



多媒体技术

MPEG 压缩编码

本章提纲

- 为什么视频能压缩
- 谁在组织视频压缩编码
- MPEG-1视频
- MPEG-2视频
- MPEG-4视频

MPEG视频（视频）

- MPEG视频是用MPEG视频标准压缩和解压缩的电视图像。现有的MPEG视频标准包括MPEG-1 Video, MPEG-2 Video, MPEG-4 Visual, H.264/MPEG-4 AVC和H.265/ MPEG-H HEVC
- 这些视频标准有许多共同之处，基本概念类似，数据压缩和编码方法基本相同，核心技术都是采用以图像块为基本单元的变换、量化、移动补偿、熵编码，在保证图像质量前提下获得尽可能高的压缩比
- 重点介绍MPEG-1，其他标准都在此基础上开发的。

为什么视频能压缩

- 视频数据之所以能够被压缩，主要是视频数据中存在大量的冗余数据，包括时间冗余、空间冗余、结构冗余、视觉冗余、知识冗余和数据冗余

➤ 时间冗余(temporal redundancy)

- 与时间相关的冗余：在某个时间间隔上出现场景相同或基本相同的连续帧时，帧与帧之间存在大量的冗余数据

➤ 空间冗余(spatial redundancy)

- 与空间位置有关的冗余：在单帧图像中，相邻像素的值常有相同或变化不大的情况，可用较少数据表达

为什么视频能压缩

➤ 结构冗余(structural redundancy)

- 图像自身构造的冗余：若从宏观上看一帧图像，有些图像存在相同或类似的结构，如用地板图案构成的图像

➤ 视觉冗余(vision redundancy)

- 与视觉系统有关的冗余：对图像的亮度变化敏感而对颜色变化不敏感，对剧烈变化区域敏感而对缓慢变化区域不敏感，对图像的亮度和颜色的分辨率都存在极限

为什么视频能压缩

➤ 知识冗余(knowledge redundancy)

- 知识是某个感兴趣领域中的实事、概念和关系
- 与知识有关的冗余：在单帧图像中含有为人熟知的知识，称为先验知识。例如，正面人头像有相对固定的结构，眼睛下方是鼻子，鼻子下方是嘴，嘴和鼻子均位于脸的中线上。这类规律性的结构往往不会改变或变化不大

➤ 数据冗余(data redundancy)

- 数据本身的冗余：视频数据本身存的冗余

为什么视频能压缩

- MPEG视频标准利用的冗余：时间冗余、空间冗余、视觉冗余和数据冗余

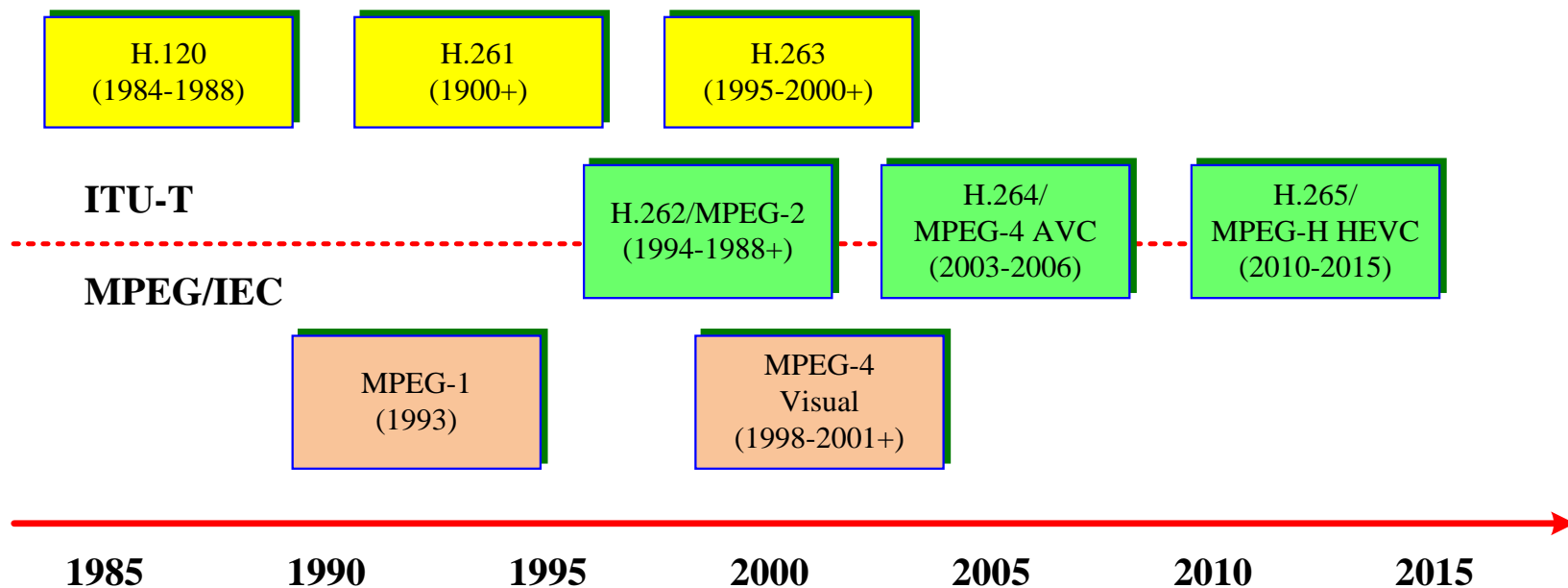
表 12-1 视像压缩利用的各种冗余信息

种类	内容	目前用的主要方法
空间冗余	像素间的相关性	变换编码，预测编码
时间冗余	时间方向上的相关性	帧间预测，移动补偿
视觉冗余	人的视觉特性	非线性量化，位分配
图像构造冗余	图像本身的构造	轮廓编码，区域分割
知识冗余	收发两端对人物的共有认识	对象编码，知识编码
其他	不确定性因素	

谁在组织视频压缩编码

- 两个著名的组织

- ITU-T VCEG(Video Coding Experts Group)专家组
- ISO/IEC MPEG(Moving Picture Experts Group)专家组
- 分别和联合开发的视频压缩标准见图1



谁在组织视频压缩编码

- 两个著名的组织：

- ISO/IEC制定的标准——MPEG-x

- MPEG标准主要用于广播电视、DVD和视频流媒体。
。例如：MPEG-1、MPEG-2、MPEG-4

- ITU-T制定的视频编码标准——H.26x

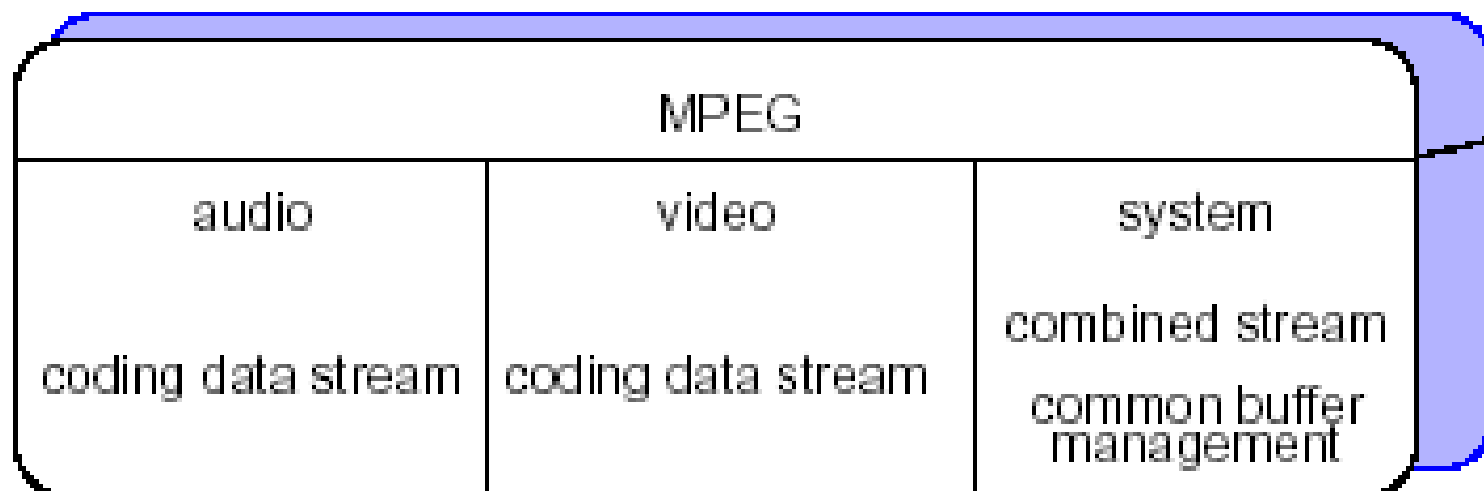
- 主要用于实时视频通信，如视频电视会议、可视电话等。例如：H.261、H.262、H.263、H.26L

谁在组织视频压缩编码

- MPEG 是运动图像专家组（Moving Picture Expert Group）的简写，其活动开始于1988年
 - 负责数字视频、音频和其他媒体的压缩、解压缩、处理和表示等国际技术标准的制定工作
 - 制定的标准推动了VCD、DVD、数字电视、高清数字电视等产业的发展
- MPEG 和 JPEG 两个专家小组都是在 ISO 领导下的专家小组，其小组成员也有很大的交叠
 - JPEG的目标：静止图像的压缩
 - MPEG的目标：运动图像的数据压缩

谁在组织视频压缩编码

- MPEG主要包括**MPEG视频**、**MPEG音频**和**MPEG系统**(视音频同步)三个部分



谁在组织视频压缩编码

- MPEG-1
 - MPEG-2
 - MPEG-4
 - MPEG-7
 - MPEG-21
- } 不涉及视频的压缩技术

谁在组织视频压缩编码

- **MPEG-1和MPEG-2是MPEG组织制定的第一代视音频压缩标准**，为VCD、DVD及数字电视和高清晰度电视等产业的飞速发展打下了牢固的基础
- **MPEG-4是基于第二代视音频编码技术制定的压缩标准**，以视听媒体对象为基本单元，实现数字视音频和图形合成应用、交互式多媒体的集成，目前已经在流式媒体服务等领域开始得到应用

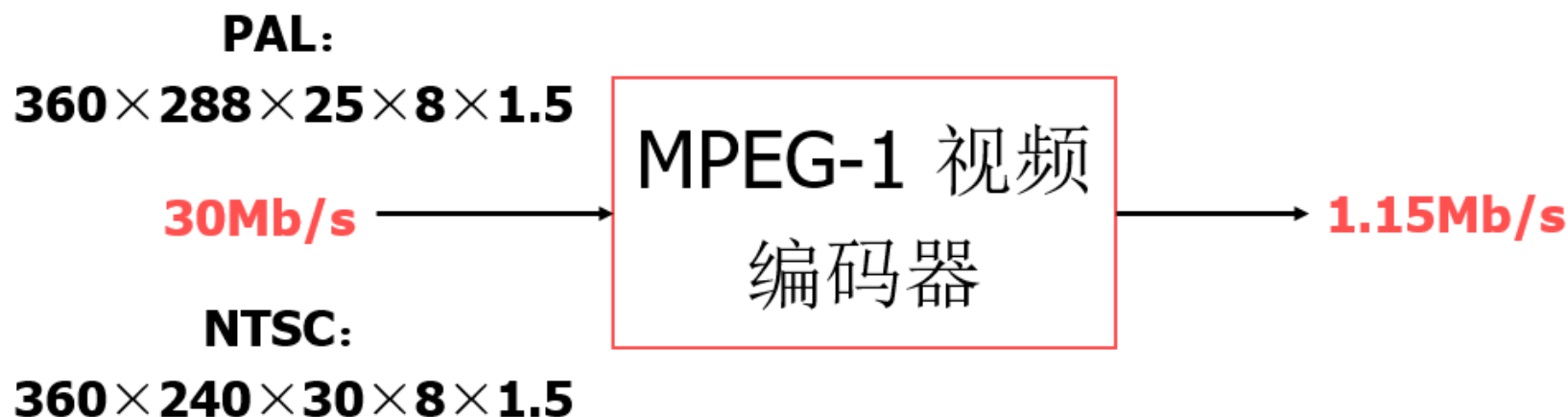
谁在组织视频压缩编码

- **MPEG-7是多媒体内容描述标准**，支持对多媒体资源的组织管理、搜索、过滤、检索
- **MPEG-21的重点是建立统一的多媒体框架**，为从多媒体内容发布到消费所涉及的所有标准提供基础体系，支持连接全球网络的各种设备透明地访问各种多媒体资源

MPEG-1视频

- MPEG-1的标准号为ISO/IEC 11172，是针对数据传输速率大约为1.5Mbps的数字存储媒体的视频及其伴音编码、解码的国际标准
- MPEG-1主要用于在CD-ROM上存储视频和音频信号。视频编码采用SIF（Source Input Format）格式
 - 分辨率： $360 \times 240 \times 30$ （NTSC制式），4:2:0的采样方式，压缩率约26:1，图像质量接近VHS
 - MPEG-1音频编码器的输入为线性PCM信号，双声道，采样频率为32kHz，44.1kHz或者48kHz，输出为32kb/s~384kb/s，可以达到CD质量

MPEG-1视频



压缩比约为 **26 : 1**

MPEG-1视频

- 1. 视频数据的压缩算法
- 2. 帧内图像I的压缩编码算法
- 3. 预测图像P的压缩编码算法
- 4. 双向预测图像B的压缩编码算法
- 5. 帧图像的编排顺序
- 6. 视频数据流的结构

1. 视频数据的压缩算法

- MPEG-1视频(MPEG-1 Video)压缩数据的基本方法
 - 在空间方向上, 采用与JPEG类似的算法去掉空间冗余数据
 - 在时间方向上, 采用移动补偿(motion compensation)算法去掉时间冗余数据
- MPEG专家组为此开发了两项重要技术
 - 定义了视频数据的结构
 - 定义了三种类型的图像

1. 视频数据的压缩算法

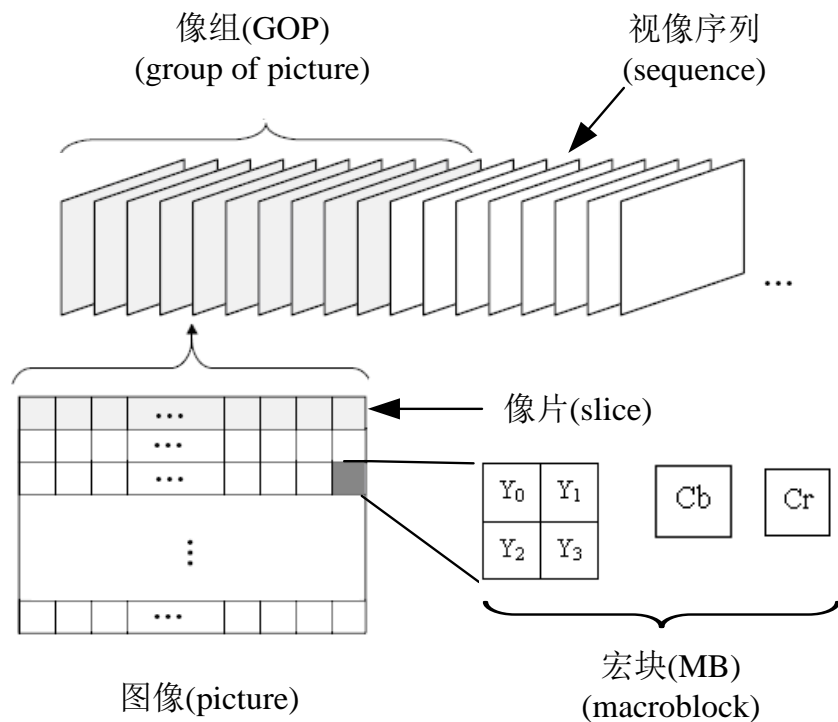
● 视频数据结构

把视频片段看成一系列静态图像的序列：

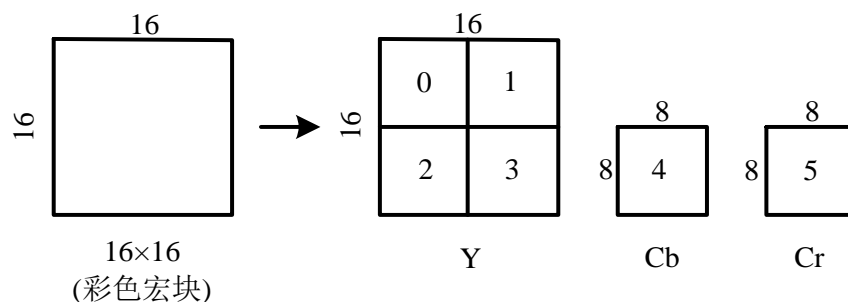
- 把视频序列分成许多像组(group of picture, GOP)
- 把像组中每一帧图像分成许多像片(slice)，每个像片由16行组成
- 把像片分成16行 \times 16像素/行的宏块(macroblock)
- 把宏块分成若干个8行 \times 8像素/行的图块(block)，
- 使用子采样格式为4:2:0时，一个宏块由4个亮度(Y)图块和2个色度图块(Cb和Cr)组成

