

# 音视频编码知识

胡 梅

2009-12-29

# 编码和文件格式

---

学习音视频相关基础知识之前，我们先来认识下编码和文件格式；

➤ **编码和文件格式（也称容器）是什么？**

文件（即容器）是既包括了视频、又包括音频、甚至还带有一些脚本的集合；

文件中视频和音频的压缩算法才是真正的编码；

对于一种文件，它的视频和音频可以分别采用不同的编码。

音视频编码及文件格式是一个很庞大的知识领域，我们在此只是对其中的相关知识做入门的介绍，让大家能够对其内容有个初步的了解；

# 主要内容

---

- ◆ 第一部分：基础知识介绍
- ◆ 第二部分：音视频编码标准与体系介绍
- ◆ 第三部分：文件格式介绍
- ◆ 第四部分：音视频格式转换工具介绍

# 第一部分：基础知识介绍

---

- ◆ 一、音视频编码的基本概念
- ◆ 二、音视频编码的原理
- ◆ 三、音视频编码技术分类
- ◆ 四、音视频文件大小计算方式

# 一、音视频编码的基本概念

---

## 1. 什么是视频编码？

所谓视频编码就是通过特定的压缩技术，将某个视频格式的文件转换成另外一种视频格式文件的方法。

视频编码的主要功能是完成图像的压缩，使数字电视信号的传输量由1Gbit/s（针对1920x1080显示格式）减少为20-30Mbit/s。

# 一、音视频编码的基本概念

---

## 2. 什么是音频编码？

自然界中的声音以及波形都非常复杂，声音其实也是一种能量波，它有频率和振幅；其中频率所对应的是时间轴线，振幅对应的是电平轴线。波是无限光滑的，弦线可看成由无数点组成。

音频编码主要是完成对声音信息的压缩。声音信号数字化后，信息量比模拟传输状态大很多，不能像模拟电视声音那样直接传输；因而需要对声音多一道压缩编码工序，即为音频编码。

## 二、音视频编码的原理

### 1. 视频编码原理简介

视频编码主要是对图像进行有效的压缩。原始视频图像数据中包含大量的冗余信息视频编码主要采取块运动估计和运动补偿技术有效的去除图像帧间冗余度，来压缩码率和带宽，实现信号有效传输的目的。

