

# 视音频压缩及处理

主讲 田捷博士

(研究员, 博士生导师)

Email: [tian@dr.com](mailto:tian@dr.com)

<http://www.digiark.com/tian>

第一部分 H. 261规范

第二部分 MPEG标准简介

第三部分 MPEG-2视频压缩标准

第四部分 MPEG-4规范介绍

第五部分 MPEG-7的发展

# 第一部分 H.261规范

§ 1.1 H.261概述

§ 1.2 H.261任务

§ 1.3 H.261图像格式

§ 1.4 H.261主要指标

§ 1.5 H.261信源编码算法

§ 1.6 H.261改进与扩展

## § 1.1 H.261概述

H.261也称 $P \times 64$ ，这是ITU-T（前称为CCITT）最早制定的关于视频编码的国际标准。考虑到ISDN以64kbps传输码率64kbps为单位，因此以 $p \times 64\text{kbps}$ 为H.261的标称码率。主要用于电视电话和电视会议电话。它提供QCIF、CIF两种图像格式。

## § 1.2 H.261任务

第一是编码算法问题。对于会议电视业务，参加会议的人员是在房间里开会，其动作量很小，仅仅是嘴巴说话的动作。确立了一种合理的、保证图像质量且为各国图像编码专家所公认的统一的算法。

第二是PCM标准的互换，确切地说是涉及PCM标准问题，编解码器工作于以 $64 \sim 1920 \text{kb} / \text{s}$ 的速率去覆盖N-ISDN或PCM一次群的通道。

第三，解决电视制式不同。PAL制式625行，每秒25帧；扫描方式是隔行扫描，每帧两场分为奇数场和偶数场，每场扫描 $625 / 2$ 行。NTSC制式为每幅图像扫描525行，每秒30帧，也是隔行扫描。

## § 1.3 H.261图像格式

为了使同一标准既能用于625 (PAL) 和525 (NTSC) 两种电视制式系统，源编码基于中间格式CIF(Common Intermediate Format)格式，但输入输出可以是数字或模拟的分量或复合信号。经转换到CIF或QCIF格式再进行源编码。CIF格式规定：亮度分量为 $352 \times 288$ 象素，色差分量为 $172 \times 144$

像素，图像尺寸的宽高比为4：3，这样，因为人眼对色度清晰度的分辨率本来就低，所以可利用这种心理视觉冗余度来减少色度像素数。亮度分量的纵向像素数恰好是两种电视制式场扫描行数的平均，因此转换也容易实现。

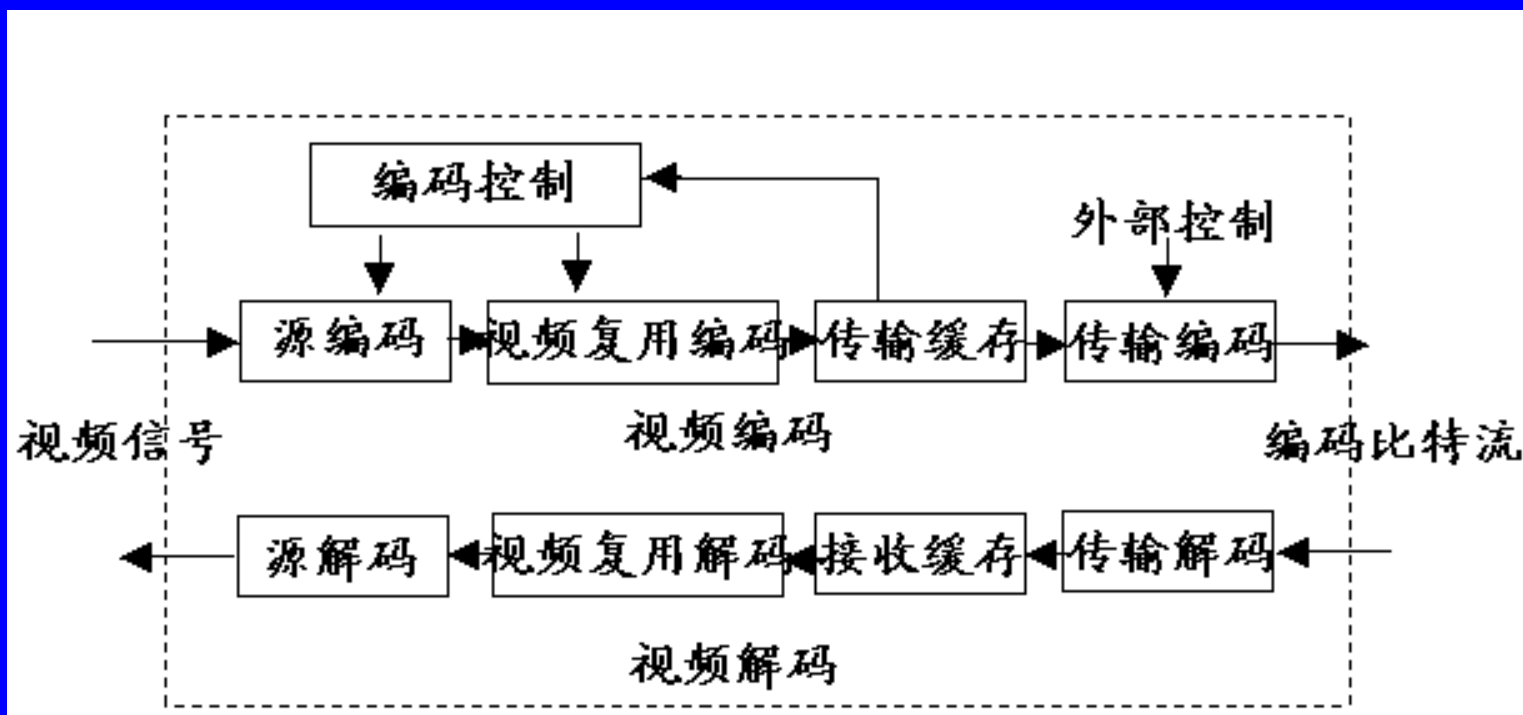
当电视电话在ISDN中传输时，若P取1或2，最高信道速率为128kbps，CIF规定的图像尺寸太大，因此定义了QCIF，亮度分量变为：176\*144，色度分量变为88\*72。所有的电视电话都必须达到QCIF要求，CIF则可以选择。

采用标准		CCIR601		ISO/MPEG-1		CCITT H.261	
参 数		PAL	NTSC	SIF		CIF	QCIF
每秒帧数		25	30	25	30	29.97	
每帧行数	Y	576	480	288	240	288	144
	Cr	288	240	144	120	144	72
	Cb	288	240	144	120	144	72
每行象素	Y	720		360		352	176
	Cr	360		180		176	88
	Cb	360		180		176	88

表1 视频编码图像格式一览表

# § 1.4 H.261主要指标

## ■ 编解码框图



图一 视频编解码框图

输入信号是CIF或QCIF格式的数字视频信号，视频编码输出端的码流进入ISDN传输网络。

源编码主要任务是对视频信号进行压缩，即DCT变换后进行量化；

图像服用复用编码任务是将每帧图像数据编码排成4个层次的数据结构，同时对DCT系数直流系数进行固定长度编码，对交流系数进行可变长度编码；

传输编码器中插入BCH正向纠错码和同步码；

编码控制器控制量化步长和编码模式，即控制帧内编码或帧间编码。

## ■ 传输速率

视频编码信号的传输速率为 $p*64$  kbps( $p=1-30$ ),即从64kbps到 1.92Mbps , 其算法必须能够实时操作 , 解码延时要短 ;

## ■ 视频输入输出信号格式

当 $P=1$ 或 $2$ 时只支持QCIF分辨率格式、每秒帧数较低的可视电话。对于64kbps码率要考虑图像用40kbps，余下16kbps安排语音；对于128kbps码率，语音可考虑16-64kbps，图像码率则为64-112kbps。时间轴分辨率为10-15帧/秒，编码单程最大延时为250ms，要考虑画面中嘴巴和语音的同步。所有编码系统都必须能处理DCIF格式，并要求有传送静止图像和图形的能力。

当 $P \leq 6$ 时,则可支持CIF图像分辨率格式（30帧每秒）的会议电视。

## ■ 误码处理方式

传输位流中包含有一个BCH吗，  
是否利用该码进行前向纠错可由解码器  
任选。

## § 1.5 H.261信源编码算法

信源编码算法的一般形式如下图所示，主要由帧间预测、帧内分块和量化组成。



首先将预测误差或输入图像划分成为 $8*8$ 的像素块。进一步，将四个亮度像块和两个在空间位置上与之行营的色差像块符合成一个 $16*16$ 的宏块（MB）。

信源编码器对每秒出现 $30000/1001=29.97$ 次的非隔行扫描画面进行运算。画面按一个来年高度分量和两个色差分量进行编码。

对于帧序列中的第一副图像或景物变换后的第一副图像，采用帧内变换编码：利用 $8*8$ 的DCT实现。各DCT系数经过线性量化、变长编码后进入缓冲器，根据缓冲器的上溢和下溢，来反馈调节量化器的量化步长，以控制视频编码位流使之与信倒速率相匹配。

帧间预测采用混合方法：利用运动补偿预测，当预测误差超过某个门限后，对误差做DCT、视觉加权量化（必要时可经过递归滤波器滤除高频噪声，以改善图像质量）及熵编码。运动矢量信息编码后也送到缓冲器中。

DCT去除空间冗余度，而使用有运动补偿的帧间预测来去除时间上的冗余。这是一个典型的帧内 / 帧间自适应预测加DCT变换的混合算法。

在帧间模式时，两个开关都打到下面。输入与经过运动补偿的前帧信号相减，差值经DCT，量化并进行可变字长编码后输出。从图中可以看到，在反馈环中有一个解码器即先反量化，反变化，然后再加上经运动补偿的前一帧，作为下一帧的预测

信号。在帧内模式时，开关都打到上面。这时输入信号直接跳过减法器，经DCT量化和VLC输出。这时运动检测和运动补偿不再起作用，只作帧内处理。

帧内编码的DCT直流系数仅用一个量化器，而交流系数则采用31个量化器。在一个宏块内，除了帧内直流系数外，所有的系数均采用一个量化器。量化器的判决电平并没有明确规定。