1. 视频数据的压缩算法

● 视频数据结构



(a) 视像数据的组织



(b) 宏块的结构(4:2:0)

1. 视频数据的压缩算法

- 三种类型的图像

三种类型图像，采用三种不同算法分别对它们压缩

- (1) **帧内图像I** (intra-picture), 简称I图像或I帧(I-picture / I-frame)

- 包含内容完整的图像，用于为其他帧图像的编码和解码作参考，称为关键帧

- (2) **预测图像P** (predicted picture), 简称P图像或P帧(P-picture / P-frame)

- 以在它之前出现的帧内图像I作参考的图像，对预测图像P进行编码就是对它们之间的差值进行编码

1. 视频数据的压缩算法

- 三种类型的图像

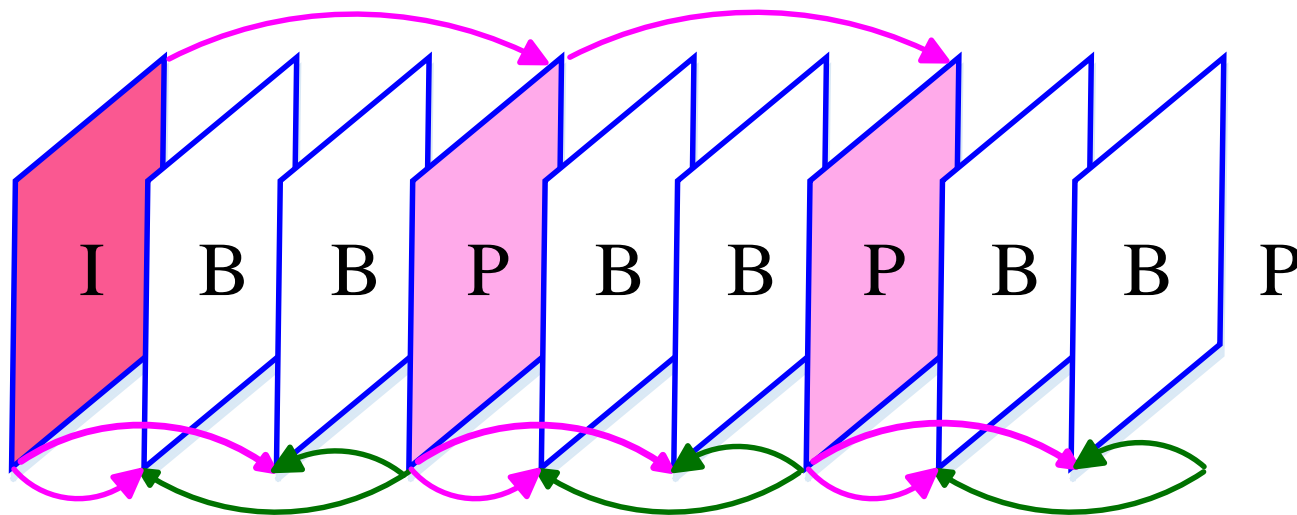
- **双向预测图像B** (bidirectionally-predictive picture), 也称

- 双向插值图像B, 简称B图像或B帧(B-picture/B-frame)

- 以在它之前和之后的帧图像(I和P)作参考的图像, 对B进行编码就是对帧内图像I和预测图像P的差值分别进行编码

1. 视频数据的压缩算法

- 三种类型的图像



1. 视频数据的压缩算法

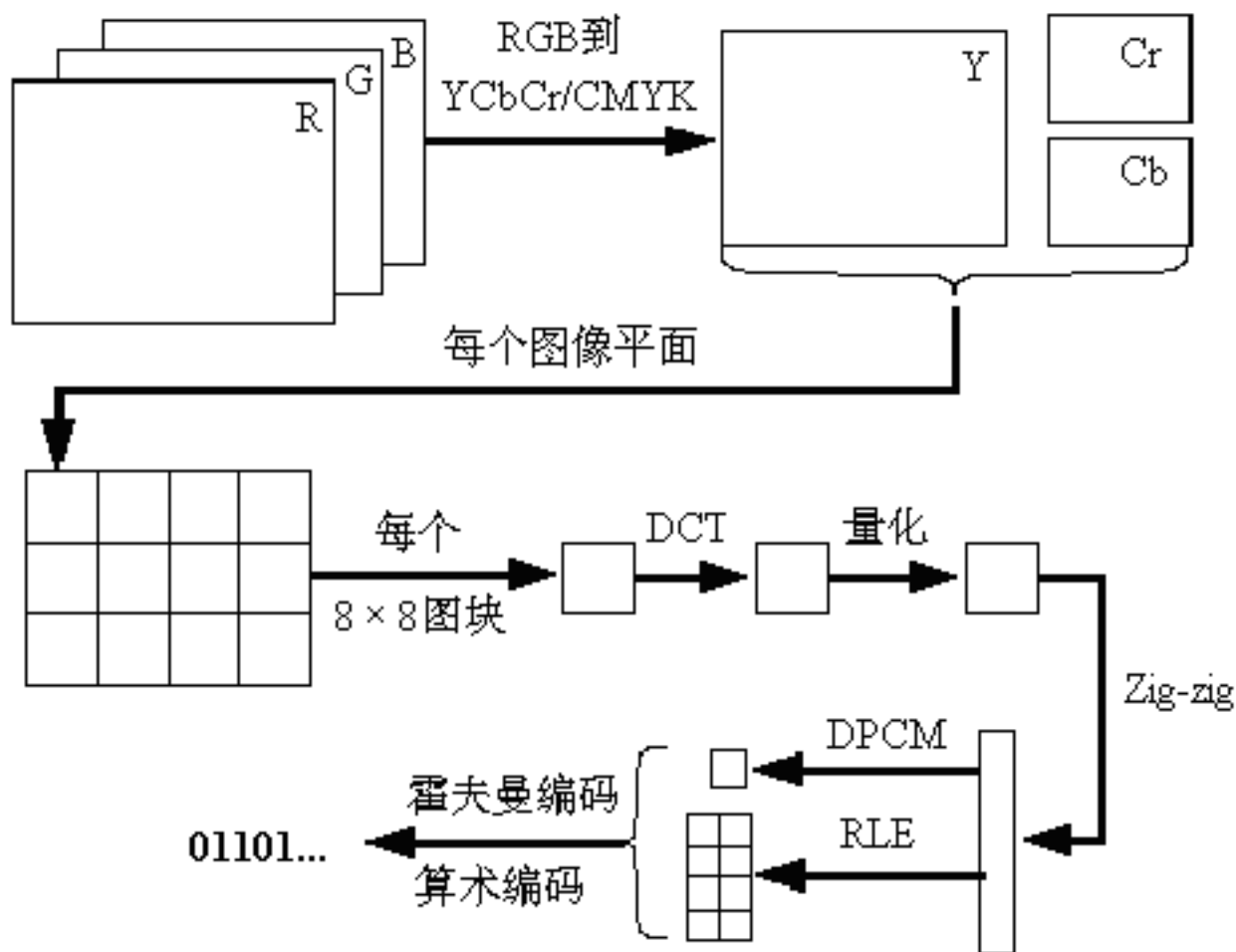
- 由于MPEG对视频信号作随机存取的重要要求，**有必要每隔一段时间传送1帧只采用帧内编码的图像作为参考图像**，使得预测环路能在某一参考帧丢失后，迅速获得另一个参考帧，从而恢复到正常工作状态
- **GOP**：两个I帧之间的图像构成一个组称为GOP (Group of Picture)

2. 帧内图像I的压缩编码算法

- 不参照过去的帧和将来的帧，采用与JPEG类似的压缩算法以减少空间的冗余数据
 - 如果视频用RGB表示，把它转换成YCrCb表示
 - 每个图像平面分成 8×8 像素的图块，对每个图块进行离散余弦变换(DCT)
 - 交流分量系数经过量化后按照Zig-zag的形状排序
 - 交流分量系数用行程长度编码(RLE)，再用霍夫曼(Huffman)编码或者用算术编码
 - 直流分量系数经过量化后用差分脉冲编码(DPCM)

2. 帧内图像I的压缩编码算法

- 帧内图像I的压缩编码算法框架



3. 预测图像P的压缩编码算法

- 算法原理

- 以宏块(MB)为基本编码单元

- 一个宏块定义为 $I \times J$ 像素，一般取 16×16

- 使用两种类型的参数表示

- 差值：当前编码宏块与参考图像宏块之差

- 宏块的移动矢量(motion vector, MV)

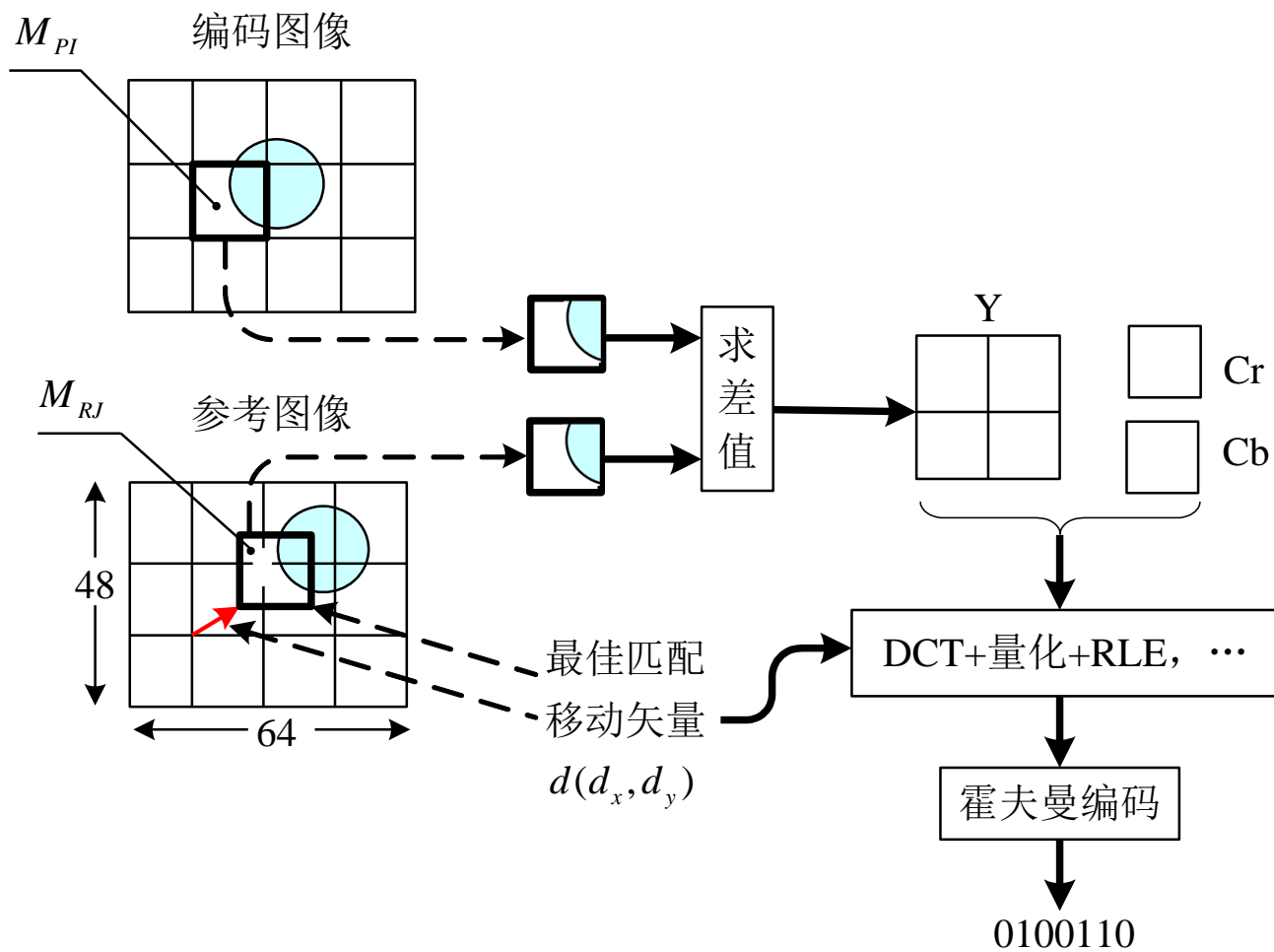
3. 预测图像P的压缩编码算法

● (1) 求解差值的方法

- 假设编码宏块MPI是参考宏块MRJ的最佳匹配块，它们的差值就是这两个宏块中相应的像素值之差
- 对所求得的差值进行彩色空间转换，然后使用4:1:1或4:2:0格式采样。对采样得到的Y，Cr和Cb分量值，仿照JPEG压缩算法对差值进行编码
- 对计算出的移动矢量也进行DCT变换和霍夫曼编码

