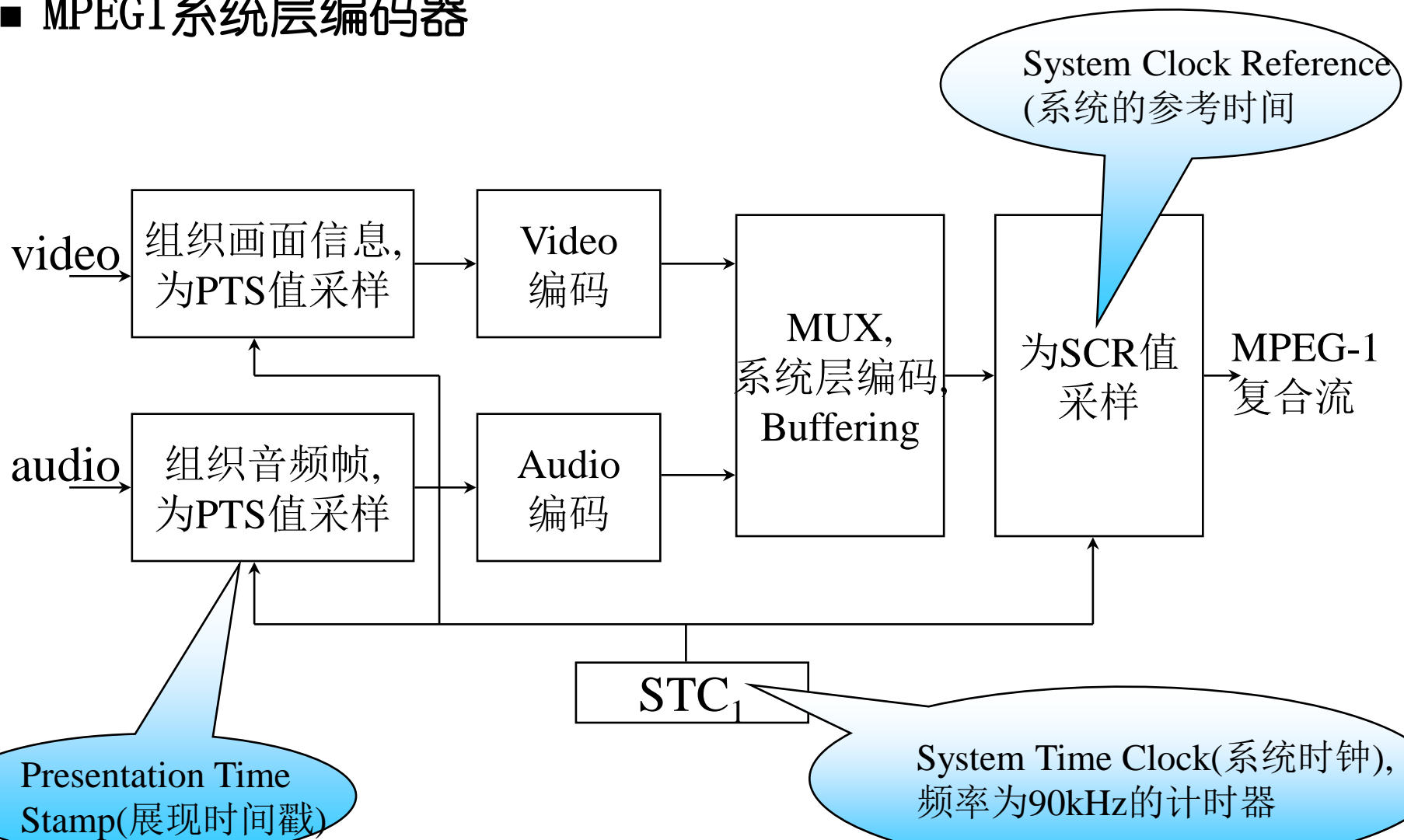
在编码前从一系列的

Multimedia technology



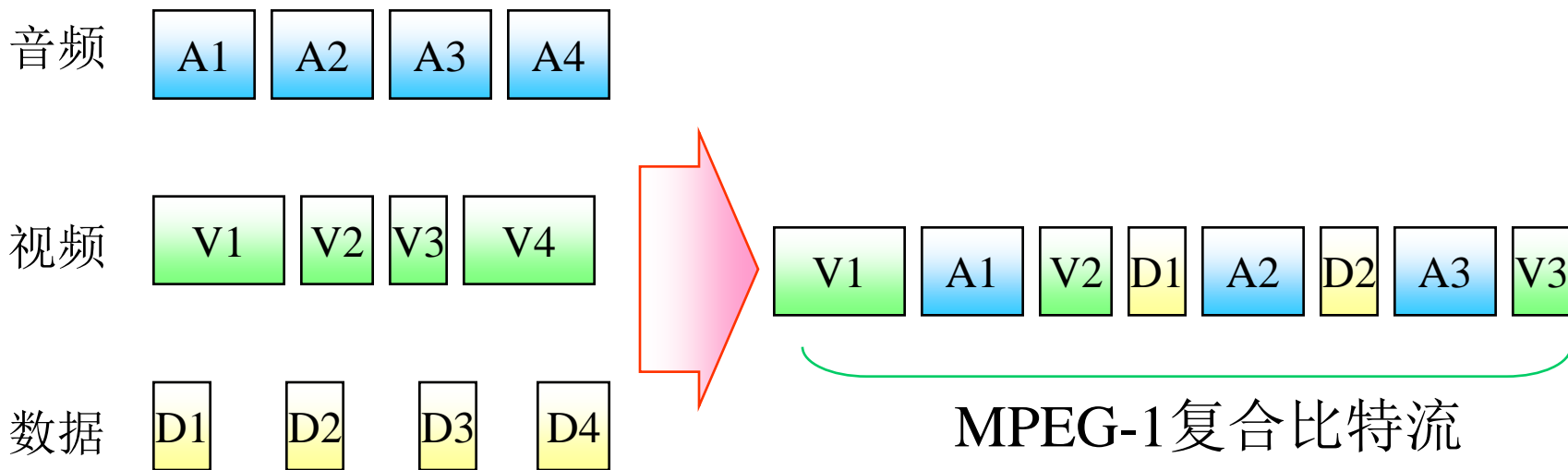
讨论：登月录象和照片是否可造假？

■ MPEG1系统层编码器

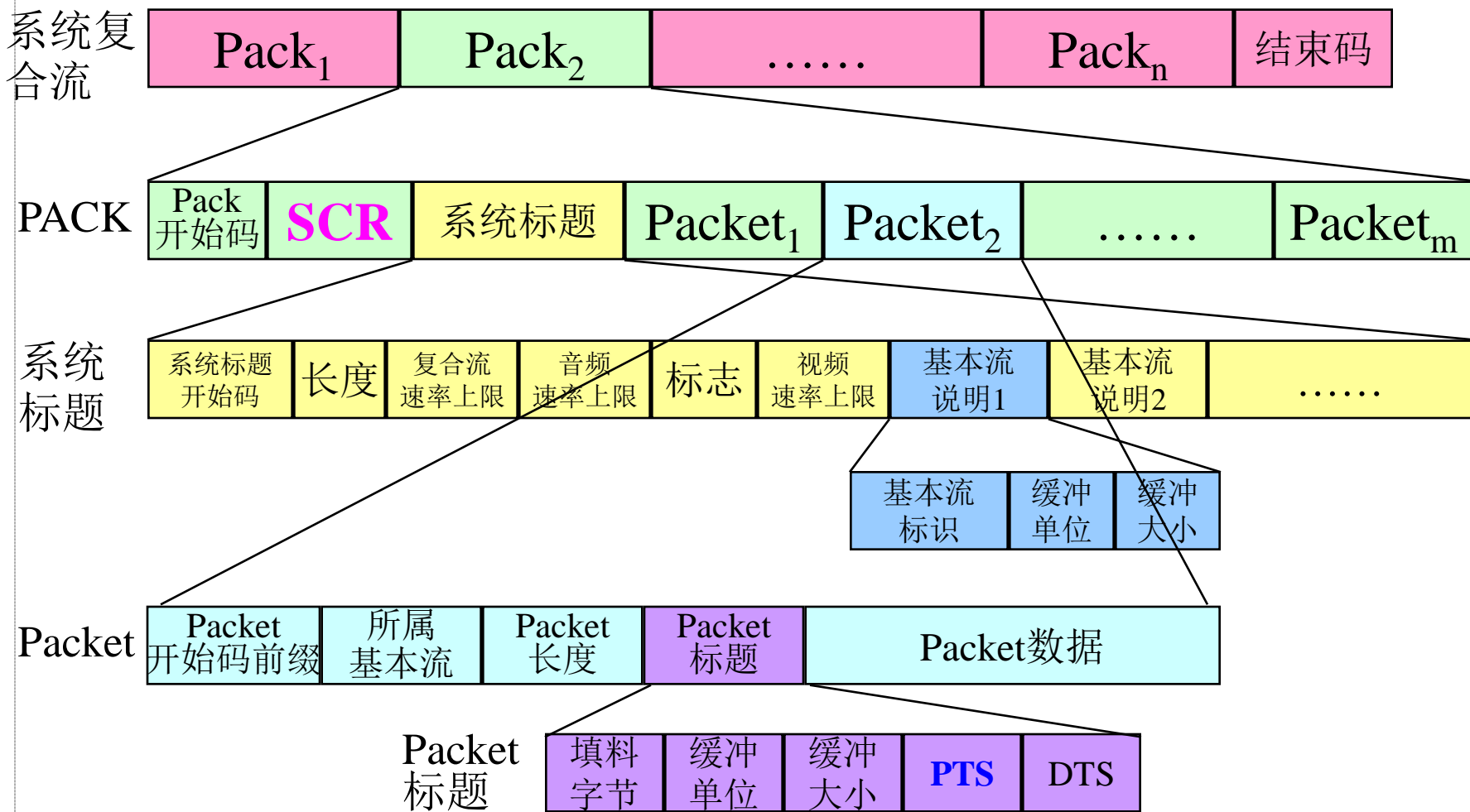