对于帧内编码模式，帧内的六个像块的系数均要传输；而在其他情况下，有宏块的类型和编码块没收来决定哪个像块要发送。量化后的系数按ZigZag方法读出后，最后再加一个像数块结束标记EOB后发送。数据读出时，按游程长度编码，最常见的方法是按连零个数(Run)和随后的系数值(Level)作可变字长编码。另一种是20比特固定长度码字，6比特退出码(Escape)，6比特连零数(Run)和8比特系数值(Level)。

运动补偿用于帧间编码；运动估值找到最佳匹配的宏矢量，计算预测误差，这是帧间编码的基本过程；滤波器采用环路滤波器，消除高频噪声。

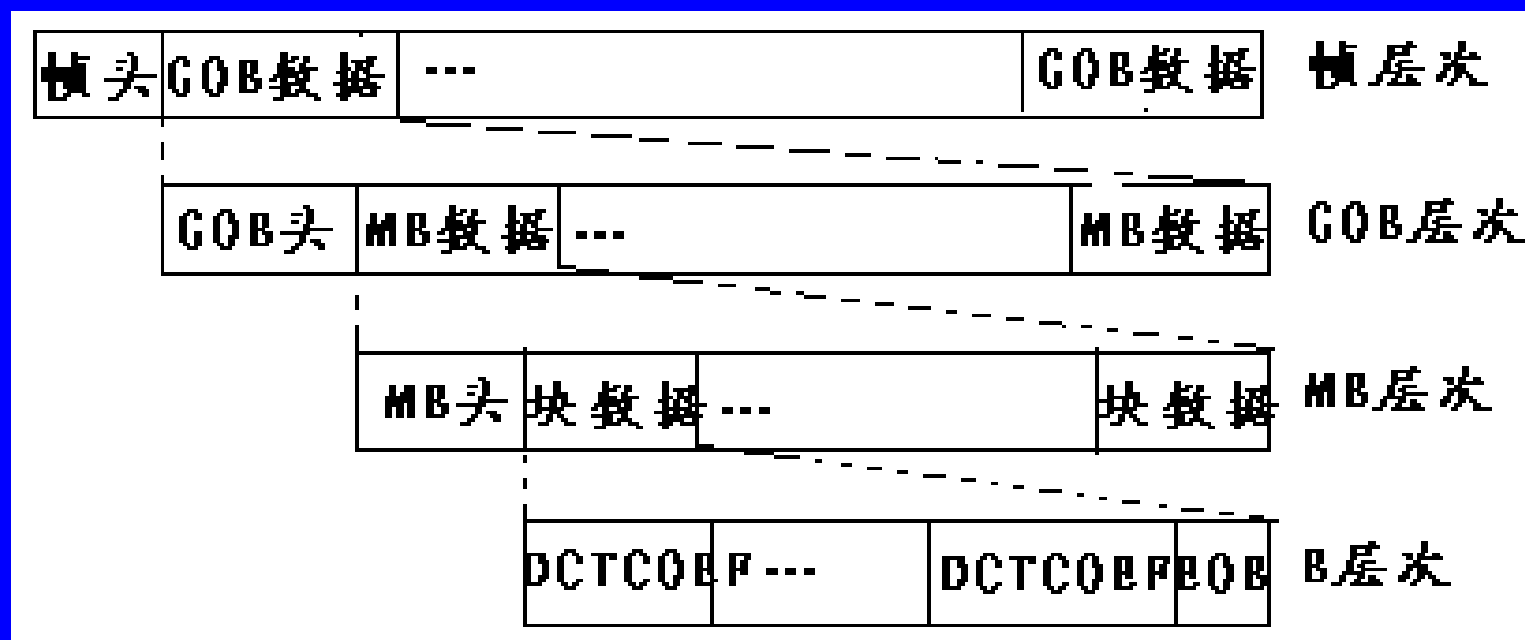
为了避免变化系数幅值的量化失真在编码器和解码器环路内引起算术溢出，插入一个限幅功能。对于重建图像，它简单地将超出0-255范围的像素值限制在这期间。

## ■ 图像复用编码

源编码后进行频分复用编码，实际上是把比特流分成图像(Picture)，像块组(GOB : Group of blocks)，宏块(MB : Macroblock)和像块(Block)，并附加相应的信息。按照CIF格式，每帧CIF图像包含12个GOB，每个GOB包含33个MB，每个MB包含4个亮度数据块和各1个Cb、Cr色度块，每个B包含8\*8象素

GOB 和MB 都是以一定的顺序存放的。

图像复用编码器把上述层次的数据按照一定的方式连列起来，就构成了1帧数据流。编码图如下：



图三 数据流结构

## ■ 传输缓冲器

传输缓冲器的目的是协调编码输出率和传输网络位率，使得充分利用网络的传输位率。当秒年米秒 位率过高时，缓冲区用做存储器，同时通知编码器提高量化步长，从而降低位率；当编码器位率过低时，缓冲器根据存储余量通知编码器降低量化步长提高位率。

## ■ 传输编码器

传输编码器的作用是传输前进行纠错编码，目的是当图像码字在传输过程中出现误差是，则在解码中予以纠正。这里使用的纠错码为BCH码（循环冗余校验码）。

## § 1.6 H.26改进与扩展

H.261在保证互通性和兼容性的基础上对视频信源编码的大框架和必要的编码传输格式做了严格的规定。但对那些与兼容性关系不大，而与图像质量息息相关的重要部分却未加限制。这样就留有充分的空间供使用者根据各自的应用需要进行充实和扩充。

ITU-T SG15于1990年成立了"ATM 视频编码专家组"，负责制定适用于B-ISDN信道ATM编码传输标准。该专家组于1993年11月与MPEG联合提出了H.262建议草案，用于数字存储介质和数字视频通信中图像信息的编码表示和解码规定。该标准向下兼容。3C业的不断发展和融合，正是这一系列的标准和国际建议超前进行技术导向并提供互连性和兼容性的保障。

归纳起来，图一的视频编解码的混合编码方法通过多种手段压缩图像序列中相关信息，主要的手段有：

- ❑ 利用二维DCT减少图像的空间域的冗余度；
- ❑ 利用运动补偿预测减少图像的时间域冗余度；
- ❑ 利用视觉加权量化减少图像“灰度域”的冗余度；
- ❑ 利用熵编码来减少图像的“频率域”的冗余度。

从而使视频图像的码率得到了较大压缩，也同时刺激了法国机上对视频编解码各模块硬件功能的开发与集成，把数字图像通信技术迅猛地推向了实用化。

# 第二部分

## MPEG标准简介

§ 2.1 MPEG-1视频标准

§ 2.2 MPEG-1音频系统介绍

§ 2.3 MPEG-1系统部分

MPEG是国际标准化组织的运动图像专家小组( Moving Picture Expert Group)制定的动态视频压缩编码国际标准，MPEG技术是近10年国际图像压缩编码技术的结晶，也是伴随信息时代应运而生的热门技术。MPEG标准：包括MPEG视频、MPEG音频和MPEG系统（视音频同步）三个部分。MPEG压缩标准是针对运动图像而设计的，可实现帧之间的压缩，其平均压缩比可达50：1，压缩率比较高，且又有统一的格式，兼容性好。

随着各种图像压缩方法的出现和应用，国际标准化组织（ISO）和国际电信联盟（ITU）致力于制定统一的规范和标准，先后推出了电视会议的H.261建议，静止图像压缩的JPEG标准，黑白二值图像压缩的JBIG标准，运动图像压缩的MPEG-1、MPEG-2、MPEG-3、MPEG-4、MPEG-7标准。

## § 2.1 MPEG-1视频标准

- ❖ 概述
- ❖ MPEG-1的应用
- ❖ MPEG-1目标
- ❖ MPEG-1影响
- ❖ MPEG-1视频压缩标准基本原理
- ❖ MPEG视频编码器源模式
- ❖ MPEG视频图像压缩算法的缺陷

## ❖ 概述

MPEG-1标准已经公布了三部分，即MPEG系统、MPEG视频、MPEG音频。MPEG-1的任务就是：将质量适当的图像包括伴音数据以可以接受的重建质量压缩到码率为1.5Mbps（其中视频约1.2Mbps，音频约0.3Mbps），并合成到单一的MPEG数据流，同时保证视频和音频内同步，MPEG数据应成为计算机数据的一种并和已有的数据（包括文字绘图等数据）在计算机内兼容，即同时显示和记录。并且把这些数据必须在现有的计算机网络和广播电视等通信网络中兼容传输。

MPEG-1采用SIF格式，图像质量优于家用VHS录像机，音频（双声道）接近CD质量。经过MPEG-1标准压缩后，视频数据压缩率为1/100-1 / 200，音频压缩率为1 / 6.5。

在制定MPEG1标准时，首先提出了一些基本原则：动像压缩与计算机应用密切结合构成多媒体终端；标准要给系统设计充分自由，质量必须适应广泛应用；努力延长标准寿命。MPEG-1采用SIF格式，帧速率为25 ~ 30帧/秒，码率为1.5Mbps（其中视频约1.2Mbps，音频约0.3Mbps），图

像质量优于家用VHS录像机，音频（双声道）接近CD质量。 MPEG-1标准只规定了编码输出和解码输入的二进制比特流的格式，因此MPEG-1的解码器就有了一定的限制，以便任何解码器都能解码任何厂家生产的MPEG数据比特流，MPEG-1的解码器应该在IDCT、反量化和运动矢量预测方面遵照MPEG-1规定的标准。

## ❖ MPEG-1的应用

MPEG的应用面十分宽广，作为动态视频压缩的国际标准，可应用于以下各方面：

1. 数字存储媒体。制作影视 VCD光碟，卡拉OK光碟，电影版交互式游戏；
2. 电影、电视节目的数字化永久存档；

3. 制作各种百科全书，应用于导游、导购、演示、培训、宣传节目等方面，并可供旅游、宾馆、房地产、教育、从事等多种领域应用。具有无损耗、低成本、易携带等优点；

4. 高质量会议的电视系统，其压缩数据能以文件的形式传送、管理和接收；

5. 图像通信和数字视频广播。

关于图像压缩有两个标准：静止图像数据压缩JPEG和电视电话会议电视图像压缩的H. 261，但是更为重要的是如何把计算机系统和广播电视系统结合起来建立一个统一的信心网络（包括各种终端，可以称之为多媒体网络）。即把多媒体中断连接起来的网络。因此，图像、伴音、存储、传输四个方面需要有一个统一的标准才能进行广泛的交流。