3. 预测图像P的压缩编码算法

● 预测图像P的压缩编码算法框架图



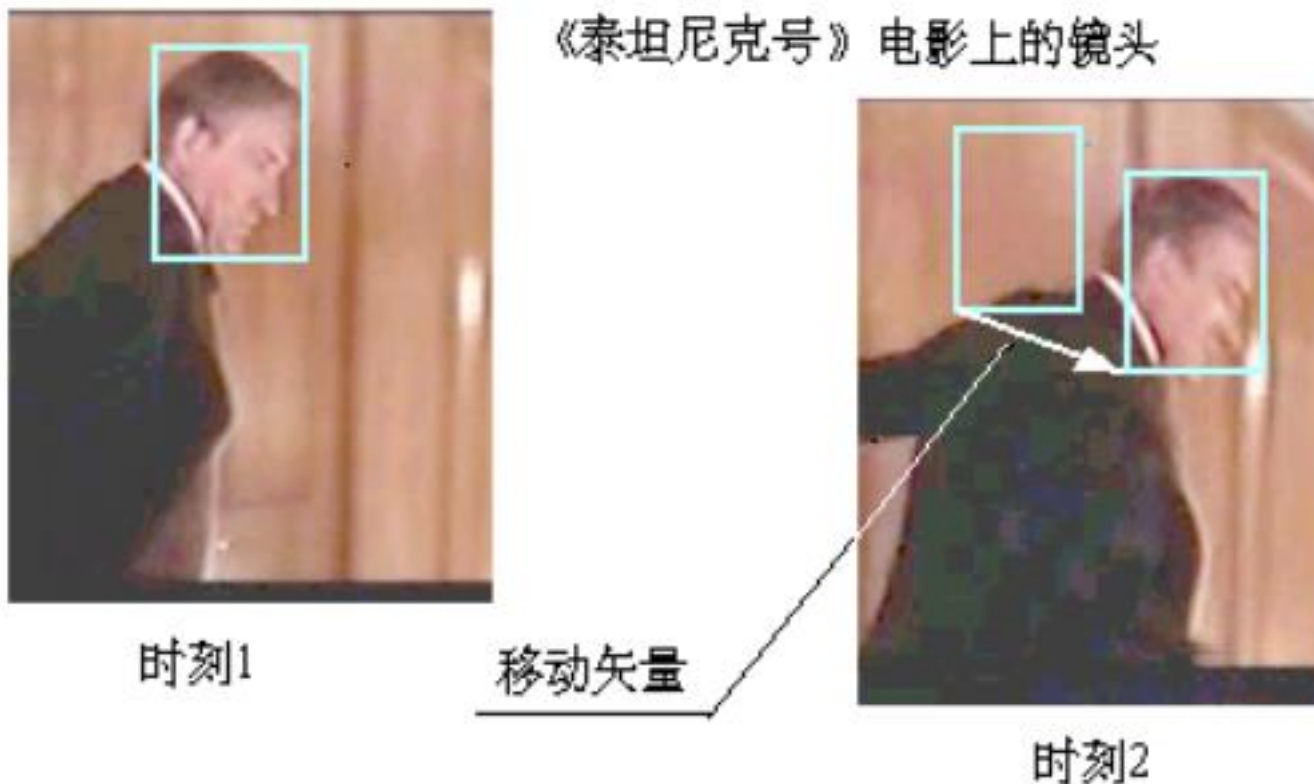
3. 预测图像P的压缩编码算法

- (2) 求解移动矢量的方法

- 在求两个宏块差值之前，需要找出预测编码图像中的编码宏块相对于参考图像中的参考宏块所移动的距离和方向，即移动矢量
- 求解移动矢量的方法见下图

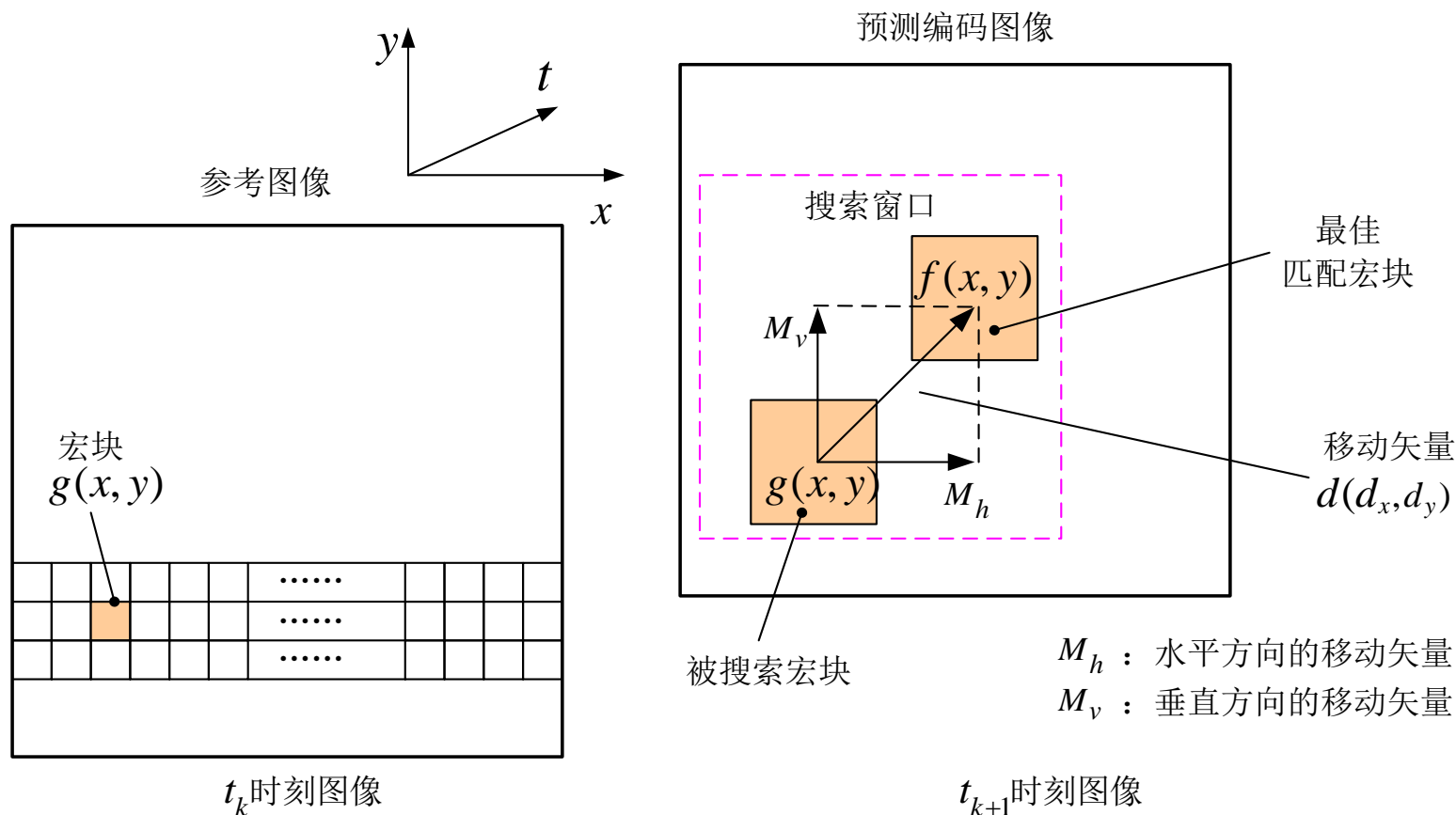
3. 预测图像P的压缩编码算法

- 移动矢量的概念和算法



3. 预测图像P的压缩编码算法

● 移动矢量的概念和算法



3. 预测图像P的压缩编码算法

● (3) 最佳匹配

➤ 编码宏块 M_{PI} (如 $f(x, y)$)是参考宏块 M_{RJ} (如 $g(x, y)$)的最佳匹配是指这两个宏块之间的差值最小

- 以绝对值(absolute difference, AE)最小作为匹配判据

$$AE = \sum_{i=0}^{15} \sum_{j=0}^{15} |f(i, j) - g(i - d_x, j - d_y)|, \quad (i = j = 16)$$

- 以均方误差(mean-square error, MSE)最小作为匹配判据

$$MSE = \frac{1}{16} \sum_{|i| \leq \frac{1}{2}} \sum_{|j| \leq \frac{1}{2}} [f(i, j) - g(i - d_x, j - d_y)]^2, \quad (i = j = 16)$$

3. 预测图像P的压缩编码算法

● (3) 最佳匹配

➤ 编码宏块 M_{PI} (如 $f(x, y)$)是参考宏块 M_{RJ} (如 $g(x, y)$)的最佳匹配是指这两个宏块之间的差值最小

- 以平均绝对帧差(mean of the absolute frame difference, MAD)最小作为匹配判据

$$MAD = \frac{1}{I \times J} \sum_{|i| \leq \frac{I}{2}} \sum_{|j| \leq \frac{J}{2}} |f(i, j) - g(i - d_x, j - d_y)|, \quad (i = j = 16)$$

其中, d_x 和 d_y 分别是参考宏块 M_{RJ} 的移动矢量 $d(d_x, d_y)$ 在 x 和 y 方向上的移动矢量

3. 预测图像P的压缩编码算法

● 搜索原理

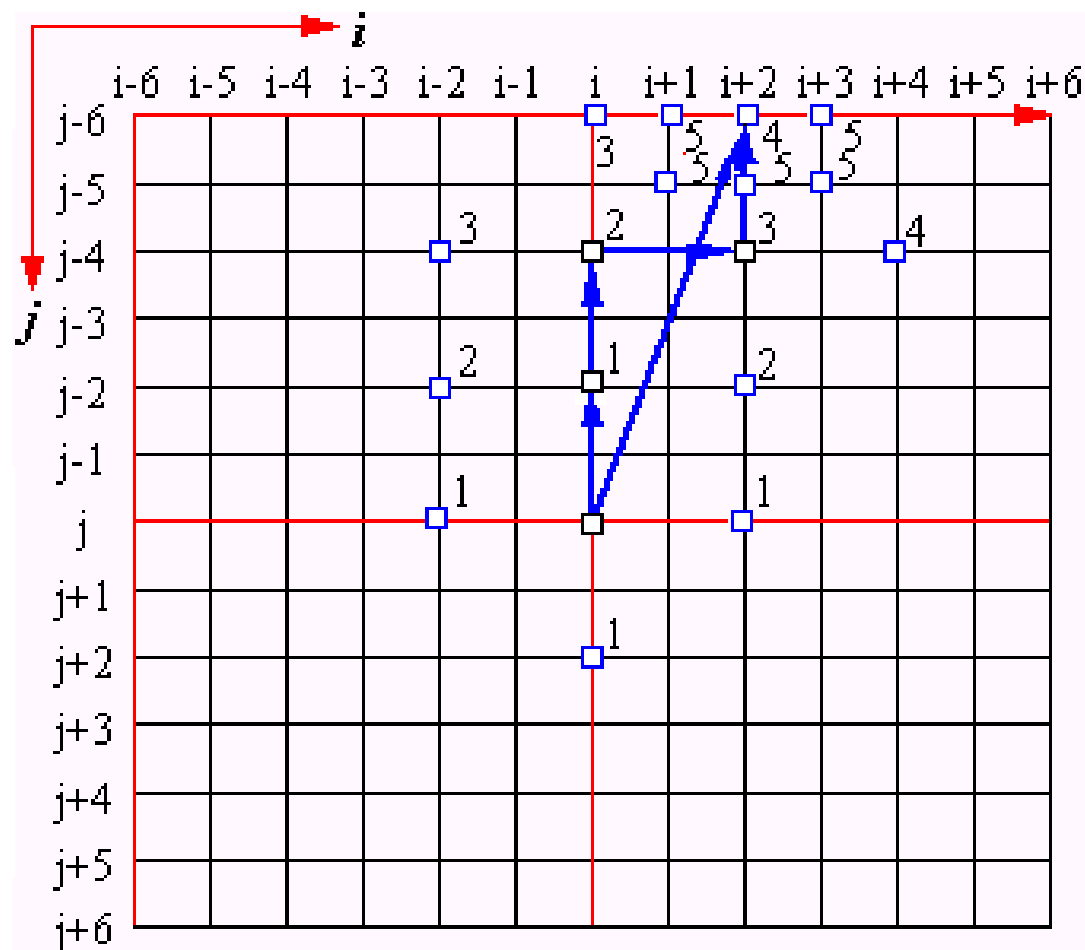
已有许多简化算法寻找最佳宏块，下面介绍其中的三种，以理解搜索算法的基本思想

➤ (1) 二维对数搜索法(2D-logarithmic search)

- 匹配判据：MSE为最小
- 搜索策略：沿着最小失真方向搜索
- 搜索方法：见下图
- 在搜索时，每移动一次，检查5个搜索点
 - 如果最小失真在中央或边界，减少搜索点之间的距离
- 步骤1, 2, ..., 5得到的近似移动矢量 d 为 $(i, j-2)$ 、 $(i, j-4)$ 、 $(i+2, j-4)$ 、 $(i+2, j-5)$ 和 $(i+2, j-6)$ ，最后得到的移动矢量为 $d(i+2, j-6)$

3. 预测图像P的压缩编码算法

● 二维对数搜索法



3. 预测图像P的压缩编码算法

● 搜索原理

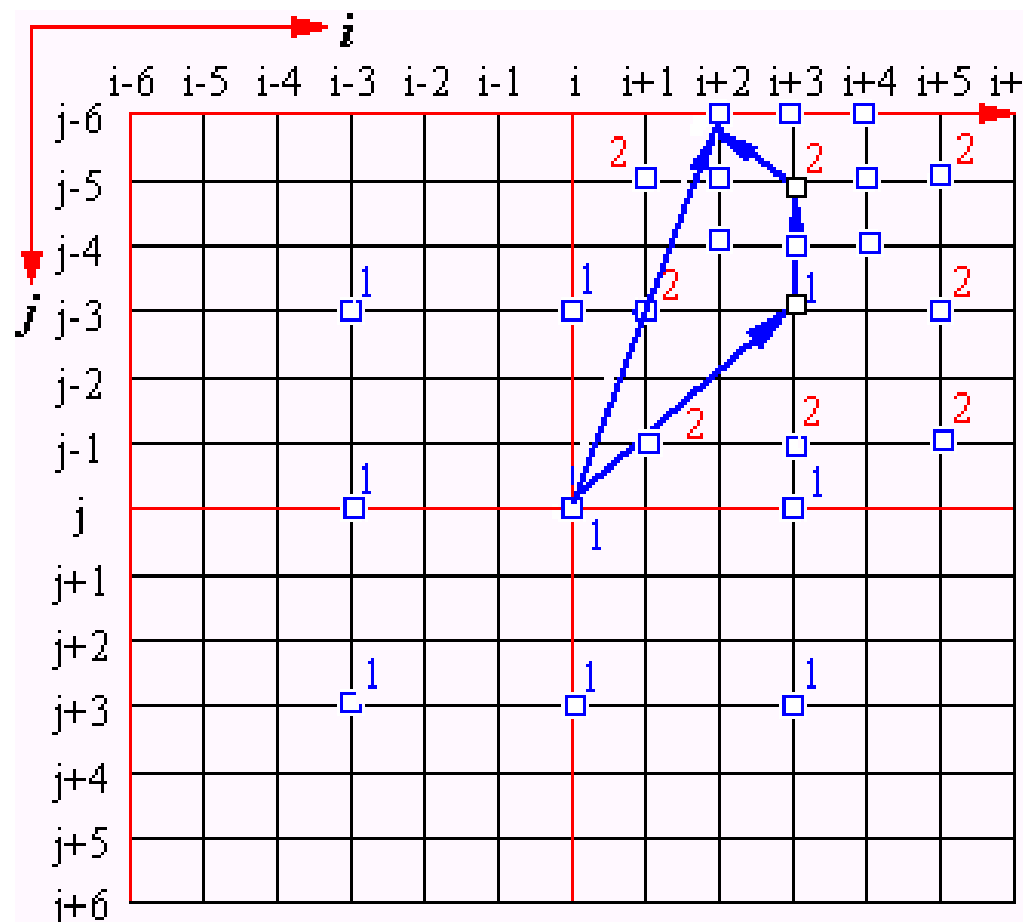
➤ (2) 三步搜索法(three-step search)

- 与二维对数搜索法很接近
- 开始时，搜索点离中心点 (i, j) 很远
- 第一步，测试8个搜索点，见下图。点 $(i+3, j-3)$ 作为第一个近似的移动矢量 d_1
- 第二步，搜索点偏离 $(i+3, j-3)$ 较近，找到的点假定为 $(i+3, j-5)$
- 第三步给出了最后的移动矢量为 $d(i+2, j-6)$