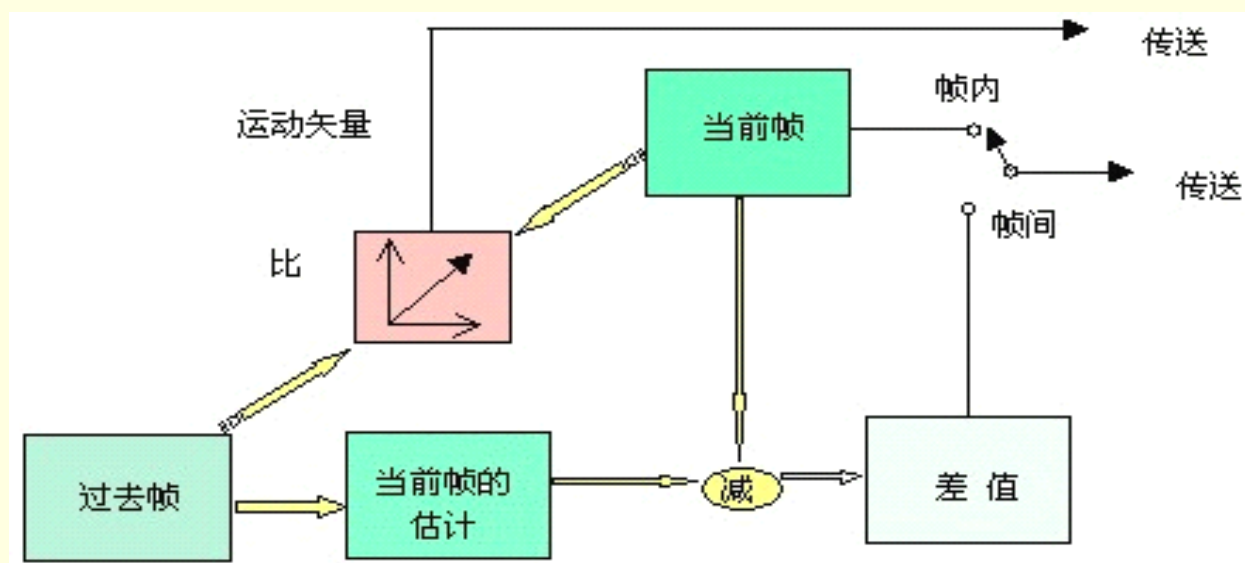


图1.1 运动处理原理图

## 二、音视频编码的原理

### 2. 视频编码模块图 (MPEG-2)

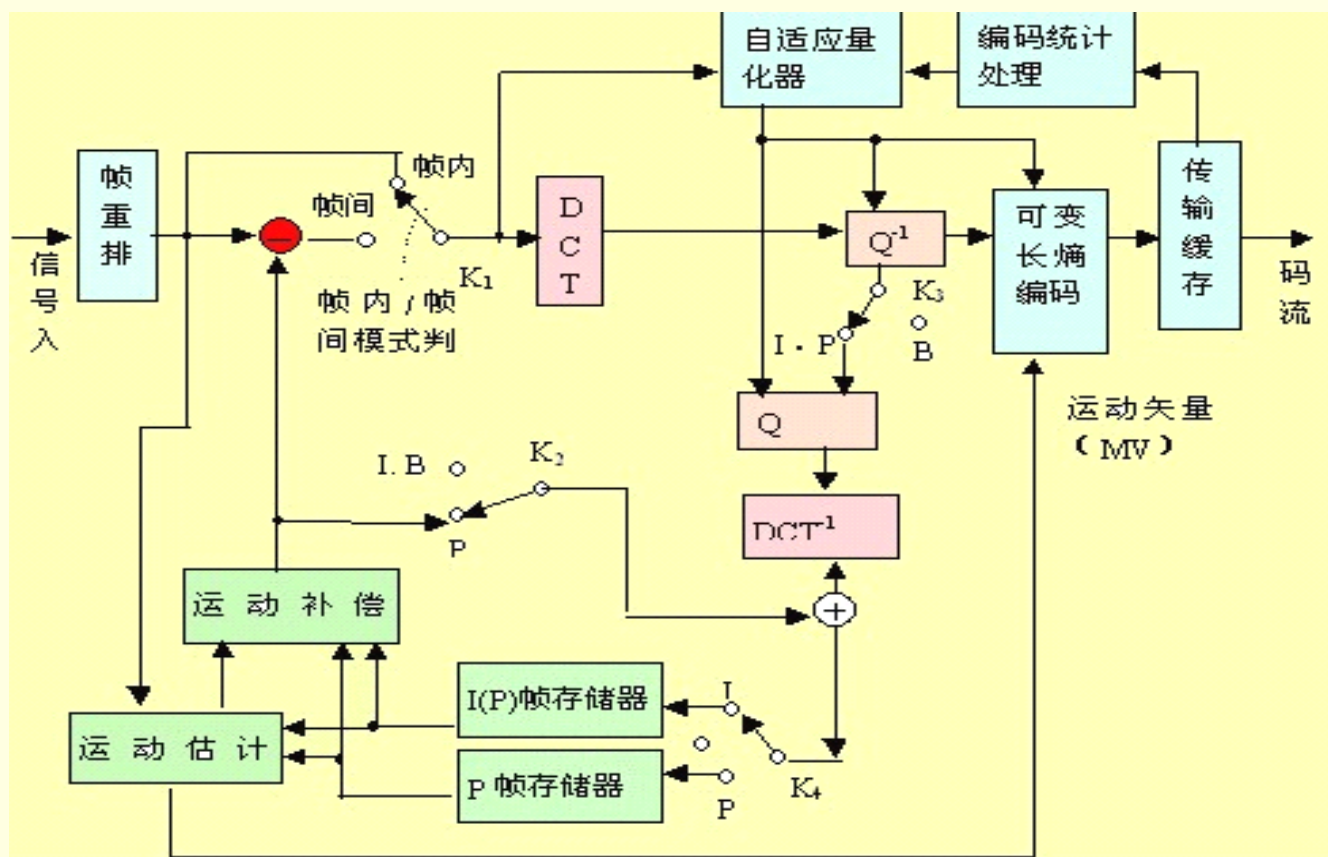


图2.1 MPEG-2视频编码原理框图



## 二、音视频编码的原理

### 3. 音频编码的原理简介

我们需要通过抽样、量化、编码三个步骤将连续变化的模拟信号转换为数字信号。抽样就是在时间上将模拟信号离散化。量化是用有限个幅度值近似原来连续变化的幅度值，把模拟信号的连续幅度变为有限数量的有一定间隔的离散值。编码就是按一定的规律把量化后的值用二进制数字表示，然后转换成二值或多值得数字信号流。

通常我们采用PCM编码，其主要过程是将语音、图像等模拟信号每隔一段时间进行取样，使其离散化，同时将抽样值按分层单位四舍五入取整量化，并将抽样值按一组二进制码来表示抽样脉冲的幅值。