## ❖ MPEG-1目标

1. 像质量方面：普遍认为应高于电视电话的图像质量，可以被大家接受的是VHS录像机的图像质量和光盘CD-ROM的放像质量
2. 存储媒体方面：光盘、数字录音带、温盘、可写磁光盘
3. 传输码率：应符合计算机的传输码率，即1-1, 5Mbps，一般以1.2Mbps为宜。
4. 通信网络：能适应多种网络，ISDN、LAN
5. 满足不对称和对称应用：一次编码和无数次解码；同时编解码

## ❖ MPEG-1影响

1. 由于MPEG-1标准对图像、伴音、存储媒体和传输指定了统一的标准，故在这些方面均能刺激产品的加速发展。
2. MPEG-1标准的实现，必将加速世界通信网络的延伸和完善，并且使世界各地的有关产品和设备兼容连网，加快建网速度，推进产品竞争，加速技术发展；
3. 统一的标准降低了成本，加快普及
4. 信息渠道的扩展、畅通和多功能化，加速了世界各方面的生产力的发展

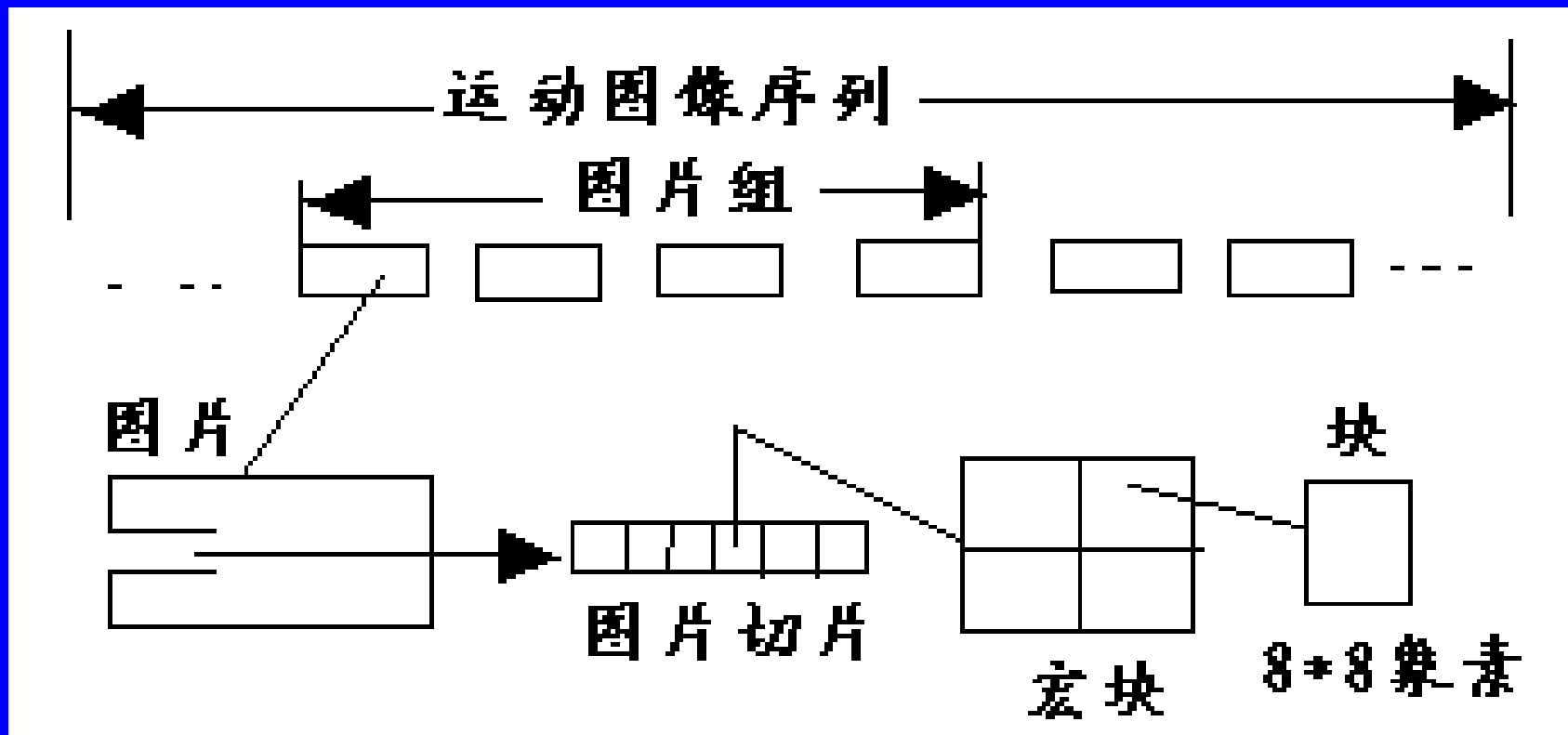
## ❖ MPEG-1视频压缩标准基本原理

MPEG-1系统与H.261建议的视频压缩算法有许多共同之处，只是在传输速率上不同。算法本身对于传输和随机存取而言是通用的，因前面的图2同样使用于MPEG标准。基本方法是：在单位时间内采集并保存第一帧信息，然后就只存储其余帧相对第一帧发生变化的部分，以达到压缩的目的。

- 一、数据流结构
- 二、编码图像的类型
- 三、视频流的组成
- 四、运动补偿

# 一、数据流结构

MPEG定义的运动图像序列即运动图像数据流包括一个表头、一组或多组图像及序列的结束标志码。图像采用SIF格式，类似与H.261建议，MPEG视频算法也使用 $16*16$ 宏块的运动补偿和 $8*8$ 像块的DCT。在数据流结构上有一个或多个连续的宏块组成所谓图像“切片”slice，切片的个数越多，误差越不容易察觉。



图四 MPEG数据体系结构

## 二、编码图像的类型

MPEG-1视频算法为了追求更高的压缩效率，更注重去除图像序列的时间冗余度，同时满足多媒体等应用的随机存取的要求，MPEG-1把视频图像序列划分为帧内图（Intra pictures）、预测图（predicted pictures）和双向图（bi-directional picture）。

1. 帧内图（I帧）仅使用类似JPEG的帧内编码，必须传送；

2. 预测图（P帧）要用过去的I或P帧进行预测，也叫前向预测。对于P帧的预测误差做有条件传送，同时还可预测B帧和下一个P帧。对P帧要做运动补偿，压缩比要高一些，但误差会扩散；

3. 双向帧：同时利用前后帧图像作为预测参考，B帧的压缩比高，由于该B帧本身不会作为预测基准，误差不会扩散。此外，利用两副图像进行双向预测的结果加以平均还有助于平滑噪声的影响。双向技术的引入是MPEG的一个最大的特色。

### 三、视频流的组成

IBBPBBPBBPBBPBBIBBPBBPBBPBBPBB

图五 典型视频流的顺序

由于双向帧的引入，使得MPEG-1视频流H.261更为复杂。为了满足不同的使用要求，MPEG-1采用了更为灵活的开放性的视频流：

1. 允许编码端自行选择I帧的使用频率和在视频流中的位置。

在那些需要保证图像的随机存取性、或要求对从感视频图像序列中截取的景物进行定位的应用场合，就希望I帧的选择有自由。典型地每秒钟I帧使用两次。

2. 允许编码端自行选择任何两帧参考图像之间的B帧数；

I、P帧间插入的B帧越多，编码器所需要的帧存储器也就越多。而用户对编码器的体积成本等因素的承受能力不尽相同，

同时，随着应用对象的不同，具体的图像序列的统计特性也会有有所不同。对于大多数景物，在参考图像之间插入两帧B帧较为适宜。

3. 编码端的视频流记录格式并不要求与典型图像显示顺序一致。

对编码器视频流记录格式的要求是使解码器的图像表示效率最高。与体地说，为重建B图像所需要的草靠帧均在响应的B帧之前发送。

## 四、运动补偿

采用运动补偿去除图像序列时间轴上的冗余度，典型地可使对P帧和B帧图像的压缩倍数较I帧提高三倍。运动步长以宏块为单位进行，包括预测和差补两种算法。

## ❖ MPEG视频编码器源模式

MPEG数字视频编码技术实质上是一种统计方法。在时间和空间方向上，视频列通常含统计冗余度。MPEG压缩技术所依赖的基本统计特性为像素之间(interpel)的相关性，这里包含这样一个设想：即在各连续帧之间存在简单的相关性平移运动。这里假定：一个特殊画面上的像素量值，可以(采用帧内编码技术)根据同帧附近像素来加以预测，或者可以(采用帧间技术)根据附件帧中的像素来加以预测。直觉告诉我们：在某些场合，如一个视频序列镜头

变化时，各附近帧中像素之间的时间相关性就很小，甚至消失。这时，该视频镜头就成为一组无相关性的静止画面的组合。在这种情况下，可采用帧内编码技术来开发空间相关性，来实现有效的数据压缩，MPEG压缩算法采用离散余弦变换(DCT)编码技术，以 $8 \times 8$ 像素的画面块为单位，有效地开发同一画面各附近像素之间的空间相关性。然而，若附近帧中各像素间具有较大的相关性时，也就是说两个连续帧的内容很相似或相同时，

就可以采用时间预测(帧间的运动补偿预测)的帧间DPCM编码技术。在多种MPEG视频编码方案中，若将时间运动补偿预测误差信息和变换码自适应地结合起来，就能实现数据的高压缩(视频的DPCM/DCT混合编码)。

- 一、二次取样和内插法
- 二、运动补偿预测
- 三、变换或编码
- 四、基本的MPEG-1帧间编码方

案

## 一、二次取样和内插法

视频编码技术在编码之前，均大量地进行了二次取样和量化。二次取样的基本概念是想减少输入视频的水平维数和或垂直维数，并在进行编码处理之前先对像素进行编号。值得注意的是：在有些应用场合，在时间方向上也对视频进行二次取样，以便在编码之前降低帧频。在接收机端，已解码图像是通过内插法来加以显示的。这一方法可以认为是一种最简单的压缩技术，这种压缩

技术利用了人眼特有的生理特性，因而去除了视频数据中含有的主观冗余度，即人眼对亮度信号的变化比对色度信号的变化更灵敏。故众多 MPEG 编码方案首先将画面分成 YUV 矢量信号(一个亮度分量和二个色度分量)；然后，相对于亮度分量，对色度分量进行二次取样，对于一些特殊应用，有一个 Y:U:V 比率(即对于 MPEG-2 标准，采用 4:1:1 或 4:2:2)。