注：本例采用MAD作为匹配判据

3. 预测图像P的压缩编码算法

- 三步搜索法



3. 预测图像P的压缩编码算法

● 搜索原理

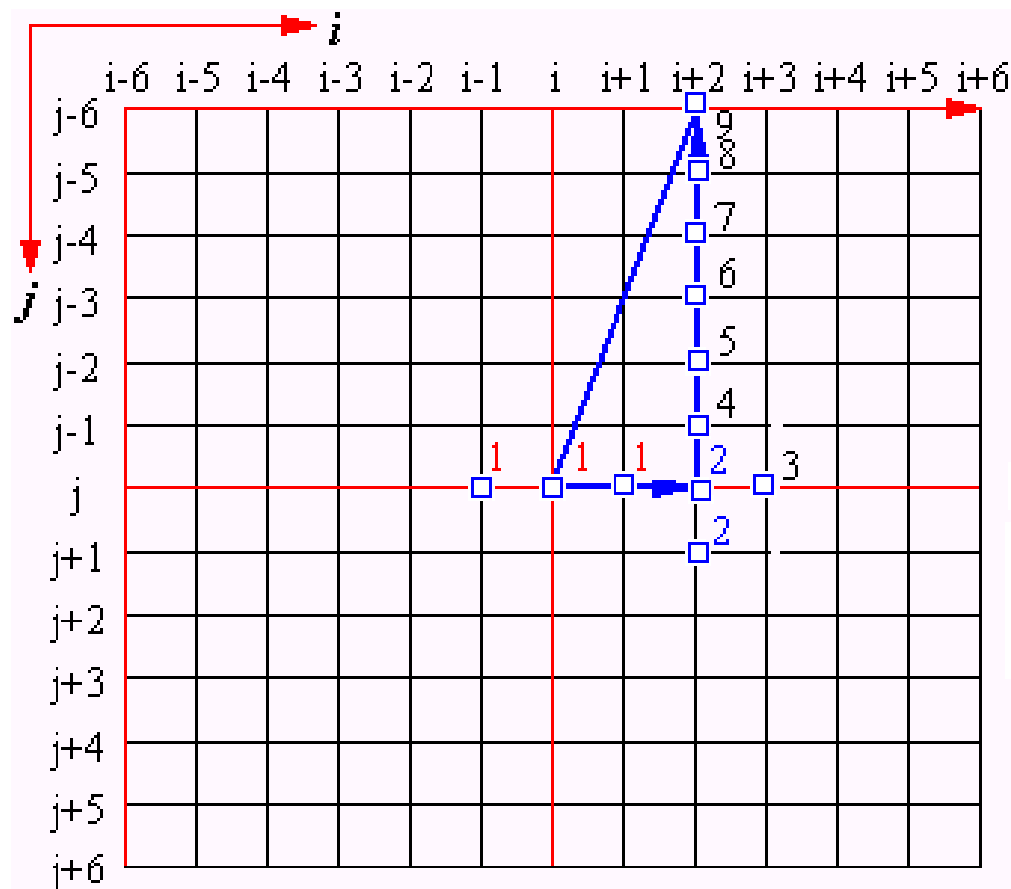
➤ 对偶搜索法(conjugate search)

- 使用MAD作为匹配判据，搜索方法下图
- 在第一次搜索时，通过计算点 $(i-1, j)$ 、 (i, j) 和 $(i+1, j)$ 处的MAD值来决定 i 方向上的最小失真
- 如果计算结果表明点 $(i+1, j)$ 处的MAD为最小，就计算点 $(i+2, j)$ 处的MAD，并从 (i, j) 、 $(i+1, j)$ 和 $(i+2, j)$ 的MAD中找出最小值
- 按这种方法一直进行下去，直到在 i 方向上找到最小MAD值及其对应的点

在整个MPEG图像压缩过程中，寻找最佳匹配宏块要占据相当多的计算时间，匹配得越好，重构图像质量越高

3. 预测图像P的压缩编码算法

● 对偶搜索法



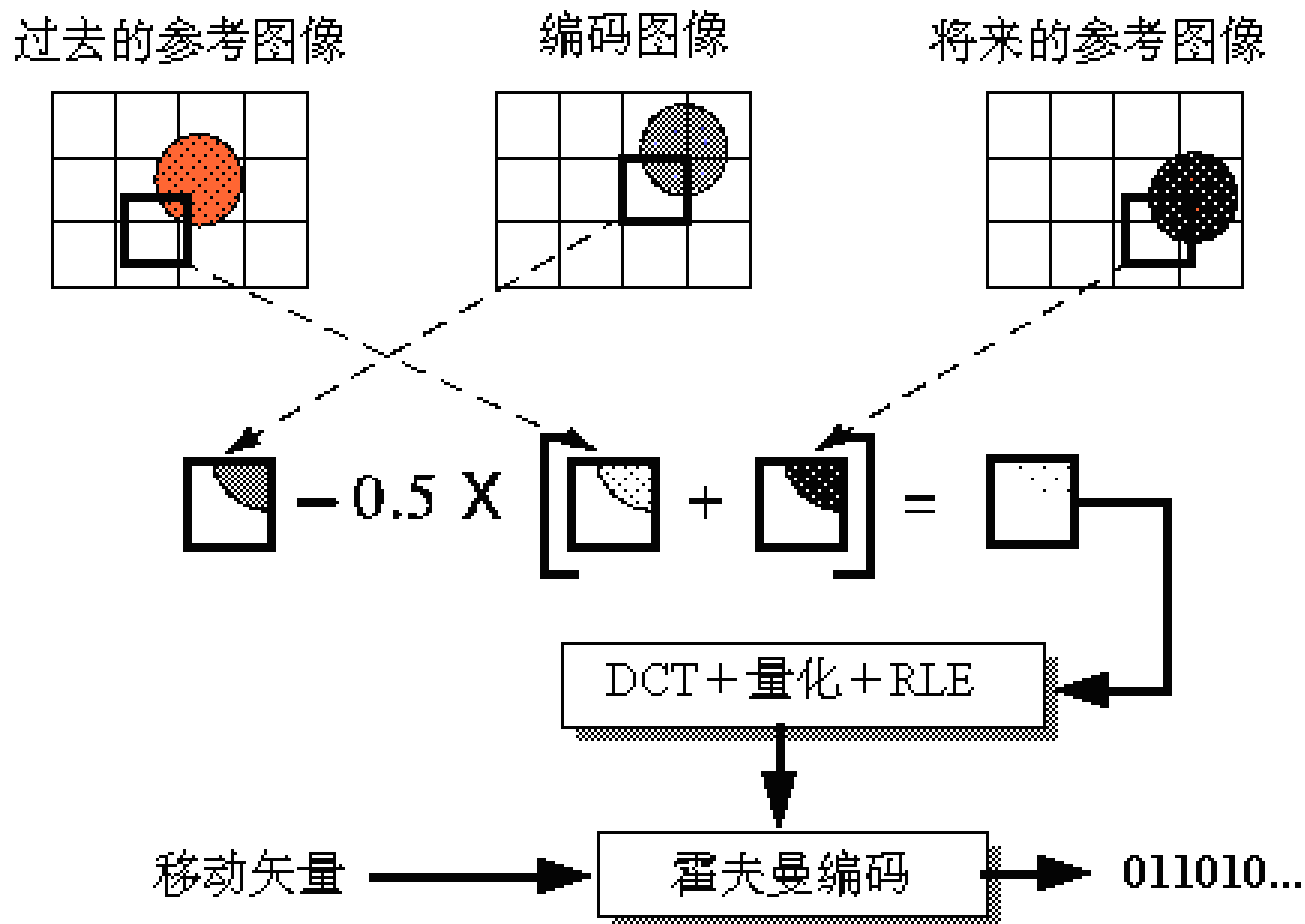
4. 双向预测图像B的压缩编码算法

- 双向预测图像B的压缩编码算法

- 对在它前后帧的像素值之差进行编码，见下图。具体计算方法与预测图像P的算法类似
- 对帧内图像I和预测图像P编码时，双向预测图像B不作為它们的参考图像，因此B不传播编码误差

4. 双向预测图像B的压缩编码算法

- 双向预测图像B的压缩编码算法框架



4. 双向预测图像B的压缩编码算法

● 三种图像压缩性能比较

- 帧内图像I、预测图像P和双向预测图像B经过压缩后的大小见下表
- 从表中可以看到，帧内图像I的数据量最大，而双向预测帧图像B的数据量最小

表 12-2 MPEG 三种图像压缩后的典型值

图像类型		I	P	B	平均
MPEG-1 CIF (1.15 Mb/s)	数据量(b/帧)	150 000	50 000	20 000	38 000
	近似压缩比	7:1	20:1	50:1	27:1

5. 帧图像的编排顺序

- 帧图像的编排顺序

- MPEG-1编码器允许选择

- 帧内图像I出现的频率和位置，如频率为2 Hz
- 在两帧图像I之间或在图像I和P之间选择图像B的数目
- 图像I、P和B的数目根据节目内容确定。例如
 - 对快速运动的图像，帧内图像I的频率可以选择高一些，双向预测图像B的数目可以选择少一些
 - 对慢速运动的图像，帧内图像I的频率可以低一些，而双向预测图像B的数目可以选择多一些，这样可保证视频的质量

5. 帧图像的编排顺序

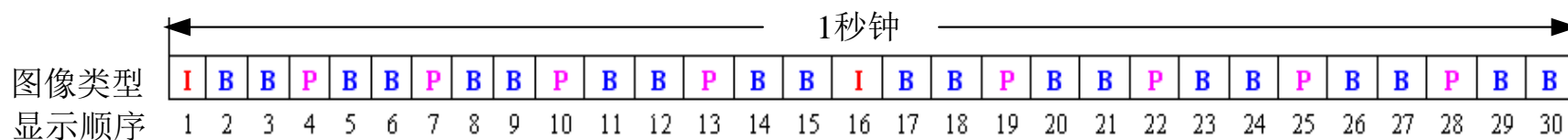
- 帧图像的编排顺序

- I、P和B的典型编排顺序见下图

- 编码参数：I的距离 $N=15$ ，P的距离 $M=3$
 - 在视频解码时，因B需I和P做参考，故在解码之前需重新组织帧图像数据流的输入顺序

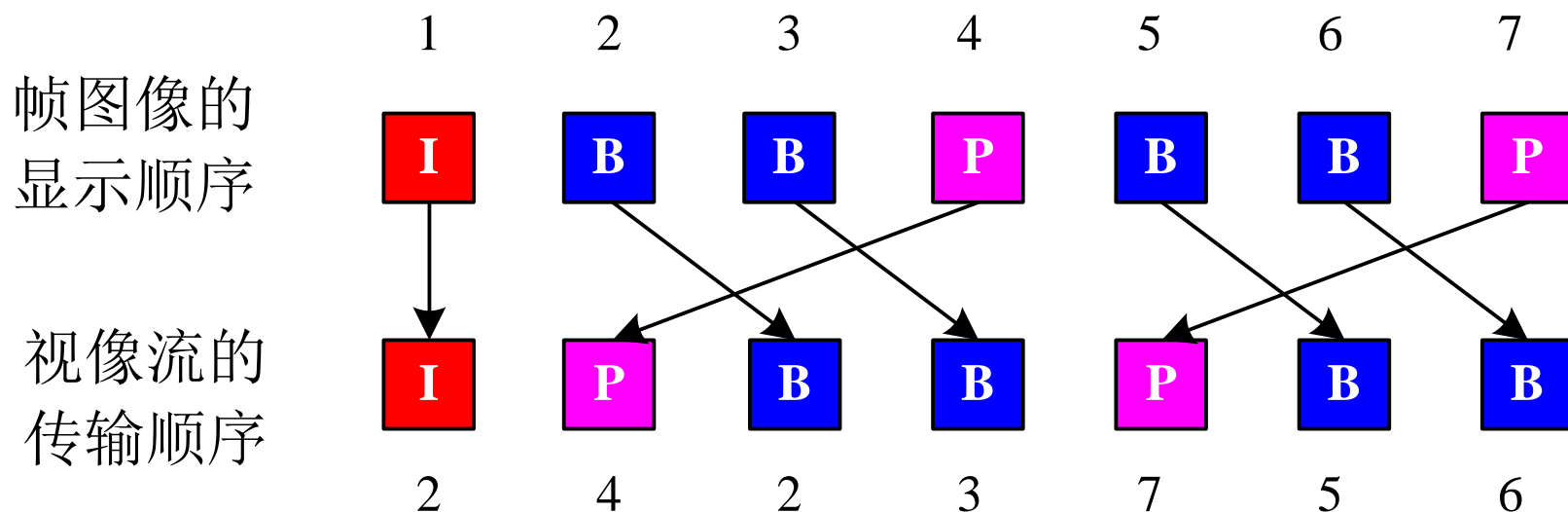
5. 帧图像的编排顺序

● MPEG帧图像的编排示例



5. 帧图像的编排顺序

- MPEG帧图像和视频流的顺序



6. 视频数据流的结构

- 视频数据流的结构

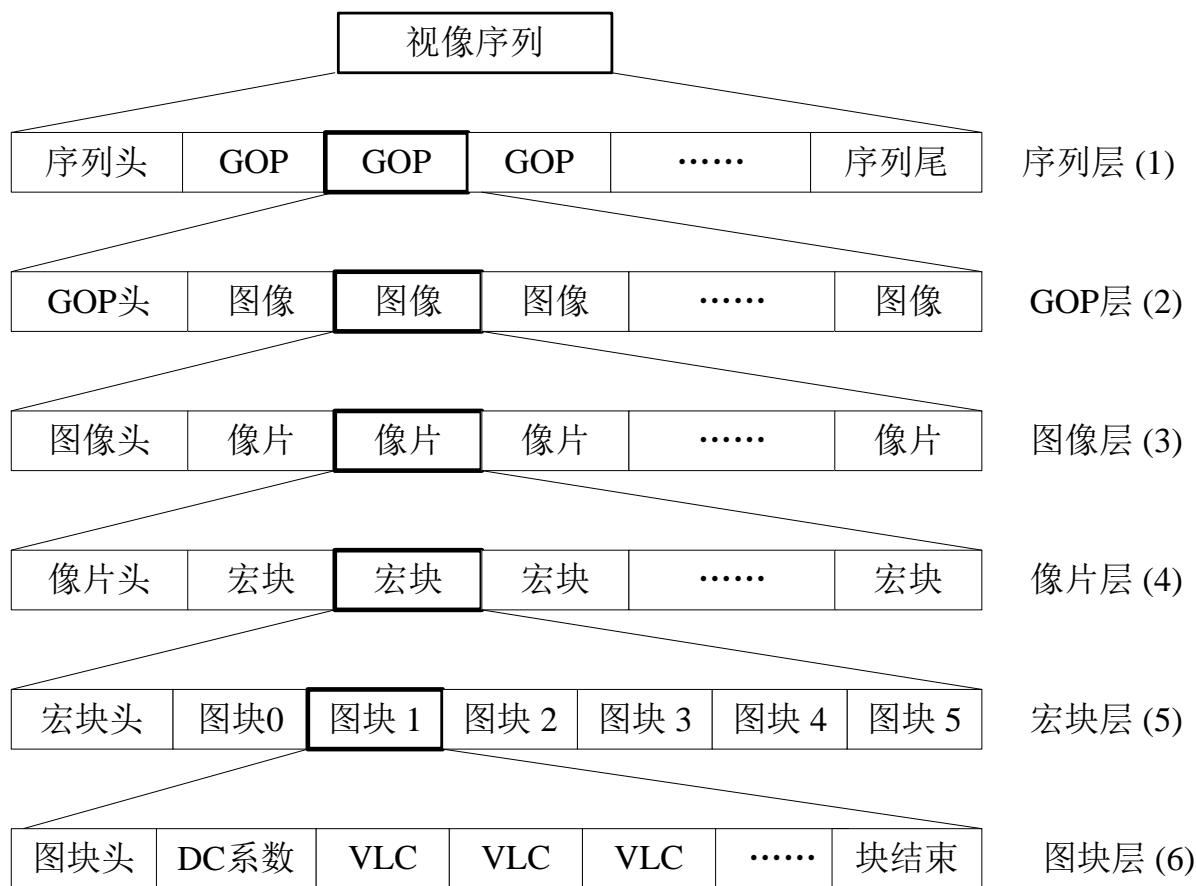
- 数据位流的组织关系到如何设计解码器，没有统一规范，设计的解码器就不能通用

- 按层次结构组织，一个视频序列(video sequence)分成6层：

- (1) 序列层(sequence)
- (2) 像组层(group of pictures, GOP)
- (3) 图像层(picture)
- (4) 像片层(slice)
- (5) 宏块层(macroblock, MB)
- (6) 图块(block)层

