

媒体大数据压缩技术

第一章 多媒体数据压缩的基本概念

1. **多媒体技术**：是指计算机综合处理多种媒体信息使多种信息建立逻辑连接，集成为一个系统并具有实时的交互性。
2. **多媒体技术的特点**：多样性、集成性、交互性。
3. **数据压缩（信源编码）**：以尽可能少的数据表示信源所发出的信号，减少容纳给定信息集合的信号空间。

所谓信号空间，指被压缩对象，通常具有三方面的含义：

- **物理空间**：如存储器、磁盘、磁带或数据存储介质；
- **时间空间**：指传输给定消息集合所需要的时间；
- **电磁频谱空间**：传输给定消息所要求的频带等。

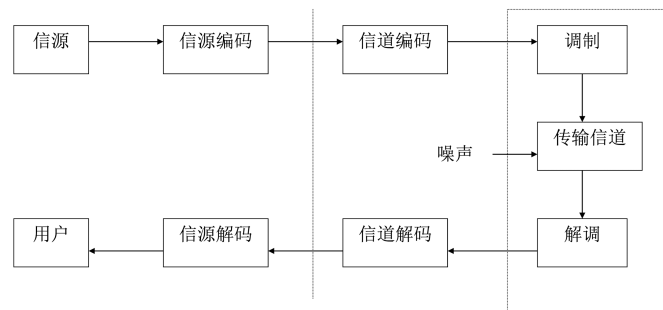
4. 数据压缩的评价指标

(1) **压缩比** ($CR = \frac{\text{原始数据量}}{\text{压缩后的数据量}}$)；

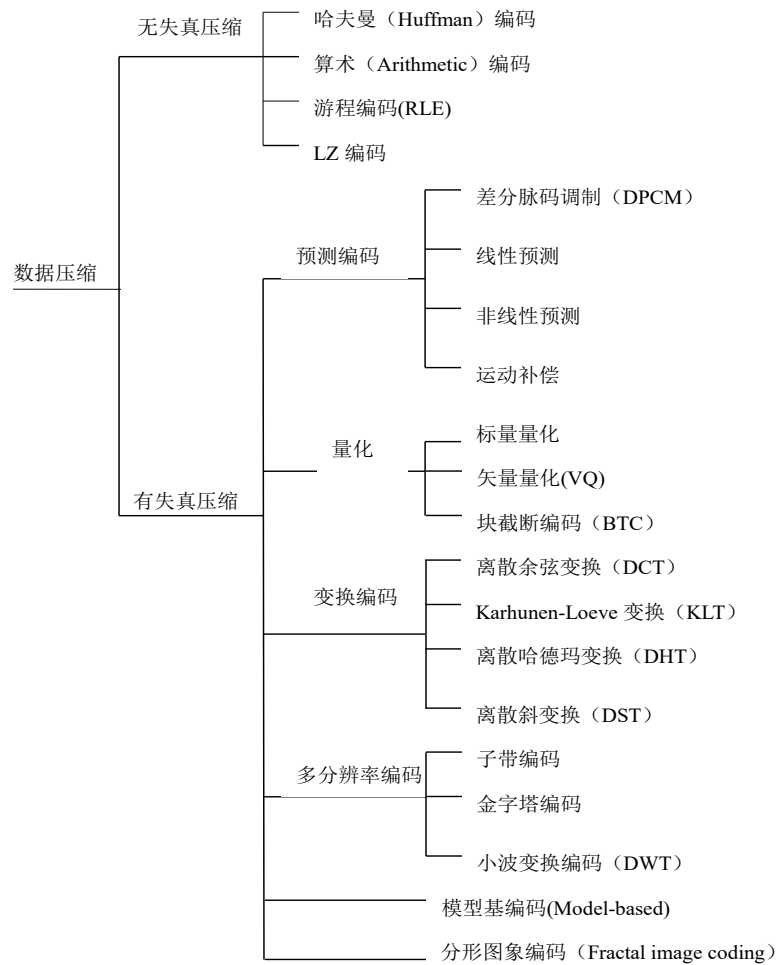
(2) **码率**（静止图像： $\text{bpp} = \frac{\text{压缩后的数据量}}{\text{原图像的像素数}}$ 视频：每秒所需的数据量）