## 二、音视频编码的原理

### 4. 音频编码模块图（PCM和MPEG）

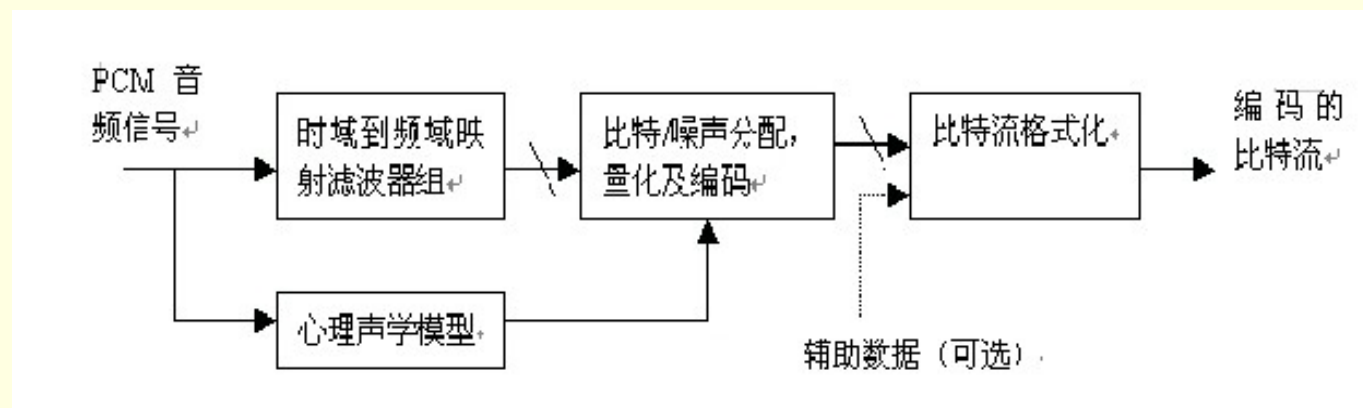


图4.1 PCM音频编码框图

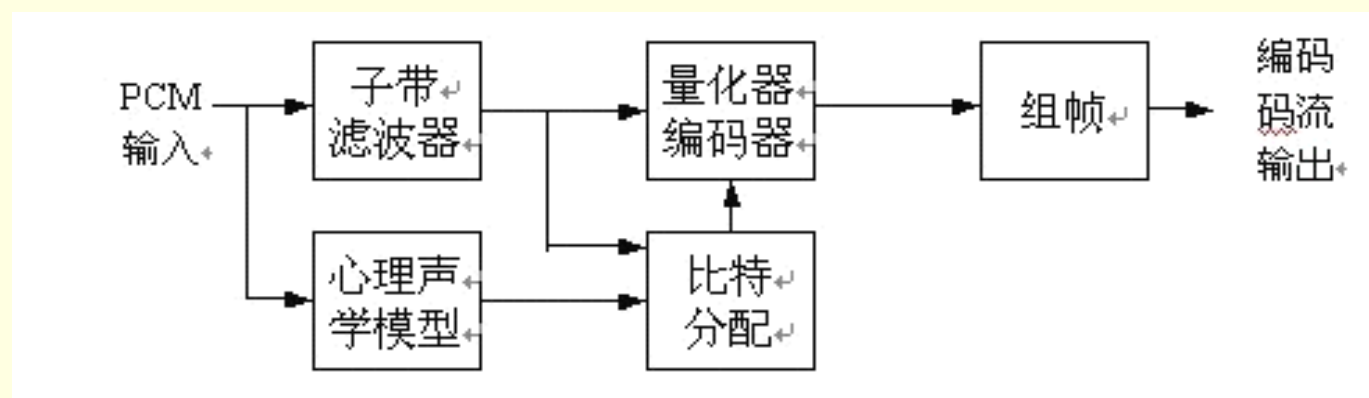


图4.2 MPEG音频解码框图

# 三、音视频编码技术分类

## 1. 视频编码技术分类

为了在Internet上有效的、高质量的传输视频流，需要多种技术的支持，其中数字视频的压缩编码技术是Internet视频传输中的关键技术之一，实际上最终用户关心的是：清晰度、存储量（带宽）、稳定性还有价格。所以采用不同的压缩技术，将很大程度影响以上几大要素。