6. 视频数据流的结构

● MPEG-1 视频数据流的结构



本章提纲

- 为什么视频能压缩
- 谁在组织视频压缩编码
- MPEG-1视频
- **MPEG-2视频**
- MPEG-4视频

MPEG-2 视频

- MPEG-2视频

- MPEG-2视频是MPEG-1视频标准的扩展版本，全面继承了MPEG-1视频数据压缩算法，增添了许多新的语法结构和算法，用于支持

- 顺序扫描和隔行扫描

- NTSC、PAL、SECAM和HDTV格式的视频

- 视频的实时传输

- 为适应各种不同的应用，MPEG-2视频标准定义了多种视频质量可变的编码方式

MPEG-2 视频

- 视频编码器和解码器

- MPEG-2视频编码器和解码器的结构框图，原理上与MPEG-1的编码和解码结构基本相同

- 编码系统

- “ME(移动估算器)”用于计算移动矢量，找出最佳匹配宏块
- “内置解码器”用于产生预测图像，它的输入包括移动矢量、量化DCT系数和用于控制数据速率的量化参数控制信号

MPEG-2 视频

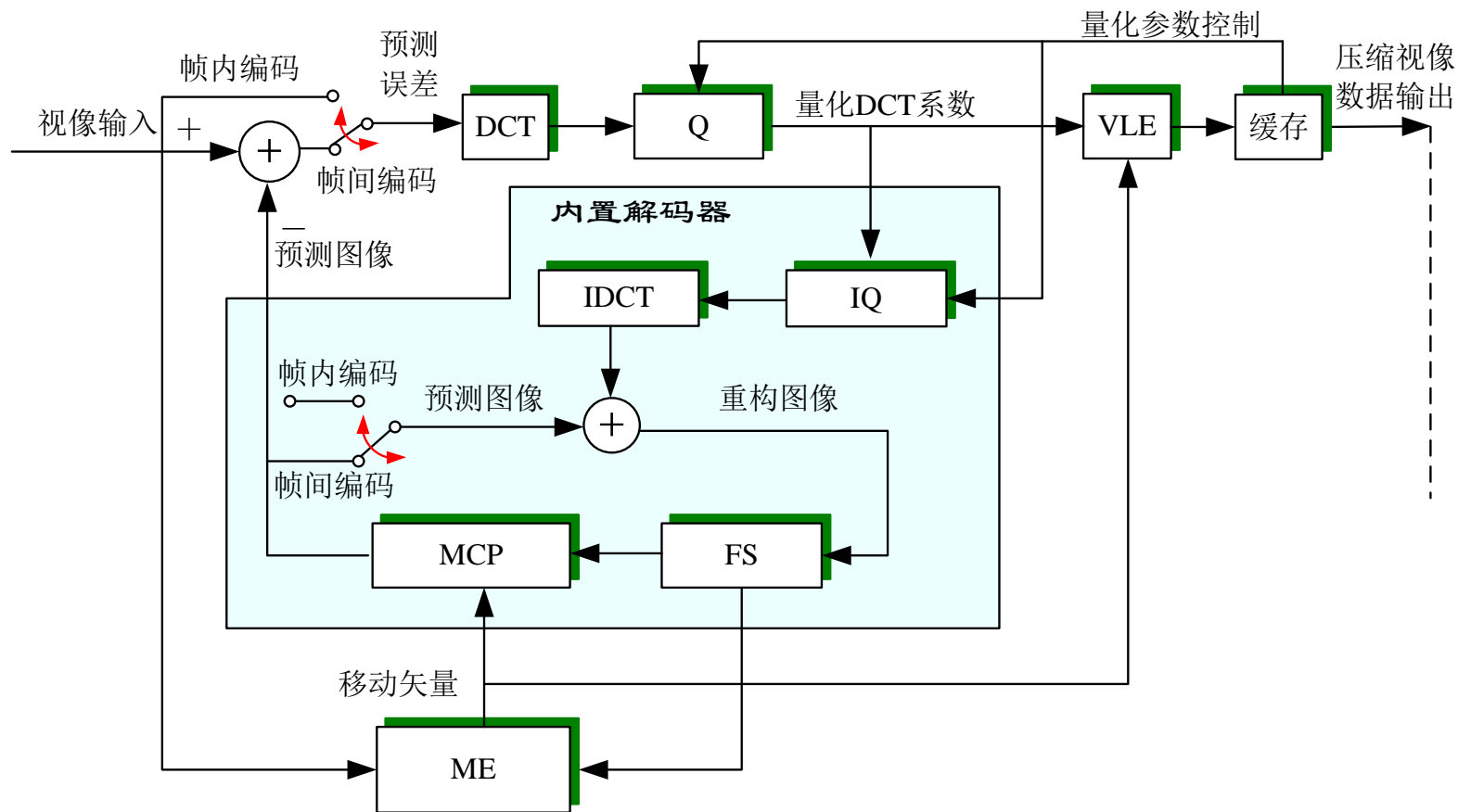
- 视频编码器和解码器

- 编码系统

- 输入视频和预测图像通过“(加法器)”产生预测误差，经过“DCT(余弦变换)”和“Q(量化)”之后送给“VLE(可变长度编码器)”，移动矢量也送到“VLE”，它们在“VLE”经过编码和复合之后送到传输媒体或存储媒体
 - “量化参数控制”信号可改变视频质量和数据速率

- 解码系统

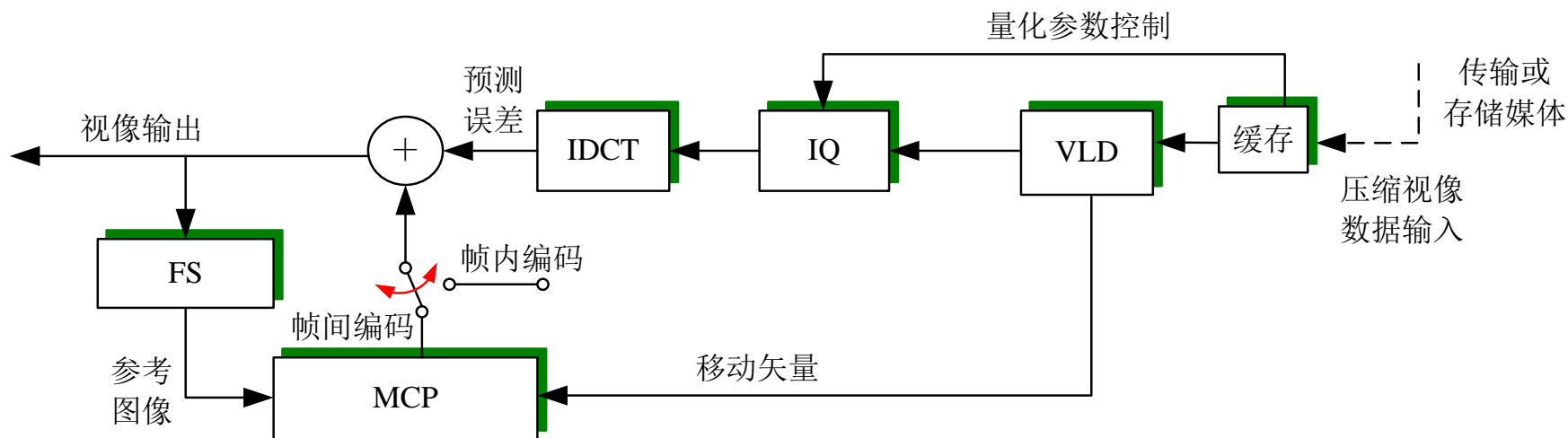
MPEG-2 视频



(a) MPEG-2编码器

MPEG-2 视频

(a) MPEG-2编码器



(b) MPEG-2解码器

Q(quantization): 量化

DCT(discrete cosine transform): 离散余弦变换
transform): 逆离散余弦变换

MCP(motion-compensated predictor): 移动补偿预测器

VLE(variable length encoder): 可变长度编码器

FS(frame memory): 帧存储器

IQ(Inverse quantization): 逆量化

IDCT(inverse cosine

ME(motion estimator): 移动估算器

VLD(variable length decoder): 可变长度解码器

MPEG-2 视频

- 视频数据位流的结构

- 结构与MPEG-1视频数据位流的结构类似

- 以子采样4:2:0为例的结构

- 一个视频序列分成G个视频组(GOP)

- 每个组包含P帧图像(picture)

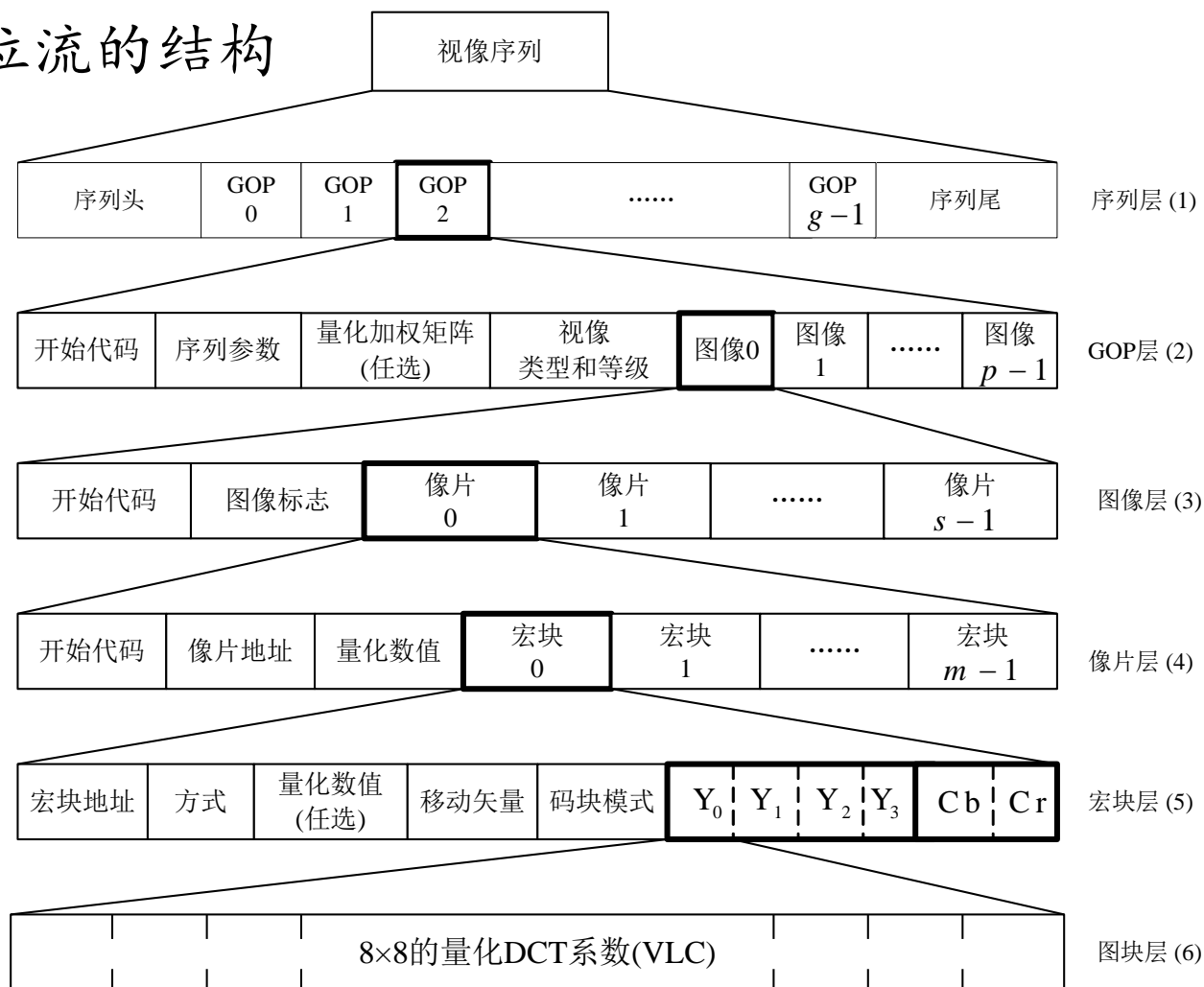
- 每帧图像分成S条像片(slice)

- 每条像片分成M个宏块(macroblock)

- 每个宏块包括4个 8×8 的亮度(Y)图块和2个 8×8 的色度(Cb, Cr)图块

MPEG-2 视频

● 视频数据位流的结构



MPEG-2 视频

- 视频质量可变编码

- 可变编码采用分层编码技术(layered coding)

- 基层编码(base-layer coding)/低层编码(lower-level coding): 编码、传输和解码可单独进行
- 增强层编码(enhancement-layer coding)/高层编码(upper-level coding): 编码、传输和解码要依赖于基层或先前的增强层