(3) **图像质量**：峰值信噪比 ($\text{PSNR} = 10 * \lg \frac{(2^N - 1)^2}{\text{MSE}}$ MSE：原图与处理图像之间的均方误差)

5. 信息传输的典型框图：



6. **信源编码**：主要解决数据表示的有效性问题。压缩、扰乱、加密，力求用最少的数据传递最大的信息量。
7. **信道编码**：解决可靠性问题，即尽量使传输过程中不出错或少出错，即使出了错也可以纠错。
8. **数据压缩的应用场合及其必要性**：
 - (1) **传输**：通过压缩发送端的原始数据，并在接收端将压缩数据解码恢复，这样可以有效地减少传输时间、增加信道带宽；
 - (2) **存贮**：在存贮时压缩原始数据，而在使用时再解压缩，这样能够大大增加存贮介质的存贮量。
9. **压缩编码的方法及其分类**：无失真压缩（原始数据可由压缩数据完全恢复出来）和有失真压缩（原始数据不能由压缩数据完全恢复出来，恢复数据只是某种失真度下的近似）。



第二章 数据压缩的理论基础

1. 无失真压缩的理论极限：信源所含有的平均信息量。
2. 自信息量：某一符号出现的概率为 p_j ，这一符号的自信息量为 $I(a_j) = -\log_m p_j$ 。M 的取值决定自信息量的单位。2-比特；e-奈特；10-哈特。

3. 信源的熵

(1) 公式

$$H(X) = \sum_{j=1}^m P_j \cdot I(a_j) = - \sum_{j=1}^m P_j \cdot \log P_j$$

(2) 基本性质：非负性、确定性、严格上凸性、极值性（最大熵定理：所有的概率分布构成的熵，以等概率时最大；当某一符号概率趋近 1 时最小。即只要不是等概率就有压缩的可能。）

(3) 对数据压缩编码的理论意义：各种可能出现的符号的自信息量平均值。是无失真压缩的理论极限、无失真压缩的最低码率，如果对信源进行编码的码率小于信源的熵，则是有失真的。

4. 联合概率： $P(a_i, b_j)$ 信源发出 a_i ，编码输出 b_j 。
5. 条件概率： $P(a_i/b_j)$ 已知编码输出 b_j 估计信源发出 a_i 的概率； $Q(b_j/a_i)$ 已知信源发出 a_i 估计编码输出 b_j 的概率。
6. 条件自信息量：

$I(a_i/b_j) = -\log P(a_i/b_j)$ ，已知编码输出 b_j 对猜测信源发出 a_i 的不确定度；

$I(b_j/a_i) = -\log P(b_j/a_i)$ ，已知信源发出 a_i 对猜测编码输出 b_j 的不确定度。

7. 条件熵：

$H(Y/X) = \sum_i \sum_j p(a_i, b_j) \cdot I(b_j/a_i)$ 表示已知编码输入集为 X 时，编码输出集为 Y 所剩余的不确定度；

$H(X/Y) = \sum_i \sum_j p(a_i, b_j) \cdot I(a_i/b_j)$ 表示已知编码输出集为 Y 时，编码输入集为 X 所剩余的不确定度。

8. 联合熵： $H(X; Y) = -\sum_i \sum_j p(a_i, b_j) \log p(a_i, b_j)$ 表示输入为 X，输出为 Y 时，整个系统所具有的不确定程度。
9. 互信息量： $I(X; Y) = H(X) - H(X/Y)$ 在已知 Y 后，所获得的关于 X 的信息量。
10. 平均失真： $D(Q) = \sum_{i=1}^m \sum_{j=1}^n P(a_i, b_j) \cdot d(a_i, b_j) = \sum_{i=1}^m \sum_{j=1}^n P(a_i) \cdot Q(b_j/a_i) \cdot d(a_i, b_j)$ ，其中 $d(a_i, b_j)$ 是某种失真的度量， $Q_D = \{Q | D(Q) \leq D\}$ 。
11. 率失真函数 $R(D)$