10.1 系统层的复合流

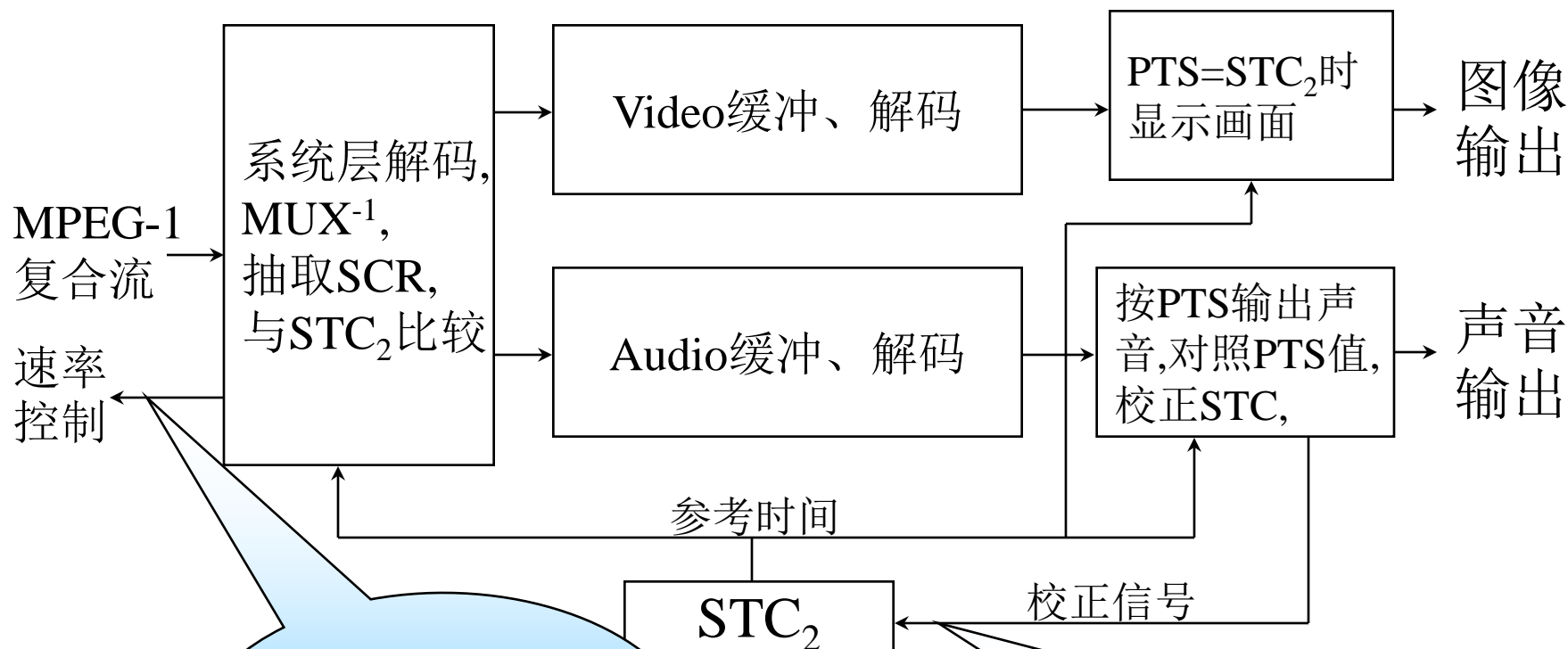
- 音频、视频流均可有多个，音频流 ≤ 32 ，视频流 ≤ 16 。
- 所有基本流均划分为“分组” (packet)，分组长度可变。
- 每个分组均有标题信息，用于指出所属的基本流。



系统复合流结构格式



10.2 系统层解码处理过程



把SCR值与STC₂比较, 发出反馈信号, 控制输入流的速度

解码器的本地STC₂值, 受控于音频PTS值, 不断进行更新与校正

ITU: H.261 → H.262 → $\begin{matrix} \text{H.263} \\ \text{H.263+} \end{matrix} \rightarrow \text{H.264}$