- 优点: 可提供不同等级的视频服务质量

缺点: 增加了编解码复杂性, 降低了压缩效率

MPEG-2 视频

● MPEG-2视频标准支持的可变编码方式

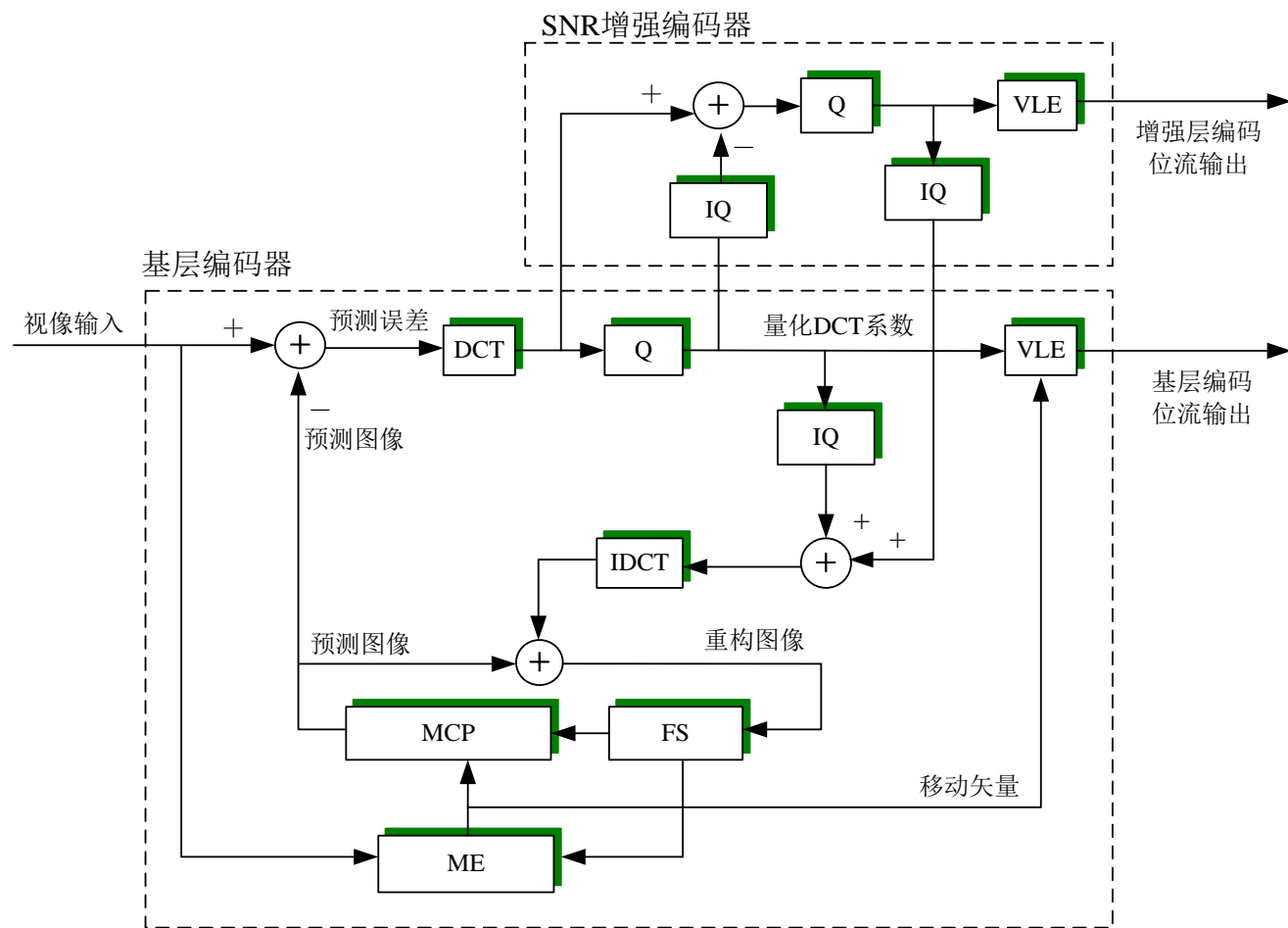
- 信噪比可变(SNR Scalability)编码：针对需要多种视频质量的应用，使用增强层编码提供比较高的信噪比
- 空间分辨率可变(Spatial Scalability)编码：针对需要同时广播多种空间分辨率视频的应用，用增强层编码提供比较高的空间分辨率
- 时间分辨率可变(Temporal Scalability)编码：针对从远程通信到HDTV以及需要有立体感视频的应用
- 数据分割(Data Partitioning)编码：针对有两个信道传输视频数据位流的应用，它将量化的DCT系数进行分割，编码后分别送到不同的信道
- 混合可变(Hybrid Scalability)编码：组合以上三种增强层编码中的任何两种编码，可获得不同性能的视频

MPEG-2 视频

- 信噪比可变编码

- 在基层编码基础上提高信噪比的技术。基层编码和增强层编码的视频有相同的空间分辨率，但提供质量不同的视频
- 在基层编码时，对DCT系数的量化比较粗，提供基本的视频质量。增强层编码对来自基层的DCT系数的量化误差进行编码，为基层的DCT系数提供精细数据，以提升视频质量

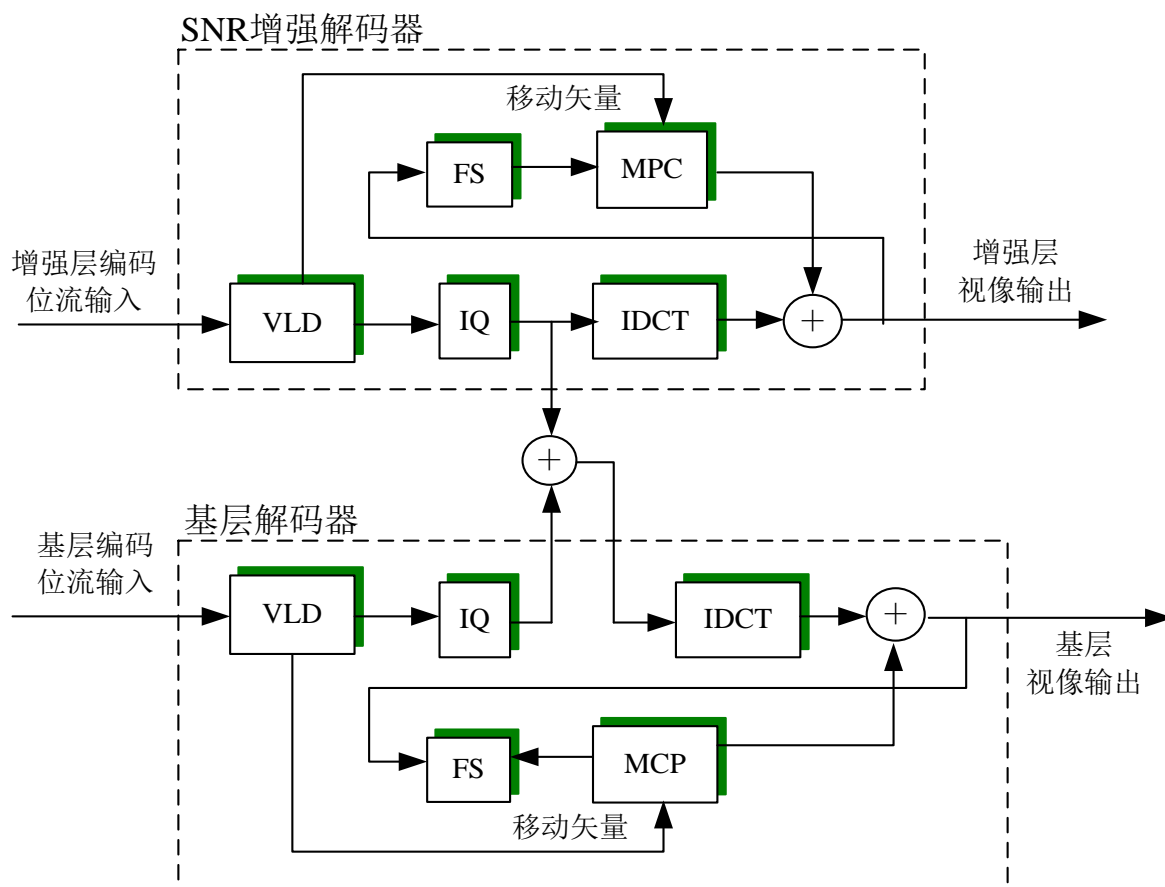
MPEG-2 视频



(a) MPEG-2 SNR可变编码器

MPEG-2 SNR可变编码/解码器结构

MPEG-2 视频



- DCT(discrete cosine transform): 离散余弦变换
- Q(quantization): 量化
- IQ(Inverse quantization): 逆量化
- IDCT(inverse cosine transform): 逆离散余弦变换
- FS(frame memory): 帧存储器
- MCP(motion-compensated predictor): 移动补偿预测器
- ME(motion estimator): 移动估算器
- VLE(variable length encoder): 可变长度编码器
- VLD(variable length decoder): 可变长度解码器
- Spatial Interpolator: 空间插值器
- Spatial Decimator: 空间抽样器
- Spatial Base Layer Encoder: 空间基层编码器

(b) MPEG-2 SNR可变解码器

MPEG-2 SNR可变编码/解码器结构

MPEG-2 视频

- 空间分辨率可变编码

- 基层编码和增强层编码的视频有不同的空间分辨率，基层提供基本空间分辨率，增强层使用来自基层的经过空间插值的视频数据

- 在解码器中可生成空间分辨率较高的视频

【例】基层编码的视频空间分辨率是 352×288 像素，经过空间插值后，增强层编码的视频分辨率可为 704×576 像素

MPEG-2 视频

- 空间分辨率可变编码

- 空间分辨率可变编码器的结构

- 空间基层编码器(Spatial Base Layer Encoder)与空间增强层编码器(Spatial Enhancement Layer Encoder)结构相同
- 空间抽样器(Spatial Decimator)/样本速率转换器(Sample Rate Converter)，用于降低当前帧的空间分辨率
- 空间插值器的输入来自空间基层编码器的重构图像，常用线性插值算法，如将 8×8 宏块生成 16×16 宏块

MPEG-2 视频

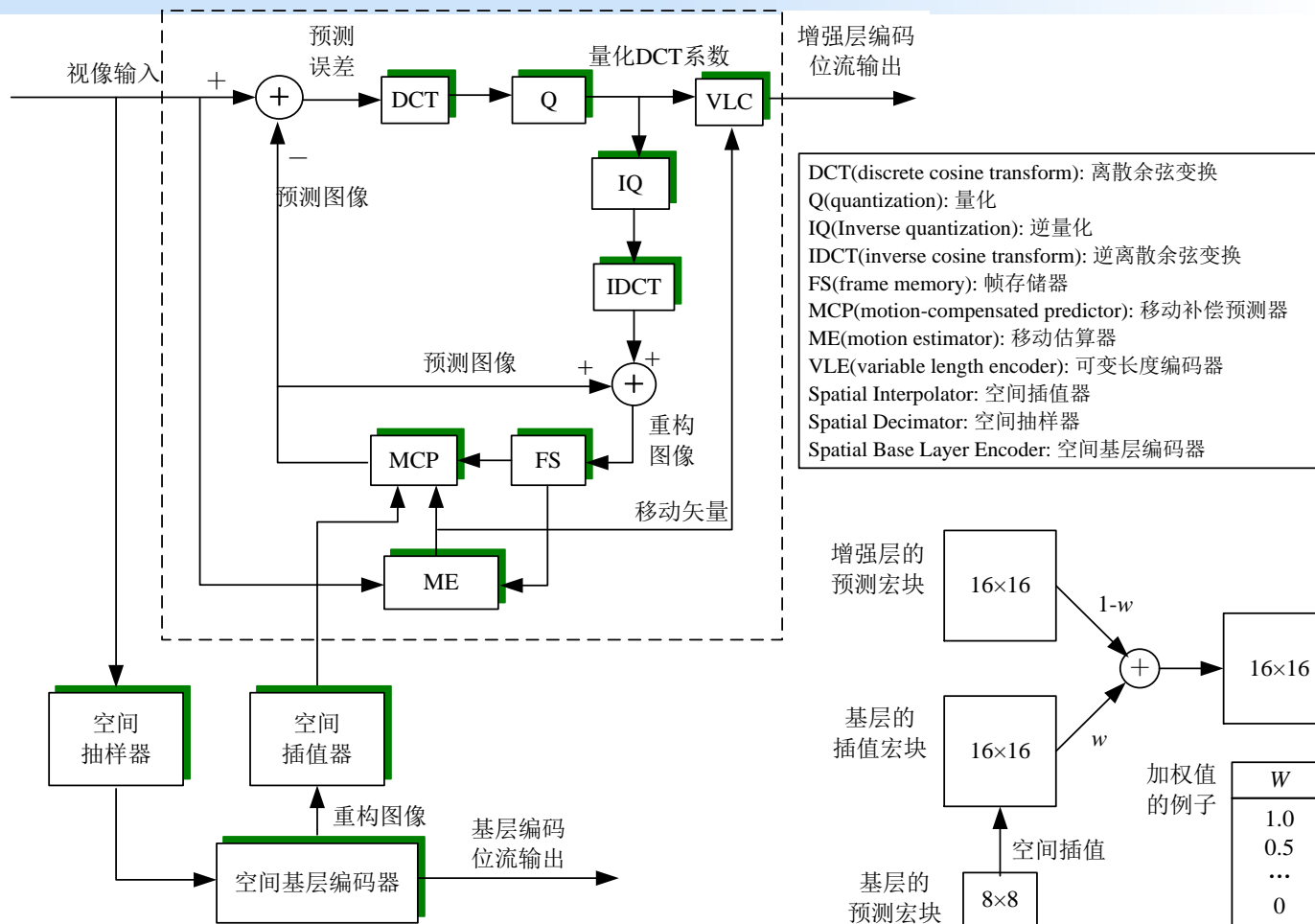
- 空间分辨率可变编码

- 增强层编码方法的示意图

- 来自基层编码器的预测宏块(如 8×8)通过空间插值后生成插值宏块(如 16×16)，插值宏块与来自增强层编码器的预测宏块(如 16×16)通过加权之后组合生成输出宏块。其中， w 称为自适应加权函数(adaptive weighting function)，取值范围为 $[0, 1.0]$

MPEG-2 视频

空间增强层编码器



(a) 编码框图

(b) 增强层编码方法

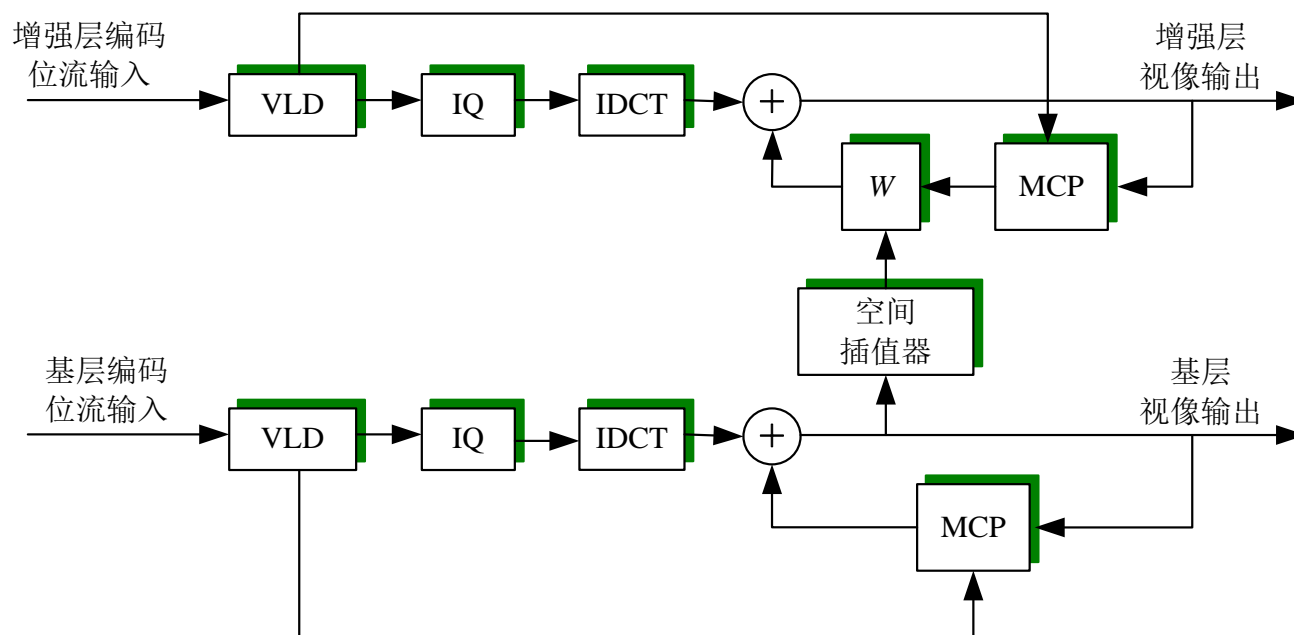
MPEG-2 空间分辨率可变编码/解码器框图

MPEG-2 视频

- 空间分辨率可变编码

- 空间分辨率可变换码器

- 工作过程与空间分辨率可变编码过程相反



(c) 解码框图

MPEG-2 视频

- 时间分辨率可变编码

- 指帧速率可变的编码，包含基层编码和增强层编码，各层的空间分辨率和颜色空间均相同
- 基层编码对帧速率较低的视频编码，增强层编码对低层的预测数据编码
- 图12-18 (a)：指派帧的方案，它将I帧和P帧指派给基层，偶数的B帧指派到增强层1，奇数的B帧指派到增强层2