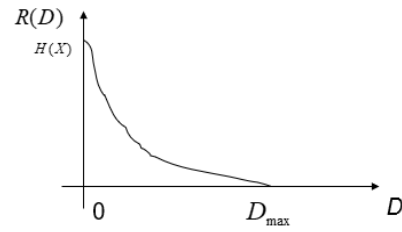
(1) 定义：研究在限定失真下为了恢复信源符号所必需的编码率，在 Q_D 范围内寻找再现信源信息最起码的平均互信息量。

$$R(D) = \min_{Q \in Q_D} I(X; Y)$$

(2) 理论意义：率失真函数是在允许失真为 D 的条件下，信源编码给出的平均互信息量的下界。当编码码率小于 $R(D)$ 时，平均失真大于 D【有失真时的信源编码的逆定理】。

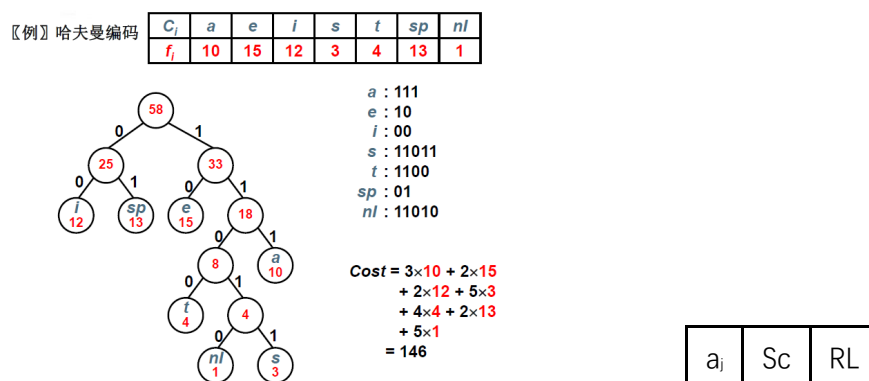
12. 率失真函数的性质:

- (1) $D < 0$, $R(D)$ 没有意义
- (2) 失真为 0 的编码记作 $R(0)$, $R(0) = H(X)$
- (3) 存在 D_{max} , 使 $D > D_{max}$, $R(D) = 0$
- (4) $0 < D < D_{max}$, $R(D) > 0$, 且是连续的下凹函数



第三章 几种常见的无失真数据编码方式

1. **哈夫曼编码**: 各码字长度按概率大小逆序排列得到的平均码长最短, 但其缺点在于: 要统计概率, 需要存储和传输码表。



2. **游程编码(RLC)**: 用一个符号值或串长代替具有相同值的连续符号, 使符号长度少于原始数据的长度。只在各行或者各列数据的代码发生变化时, 一次记录该代码及相同代码重复的个数, 从而实现数据的压缩。用二进制记录串的字符、串的位置以及串的长度就能换算出原来的数据流。游程长度编码(RLC)就是用二进制码字给出上述信息的一类方法。

3. **MH/MR 编码**: 一种针对于二值图像的压缩方法

(1) **MH 编码**: 改进的哈夫曼编码, 使用串长的出现的概率编成哈夫曼码表, 结尾码和组合基于码结合。

(2) **MR 编码**: 按行扫描, 若干行分为一块, 一块的第一行(参考行)是 MH, 后几行(编码行)是 MR 编码。按照迁移像素 a_1 与 b_1 的相对位置, 分成通过模, 水平模和垂直模。

4. **算术编码**

(1) **原理**: 算术编码就是将被编码的信息表示成实数 0 和 1 之间的一个间隔, 信息越大, 编码表示它的间隔就越小, 表示这一间隔所需要的二进制位就越多。

(2) **特点**: 算术编码分为固定方式和自适应方式两种编码, 选择不同的编码方式直接影响到编码效率; 算术编码的方法不必预先统计概率, 适合于无法知道信源字符概率分布的情况; 当信源概率比较接近时, Huffman 编码的效率不高, 算术编码要好于 Huffman; 算术编码的实现方法复杂于 Huffman 方法。