VCEG+MPEG JVT

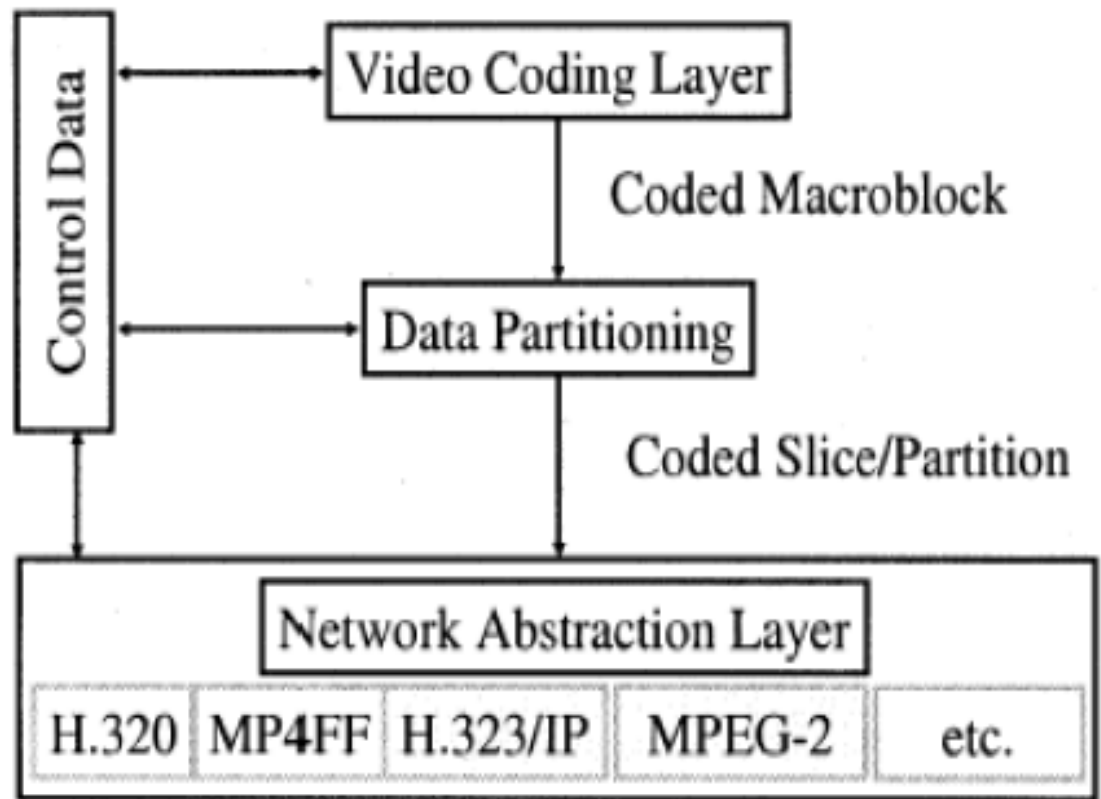
ISO/IEC: MPEG1 → MPEG2 → MPEG4 (Part2)
→ MPEG4(Part10) → MPEG7 → MPEG21

- H. 264/AVC 是ITU-T VCEG 和ISO/IEC MPEG共同开发的视频处理标准。H. 264在1997年ITU的视频编码专家组(Video Coding Experts Group)提出时被称为H. 26L, 在ITU与ISO合作研究后被称为MPEG4 Part10高级视频编码(MPEG4 AVC)或H. 264联合视频组(JVT, Joint Video Team)。
- H. 264/AVC相对以前的编码方法, 在图像内容预测方面提高编码效率, 采用可变块大小运动补偿、1/4采样精度运动补偿、加权预测等算法, 改善了图像质量。H. 264/AVC具有强大的纠错功能和各种网络环境传输的适应性。
- 编解码流程主要包括5个部分:
 - 帧间和帧内预测(Estimation)
 - 变换(Transform)和反变换
 - 量化(Quantization)和反量化
 - 环路滤波(Loop Filter)
 - 熵编码(Entropy Coding)。

