# 第一章 概 论

## 主要内容**：**

### 1.1 多媒体技术的概念

### 1.2 多媒体技术的发展历程

### 1.3 多媒体技术的研究内容

### 1.4 多媒体技术的应用及发展前景

# 1.1 多媒体技术的概念

## 什么是媒体?

## 媒体(Media)是信息表示和传输的载体。

## 媒体有哪些类型?

## CCITT（国际电话电报谘询委员会）对媒体分类如下：

### **感觉媒体**: 指能直接作用于人的感官，使人能直接产生感觉的一类媒体（文字、图形、图像、音频、视频、动画）

### **表示媒体**: 是为了在计算机和网络中加工、处理和传输感觉媒体而人为研究、构造出来的一种媒体（各种编码：如图像编码（JPEG,MPEG）,文字编码（ASCII码、GB2312等））

## CCITT（国际电话电报谘询委员会） 目前已被ITU（国际电信联盟取代）。

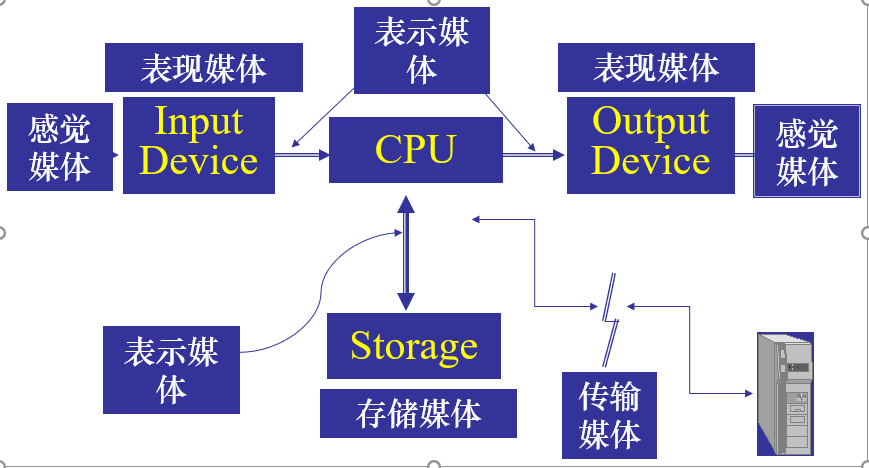
# 1.1 多媒体技术的概念

### **表现媒体**: 指感觉媒体和用于通信的电信号之间转换用的一类媒体，分为输入表现媒体和输出表现媒体（即鼠标、键盘等输入设备和显示器、扬声器等输出设备）

### **存储媒体**: 用于存放表示媒体(感觉媒体数字化后代码)以便计算机随时处理、加工和调用信息编码（如硬盘、软盘、磁盘、光盘、ROM及RAM等）

### **传输媒体**: 是将媒体从一处传送到另一处的物理载体 （如电缆、光缆等。）。

# 计算机与媒体



## 比较公认的“多媒体技术”定义：

## 所谓多媒体技术(Multimedia Computing)就是利用计算机交互式综合处理多种媒体信息──文本、图形、图象和声音、音频、视频、动画,使多种信息建立逻辑连接,集成为一个系统并具有交互性。

## 简而言之,多媒体技术就是利用计算机综合处理声、文、图、音频、视频、动画信息的相关软硬件技术,具有**集成性、实时性和交互性**。

# 多媒体技术的特点

## **集成性:** 媒体信息集成，表现媒体设备的集成

## **实时性：**声音、视频、动画等媒体是强实时的；提供时基媒体实时处理的能力

## **交互性：**与家用声像电器区别的关键特征，用户不能通过家用声像电器介入媒体内容

# 多媒体技术相关学科

## 它是多个相关学科综合发展的产物

### 微电子技术

### 计算机技术

### 通信技术

### 等

## 两种消费产业驱动：计算机、电视

## 人人交互、人机交互发展的必然

# 1.2 多媒体技术的发展历程

## 启蒙发展阶段(80’)

## 标准化阶段(90’)

## 应用普及(2000-)

# 一、启蒙发展阶段

## 多媒体技术最早起源于八十年代中期。

## **1984，Apple Macintosh计算机**

## 图形用户界面(GUI)：位映射、窗口、图符等技术;

## 鼠标配合GUI使用, 大大方便了用户的操作。

## **1985，Microsoft推出Windows**

## 一个多任务的图形操作环境;

## 使用鼠标驱动图形菜单;

## 一个用户界面友好的多层窗口操作系统。

## **1985, 美国Commodore Amiga系统**

## 世界上第一台多媒体计算机;

## CPU: Motorola M68000微处理器;

## 3个专用芯片: 图形处理Agnus 8370, 音响处理Paula8364, 视频处理Denise 8362。

## 专用的操作系统: Amiga DOS, 它能处理多任务, 并具有下拉菜单、多窗口、图符等功能。