视频编码技术主要是压缩原视频图像数据中的冗余信息（分为时域和空域冗余信息），通常采用了帧内图像数据压缩技术、帧间图像数据压缩技术和熵编码压缩技术；

### 1.1 时域冗余信息

去除时域冗余信息使用的是帧间编码技术，它包括三个部分：运动补偿、运动表示、运动估计；

#### 1.1.1 运动补偿

运动补偿是通过先前的局部图像来预测、补偿当前的局部图像，是减少帧序列冗余信息的有效方法。

# 三、音视频编码技术分类

## ✚ 1.1.2 运动表示

不同区域的图像需要使用不同的运动矢量来描述运动信息。运动矢量通过熵编码进行压缩。

## ✚ 1.1.3 运动估计

运动估计是从视频序列中抽取运动信息的一整套技术。

注：通用的压缩标准均使用的是基于块的运动估计和运动补偿；

## 1.2 空域冗余信息

去除空域冗余信息使用的是帧内编码技术和熵编码技术，包括三个部分：变换编码、量化编码、熵编码。

## ✚ 1.2.1 变换编码

变换编码是将空域信号变换到另一正交矢量空间，使其相关性下降，数据冗余度减小。

## ✚ 1.2.2 量化编码

经过变换编码后，会产生一批变换系数，对这些系数进行量化，使编码器的输出达到一定的位率，这一过程称为量化编码。

## ✚ 1.2.3 熵编码

熵编码是对变换、量化后得到的系数和运动信息，进行进一步的压缩。

## 三、音视频编码技术分类

下面我们将主要介绍目前常见的几种编码技术方案；

### 1) MPEG-2

MPEG-2图像压缩的原理是主要利用了图像中的两种特性：空间相关性和时间相关性。另外它综合采用了运动补偿的帧间预测、空间域离散余弦变换、自适应量化和可变长编码的混合编码。同时它开始用半像素精度运动矢量搜索，引入了“帧”和“场”的编码方法，和MPEG-1向下兼容；**主要针对存储媒体、数字电视、高清晰等应用领域**，分辨率为：低（352\*288）中（720\*480）次高（1440\*1080）高（1920\*1080）。

### 2) MPEG-4

采用基于对象的编码理念，即在编码时将一幅景物分成若干在时间和空间上相互联系的视频音频对象，分别编码后，再经过复用传输到接收端，然后再对不同的对象分别解码，从而组合成所需要的视频和音频。实现了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。

**主要特点是支持交互性、高压缩比及能用存储性；**

## 三、音视频编码技术分类

### 3) H. 264/AVC

主要包括2个层次：视频编码层和网络抽象层。前者主要致力于有效的表示视频内容，后者格式化VCL视频表示，提供头部信息，适合多种传输和存储媒体。采用了全新的编码技术：帧内预测，可变大小的图像分块，多预测参考帧，1/4和1/8像素精度的运动估计，残差图像的整数变换编码等。由于其压缩效率高、容错能力好、网络适应能力强等特点，能适用于不同网络中的视频传输。

### 4) Windows Media Video

采用了MPEG4的压缩算法，所以压缩率和图像的质量都很不错，由于Windows Media video是微软的杰作，所以Windows系统对Windows Media Video给予了很好的支持。

### 5) RealVideo

采用REAL公司可变编码率的REAL格式的视频编码技术。它能在保持文件较小的条件下获得较高编码率的视频质量。主要定位在视频流应用方面的，不太适合专业场合，在网络和娱乐场合占有不错的份额。

## 三、音视频编码技术分类

### 6) Sorenson Video

Sorenson Media公司随 QuickTime5发布的编码器，采用的是sorenson公司的压缩算法，质量很不错，已经成为QuickTime 的标准视频编码，网络上大部分电影预告片都采用这种编码。

### 7) Intel Indeo video

由Intel架构实验室开发，常见的视频编码，主流的有4.5和5.10两种，质量比Cinepak和R3.2要好，可以适应不同带宽的网络，但必须有相应的解码插件才能顺利地将下载作品进行播放。

### 8) Dvix和xviD

Dvix主要是将影像部分以MPEG4来压缩，音效部分则以MP3来压缩；采用了空间和时间压缩技术。在档案大小上占了相当大的优势，画质部分的表现也很不错。

XviD是目前世界上最常用的视频[编码解码器](#)，是第一个真正开放源码的，它是基于Open DivX编写的。Xvid支持多种编码模式，量化方式和范围控制，运动侦测和曲线平衡分配等众多编码技术。

# 三、音视频编码技术分类

## 2. 音频编码技术分类

数字音频信息压缩主要是依据音频信息自身的相关性以及人耳对音频信息的听觉冗余度。音频信息在编码技术中通常分成两类来处理，分别是语音和音乐，各自采用的技术也存在着差异。

### 2.1 语音编码技术

将模拟语音信号转变为数字信号在信道中传输。语音编码的目的是在保持一定算法复杂程度和通信时延的前提下，占用尽可能少的通信容量，传送尽可能高质量的语音。它可分为波形编码、参量编码和混合编码三大类。