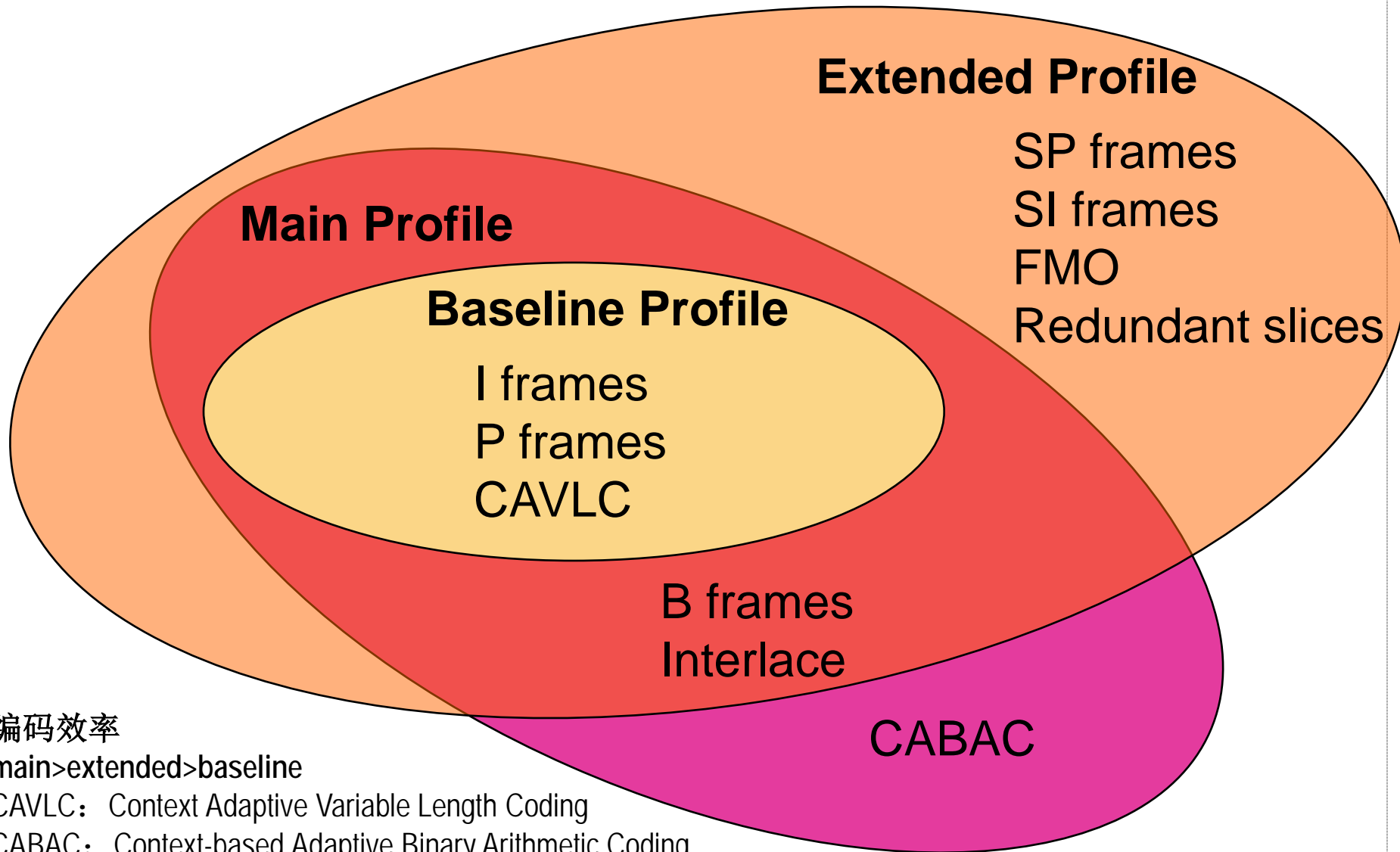
H. 264/AVC Structure

■ Consists of two layers:

- VCL—视频编码层
- NAL—网络提取层



H. 264 Profiles



编码效率

main>extended>baseline

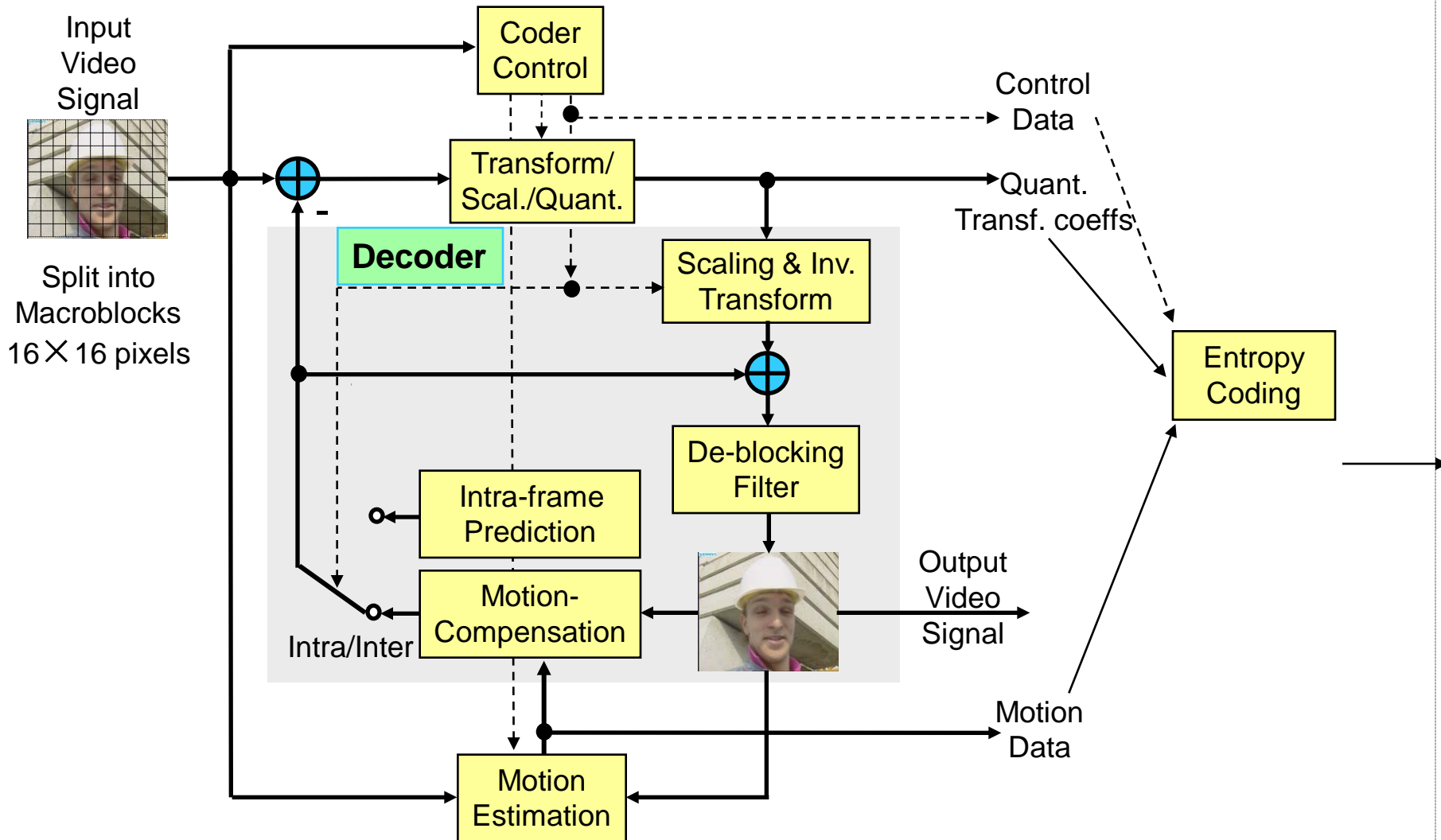
CAVLC: Context Adaptive Variable Length Coding