5. **LZ 编码(基于字典的无失真编码)**

(1) **基本原理**: LZ 编码是一种基于字典的编码形式, 分为字典窗口和缓冲区窗口, 读入缓冲区后将字符和字典窗口中的进行匹配, 记录缓冲区中匹配的位置、长度以及匹配段后的第一个字符。

(2) **实现方法**

- A. 首先定义滑动文本窗;
- B. 初始化文本框, 将码流读入缓冲区;
- C. 将 Buffer 中的内容和字典窗口中内容进行比较, 如果有匹配部分则按 (S_k, L_k, T_k) (位置, 匹配段长度, 匹配段后的第一个符号) 输出代码, 尚未出现的字符编码为 $(0, 0, T_k)$;
- D. 编码后的数据量进入字典窗口中, 将 L_k+1 个新字符读入缓冲区; 回到 C 直到结束。

(3) **商业压缩软件**: PKZIP, WinZIP, gzip, Arj

第四章 量化编码

1. **量化**：给定幅值 x ，将 x 的值域划分成为 K 个互不重叠的间隔 S_k ， $S_k = \{x | x_k < x < x_{k+1}\}, k=1,2,3,\dots,K$ ； x 落入 S_j ，则用 y_j 来表示。
2. **标量量化**
 - (1) 均匀量化：量化步长相等
 - (2) 非均匀量化：对大幅值信号的量化步长大，小幅值的信号量化步长小。
3. **矢量量化**
 - (1) **原理**：利用样值之间的相关性。使用 N 个相邻幅值的相关性，将 N 维空间分割成为一个一个的蜂窝。
 - (2) **码书设计**

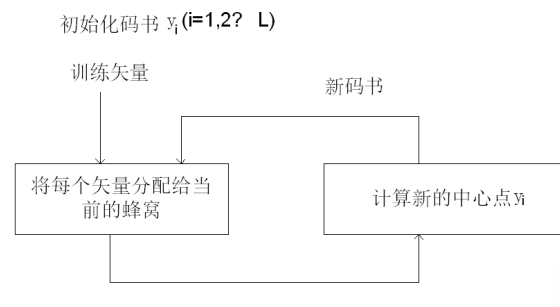
将矢量空间转换成蜂窝，使得平均量化误差最小。

量化误差 $\vec{q} = \vec{x} - \vec{y}_i$

$$d(x, y_i) = \frac{1}{N} \sum_{k=1}^N (x_k - y_{i,k})^2$$

$$\text{平均误差 } D = \sum_{i=1}^L \int_{c_i} d(\vec{x}, \vec{y}_i) p(\vec{x}) d\vec{x}$$

LBG 编码



第五章 预测编码

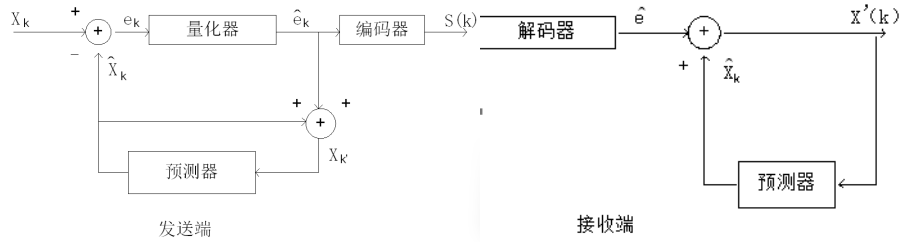
1. DPCM（差分预测编码调制）

随着已知信号的不断增多，猜中后面的一个符号越容易，知道的越多，后一个符号具有的不确定度越小。信号为零均值时，误差信号的方差小于原始信号。故选择传输差值信号而非直接传输原始信号能降低传输的码率。

2. 差分脉冲编码调制的编码和解码

X_K 是某信号的实际值， \hat{X}_K 是对该信号的预测值， e_K 是预测误差， \hat{e}_K 为量化预测误差， $X'_K = \hat{e}_K + \hat{X}_K$ ，得到当前信号的重建值，作为下一个信号的预测值。数个预测出的失真，完全由量化器产生。如果没有量化器，则 DPCM 可应用于无失真编解码中。