## **1986,荷兰Philips/日本Sony CD-I**

## CD-I: 交互式紧凑光盘系统;

## 对大容量存储设备光盘的发展产生了巨大影响。

## 公布了该系统采用的CD-ROM的数据格式,经ISO认可成为国际标准;

## 大容量光盘的出现为存储表示声音、文字、图形、视频等高质量的数字化媒体提供了有效的手段。

## **交互式视频技术**

## 1983年, 美国RCA研究中心就着手研制交互式数字视频系统DVI, 后来RCA把DVI卖给GE公司。

## 1987年, Intel又从GE买到这项技术, 于1989年初把DVI技术开发成为一种可普及商品。

## 随后又和IBM合作,在Comdex/Fall’89展示会上推出Action Media 750多媒体开发平台。

## 1991年, Intel和IBM合作又推出了改进型的Action MediaⅡ, 在扩展性、可移植性、视频处理能力等方面均大大改善。

# 二、标准化阶段（90-）

## 1990年10月, 在微软公司召开多媒体开发工作者会议上提出MPC1.0标准 (基于286－386)。

## 1993年由IBM, Intel等数十家软硬件公司组成的多媒体个人计算机市场协会(MPMC)发布了多媒体个人机的性能标准MPC2.0 (486)。

## 1995年6月, MPMC又宣布了新的多媒体个人机技术规范MPC3.0 (586－75)。

## 1996年后, 新的PC机均支持基本多媒体功能

# 数字化图象压缩国际标准

## 多媒体技术分的关键是多媒体数据的压缩，不经压缩的多媒体数据（图像、音频、视频）等需要占用大量的存储空间，同时也为多媒体数据的网络传输带来困难，

## 如：

## CD音频：近80分钟的音频，需要近700MB的存储空间；

## 原始视频：分辨率1280x720，24位真彩色，每秒25帧的无声视频，一分钟未压缩视频约需4GB的存储空间。

# 数字化图象压缩国际标准

## JPEG标准（图像）

## 1991通过，成为ISO/IEC10918 JPEG; JPEG2000（适应Internet的需要）。

## MPEG标准（视频）

## MPEG-1(1.5Mbps,1992，VCD); MPEG-2(大于1.5Mbps，DVD); MPEG-4(甚低速率，网络视频); MPEG-7(多媒体检索); MPEG-21(多媒体应用框架)

## ITU H.26X标准

## H.261(P×64Kbps); H.262/3/4（面向不同速率要求的通讯，如可视电话、视频会议）

## AVS标准（中国国家音视频编码标准工作组）

# 数字化音频压缩标准

## MPEG-1,2,4音频标准，除此之外还有ITU的一些标准

## 16Kbps ITU 标准化方案G.728

## 64Kbps ISDN可视电话,语音16K,图象48K

## 32Kbps ITU标准化方案G.721标准

## 最初面向卫星通信, 长距离通信以及信道价格很高线路的语音传输。目前包括电视会议语音编码, 数字录音电话及高质量语音合成器等。

## 64Kbps ITU标准化方案G.722

## 面向7KHz带宽以语音和音乐为对象的标准化编码方案,面向高质量语音通信会议。具有3种工作模式即64Kbps,56Kbps和48Kbps。

## CD-ROM（主要存储数据）

## CD-DA（激光数字唱盘、即音乐CD）

## CD-I（交互式光盘，用于数字出版物，如多媒体教程）

## VCD（MPEG1视频）

## DVD（MPEG2视频）

# 三、应用普及阶段

## 90年代开始出现了大量的多媒体应用系统（单机）

## 2000以后随着网络的普及，网络多媒体应用大量出现

## 应用需求是驱动多媒体技术发展的主要动力

# 1.3 多媒体技术的研究内容

## 数据压缩/解压算法与标准（2-3）

## 多媒体数据存储技术 （4）

## 多媒体处理器(系统)（4）

## 多媒体系统软件与开发工具 （5）

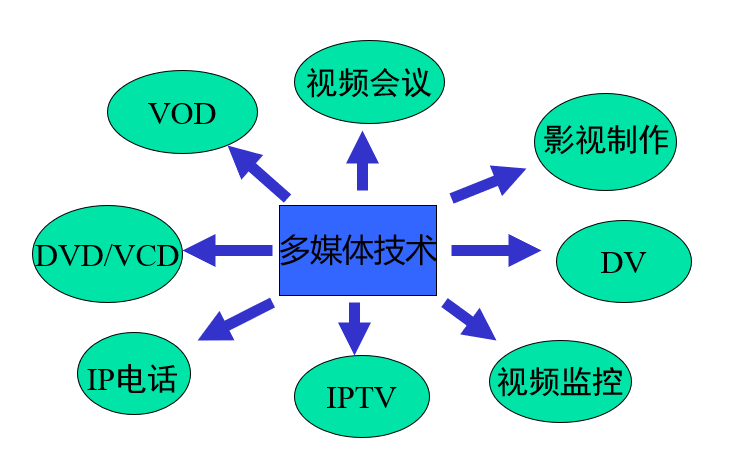