CABAC: Context-based Adaptive Binary Arithmetic Coding

■ Changes:

- Motion Compensation (运动补偿:多参考帧, 最多15帧;可变块大小, 树形结构运动补偿)
- Intra Prediction (帧内预测:亮度intra 4×4 有9种预测模式, intra 16×16 有4种预测模式;色度 8×8 有4种预测模式)
- Transform (变换编码:对 16×16 的亮度分量的16个 4×4 块进行DCT变换后, 组成一个 4×4 的luma DC块, 进行 4×4 的 Hadamard变换。对 8×8 色度分量的4个 4×4 块进行DCT变换后, 也将每个 4×4 块的DC系数提取出来, 组成一个 2×2 chroma DC块, 对其进行 2×2 的Hadamard变换。)
- Deblocking Filter (去块滤波)
- Entropy Coding (熵编码, mode = 0 时, 选择 Context Adaptive Variable Length Coding, CAVLC编码; mode = 1 时, 选择 Context-based Adaptive Binary Arithmetic Coding, CABAC编码)

Basic Macroblock Coding Structure



Deblocking Filter (去块滤波)

■ 出现方块效应的原因：

- 变换系数的量化过程相对粗糙，因而反量化恢复的系数有误差，造成边界上视觉不连续。
- 运动补偿，匹配不可能绝对准确，在复制块的边界上数据不连续。



Deblocking Filter (去块滤波)



1) Without Filter



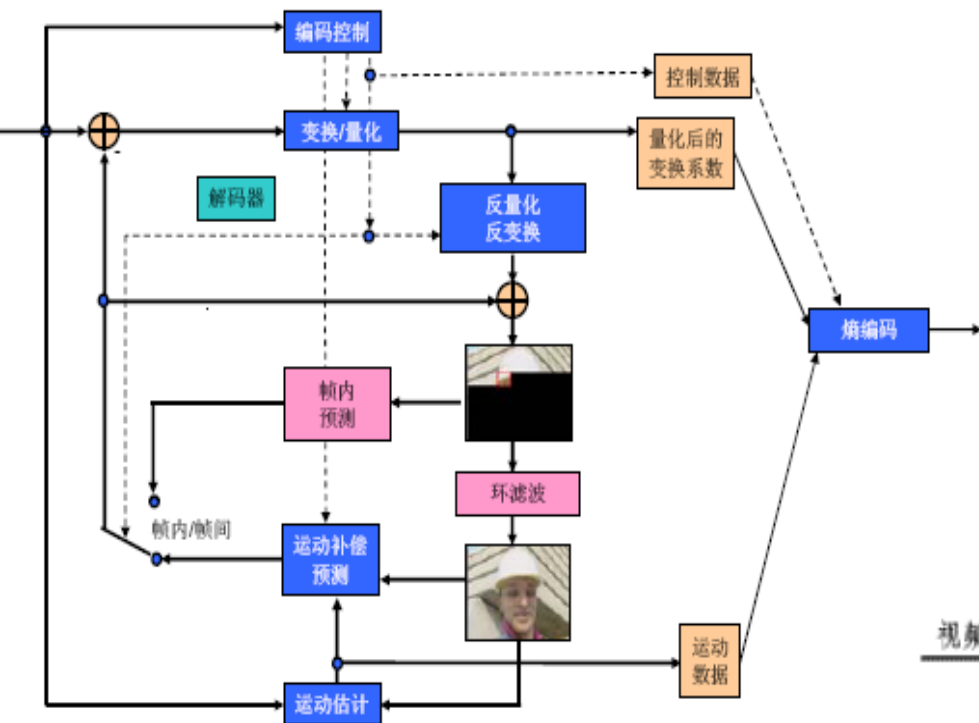
2) with H264/AVC Deblocking

- **数字音视频编解码技术标准工作组（AVS）**于2002年6月**国家信息产业部科学技术司**批准成立，AVS是我国具备自主知识产权的**第二代信源编码标准**。
- **AVS任务**：面向信息产业需求，联合国内外企业和科研机构，制订数字音视频的压缩、解压缩、处理和表示等共性技术标准，为数字音视频设备与系统提供高效经济的编解码技术，**服务于高分辨率数字广播、高密度激光数字存储媒体、无线宽带多媒体通讯、互联网宽带流媒体**等重大信息产业应用。（<http://www.avs.org.cn>）