## 二、运动补偿预测

运动补偿预测减小帧间的时间冗余度；并作为用于时间DPCM编码的预测技术，这一工具在MPEG-1和MPEG-2视频编码标准中得到广泛应用。运动补偿概念是以对视频帧间运动的估算为基础的，也就是说，若视频镜头中所有物体均在空间上有一位移，那么用有限的运动参数(如对于像素的平移运动，可用运动矢量来描述)来对帧间的运动加以描述。在这一简单例子中，一个来自前

编码帧的运动补偿预测像素，就能给出一个有效像素的最佳预测。通常，预测误差和运动矢量均传送至接收机。由于一些运动矢量之间的空间相关性通常较高，有时可以这样认为：一个运动矢量代表一个相邻像素块的运动。为了做到这一点，画面一般划分成一些不连接的宏块(在 MPEG-1和MPEG-2标准中一个宏块为 $16 \times 16$ 像素)，对于每一个这样的宏块，只对一个运动矢量进行估算、编码和传送。在MPEG压缩算法中，运动补偿

预测技术用来减少帧间的时间冗余度，只对预测误差画面(原始画面与运动补偿预测画面之间的差别)加以编码。总的来说，由于采用依据于前一个编码帧的预测，待编码的运动补偿帧间误差图像中像素之间的相关性就差了。

### 三、变换或编码

变换编码的目的在于去掉帧内或帧间误差图像内容的相关性，然后对变换系数进行编码，不是对画面的原始像素进行编码。为此，输入画面被分成不连接的b像素的画面块(即 $N \times N$ 像素)。由 $8 \times 8$ 像素组成的画面块的离散余弦变换(DCT)已成为一种最佳的变换，用于静止画面和视频编码。事实上，由于基于DCT的方法具有较高的抗相关性能，并能获得快速DCT算法，适用于实时应用，已在大多数画面和视频编码标准中加以使用。VLSI技术的运行速率适合较广泛的视频应用范围，故已商业化。

变换编码的主要目的是使尽量多的变换系数足够的小，同时，应尽量减少系数之间的统计相关性，目的在于减少对剩余系数进行编码所需的比特数量。一般来说，为了获得画面像素块的有用的近似再现，只要将少量DCT系数传送给接收机就行。然而，那些最高有效位 DCT系数集中在左上角(直流DCT系数)，且随着距离的增加，系数的有效牲就逐步下降。这意味着：与直流的系数相比，交流DCT系数在画面像素块再现时

的重要性就差一些。采用运动补偿预测，DCT变换的结果是，使DCT定义域中的时间DPCM信号实现简单的再现。从频率角度考虑，支流DCT系数与较低的空间频率有关，而交流DCT系数与较高的频率有关。这一特性在MPEG编码方案中被加以应用，以去除画面数据中所包含的主观冗余度，这一切是以人类视觉系统标准为基础的，即跟较高空间频率有关的再现误差比较，观众对较低空间频率的再现误差更加敏感，故在给定比特

率情况后，为了要改进解码画面的视觉质量，往往根据视觉(感觉量化)对系数进行频率自适应加权(量化)。

时间运动补偿预测和变换域编码的结合，被认为是MPEG编码标准的关键点。MPEG编码算法通常又叫作基于画面块的DPCM/DCT混合算法。

## 四、基本的MPEG-1帧间编码方案

基本MPEG-1(及MPEG-2)视频压缩技术的基础为：宏模块结构、运动补偿及宏模块的有条件再补给。

MPEG-1 编码算法以帧内编码模式(I图像)对视频序列的第一帧进行编码。每一个下一帧采用帧间预测法(P图像)进行编码。视频序列中的每一个彩色输入帧被分割成多个非重迭的"宏模块"，每一个宏模块包含4个亮度块(Y1, Y2, Y3, Y4)及两个色度块

(U, V), 这些数据块来自于亮度带和共址的色度带。在Y:U:V亮度与色度像素之间的取样比为4:1:1 以最近的前一帧为基础, 采用运动补偿预测法来对P图像加以编码。每一帧被分割成不连接的"宏模块"(MB)。对于每一个宏模块, 对有关4个亮度块(Y<sub>1</sub>, Y<sub>2</sub>, Y<sub>3</sub>, Y<sub>4</sub>)和两个彩色块(U, V)的信息均加以编码。每个包含8×8个像素。视频序列第1帧(I图像)以帧内(INTRA) 模式加以编码不参照任何一个过去帧或未来帧。在编码

器处，DCT 被加到每一个 $8 \times 8$ 亮度块和色度块上，在DCT变换后，该64个 DCT系数中的每一个系数量化(Q)，在宏模块中被用来对 DCT系数进行量化的量化器步长，传送给接收机。量化之后，直流系数(DC系数)的处理方法跟交流系数 (AC系数)的处理方法是不一样的，DC系数表示分量模块的平均亮度，可用微分DC预测法对DC系数加以编码。交流系数及它们位置的非零量化器值被折线扫描，并采用可变长度编码 (VLC)表对其进行扫描宽度熵编码。

为了对P图像进行编码，前面的I图像的第N-1帧，存储在设置在编码器和解码器内的帧存储器中。在宏模块中执行运动补偿(MC)-对于即将进行编码的那个宏模块，在第N帧与第N-1帧之间，仅对一个运动矢量进行估算。这些运动矢量被编码及传送至接收机。然后将 $8 \times 8$ DCT加入到包含在该宏模块内的每一个 $8 \times 8$ 块中，接着是对DCT系数进行量化(Q)，并进行扫描编码和熵编码(VLC)。对于帧内每一个宏模块来讲，量化

步长(SZ)是可以调整的，以便获得给定的目标比特率，并避免缓冲器出现溢流和下溢现象。 解码器采用反向处理，以便在接收机中再生第N帧的一个宏模块。对包含在视频解码器缓冲器(VB)中的可变长度字(VLD) 进行解码之后，就能再现(Q和DCT-1操作)预测误差像素值。

## ❖ MPEG视频图像压缩算法的缺陷

MPEG-1算法理论虽然已经非常成熟，但是在技术上却很难克服噪声、雪花和“鬼影”对压缩图像质量的影响，对于劣质的视频信号源，MPEG算法会将噪声和缺陷放大，使图形图像进一步恶化，这就使得MPEG压缩方法对视频节目源有一定的要求。MPEG压缩标准的压缩比可达200：1。这意味着99%的数据要被丢掉。因此MPEG压缩法将导致图像细节信息的丢失而产生压缩失真，

## ❖ MPEG视频图像压缩算法的缺陷

使压缩图像的质量有所下降。对于快速变化的背景，叶子沙沙作响的树林。波纹荡漾的水面（特别是有太阳光反射的水面）。MPEG压缩算法有时反到会产生负面影响，使压缩后的画面出现马赛克（小方块）现象，MPEG压缩算法对那些低比度、光线柔和、边缘变化缓慢的视频图像具有最佳的压缩效果。

## § 2.2 MPEG-1音频系统介绍

- ❖ MPEG音频系统系统基本框架
- ❖ MPEG-1编码参数标志信息
- ❖ MPEG-1音频编码器简介

MPEG建议的第三部分是音频的压缩标准，具体规定了用于数字存储媒体的高质量音频的编码与解码表示，适应的采样率包括32kHz、44.1kHz和48kHz。编码器的输入和解码器的输出与现行的CD、DAT等媒介所用的PCM标准兼容，也是目前mp3遵循的标准。

MPEG音频标准描述了一个二进制数据格式和一个解码变换函数，从1991年开始，MPEG音频压缩算法已经成为国际上公认的标准。该标准定义了比特流语法和解码规范。它的开放型的结构便于以后的扩展，也能很好地满足应用程序的要求，这种灵活性使得基于MPEG的音频系统总能与世界上最先进的技术同步，另外，解码系统也是通用的，这样产生的数据就不在依赖与生产厂家。

MPEG 音频系统把流入/流出串行口的二进制数据流描述为帧，不同厂家的系统能读和解码任何MPEG文件，当前的标准并未提供磁盘记录系统文件结构的标准。

MPEG压缩比是可调整的，它适应于单声道、双声道、立体声、联合立体声，提供三层结构的编解码。MPEG编解码能依照用户应用程序要求，可编程地提供辅助比特位，这是和解码器相独立的，对与传输特定信息是非常有用的。

利用该技术可将声音以1:12的高压缩比压缩成更小的文件同时保持相当逼真的效果。比如,一首容量为30M的CD质量的歌曲,压缩成MP3格式后仅为2M多.它与JPG文件一样是一种有损压缩,文件被压缩得越小失真越大.以前要制作MP3格式的文件需要使用昂贵的硬件,随着机器性能的飞速提高,如今普通的奔腾微机足以胜任MP3文件的制作。

## ❖ MPEG音频系统系统基本框架

MPEG音频系统的核心是利用了心理声学模型。研究表明，如果不利用掩蔽效应（人的说话间歇会听到噪音，而在说话时则听不到噪音），要想在96kbps的码率上达到CD的音质是不可能的。因此，为了取得良好的压缩效果，MPEG-1标准强调对于人的听觉生理和心理特性的利用，在基本系统中包含了一个心理声学模型。编码器的算法并

未标准化，它可利用诸如估算听觉掩蔽阈、量化、标度化等各种手段，但输出必须保证使符合本标准的解码器能够解读的声音。为此MPEG-1引入了级（layer）的概念。级数越高，性能越好。

各级（LAYER）采用的编码算法如下：

Audio Layer I ——包含将数字音频输入分解成32个子带得计本映射、将数据按一定的格式分块的固定分割、确定自适应比特分配的心理声学模拟、以及使用块压扩和格式化的量化；