#### ✚ 2.1.1 波形编码

波形编码是在时域上进行处理，主要是为了使重组的语音波形保持原始语音信号的形状，它将语音信号作为一般的波形信号来处理。



## 三、音视频编码技术分类

### ✚ 2.1.2 参量编码（声源编码）

利用语音信息产生的数学模型，提取语音信号的特征参量，并按照模型参数重构音频信号。

### ✚ 2.1.3 混合编码

将波形编码和参量编码组合起来，结合各自的长处，保持波形编码的高质量和参量编码的低速率，在4-16Kbit/s速率上能够得到高质量的合成语音，适合于数字移动通信的语音编码技术。

## 2.2 音乐编码技术

### ✚ 2.2.1 自适应变换编码（频域编码）

利用正交变化，把时域音频信号变换到另外一个域，通常采用自适应比特分配和自适应量化技术来对频域数据进行量化。

# 三、音视频编码技术分类

## ✚ 2.2.2 心理声学模型

对信息量进行压缩，同时使失真尽可能不被察觉，利用人耳的掩蔽效应即较弱的声音会被同时存在的较强的声音所掩盖，使得人耳无法听到。

## ✚ 2.2.3 熵编码

对于一串由许多数值构成的数据来说，如果其中某些值经常出现，另外一些值极少出现，则可以对其进行压缩，即对常出现的数值用短的码组表示，不常出现的数值用长的码组表示；这样最终用于表示这一串数据的总码位相对于用定长码组来表示的码位而言得到了降低；这就是熵编码的思想；常见的熵编码算法有Huffman编码、shannon编码。目前大多采用Huffman编码，即一个数值的编码长度与此数值出现的概率尽量成反比；

## 三、音视频编码技术分类

下面我们将主要介绍目前常见的几种编码技术方案；

### 1) MPEG Audio Layer 3

MP3采用了名为“感官编码技术”的编码算法：编码时先对音频文件进行频谱分析，然后用过滤器滤掉噪音电平，接着通过量化的方式将剩下的每一位打散排列，最后形成具有较高压缩比的MP3文件。能够在音质丢失很小的情况下把文件压缩到更小。

### 2) FLAC

采用了无损压缩技术，使用了同步代码和CRC校验码，不会破坏任何原有的音频资讯。被编码的音频(PCM)数据没有任何信息损失，解码输出的音频与编码器的输入的每一个字节都是一样的。

### 3) AAC

采用了临时噪声重整、后向自适应线性预测、联合立体声技术和量化哈夫曼编码等最新技术，通过结合其他的功能来提高编码效率,同时支持多达48个音轨、15个低频音轨、更多种采样率和比特率、多种语言的兼容能力、更高的解码效率。

## 三、音视频编码技术分类

---

### 4) Ogg Vorbis

主要算法是利用MDCT（改进离散余弦变换）技术，MDCT先将整体图像分成 $N \times N$ 像素块，然后对 $N \times N$ 像素块逐一进行DCT变换。是一种灵活开放的音频编码，能够在编码方案已经固定下来后还能对音质进行明显的调节和新算法的改良。可以不断导入新技术逐步完善。

### 5) Windows Media Audio

采用的压缩技术与MP3压缩原理近似，但它并不删减大量的编码，音质要强于MP3格式，更远胜于RA格式，是以减少数据流量但保持音质的方法来达到比MP3压缩率更高的目的，WMA的压缩率一般都可以达到1：18左右。

## 三、音视频编码技术分类

---

### 6) PCM

通过抽样, 对样值幅度量化, 编码三个步骤将模拟信号转换为数字编码。可以提供话音、图象传送、远程教学等业务, 特别适用于对数据传输速率要求较高, 需要更高带宽的用户使用。

### 7) RA (RealAudio)

采用的SureStream (自适应流) 技术是RealNetworks公司具有代表性的技术。它通过RealServer将A/V文件以流的方式传输, 然后利用SureStream方式, 根据客户端不同的拨号速率 (不同的带宽), 让传输的A/V信息自动适应带宽, 并始终以流畅的方式播放。

## 四、音视频文件大小计算方式

编码率/比特率 (kbps/千位每秒) 直接与文件体积有关, 且编码率与编码格式配合是否合适, 直接关系到视频文件是否清晰。

例如800Kbps其中, 1K=1024 1M=1024K, b为比特 (bit) 就是电脑文件大小的计量单位, 1KB=8Kb, 区分大小写, B代表字节(Byte), s为秒, p为每 (per)。