量化误差： $X_K - X'_K = X_K - (\hat{X}_K + \hat{e}_K) = e_K - \hat{e}_K = q_k$



3. 最佳线性预测器

设集合 $\{X\}$ 表示信号序列， K 时刻的信号值 X_K 用过去 N 个信号值的线性组合来预测。

$$\hat{X}_K = \sum_{i=1}^N a_i \cdot X_{K-i}$$

$$e_K = X_K - \hat{X}_K$$

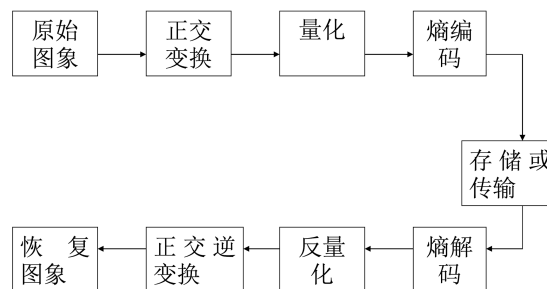
$$\sigma_e^2 = E \{ (X_k - \hat{X}_k)^2 \}$$

$$\frac{\partial \sigma_e^2}{\partial a_i} = 0 \quad i = 1, 2, \dots, N$$

$$\begin{bmatrix} R(0) & R(1) & \dots & R(N-1) \\ R(1) & R(0) & \dots & R(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(N-1) & R(N-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} R(0) \\ R(1) \\ \vdots \\ R(N-1) \end{bmatrix}$$

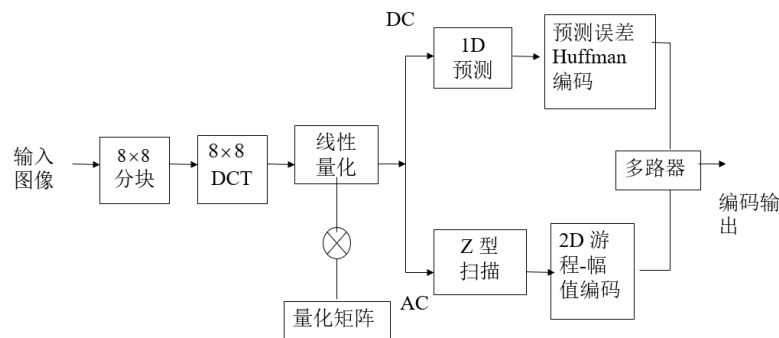
第六章 变换编码

1. **线性变换**：处理后的输出量是输入图像的像素线性组合而成。
2. **正交变换**：线性变换的特殊形式。可逆且正交。
3. **正交变换性质**
 - (1) 正变换后反变换回来熵保持，不丢失信息
 - (2) 能量保持（像素点的值的平方和在变换前后保持不变）
 - (3) 尽管能量总和保持不变，但能量分布变了，原来能量分布均匀，变换后低频成分多，高频成分少。（低频能量多，高频能量少）
 - (4) 去相关性，将高度相关的空间样值变为相关性较弱的变换系数。
4. **正交变换编码过程**



5. **2DDFT（傅里叶变换）**：变化结果是复数，处理不便。
6. **2DDCT（余弦变换）**：全实数运算，处理容易；去相关能力强；占用机时较多。
7. **JPEG**

(1) JPEG 基本系统压缩过程框图



(2) 各部分功能

- A. 首先将原始图像分割成 8×8 的子块，对于彩色图像，它要求 YUV 4: 2: 2 格式。
 - B. DCT：为消除直流电平的影响，首先将原始图像的所有像素减去 128，再进行 DCT 变换进行去相关性。
 - C. 线性量化：进行四舍五入取整，是图像质量下降的主要原因，JPEG 给出了量化系数矩阵表，在量化表设计时考虑了人眼的视觉特性。
 - D. 编码输出：
 - ①对于直流系数预测，预测后进行熵编码；
 - ②对于交流 AC 系数，首先对交流系数进行 Z 型扫描，编码以游程-幅值，Huffman 编码的形式完成。
8. **基于 DCT 的扩展过程与基本过程不同点**
 - ①源图像可以 12bit/pixel
 - ②可采用算术编码
 - ③累进操作模式
 9. **层次过程**
 - ①把原图像的分辨率按 2 的倍数降低