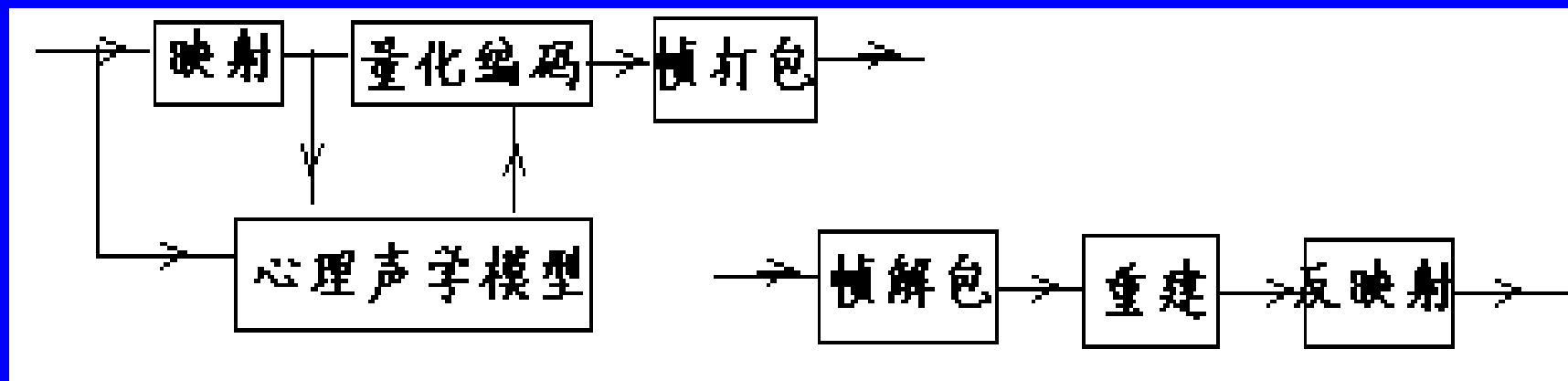
Audio Layer II ——提供对比特分配、标度因子和取样附加码，使用了差帧；

Audio Layer III ——基于一个混合滤波器组将频率分辨率提高，增加了非均匀量化器、自适应分割以及对于量化值的熵编码。

对联合立体声的编码可以作为一个附加特征添加到任何一级中。

Audio Layer III音频压缩方案可以解码Audio Layer I、Audio Layer II的音频流。Audio Layer III是当前ITU-R推荐在60kbps使用的唯一的音频压缩规范，Audio Layer III广泛地用于ISDN网络和电视应用、数字卫星广播。

MPEG-1音频系统工作原理：



编码器

解码器

图七 MPEG-1音频系统编解码基本系统简图

**映射**：产生输入音频数据流的滤波或亚采样表示。映射输出样本既可以叫做子带取样也可以叫做变换后的子带取样（例如在 Layer III）；

**心理声学模型**：产生一串数据，用来控制量化和编码，这些数据随具体编码器的实现而不同。一个可能的量化器控制方案是利用掩蔽阈的估值来实现；

**量化编码**：把经过映射的输入样本转换成一串编码符号。

**帧打包**：把其他模块的输出数据编排成实际的位流，还可根据需要加入其他信息（纠错信息、同步信息等）；

基本的解码器如图所示：它首先将输入的位流解包，如果编码器采用了差错控制措施，就接着调用解码模块进行误码检测。从解包后的位流数据中几恢复出了不同的信息段。重建模块则恢复出一组映射样本的量化值，将其通过反隐射模块即可变换成均匀的PCM信号。解码器所需要的与具体引用有关的全部信息，就可以按MPEG-1的规定格式从编码流中提取。

各种数据流例如已经编码的视频、音频、同步数据、系统数据、辅助数据等都可以共同保存在同一存储截止中，但如果对音频的编辑点和寻址点包出一致进行限制，那么从符合数据流中对音频进行编辑就更方便些。

## ❖ MPEG-1 编码参数标志信息

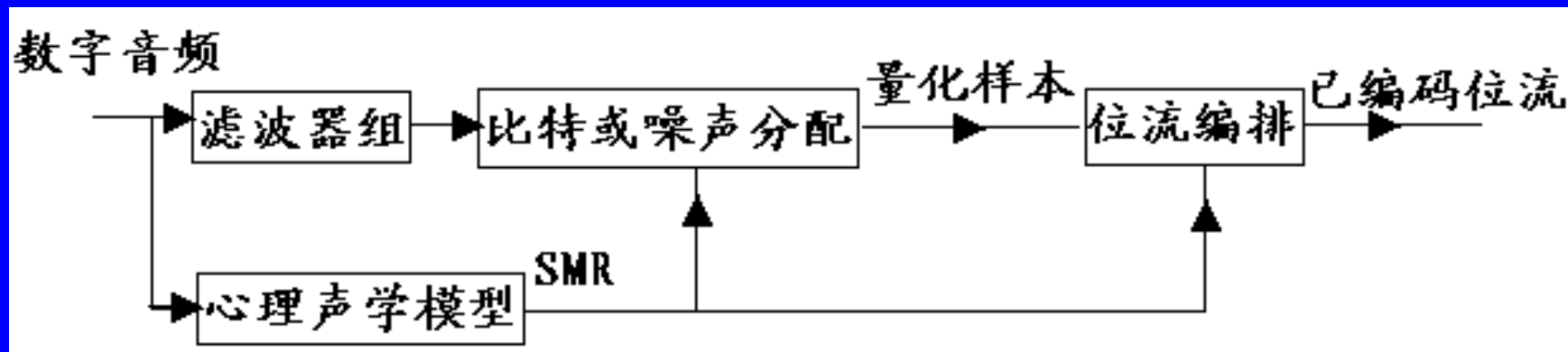
MPEG-1 码流中包含了丰富的信息，设计到关于数据格式以及其它方面许多席位而具体的规定。32 比特音频帧的头信息包括了系统的同步与状态信息，而且对三个编码级别都适用。解码器可从中提出必要的信息，以决定解码算法的选择以及解码器的重新设置。

## 表2 音频帧头格式

功能	比特数	格式与含义
同步字	12	1111 1111 1111
算法标识	1	是否为 MPEG-1 音频
级别标识	2	三个级别中的哪一级别
纠错标识	1	是否添加了误码保护冗余
码率标识	4	分别定义了 15 个码率
取样频率	2	44.1 kHz、48 kHz、 32 kHz
缓冲标识	1	帧内是否包含将平均码率调至取样率的附加 Slot
专用比特	1	保留位
模式识别	2	立体声、联合立体声、双声道、单声道
模式扩充	2	仅用于联合立体声
版权识别	1	有无版权保护
加重标识	1	指示应采用那种去加重措施

## ❖ MPEG-1 音频编码器简介

MPEG-1 音频算法是一个心理声学算法，其主要部分如图七所示：



图八 心理声学编码算法主要部分

- 一、滤波器组
- 二、心理声学模型
- 三、比特或噪声分配
- 四、位流编组器

# 一、滤波器组

完成从时间到频率的映射。MPEG-1音频算法使用两种滤波器组，每组提供一个特定的时间-频率映射并被临界取样（即分析域的样本和时间域的一样多）。滤波器组为编码器提供了基本的频率分割，而对于解码器则用做重建滤波器。滤波器组的输出样本被量化。在第I级和第II级，滤波器组使用了32个子带，每个子带内有12或36个样本编为一组处理；而在第III级，滤波器组的分辨率有信号有关，可以有 $6*32$ 或 $18*32$ 个频率样本。如果采用 $6*32$ 个样本，则每一频带的三组样本分别量化；

## 二、心理声学模型

计算出滤波器组每一频带内的“刚好可察觉的噪声级”，用语比特或噪声分配以确定实际的量化器和量化步长。MPEG-1建议了两种可选的心理声学模型，但第I、II级采用第一种模型；而第III级采用第二种模型。两种模型的最终输出都是信号掩蔽比（SMR）。第III级输出的是一组频带的SMR。

### 三、比特或噪声分配

分配器监视着滤波器组的输出样本和心理声学模型输出的SMR，据此调节比特分配或噪声分配。以便同时满足比特率要求和掩蔽要求。分配器输出都是一组量化参数和量化样本。

## 四、位流编组器

将滤波器组的量化输出、比特分配或噪声分配以及其他所需的边信息编码，并以高效的方式按一定固定的格式编组这些信息。每一级所用的编组器都不同。第I、II级对每一子带样本都使用固定PCM编码，唯一的差别就是第III级的量化样本可以成组编排，对量化的频率样本采用变长的霍夫曼编码方法，效率高一些，复杂程度也相应地高一些。

此外，MPEG-1还建议在编码编码器输入端包括一个截止频率为2-10Hz的高通滤波器，以避免最低频带不必要的高比特率要求并提高总的音频质量。

## § 2.3 MPEG-1系统部分

MPEG-1系统是把编码后的视频和音频数据流合并到一个单一的数据流的技术规范。该规范提出了视频和音频完全同步的建议，有利于信号的存储和不同传输介质中更远距离的传输。

系统编码部分的输出码流包括了实现把编码后的视音频数据流同步系统级功能所需要的必需的和足够的信息，同时，系统编码管理编码数据缓冲区，防止出现上溢和下溢现象。

编码层规定了多种数据格式，允许多个视音频数据流在数据流中可以以多路技术存在，也可以单独存在。

MPEG系统编码的基本原则是利用时间标记，这些时间标记规定了视音频的解码和显示时间以及多路编码数据在解码器端的接受时间，所有的都以90khz的系统时钟为基准。这种方法使得在解码器设计、流的个数、多元数据包的长度、视频图片的速率、音频采样速率、编码数据流速率、数字存储介质和整个网络的性能等诸多方面有很大的灵活性，同时也支持可变的数据速率操作。

MPEG系统规范中对解码器系统定了一个参考模型，这个参考模型对编码参数作了限制，也对解码器提出了一定的要求。另外，还有一些可选的限制项对一般解码工业生产提出了一个基本框架。MPEG系统具有很强的兼容性，它的设计便于将来在视音频以及超媒体编码方面的扩展。

# 第三部分

## MPEG-2视频压缩标准

§ 3.1 背景介绍

§ 3.2 概述

§ 3.3 MPEG-2的视频体系

§ 3.4 MPEG-2分级编码

§ 3.5 MPEG-2 与MPEG-1的比较

## § 3.1 背景介绍

MPEG第二阶段的国际标准化活动开始于1990年7月，同时ITU-TS也开始了采用异步转移模式ATM上音视频编码建议H.262的制定工作，并决定与MPEG-2联合开展工作。

1993年的下半年，美国的“高级电视联盟”和欧洲的数字视频广播DVB计划已先后决定将MPEG-2建议用于自己的HDTV广播；日本邮电数字广播研究组在阶段性研究报告中也建议采用OFDM传输方式和MPEG-2压缩方法。ISO正式于1993年11月的汉城会议通过了有关MPEG-2的有关建议。