以800kbps来编码: 表示经过编码后的数据每秒钟需要用800K比特来表示。

➤ 完整的视频文件由音频流和视频流2部分组成, 音视频使用的是不同的编码率, 一个视频文件的最终技术大小的编码率是**音频编码率+视频编码率**。

**输出文件大小计算方式:**

$$(\text{音频编码率 (Kbit)} / 8 + \text{视频编码率 (Kbit)} / 8) \times \text{影片总长度 (秒为单位)} = \text{文件大小 (MB为单位)}$$

例: 有一个1.5小时 (5400秒) 的影片, 希望转换后文件大小刚好为700M

计算方法如下:  $700 \times 8 \div 5400 \times 1024 \approx 1061 \text{Kbps}$

意思是只要音频编码率加上视频编码率之和为1061Kb, 则1个半小时的影片转换后文件体积大小刚好为700M。

故: **压缩后的文件大小主要是与编码率有关, 而音视频编码率具体如何设置,**

## 四、音视频文件大小计算方式

就看选择怎样的编码格式了，不同的编码格式压缩率是不一样的。

■ 当然未经压缩文件的计算公式又不同：

声音数据量（位/秒）：采样率（Hz）\*采样大小（bit）\*声道数

文件总字节：时间\*码率/8（KB）

其中，数据量即码率，单声道的声道数为1，立体声的声道数为2。

图象数据量（位/秒）=（画面尺寸\*彩色位数（bit）\*帧数）

文件总字节：数据量\*时间/8

例：计算2分钟, 25帧/秒, 640×480分辨率, 24位真彩色数字  
视频的不压缩的数据量；

数据量 = 640x480x24\*25=184320Kb=180Mb,

2分钟文件字节 = 180Mb\*120/8=2700000KB=2636.7MB=2.57GB;