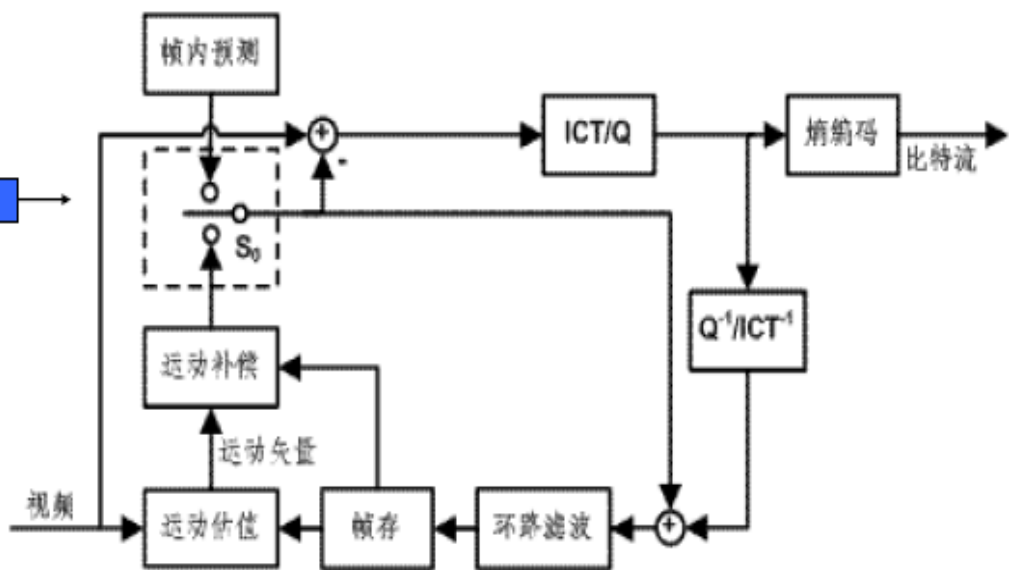
- AVS支持的芯片、软件、整机、系统、运营、服务等庞大的产业链。

AVS部分	国家标准计划号	小组草案 (WD)	工作组 草案(CD)	最终草案(FCD)	标准送 审稿 (FD)	国家标准 (GB)
AVS1-P1 (系统-广播)	20051304-T-339	2003.10	2003.12	2004.4	2004.8	
AVS1-P2 (视频-基准)	20032265-T-339	2003.10	2003.12	2004.4	2004.8	2006.2
AVS1-P2 (视频-增强)	20032265-T-339	2005.9	2006.3	2006.9	2006.12	
AVS1-P3 (音频-双声道)	20051305-T-339	2004.12	2005.3	2005.12	2006.4	
AVS1-P3 (音频-5.1声道)	20051305-T-339	2005.9	2005.12	2005.12	2006.4	
AVS1-P3 (移动)	20051305-T-339	2006.12	2007.6	2007.9	2007.12	
AVS1-P4 (一致性测试)	20051306-T-339			2006.3		
AVS1-P5 (参考软件)	20051307-T-339			2006.3		
AVS1-P6 (DRM)	20051308-T-339	2005.3	2005.12	2006.6		
AVS1-P7 (视频-移动1)	20051309-T-339	2005.3	2005.6	2005.9	2006.4	
AVS1-P8.1 (系统-IP)	TBD	2005.3	2005.9	2005.12	2006.3	
AVS1-P8.2 (系统-IP)	TBD	2006.1	2006.3		2006.3	
AVS1-P9.1 (文件格式)	TBD	2005.3	2005.9	2005.12	2006.3	
AVS1-P9.2 (文件格式)	TBD	2006.1	2006.3	2006.3	2006.3	

- AVS-P2是基于时空的预测与补偿、空域变换和基于统计的熵编码的混合编码。8X8块采用带PIT的二维整数余弦变换(ICT),环路滤波去除块效应,基于上下文的2D_VLC编码块变换系数。