## § 3.2 概述

- ❖ MPEG-2基本要求
- ❖ 更高要求
- ❖ 主要任务
- ❖ MPEG-2所表现出的优势

## ❖ MPEG-2基本要求

- ✓ 视频图像格式采用720\*480像素，与CCIR-601规定的演播室用数字电视像素数相同，是MPEG-1的四倍；
- ✓ 传输码率为4-10Mbps，是MPEG-1传输码率的四倍；
- ✓ 兼容MPEG-1；
- ✓ 主要用于卫星广播和有线电视领域、DBS (Direct Broadcast Satellite)、CATV (Cable Television)、DigiCipher、HDTV

## ❖ 更高要求

- ✓ 具有更低的成本
- ✓ 快速高效的操作性，可以更好的利用人力资源
- ✓ 对于节目制作和编辑，具有更大的创造性和灵活性
- ✓ 可通过辅助和不同的通道进行传送
- ✓ 在国际范围内，节目的相互交换性
- ✓ 从经济的低比特率到高清晰度图像质量的可扩展性

## ❖ 主要任务

制定通用的活动图像以及伴音的编码标准，以适应更为广泛的应用。

## ❖ MPEG-2所表现出的优势

### ➤ 开放的标准

它使所有使用者，既可获得高标准的质量和可操作性，而且并不依赖于任何一个厂商。

### ➤ 与外部网络的兼容性

由于考虑压缩方案的潜力和从前期到传输的简化，广播电视业界决定为电讯和卫星链路采用与MPEG2 MP@ML兼容的MPEG2 4:2:2P@ML。

## ➤ 技术成本低

由于从民品应用，到广播级和高清晰度都将采用MPEG2编码的现实，以及MPEG2 4:2:2编码器已经可以从一些有实力、独立的芯片制造商那里得到。这样，较低的元器件成本，使得在广播领域采取MPEG编码技术将会更加具有吸引力。

## ➤ 产品成员间的互操作性和灵活性

灵活的比特率允许在无转换损失的前提下，存储一定范围内的信号源，并同时具备优化存储效率和由用户确定质量等级的特点。解码器的灵活性和比特率转换的智能化能够允许不同的比特率和不同MPEG家族成员共存在一个环境下，并能够进行无缝连接。

## ➤ 比特流及网络的兼容性

一个能够在文件传输和网络协议支持下，在包括磁带、磁/光盘等存储介质上进行存储，必需有一个先进的数字基础结构。这个基础结构应是可扩展的，以适应未来低成本的非广播档，高质量的广播档以及未来高清晰度的需要。为了达到这些MPEG世界的目标，从民品到广播级，甚至更高领域的开发工作也在一直进行中。

## ➤可扩展性-由低到高的比特率

无论在高清晰度还是在低比特率方面，MPEG2都具有可扩展性。既可以在低比特率下进行编辑和图像处理，又可以实现经济高效的扩展GOP。所有这些都共存于一个家族内，低成本和高质量的压缩技术还会不断涌现出来。

## ➤大量工业厂家的支持。

## § 3.3 MPEG-2的视频体系

MPEG-2系统要求必须保证MPEG-1系统向下兼容，但为了满足从广播、通信、计算机到家用电子产品这样广阔的需求以及欧美不同的HDTV体系，制定了分辨率从低（会议电视/电话，H.261建议）、基本（MAIN，数字电视）、到高（HDTV）的四级（level,指分辨率）五类（profile,压缩比）共11种单独的技术规范集。

## 表3 四级图像标准

级别	最大采样维数 帧频	象素/秒	最大比 特率	规范
低级 low	352x240x30	3.05 M	4 Mb/s	CIF, 磁带设备
基本 main	720x480x30	10.40 M	15 Mb/s	CCIR 601, 电视
高 1440 High1440	1440x1152x30	47.00 M	60 Mb/s	4x 601, consumer HDTV
高级 high	1920x1080x30	62.70 M	80 Mb/s	production SMPTE 240M std

## 表4 MPEG-2中的类profile

类名	描述
简单 simple	与 main 相同，只是没有 B 帧图片，应用于 CATV
基本 main	多数解码芯片标准，CATV、卫星，有 95% 的用户
信噪比可调 main+	SNR 可调的 main
空间可调 main+	空间可调 main
增强 next	4:2:2 宏块的 main+

在下表中，四种技术规范是最为重要的：

表5 MPEG-2的图像规范

分辨率 ↑	种类		信噪比可调	空间可调	增强型
	级别	简单型			
高级	不用	<u>MP@HL</u>	不用	不用	HP@HL
高 1440	不用	<u>MP@H1440</u>	不用	<u>SSP@H1440</u>	HP@H1440
基本	<u>SP@ML</u>	<u>MP@ML</u>	<u>SNP@MP</u>	不用	HP@ML
低级	不用	<u>MP@LL</u>	<u>SNP@LL</u>	不用	不用

支持的压缩算法 →

➤ 基本级的基本型（Main Profile at Main level简写为MP@ML）。

将不同领域的共性应用集于一身，虽然基本的解码器功能实质上与MPEG-1相同，但图像质量要比现有的电视信号要好，而且还可用于数字影碟以及其他应用；

➤基本级的简单型（ Sample Profile at main level简写为SP@ML）

将用于数字电缆电视以及数字盒式录像机。对于家用而言，安装 要比 要便宜的多，因为 编解码算法对于帧间预测没有使用双向帧，帧存储器也比采用 要节约一半左右；

➤高级的基本型（Main Profile at high level 简写为MP@HL）

用于处理HDTV信号。这一类型的解码器要在80M 的数据率下工作；

➤高-1441级的空间可调型（Spatial scalable profile at high 1440 level简写为SSP@1440）

这一类型的显著的特征是包括了可调节功能（Scalable）。

## § 3.4 MPEG-2分级编码

在MPEG-2视频规范中，同一类(profile)中不同级别间的图像分辨率和视频码率相差甚大。为了保持解码器的向上兼容性，MPEG-2采用了分级编码(Hierarchical coding)。两类不同的分级编码方法：

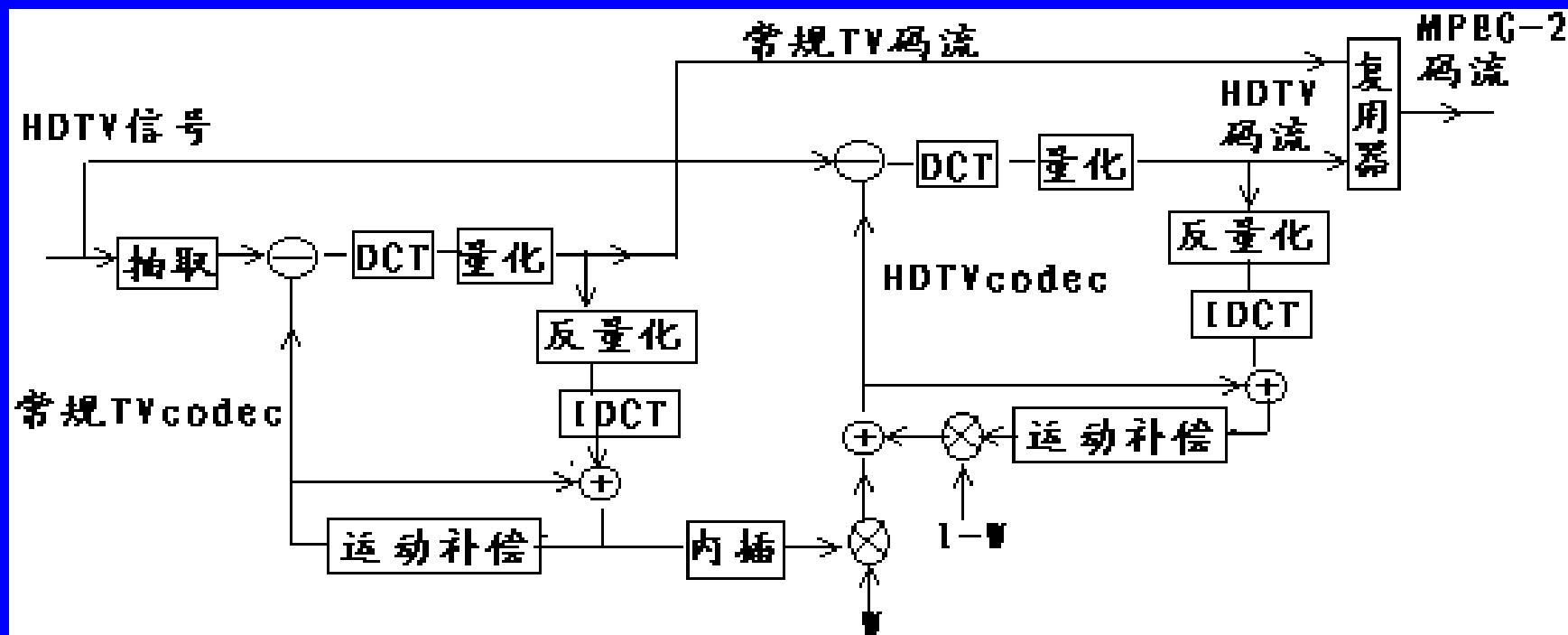
## ➤信噪比可调型

分级改变DCT系数的量化阶；

## ➤空间可调型

利用对象素的抽取（downsample）和内插(upsample)来实现不同级别的转换。

按空间可调型的MPEG-2编码器的框图如图：



图九 空间可调型的MPEG-2编码器框图

对于常规的电视的编码，它与通用的活动图像编码框图在原理上是没有区别的，带对于HDTV信号的编码，从表面上看，似乎分成了两级，但实质并没有改变。

## § 3.5 MPEG-2 与MPEG-1的比较

### ❖ 相同点

- 都是对活动图像的压缩编码建议规范；
- 都是采用了DCT变换和Huffman的编码及运动估计。

## ❖ 不同点

### ➤ 应用范围不同：

MPEG-1(ISO/IEC1172)的码率约为1.5Mbps,广泛用于VCD、DAB、因特网上的各种视音频存储及电视节目的非线性编辑中。MPEG-2(ISO/IEC13818)的码率达10Mbps,已广泛用于数字电视广播(DVB)、高清晰度电视(HDTV)、DVD以及下一代电视节目的非线性编辑系统及数字存储中。MPEG-2对于电视广播数字化起到了举足轻重的作用。

## ➤ 图像格式和码率不同：

图像格式 at 码率	图像类型			
	I 帧	P 帧	B 帧	平均
MPEG-1 SIF @ 1.5 Mbps	150,000	50,000	20,000	38,000
MPEG-2 601 @4Mbps	400,000	200,000	80,000	130,000