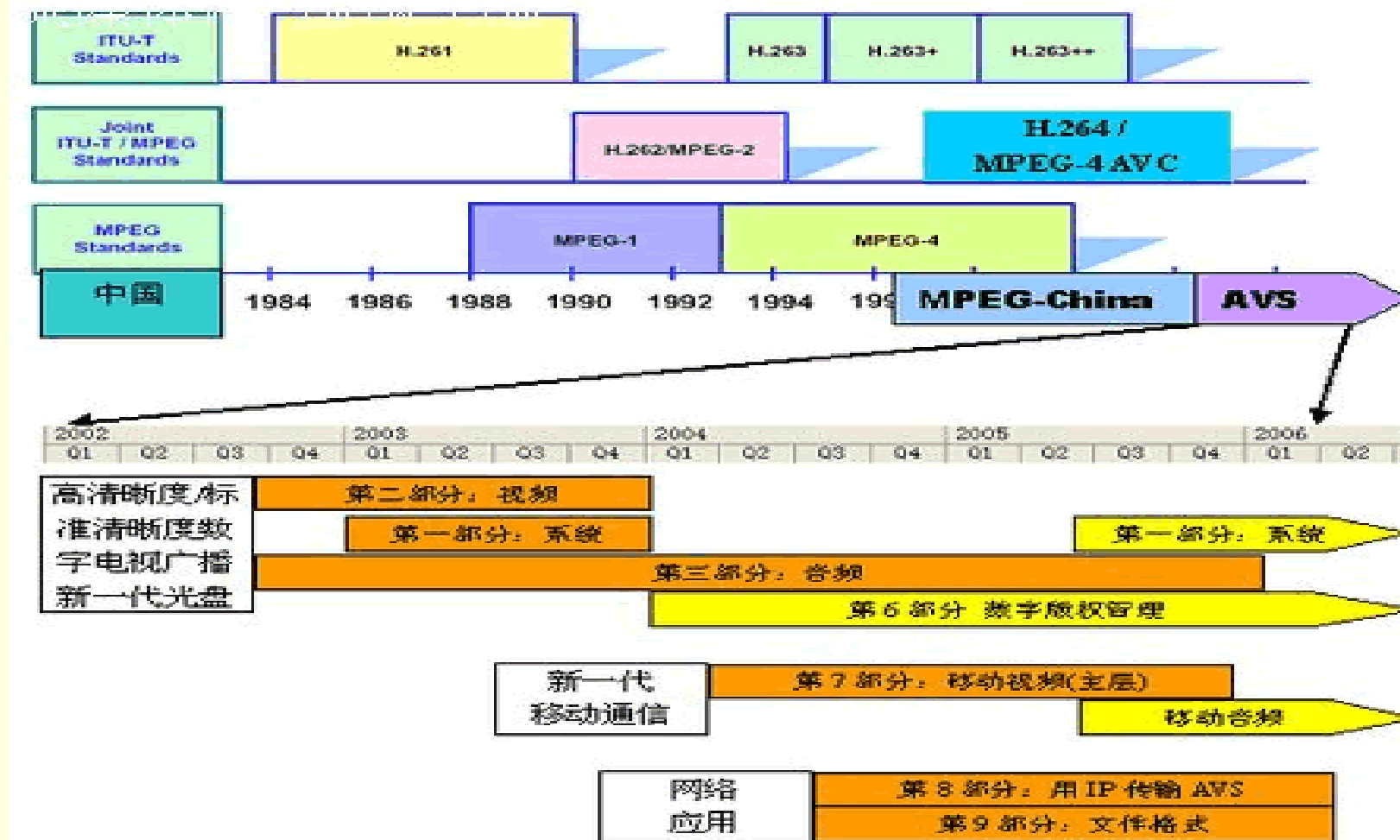
## 第二部分：编码标准与体系介绍

---

- ◆一、音视频编码标准
- ◆二、常见编码体系介绍



# 一、音视频编码标准



音视频编码标准的发展历史

# 一、音视频编码标准

## 1. 常见音视频编码标准介绍

国际上音视频编解码标准主要两大系列：

- ISO/IEC JTC1 制定的 MPEG 系列标准；
- ITU-T 针对多媒体通信制定的 H. 26x 系列视频编码标准和 G. 7 系列音频编码标准。

国内自主制定的音视频编码技术标准：

- AVS-《信息技术先进音视频编码》系列标准；

目前音视频产业可以选择的信源编码标准有四个：

MPEG-2、MPEG-4、MPEG-4 AVC (简称 H. 264，也称 JVT、AVC)、AVS。

- ◉ MPEG-2标准：1994年11月成为国际标准（ISO/IEC 13818），这是一个适应性广的动态影像和声音编码方案。MPEG-2 可用于数字通信、存储、广播、高清晰度电视等的压缩编码。例 DVD、数字电视广播均采用 MPEG-2 标准。
- ◉ MPEG-4标准：主要是为了低带宽应用的需要和交互式图形应用（如游戏等）、交互式多媒体（WWW 等内容分发和访问技术）的快速发展而制定的一种编码标准。

# 一、音视频编码标准

- **H. 264标准**：是由ITU-T的VCEG(视频编码专家组)和ISO/IEC的MPEG(活动图像编码专家组)联合组建的联合视频组(JVT: joint video team)提出的一个新的数字视频编码标准。
- **AVS标准**：是我国具备自主知识产权的第二代信源编码标准。 AVS标准包括系统、视频、音频、数字版权管理等四个主要技术标准和一致性测试等支撑标准。特点：性能高、成本低、复杂度低、专利授权模式简单；
- **MPEG-1标准**：用于数字存储媒体速率为1.5Mbps的运动图像及其伴音的压缩编码。
- **MPEG-7与MPEG-21标准**：MPEG-7是面向多媒体信息搜索、过滤、管理和处理的内容表达标准，2001年7月成为国际标准。正在制定的MPEG-21的重点是多媒体框架，为与多媒体内容递交相关的所有已开发的和正在开发的标准提供基础体系。

# 一、音视频编码标准

## 2. 主流音视频编码标准的应用

当前音视频编解码技术发展的理想方向，应当综合现有压缩编码技术，制定全球统一标准，使信息管理系统具有普遍的互操作性并确保未来的兼容性。但现实情况却是：音视频编解码技术经过10多年的发展，在该领域存在多种互不兼容的标准，在未来相当长一段时间内这一局面仍将继续维持。

- **MPEG-2标准**：第一代音视频标准，由于建立时间长、技术相对成熟、设备易用而且可靠性高，因而应用范围最广、影响最大。高清数字电视芯片仍以MPEG-2为主导。
- **MPEG-4标准**：主要应用于视像电话(Video Phone)，视像电子邮件(Video Email)和电子新闻(Electronic News)等领域；移动多媒体以此为主导；
- **H. 264标准**：目前多面向网络传输领域，如IP机顶盒应用；
- **AVS标准**：AVS视频标准主要面向高清晰度和高质量数字电视广播、数字存储媒体、网络电视、视频通信等相关应用；