- ②把降低的子图像采用基本过程编码
- ③将压缩数据解码，重建低分辨率图像，使用插值、滤波对其内插，幅度水平和垂直分辨率
- ④对二者的差值进行基本过程编码
- ⑤重复

10. 分形图像编码

(1) 原理：分形的方法是把一幅数字图像，通过一些图像处理技术将原始图像分成一些子图像，然后在分形集中查找这样的子图像。分形集存储许多迭代函数，通过迭代函数的反复迭代，可以恢复原来的子图像。

(2) 特点：压缩比高，压缩后的文件容量与图像像素数无关，在压缩时时间长但解压缩速度快。

11. 方块效应

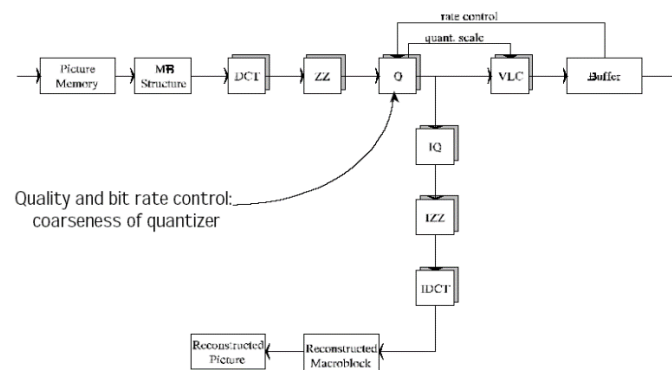
(1) 原因：JPEG 过程在进行分块后对每一块进行 DCT 变换，高压缩比时，线性量化的量化步长较大，量化后将一些高频系数变成了 0，反变换后，成为了全灰的，能看到明显的方块效应。

(2) 改进方法：用小波变化代替余弦变换、降低量化步长、分块 overlap、后处理。

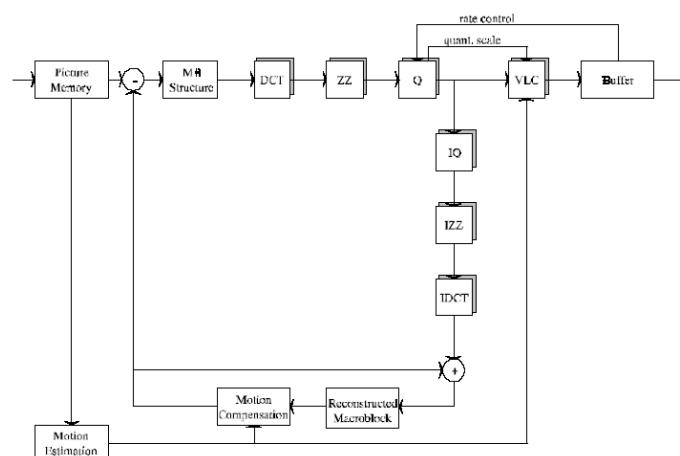
第七章 序列图像编码

1. **运动估计**: 对于当前帧的某块, 在已编码的前一帧中找到对应块的过程, 两块的位置差被称为运动矢量。块匹配方法: 全局搜索、N 步寻找法、金字塔
2. **MPEG**: 适用于多媒体应用的视频及音频的压缩标准。
3. **层次化语法结构**: 若干帧作为一个图片组 (GOP), 作用在于能够不按顺序独立进行压缩和解压 → 将每一个 GOP 分成切片, 在切片的头和尾加入一些用于同步的码字, 以确保当前的错误不会传到下一个切片 → 每个切片分成若干个宏块 (Micro Block 16*16), 宏块是运动估计的最小单元 → 在做 DCT 变换时, 将宏块分成规定 (8*8) 的小块。
4. **GOP**