MPEG-1规范中建议了MPEG-1码流的标准，即CPB(Constrained Parameters Bitstream),其中定义了一系列的关于采样率、比特速率（与具体的工业硬件实现有关）和帧大小（CPB中限制：当帧频率小于等于25HZ时，帧的大小为每帧396个宏块即101376个像素；当帧频率小于等于30HZ时，帧的大小为每帧330个宏块即84480个像素。

典型的MPEG-1视频图像为

352\*240\*30fps或352\*288\*25fps ) 等限制。虽然MPEG-1G中规定了帧的大小，但从原则上讲，MPEG-1可以处理任意大小的帧，只是不满足CPB标准的图像数据流有可能不被一些厂家的解码器识别。而MPEG-2中对帧的大小有严格的规定，帧的大小必须是16的整数倍。

## 视频压缩算法上的主要不同之处：

- MPEG-2中，非线性宏块的量化使得数据流的动态范围从MPEG-1的1-31增加到0.5-56；
- MPEG-2对DCT交流系数的扫描用交叉扫描代替了MPEG-1中的直线扫描，从而增加了熵编码的效率和性能；
- MPEG-2视频系统中对DCT的操作采用两种可选模式：交叉编码和渐进编码；

- MPEG-2视频系统的系数范围由MPEG-1的-127-127扩展到-2043-+2043；
- 对DCT直流系数的预测编码方式不同；
- 整个编码系统采用分级编码，因此使得MPEG-2对不同类型和不同的分辨率的图像的处理有不同的方法；

- EOB码的不同应用：MPEG-1EOB码占2比特位，意味着每帧平均仅有3或4个非零的交流系数；，而MPEG-2的帧内帧（I帧）中EOB码占4比特位，意味着每帧中非零的交流系数的个数是9到16个；
- 霍夫曼表；MPEG1中的VLC表不在是平常意义的霍夫曼表，但更象G3中的表；MPEG2中增加了INTRA表。

- MPEG-2在MPEG-1的基础上又增加了处理隔行扫描信号、分级编码和抗错能力等功能。
- 应用要求不同：

视频类	典型解码器晶体管个数	总共的 DRAM	DRAM 总线宽度/速度
MPEG-1 CPB	0.4 to .75 million	4 Mbit	16 bits @ 80 ns
MPEG-1 601	0.8 to 1.1 million	16 Mbit	64 bits @ 80 ns
MPEG-2 MP @ML	0.9 to 1.5 million	16 Mbit	64 bits @ 80 ns
MPEG-2 MP@High1440	2 to 3 million	64 Mbit	N/A

# 第四部分

## MPEG-4规范介绍

§ 4.1 MPEG-4引入背景

§ 4.2 MPEG-4目标

§ 4.3 MPEG-4视频编码功能

§ 4.4 MPEG-4解决方案

§ 4.5 MPEG-4特点

§ 4.6 总结

## § 4.1 MPEG-4引入背景

国际上从1988年起，先后公布了有关视频压缩的标准主要有两大方面：

1) 用于连续色调静止图像的"JPEG"和用于会议电视、可视电话的"H.261"。

2) 用于活动图像的MPEG-1和MPEG-2。应该说：这些标准极大地推动了相关产业的发展。纵观这两个标准的关键技术，都是采用了DCT变换和Huffman的编码及运动估计。但MPEG-2在MPEG-1的基础上又增加了处理隔行扫描信号、分级编码和抗错能力等功能。

由于3C业的迅速融合，新业务的需求不断涌出。例如：因特网的多媒体应用；多媒体移动通信；多媒体创作工具；多媒体数据库；交互式视频购物和视频游戏；以及远程监控和医疗等。而MPEG-1和MPEG-2采用的基于“帧”和基于“块”的压缩算法，不能支持表征图像内容的数据结构。而这些又正是许多交互式应用所必需的。其次，是当“码率很低时”会产生严重的“方块效应”和“动作失真”，而“低码率”又是多媒体移动通信所必需的。因此，既然预见到了3C融合所出现的需求，就必须努力克服MPEG-2的不足。

目前使用最为广泛的通信网络是公用电话网和城市蜂窝移动通信网，这两类电话网只有有限的频带宽度容量，他们只能传递甚低速率的数字信号。随着时代的前进，移动通信业务和个人通信业务的信息业务，要求从普通话音扩展到多媒体业务，即在移动环境下提供声音、数字、数据、图形、和视频等信息媒体，使用户在移动通信中进行生动、丰富、有效的多媒体信息交流。显然，实现这一美好愿望的关键是在甚低速率下音频视频压缩问题。

为此，国际标准化组织（ISO）所属的活动专家组（MPEG）从1993年7月开始又制定出一种新的压缩标准---MPEG-4，并于1999年3月宣布了第一个版本，已成为国际标准ISO/IEC14496。1999年12月将推出第2个版本。

## § 4.2 MPEG-4目标

- ❖ 专门用于64kbps以下的甚低速率下的音视频编码；
- ❖ 不仅适应与移动通信和个人通信业务，而且也使用于固定公用通信网和电视电话；
- ❖ 使用于窄带多媒体通信等广泛的应用；
- ❖ 实现基于内容的压缩编码，具有良好的兼容、伸缩性、可靠性。

# § 4.3 MPEG-4视频编码功能

功能	描述	应用举例
<b>基于内容的交互性</b>		
基于内容的多媒体数据存取工具	MPEG-4 应该基于音频视频内容而提供数据的存取，并利用不同的存取工具例如索引、超链接、查询、浏览、上载、下载、删改等	<ul style="list-style-type: none"> <li>■ 从联机库中检索信息，并漫游数据库</li> </ul>
基于内容的管理和数码流的编辑	提供一描述语言（MSDL—）及编码方案，以支持基于内容的管理和数码流的编辑。MSDL 相当灵活便于未来扩展	<ul style="list-style-type: none"> <li>■ 交互式家庭购物</li> <li>■ 家庭电影创作和编辑</li> <li>■ 数字效果（淡入淡出）</li> </ul>
自然的和合成的混合编码	将合成的景物与自然景物进行组合，支持对自然和合成的音频视频数据进行编码和管理的能力，以及解码器可控制的方法，从而利用普通的音频和食品数据创作合成数据实现交互性	<ul style="list-style-type: none"> <li>■ 创作动画和合成声音</li> <li>■ 图形和声音可以从不同观察角度进行描述</li> </ul>
改进时间域的随机存取	在有限的时间以良好的分辨率，随机地从某个音频视频序列存取其中一部分	<ul style="list-style-type: none"> <li>■ 通过遥控终端和有限容量的媒体实现随机存取</li> <li>■ 在序列中对单个 AV 目标“快进”</li> </ul>

功能	描述	应用举例
<u>压缩</u>		
改进编码效率	对于所瞄准的特定应用,与现存的和正在形成的标准相比,MPEG-4 提供主观上更好的音频/视频质量	<ul style="list-style-type: none"> <li>■ 音频视频在窄信道的有效传输</li> <li>■ 音频和视频数据在有限容量媒体中有效存储</li> </ul>
多路并存的数据码流的编码	提供对景物的多视角/多声迹的有效编码,以及单元数据流之间良好同步;充分利用同一景物的多视角和多声迹点中的冗余度,允许采用联合编码方案,既能与普通音频和视频兼容有不受制于其他限制	<ul style="list-style-type: none"> <li>■ 多媒体娱乐</li> <li>■ 训练和飞行模拟</li> <li>■ 多媒体演示和教育</li> </ul>

功能	描述	应用举例
<u>通用存取</u>		
差错环境的鲁棒性	提供对差错坚韧的能力,允许访问不同的无线和有线网络以及存储媒体。在恶劣的差错环境下,对于低数码率的应用,提供非常好的鲁棒性	<ul style="list-style-type: none"> <li>■ 与移动终端通信</li> <li>■ 从遥远地点采集音频视频信号</li> </ul>
基于内容的可分级性	提高实现内容质量具有良好粒度的能力,以及复杂度面可分级的能力	<ul style="list-style-type: none"> <li>■ 用户能选择或自动选择景物中各物体的解码质量</li> <li>■ 能按不同的内容等级、尺度、分辨率及质量</li> </ul>

## § 4.4 MPEG-4解决方案

MPEG-4的视频编码部分为了实现这些丰富多彩的多媒体应用，以算法和工具的形式为下列功能的实现提供了解决方案：

- ❖ 图像和视频的有效压缩；
- ❖ 2D和3D网格纹理映射图(用于合成图像编码)的有效压缩；

- ❖ 隐含(implicit)的2D网格的有效压缩；
- ❖ 控制网格运动的节点时变几何数据流的有效压缩；
- ❖ 各种视频对象的有效存取；
- ❖ 对图像和视频序列的扩展操纵；
- ❖ 基于内容的图像和视频编码；
- ❖ 纹理、图像和视频基于内容的伸缩性；
- ❖ 视频序列中时域、空间及质量的伸缩性；
- ❖ 易错环境下的鲁棒性。

上述的这些功能大部分与基于内容的创作、发布和存取有关。

此外，MPEG-4对合成视频对象的支持同样引人注目：MPEG-4可对合成的面部与人体进行参数化描述；对面部与身体活动信息以参数化的数据流进行描述；支持具有纹理映射功能的静态/动态网格编码；支持视点有关应用(View Dependent Application)中的纹理编码。 MPEG-4视频

编码标准将支持MPEG-1、MPEG-2中的大多数功能，提供不同视频标准源格式、码率、帧频下矩形图像的有效编码，同时也将支持基于内容的图像编码。

在这一功能集的底层是VLBV核心(VLBV：Very Low Bit Rate Video)。它为5-64kbits/s视频操作与应用提供算法与工具，支持较低的空间分辨率(低于 $352 \times 288$ 像素)和较低的帧频(低于15Hz)。

## VLBV核心支持的专用功能包括：

- (a) 实时多媒体应用：支持矩形图像序列的有效编码，具有编码效率高、高精度、高容错度、低延时等特点。
- (b) 多媒体数据库应用：支持多媒体数据库的存储、随机存取以及FF/FR(快进/快退)等功能与操作。 MPEG-4在HBV(HBV: High Bit Rate Video，范围在64kbits/s-4Mbits/s之间)同样支持