## 二、常见编码体系介绍

音视频编码方案有很多，用百家争鸣形容不算为过，具体详细编码方案介绍请参见附表【[音视频编码和文件格式列表.xls](#)】

下面我们主要介绍目前常见的音视频编码标准系列：

➤ **MPEG系列（由ISO下属的MPEG[运动图象专家组]开发）**

**视频编码：**Mpeg1（VCD使用）、Mpeg2（DVD使用）、Mpeg4（DVDRIP使用）、MPEG4 AVC

**音频编码：**MPEG Audio Layer 1/2、MPEG Audio Layer 3（mp3）、MPEG-2AAC、MPEG-4 AAC等等。

注意：DVD音频没有采用Mpeg的。

➤ **H. 26X系列（ITU[国际电传视讯联盟]主导，侧重网络传输，仅视频编码）**

**视频编码：**H261、H262、H263、H263+、H263++、H264（与MPEG4 AVC合作的结晶）。

## 二、常见编码体系介绍

### ➤ 微软**windows media**系列

视频编码: Mpeg-4 v1/v2/v3 (基于MPEG4, DIVX3的来源) Windows Media Video V1/V2/7/8/9/10

音频编码: Windows Media audio v1/v2/7/8/9

### ➤ Real Media系列

视频编码: RealVideo G2 (早期)、RealVideo 8/9/10;

音频编码: RealAudio cook/sipro (早期)、RealAudio AAC/AACPlus等;

### ➤ QuickTime系列: (是一个平台, 有很多编码器)

视频编码: Sorenson Video3 (用于QT5)、Apple MPEG-4、Apple H.264;

音频编码: QDesign Music 2、Apple MPEG-4 AAC;

### ➤ Indeo Video系列 (仅视频编码)

视频编码: Indeo video R3.2、Indeo video 4.5、Indeo video 5.10

# 第三部分：文件格式大全

---

## 1. 主流视频文件格式

视频是现在电脑和多媒体系统中的不可缺少的一部分。为了适应储存视频的需要，人们设定了不同的视频文件格式来把视频和音频放在一个文件中，以方便同时回放。详细的视频文件格式请查阅附表

【[音视频编码和文件格式列表.xls](#)】

# 第三部分：文件格式大全

---

## 2. 主流音频文件格式

音频文件格式专指存放音频数据文件的格式。

一般获取音频数据的方法是：采用固定的时间间隔，对音频电压采样量化，并将结果以某种分辨率存储。（例如：CDDA每个采样为16比特或2字节）采样的时间间隔可以有不同的标准，如CDDA采用每秒44100次；DVD采用每秒48000或96000次。因此，采样率，分辨率和声道数目（例如立体声为2声道）是音频文件格式的关键参数。尽管一种音频文件格式可以支持多种编码，例如AVI文件格式，但多数的音频文件仅支持一种音频编码。



## 第三部分：文件格式大全

---

有两类主要的音频文件格式：

- 无损格式，例如WAV，PCM，FLAC，APE，TAK，WavPack (WV)
- 有损格式，例如MP3，WMA，OGG，AAC

详细的音频文件格式请查阅附表【[音视频编码和文件格式列表.xls](#)】

## 第四部分：音视频格式转换工具介绍

文件转换（实际上也是编码转换），现在我们介绍几款比较实用的音视频格式转换工具；

### 1. 音视频转换工具的介绍

#### ■ **Xilisoft Video Converter**（音视频转换专家）

是一款多媒体转换工具，可以在DVD、VCD、SVCD、MOV、RM、AVI、MPEG、WMV等格式（支持的格式多达30种）间相互转换，你可以指定视频和音频编码器、设置各项输出参数，也可以在各音频格式如mp3、ogg、wav之间相互转换。其重要之处在于：可以从视频文件转换出任意流行的音频文件，诸如MP2、MP3、AC3、RA、M4A等。此工具支持多线程，可以同时进行多个文件的转换，所有的转换过程都非常简明快捷。

**使用方法：**选择文件—>添加文件—>预置方案（选择需转换的编码方式和格式）—>预置选项（可设置）—>输出目录—>点击工具栏中“转换勾选项目”即可；

## 第四部分：音视频格式转换工具介绍

### ■ Color7 video converter7.8

支持各种视频、音频转换，也可以刻录文件，界面漂亮速度快；

**使用方法：**选择输出文件名—>设置输出目录—>选择配置文件（目标文件）—>点击转换；

转换文件格式的同时，编码也相应转换了；

### 2. 音视频文件的编码以及详细信息查看

关于查看音视频文件的编码方式、文件格式、文件大小、采样率等信息；各媒体播放软件如：视频（Windows Media Player）音频

（TTPlayer）中选择查看文件属性即可查看到以上相关信息；或者是在播控界面也可以看到相关信息滚动；

同时也可以通过音视频转换工具来查看相关信息；

# 小结

---

此课件主要从音视频编码的基础知识、编码体系、文件格式等几个方面对音视频编码相关知识进行介绍，且对编码方案以及文件格式进行了详细的整理！（见附表：【音视频编码和文件格式列表.xls】）



THANK YOU!