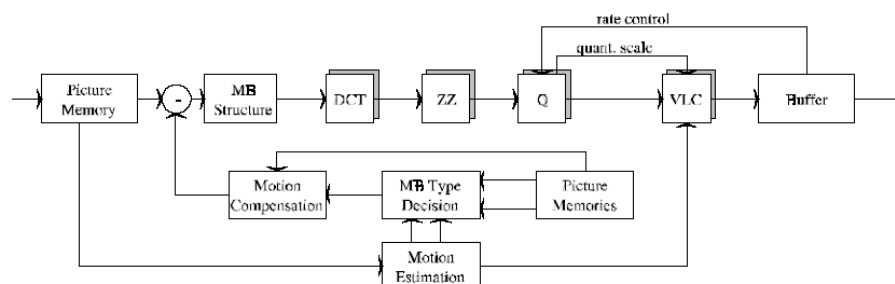
I 帧: 单独编码; 允许解码器随机解码; 作为其他图片编码的参考帧。

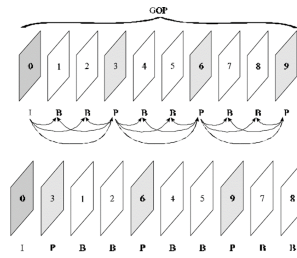


P 帧: 使用混合编码器编码; 作为其他图片编码的参考帧。



B 帧: 使用扩展的混合编码器; 从不作为其它图片编码的参考。





GOP 结构对延时的影响：

(IIIIIIIIII) 仅帧内编码：不参考其他帧，仅使用本帧的信息进行编码

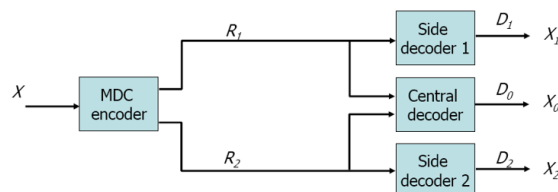
(IPPPPPP) 存在延时，要通过上一帧使用运动预测的方式进行帧间预测编码。此时上一帧的错误也会传递给下一帧。

(IBBPBBPBBP) I 为帧内编码，单独编解码，不参考其他帧；P 为预测帧，以 I 帧或前一个 P 帧为基础；B 帧为双向预测帧，传送其与 I 帧或 P 帧之间的差值信息，所以 B 帧需要得到其前后的 I 帧和 P 帧来进行预测。

第八章 分布式/多描述图像视频编码

1. 多描述编码（MDC）

（1）原理：信源通过多描述编码器生成多个描述（比特流），分别在多个独立的信道上传输，如果只收到其中某一个描述，则恢复较低质量但可接受的重建信号，如果收到两个描述，则得到高质量的中心路重建信号。



（2）优缺点

优点：提高数据可靠性，较高的鲁棒性

缺点：编码效率有所损失，即中心路（decoder0）和单路（上图 decoder1 和 decoder2）的编码效率不会同时达到最优。

（3）R-D 区域

$$D_i \geq \sigma^2 2^{-2R_i}, \text{ for } i = 1, 2$$

$$D_0 \geq \sigma^2 2^{-2(R_1+R_2)} \cdot \gamma_D$$

$$\text{其中, } \gamma_D = \begin{cases} 1, & \text{if } D_1 + D_2 > \sigma^2 + D_0, \\ \frac{1}{1 - \left(\sqrt{(1-D_1)(1-D_2)} - \sqrt{D_1 D_2 - 2^{-2(R_1+R_2)}} \right)^2} (> 1), & \text{otherwise.} \end{cases}$$

（4）实现方法和应用场合

① 前处理法：抽样法，多描述相关变换法；

② 嵌入到编码方案中：多描述标量量化法，多描述矢量量化法；

③ 后处理法：基于纠错码的多描述编码。

2. **分布式信源编码（DSC）：**单独编码联合解码，所以编码简单解码复杂，利用纠错码实现信源的编解码，具有一定的抗干扰能力。在同等码率下，在有无失真的情况下，DSC都能够获得与传统联合编码一样的压缩性能。
3. **分布式视频编码（DVC）：**将奇数帧和偶数帧分别作为编码帧和关键帧（边信息）。