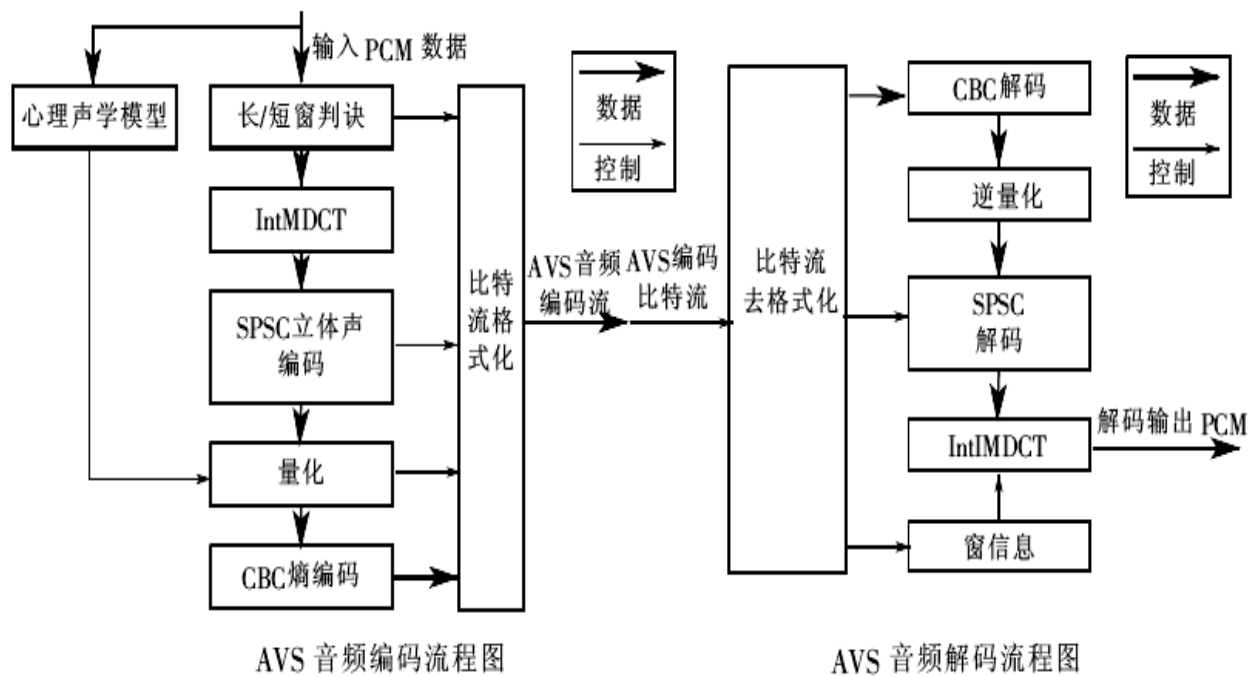


典型视频编码框架



AVS 视频编码器框图

- AVS音频编码模式策略，分高质量、高复杂度的Main Profile模式和可分级码率和编码质量的Scalability两种。编/解码的运算复杂度和编码质量高于MPEG AAC LC Profile.
- 支持8~96KHz采样的单/双声道PCM音频输入，编码器编码后输出每声道码率为16~96Kbps，CR=10-16:1

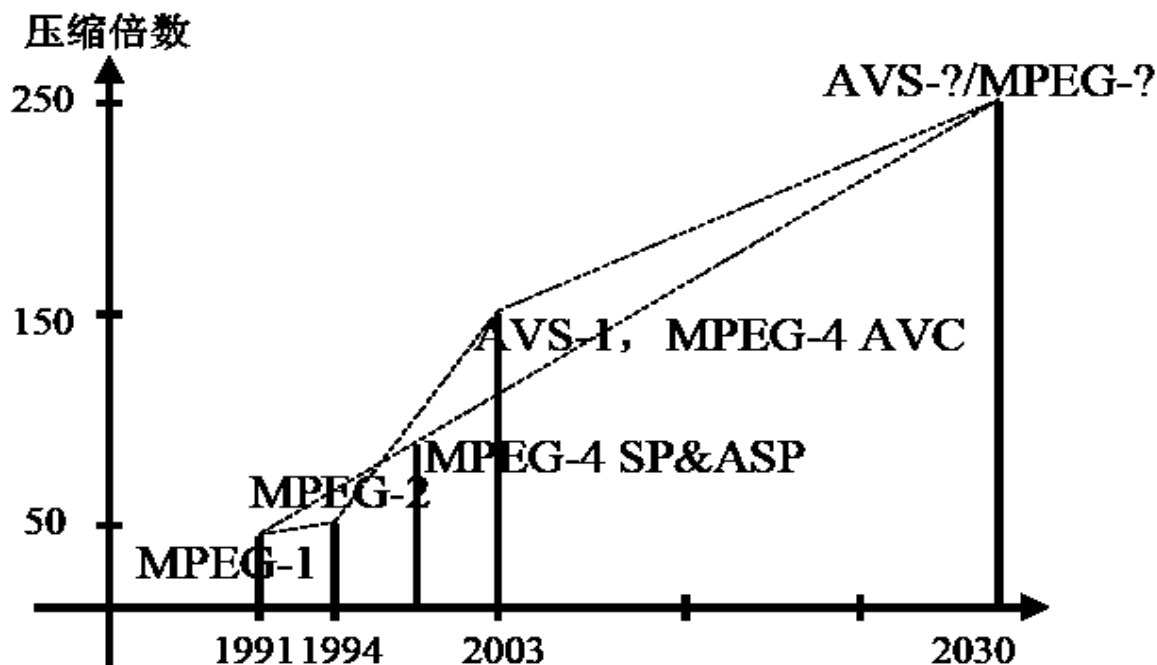


■ IntMDCT整数改进 DCT

■ SPSC(Square Polar Stereo Coding)立体声编码

■ CBC(Context-dependent Bitplane Coding)熵编码

- AVS编码复杂度是AVC的30%, 解码复杂度相当于AVC的70%。即编码性能基本相当, 实现复杂度明显降低。



AVS视频编码系统结构和编码效率

AVS演示

- 数据压缩的基本概念
- 各种压缩技术的分类
- 预测编码的基本原理
- 变换编码的基本原理(难点)
- 统计编码的基本原理
- 声音压缩标准
- 图像压缩标准(重点)
- 视频压缩标准(重点)

课外作业：检索有关AVS、AVC、MPEG-4和MPEG-7技术资料
写出报告