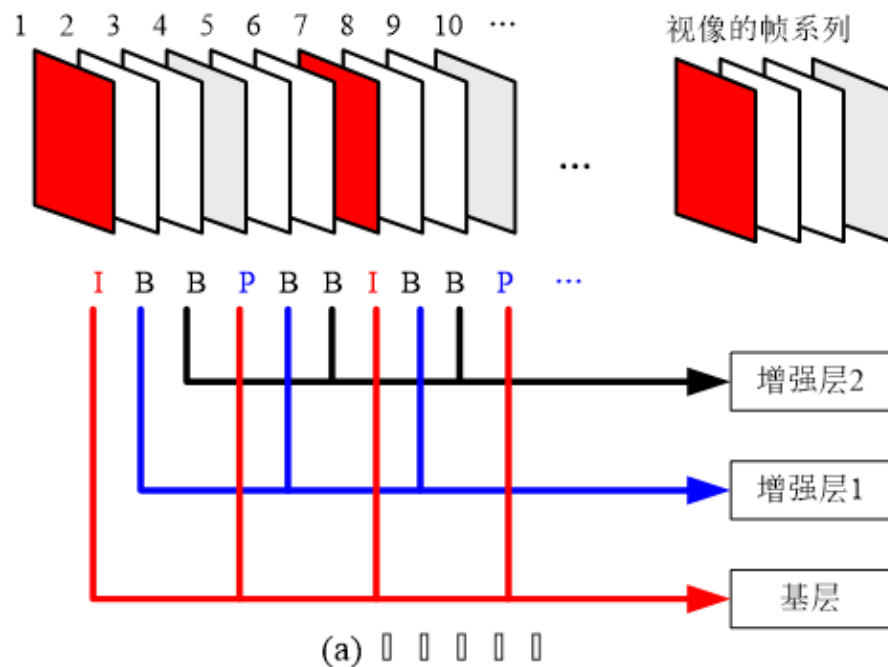
MPEG-2 视频

➤ 图12-18 (b): 3部分组成的编码系统

- 时间多路分解器(temporal demultiplexer): 将输入视频分解成2个系列, 分别送到基层和增强层编码器, 帧速率通常为输入的1/2
- 时间基层编码器(temporal base layer encoder)
- 时间增强层编码器(temporal enhancement layer encoder)

➤ 12-18图(c): 解码器框图

- 组合基层和增强层的解码数据, 生成较高帧速率视频
- 可达满帧速率(full frame rate), 如30帧/秒, 60帧/秒



- 时间多路分解器(Temporal Demultiplexer)
- 时间多路复合器(Temporal Multiplexer)
- 时间增强层编码器(Temporal Enhancement Layer Encoder)
- 时间增强层解码器(Temporal Enhancement Layer Decoder)
- 时间基层编码器(Temporal Base Layer Encoder)
- 时间基层解码器(Temporal Base Layer Decoder)

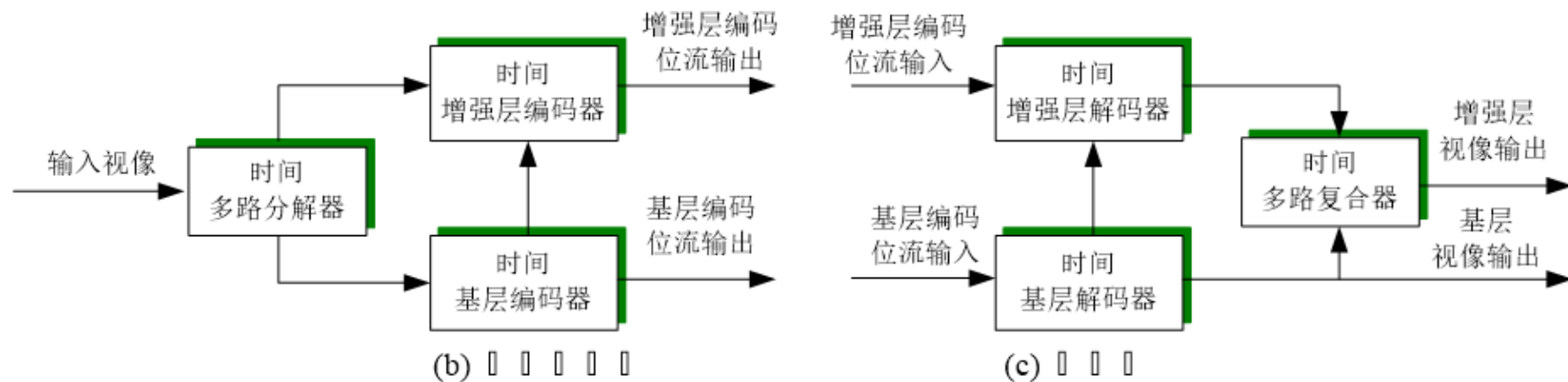


图12-18 MPEG-2 时间分辨率可变编码器框图[12]

MPEG-2 视频

- 数据分割编码

- 针对有两个信道传输视频数据位流的应用，将量化的DCT系数分割成两部分，编码后分别送到不同的信道
- 关键的视频数据(如在数据位流中的开始代码、移动矢量、频率较低的DCT系数)在性能比较好的信道上传输，不影响大局的数据(如频率较高的DCT系数)可在性能稍差的信道上传输

MPEG-2 视频

- 数据分割编码

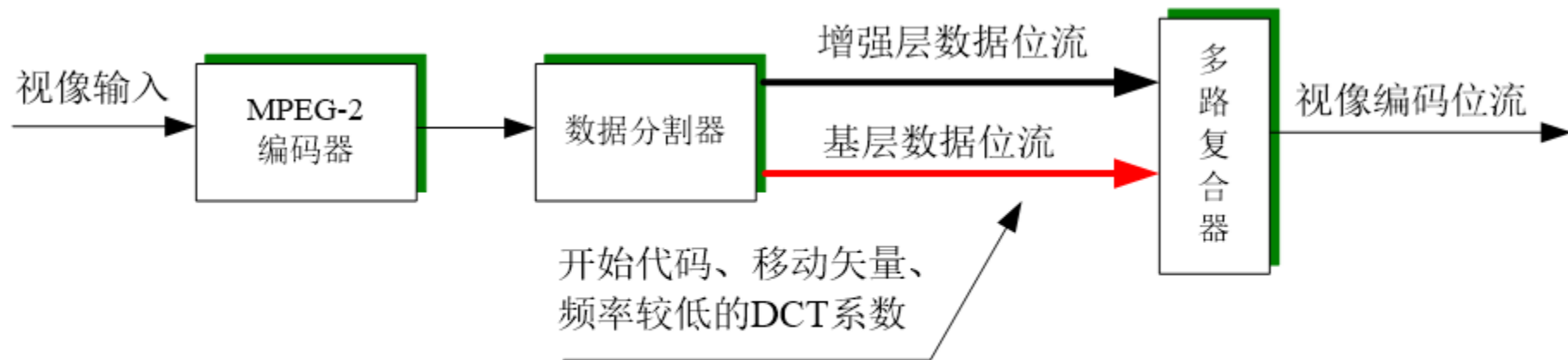


图12-19 MPEG-2数据分割编码器结构框图

本章提纲

- 为什么视频能压缩
- 谁在组织视频压缩编码
- MPEG-1视频
- MPEG-2视频
- **MPEG-4视频**

MPEG-4 视频

- MPEG-4 Visual是什么

- 1999年发布的可视对象编码标准，第3版修改于2009年，ISO/IEC 14496-2 Part 2
- 开发初衷：针对不同应用，提供数据率小于64 kbps、64~384 kbps和0.384~4 Mbps的视频压缩编码，用于自然对象、合成对象以及合成-自然对象混合编码(SNHC)
 - 自然对象编码：形状编码(shape coding)、纹理编码(texture coding)、移动编码(motion coding)和精灵编码(sprite coding)
 - 合成对象编码：图形、人脸面部活动，身体动作等

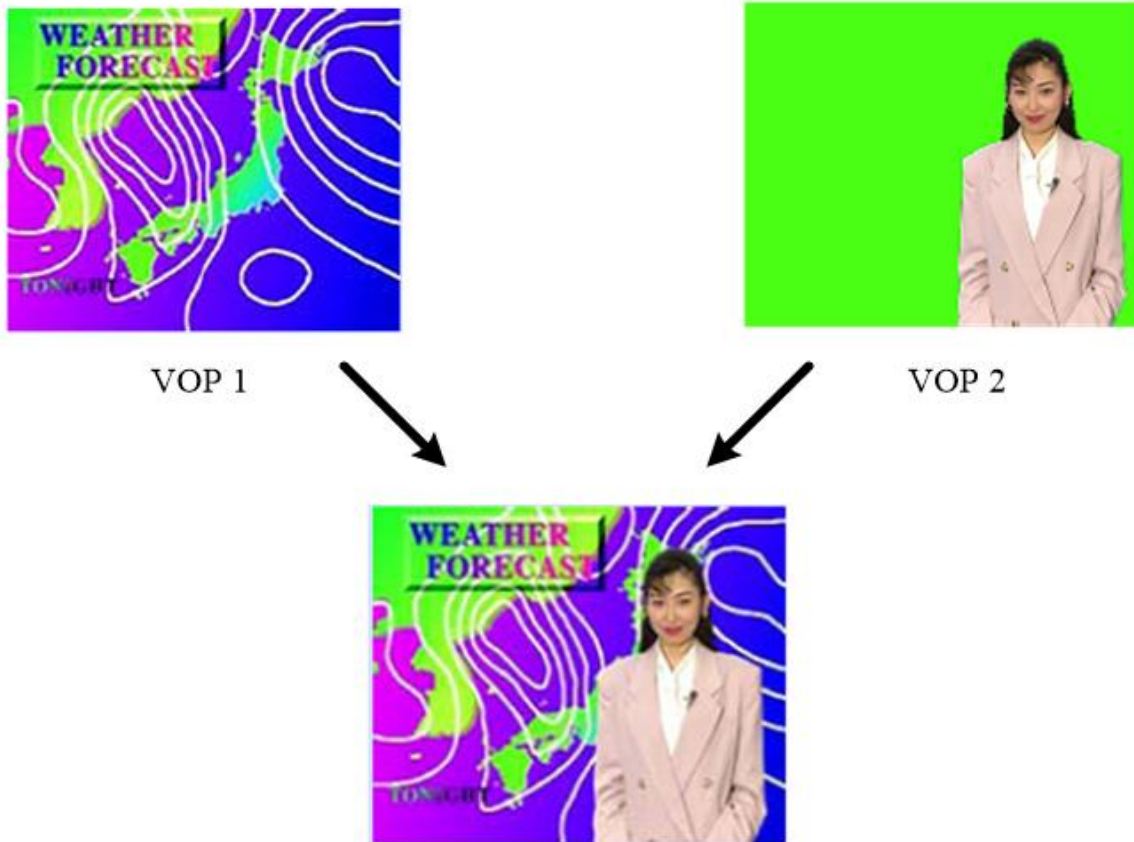
MPEG-4 视频

- 视频对象编码与解码概要

- 1. 视频对象平面的概念

- MPEG-4 Visual标准使用自然可视对象(natural visual object)和合成可视对象(synthetic visual object)代替MPEG-1和MPEG-2使用的帧(frame)
 - 把在给定时刻的视频对象的画面称为视频对象(平)面(video object plane, VOP), 见下图中的VOP 1、VOP 2和合成VOP
 - 对任意形状的编码难度较大, 采用折衷的办法, 用矩形把对象框起来, 如同MPEG-1,-2的帧, 以继承以前的编码技术和开发新的编码技术

MPEG-4 视频



视频对象平面(VOP)示图

MPEG-4 视频

- 视频对象编码与解码概要

2. 视频对象平面的结构

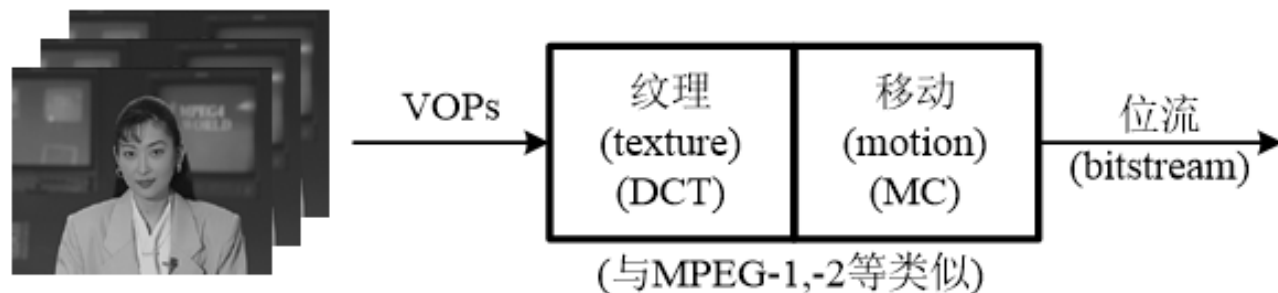
➤ 视频对象平面(VOP): 为支持MPEG 1, 2提供的功能, 包括各种分辨率、帧速率和隔行扫描图像的编码, 把视频帧当作视频对象平面(VOP)对待, 其画面当作纹理(texture)对待。于是MPEG-1, -2视频编码被认为由纹理编码和移动编码组成, 并将VOP编码器称为MPEG-4核心编码器(MPEG-4 core coder), 见图12-21(a)

MPEG-4 视频

- 视频对象编码与解码概要

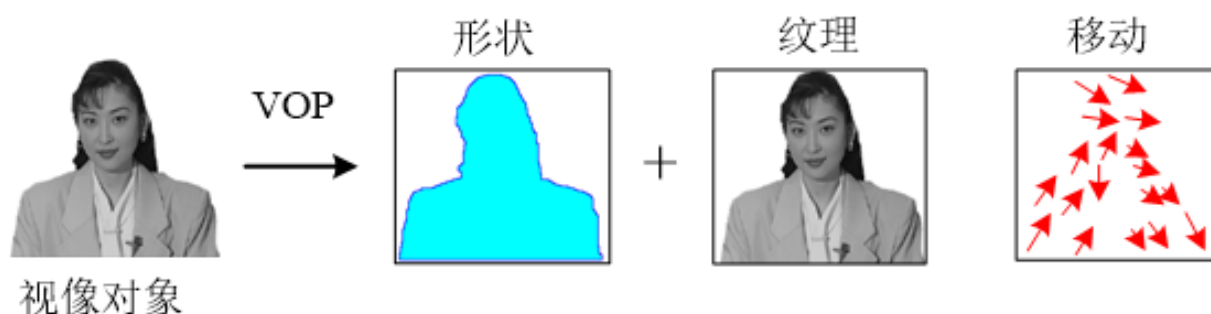
2. 视频对象平面的结构

- 可视对象：指使用分割算法从场景中抽取的单独的物理对象，它的VOP在空间上用其形状和纹理描述，在时间上用移动描述，见图12-21(b)。于是VOP编码被认为由形状编码、纹理编码和移动编码组成，并将它称为扩展MPEG-4核心编码器(extended MPEG-4 core coder)，见图12-21(c)