上述功能，但它支持较高的空间与时间分辨率。其输入可以是ITU-R 601的标准信号，因此其典型应用为数字电视广播与交互式检索。未来的多媒体应用呼唤全新的交互方式，以满足用户的需求。传统的交互存取方式中，在传统的交互过程中，用户得到的场景是制作人员事先编排好的，用户只能对音视频序列进行简单的回放。

MPEG-4提出了基于内容（Content-based）的存取概念，努力使用户根据制作者设计的具体自由度，与场景进行交互。用户不仅可以改变场景的视角，在其中尽情畅游，还可以改变场景中物体的位置、大小和形状，或对该对象进行置换甚至清除。用户将从这些简便、灵活的交互过程中获得的丰富的信息和极大的乐趣。

## § 4.5 MPEG-4特点

- ❖ 从矩形帧到VOP(Video Object Plane: 视频对象面)
- ❖ 基于VOP的视频编码

## ❖ 从矩形帧到VOP(Video Object Plane: 视频对象面)

传统图像编码方法依据信源编码理论的框架，将图像作为随机信号，利用其随机特性来达到压缩的目的。这种方法本身未能考虑信息获取者的主观意义与主观特性，未能考虑事件本身的特性如具体含义、重要性以及后果等等。但正是由于信源编码理论的限定使得传统的图像编码具有较

高的概括性和综合性，而基于矩形帧编码的传统视频编码标准如H.261、MPEG-1/MPEG-2在实际应用中也获得了巨大成功。然而MPEG-4并不满足于此，它的目标在于采用现代图像编码方法，利用人眼的视觉特性，抓住图像信息传输的本质，从轮廓--纹理的思路出发，支持基于视觉内容的交互功能。而实现基于内容交互功能的关键在于基于视频对象的编码，为此MPEG-4引入了视频对象面VOP的概念。

在这一概念中，我们根据人眼感兴趣的一些特性如形状、运动、纹理等，将图像序列中每一帧中的场景，看成是由不同视频对象面VOP(Video Object Plane)所组成，而同一对象连续的VOP称为视频对象VO(Video Object)。VO可以是视频序列中的人物或具体的景物。

VOP的编码包括对运动(采用运动预测方法)及纹理(采用变换编码方法)的编码，其基本原理与H.261和MPEG-1/MPEG-2极为相似。由于MPEG-4基于内容图像编码方法，VOP具有任意形状，因此要求编码方案可以处理形状(Shape)和透明(Transparency)信息，这同只能处理矩形帧序列的现有视频编码标准形成了鲜明的对照。在MPEG-4中，矩形帧被认为是VOP的一个特例，这时编码系统

不用处理形状信息，退化为类似于H.261、MPEG-1/MPEG-2的传统编码系统，同时也实现了与现有标准的兼容。从矩形帧到VOP，MPEG-4顺应了现代图像压缩编码的发展潮流，即从基于像素的传统编码向基于对象和内容的现代编码的转变。

## ❖ 基于VOP的视频编码

VOP编码器由两个主要部分组成：形状编码和纹理、运动信息编码。其中纹理编码、运动预测和运动补偿部分在原理上同现有标准是一致的。值得注意的是形状编码，这是图像编码标准中第一次引入形状编码技术。

为了支持基于内容的功能，编码器可对图像序列中具有任意形状的VOP进行编码。尽管如此，编码器内的机制都是基于 $16 \times 16$ 像素宏块(Macroblock)来设计的，这不仅是出于与现有标准在兼容问题上的考虑，而且是为了便于对编码器进行更好的扩展。VOP被限定在一个矩形窗口内，称之为VOP窗口(VOP Window)，窗口的长、宽均为16的整数倍，同时保证VOP窗口中非VOP的宏块数目最少。

标准的矩形帧可认为是VOP的特例，在编码过程中其形状编码模块可以被屏蔽。系统依据不同的应用场合，对各种形状的VOP输入序列采用固定的或可变的帧频。编码算法仍是基于块( $8 \times 8$ 像素)的预测-变换混合编码。对VOP的编码采用帧内(Intra)编码与帧间预测编码相结合的方法，帧内编码中对DCT变换的DC、AC因子进行有效的预测。帧间预测分为前向预测和双向预测。对于甚低码率( $<64\text{kbits/s}$ )下的应用，由于方块效应较明显，所以需用除方块滤波器进行相应处理。

- 一、形状编码
- 二、运动信息编码
- 三、纹理编码
- 四、分级编码

# 一、形状编码

MPEG-4引入了形状信息的编码。VO的形状信息有两类：二值形状信息和灰度形状信息。

二值形状信息用0、1来表示VOP的形状，0表示非VOP区域，1表示VOP区域。二值形状信息的编码采用基于运动补偿块的技术，可以是无损或有损编码。

灰度形状信息用0~255之间的数值来表示VOP的透明程度。灰度形状信息的编码采用基于块的运动补偿DCT方法(同纹理编码相似)，属于有损编码。

目前的标准中采用矩阵的形式来表示二值或灰度形状信息，称之为位图(或阿尔法平面)。实验表明，位图表示法具有较高的编码效率和较低的运算复杂度。但为了能够进行更有效的操作和压缩，在最终的标准中可能出现另一种表示方法，即借用高层语义的描述，以轮廓的几何参数进行表征。

## 二、运动信息编码

MPEG-4采用运动预测和运动补偿技术来去除图像信息中的时间冗余成分，而这些运动信息的编码技术可视为现有标准由向任意形状的VOP的延伸。

VOP的编码有3种模式，即帧内(Intra-frame)编码模式(I-VOP)，帧间(Inter-frame)预测编码模式(P-VOP)，帧间双向(Bidirectionally)预测编码模式(B-VOP)。

在MPEG-4中运动预测和运动补偿可以是基于 $16 \times 16$ 像素宏块的，也可以是基于 $8 \times 8$ 像素块的。为了能适应任意形状的VOP，MPEG-4引入了图像填充(Image Padding)技术和多边形匹配(Polygon Matching)技术。图像填充技术利用VOP内部的像素值来外推VOP外的像素值，以此获得运动预测的参考值。多边形匹配技术则将VOP的轮廓宏块的活跃部分包含在多边形之内，以此来增加运动估值的有效性。

此外，MPEG-4采用8参数仿射运动变换来进行全局运动补偿；支持静态或动态的SPRITE全局运动预测，对于连续图像序列，可由VOP全景存储器预测得到描述摄像机运动的8个全局运动参数，利用这些参数来重建视频序列。

### 三、纹理编码

纹理编码的对象可以是帧内编码模式的I-VOP，也可以是帧间编码模式B-VOP或P-VOP运动补偿后的预测误差。

编码方法基本上仍采用基于 $8 \times 8$ 像素块的DCT方法。在帧内编码模式中，对于完全位于VOP内的像素块，则采用经典的DCT方法；对于完全位于VOP之外的像素块则不进行编码；对于部分在VOP内，部分在VOP外的像素块则首先采用图像填充技术来获取VOP之外的像素值，之后再继续进行DCT编码。帧内编码模式中还将对DCT变换的DC及AC因子进行有效的预测。

在帧间编码模式中，为了对B-VOP和P-VOP运动补偿后的预测误差进行编码，可将那些位于VOP活跃区域之外的像素值设为128。此外，还可采用SADCT(Shape-adaptive DCT)方法对VOP内的像素进行编码，该方法可在相同码率下获得较高的编码质量，但运算的复杂程度稍高。变换之后的DCT因子还需经过量化(采用单一量化因子或量化矩阵)、扫描及变长编码，这些过程与现有标准基本相同。

## 四、分级编码

很多多媒体应用需要系统支持时域、空间及质量的伸缩性，分级编码就是为了实现这一目标。例如，在远程多媒体数据库检索及视频内容重放等应用中，分级编码的引入使得接收机可依据具体的信道带宽、系统处理能力、显示能力及用户需求进行多分辨率的解码及回放。接收机可视具体情况对编码数据流进行部分解码，以

获得： 较低的解码复杂度，同时也意味着较低的重建图像质量，较低的空间分辨率，较低的时间分辨率，即帧率；相同空间分辨率及帧率条件下，较低的重建图像质量。

## MPEG-4中通过视频对象层

VOL(Video Object Layer)的数据结构来实现分级编码。每一种分级编码都至少有2层VOL，低层称为基本层，高层称为增强层。空间伸缩性可通过增强层强化基本层的空间分辨率来实现，因此在对增强层中的VOP进行编码之前，必须先对基本层中相应的VOP进行编码。同样对于时域伸缩性，可通过增强层来增加视频序列中某个VO(特别是运动的VO)的帧率，使其与其余区域相比更为平滑。

## § 4.6 总结

总之，MPEG-4采用的基于内容和对象的编码方法，适应了当前通信、广播、计算机3C界的要求，具有很强的发展潜力，并于1999年1月正式成为国际标准(标准号为ISO/IEC 14496)，在1999年11月的后继版本中增加可变形、半透明视频对象和工具，以进一步提高效率，后继版将与第一版反向兼容。

在过去的几年中，MPEG-1/MPEG-2和VCD/DVD的成功与蓬勃发展表明：符合MPEG标准的产品与应用具有强大的生命力。MPEG在压缩编码标准化、商业化方面所获得的巨大成功也使人们感到：MPEG总是能够兼顾技术发展潮流与实际应用需求，紧扣商业应用的脉搏，这使得它的活动变得越来越引人注目而又激动人心。MPEG-4视频编码技

术采用了现代图像编码方法，利用人眼视觉特性，从轮廓--纹理的思路出发，支持基于内容和对象的编码，支持基于内容的交互功能。MPEG-4视频编码正在完成从基于像素的传统编码向基于对象和内容的现代编码的转变，它代表了新一代智能图像编码，必将对未来图像通信机制产生深远的影响。

# 第五部分

## MPEG-7的发展

基于对象编码的MPEG-4把支持基于内容的检索作为其目标之一，但这种支持有限的。为了克服MPEG-4的不足，MPEG启动了新的项目 MPEG-7。它的目标是建立一种多媒体内容描述接口（Multimedia Content Description Interface），支持多媒体信息基于内容的高效快速检索。他将确立各种类型多媒体信息的标准描述方法。

MPEG-7建立在MPEG-4的基础上，期望用很少的特征就能对信息内容进行检索。例如对图形，只要画出很少几条线就可以找到包括该特征的相应图形、商标等。预计到2001年MPEG-7才能形成标准。