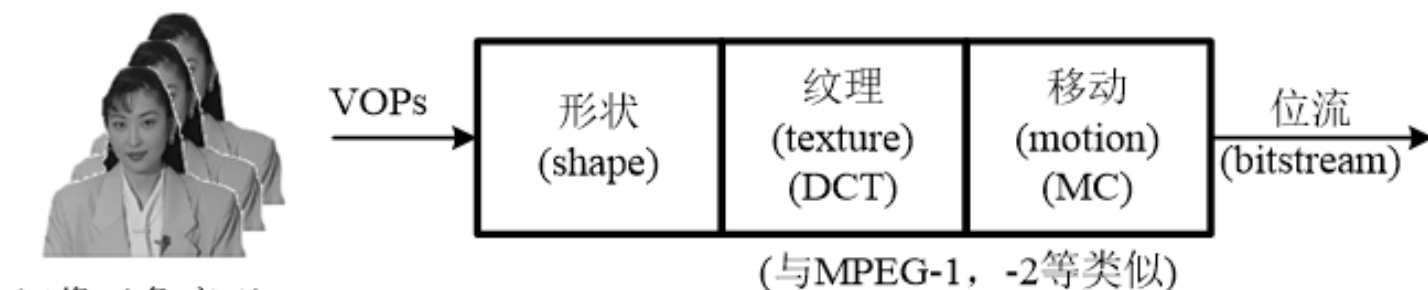


视像对象序列

(a) MPEG-4核心编码器



(b) 视像对象平面的结构



视像对象序列

(c) 扩展MPEG-4核心编码器

图12-21 视像对象平面的结构与编码器

MPEG-4 视频

- 视频对象编码与解码概要

3. 视频对象平面的类型

➤ 与MPEG-1, -2相应的类型

- I-VOP(帧内视频对象平面)
- P-VOP(预测视频对象平面)
- B-VOP(双向预测视频对象平面)

MPEG-4 视频

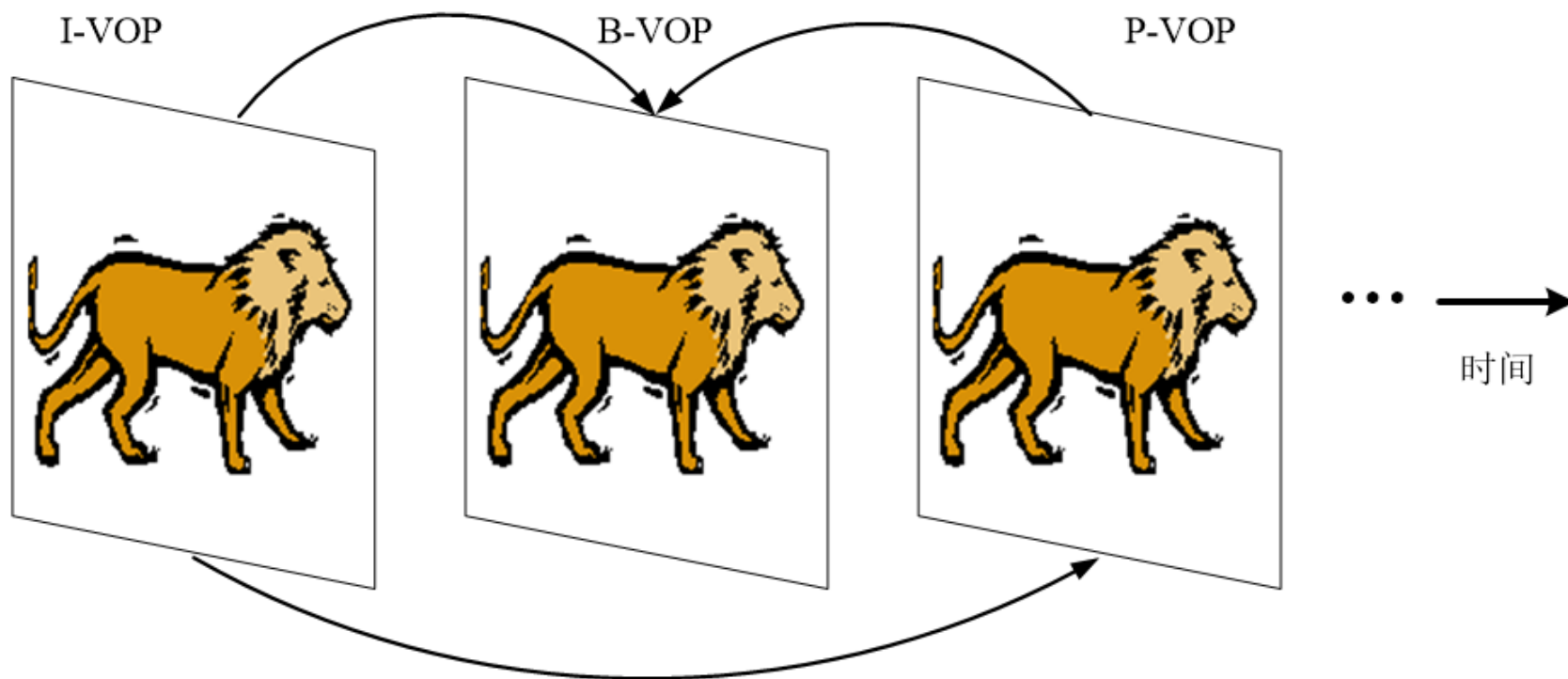


图12-22 VOP的类型

MPEG-4 视频

- 视频对象编码与解码概要

3. 视频对象平面的类型

➤ 两种精灵视频对象平面

- S-VOP(sprite-VOP): 对从静态精灵对象或参考VOP中获得的信息进行编码的图像。精灵(sprite): 指可由用户管理并可在屏幕上独立移动的图像, 广泛应用于动画序列和电视游戏
- S(GMC)-VOP: 使用全局移动补偿(global motion compensation, GMC)技术得到的预测编码图像。如对向左或向右移动的精灵对象, 相邻VOP之间的每个宏块(MB)使用完全相同的移动矢量(MV)

MPEG-4 视频

- 视频对象编码与解码概要

4. 编码器和解码器结构

➤ 图12-23：MPEG-4 Visual的编码器示意框图

- 虚线右边：背景图像对象平面(VOP1)、前景图像对象平面(VOP2)和前景文字对象平面(VOP3)相互独立编码
- 虚线左边：视频对象分割或称视频对象抽取不在MPEG-4 Visual标准范围

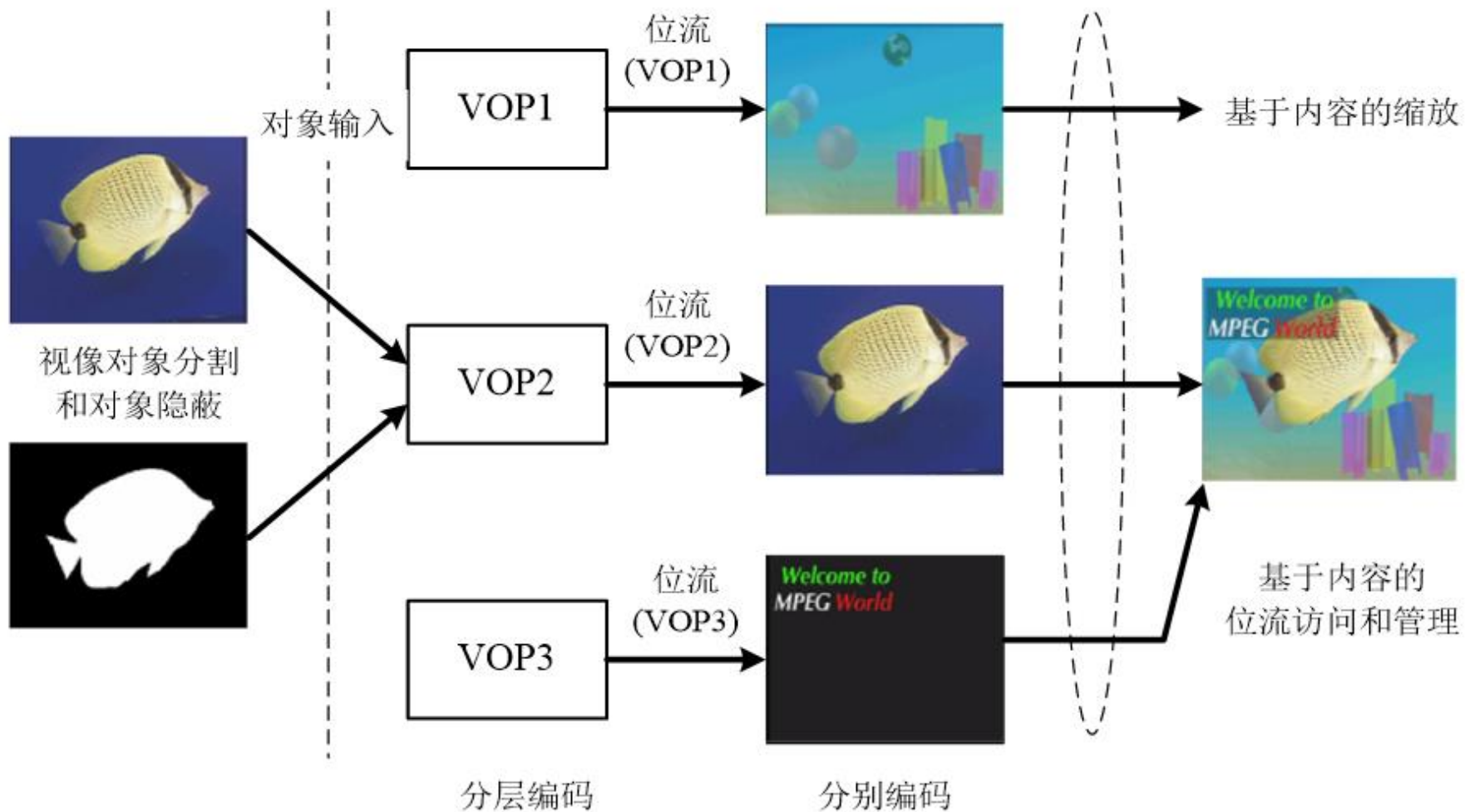


图12-23 MPEG-4 Visual编码器示意框图

MPEG-4 视频

- 视频对象编码与解码概要

4. 编码器和解码器结构

➤ 图12-24：MPEG-4 Visual的解码器示意框图

- 场景可由多个视频对象组成，每个对象用其空间和时间方向上的形状、纹理和移动参数刻画
- 输入解码器的视频位流经多路分解，得到形状、纹理和移动位流，通过各自的解码器解码，可得到重构的VOP
- 通过编排器产生有背景和前景的视频输出

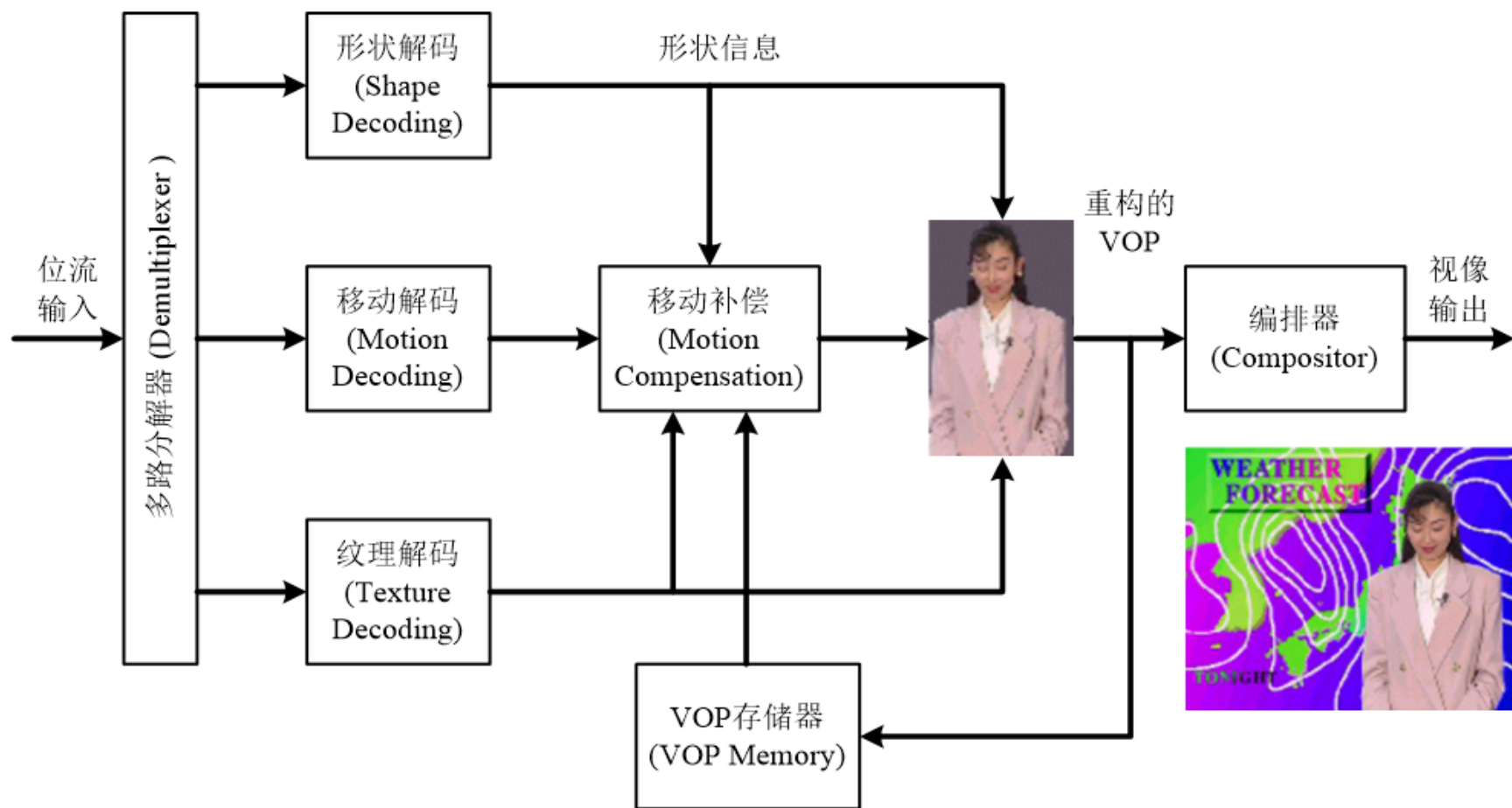


图12-24 MPEG-4 Visual解码器示意框图

MPEG-4 视频

- 可视对象的层次结构

- MPEG-4标准以视听对象为中心，便于互动应用和直接访问场景内容，将可视对象序列分成多个层次，见图12-25
- 可视对象序列(visual object sequence, VS): 构成场景的可视对象的有序集合，可含2-D、3-D的自然对象或合成对象
- 视频对象(video object, VO): 场景中的视频对象，最简单的对象是矩形的帧，也可以是任意形状的对象或场景的背景

MPEG-4 视频

- 可视对象的层次结构

- 视频对象层(video object layer, VOL): 对象可变分辨率表示方法, 单层表示分辨率不变, 多层表示空间分辨率或时间分辨率可变
- 视频对象平面组(group of video object planes, GOV): 多个视频对象平面的组合
- 视频对象平面(video object plane, VOP): 每个对象在时间方向上每次采样得到的画面, 并且单独编码

可视对象序列
(Visual Object Sequence)

视像对象
(Video Object)

视像对象层
(Video Object Layer)

视像对象面组(Group of
Video Object Planes)

视像对象平面
(Video Object Plane)