## 多媒体内容管理（6）

## 超文本和WEB技术（7-8）

## 多媒体网络 （9-10）

## 多媒体通信系统（9-10）

## 多媒体系统模型及典型应用（8，11）



# 1.4 多媒体技术的应用及前景

## 典型应用领域

#### 教育和培训

#### 咨询和演示

#### 娱乐和游戏

#### 管理信息系统(MIS)

#### 视频会议系统

#### 计算机支持协同工作

#### 视频服务系统

## 发展动向：个人化、网络化、智能化

# 第二章 多媒体数据编码基础

## 数字音频编码

## 数字图像编码

## 数字视频编码

## 常用的数据压缩技术

## 多媒体数据转换

# 2.1 数字音频编码

# **音频的基本特性**

## 声音是由振动的声波所组成（模拟音频），在任一时刻*t*，声波可分解为一系列正弦波线性叠加：

## *f*(*t*)= Σ*An*sin(*nωt*+*φn*)

## 其中, ω称为基频或基音，它决定声音的尖锐程度；*nω*称为*ω*的*n*次谐波分量或称为泛音，与声音的音色（是否有噪音）有关; *An*是振幅,表示声音的强弱（高低）；*φn*是*n*次谐波的初相位。

# 音频数字化

## 以数字方式表示声音。计算机是用声卡(包含ADC和DAC)来录制与播出数字化声音的。计算机对模拟的声音主要通过采样、量化、编码等步骤产生一系列数字化的声音数据。

## 事实上, 声波按频率可分为4类, 其中多媒体系统仅处理人类的听力所接受的频率范围的声音，我们称之为音频, 这个频率范围的声波称之为声音信号。

# 声音的频率分类



## 通过规则时间间隔测出音波振动幅度从而产生一系列声音数据。这种测出数据方法称之为**采样**,一秒内采样次数叫**采样率**。采样率需要满足奈奎斯特采样定理：采样频率应是信号频率的2倍以上。

## 采样的离散音频数据要转换成计算机能够表示的数据范围,这个过程称之为**量化 。**

## **量化后数字音频存储量计算公式**

## **音频数据存储量(字节)=采样率(Hz)×量化位数(位)×声道数×音频长度(秒)/8**

# MIDI （音乐设备数字接口）

## 1980年制定一项工业标准,目的是让音乐可以在不同的音乐设备间传输。

## MIDI提供了计算机外部的（数字）电子乐器与计算机之间的连接器接口。这种连接接口定义物理连接及电子乐器之间沟通的协议。

## MIDI也定义音频的形态与存储的方法。MIDI音频是以消息的方式而非波形的方式组成。

## MIDI有三种连接器(In、Out、Thru)。

## In为输入, Out为输出,而Thru是用来扩充MIDI与其它设备连接的。

## 计算机可以连接电子琴、电吉他、电萨克斯等。

## 音序器：计算机中记录、编辑、播放MIDI的程序。

## 音序器可以将音乐等声音以一种序列来储存。所谓序列便是一连串的音符加上系统事件的命令。

## MIDI适配器是用来改变频道、路径与按键的。当电子琴的键盘与一般的MIDI规格不一致时可以经由适配器来修正使两者一致。

# MIDI音频文件

## MIDI音频文件是一串时序命令,它记录乐器的行为。

## 命令消息分为频道消息(频道声音消息、频道模式消息)和系统消息(系统实时消息、系统通用消息与系统专用消息)。它是以某种乐器的发声为其数据记录的基础。

## 它的文件占用很少存储器空间,且可以做细部的修改，如修改节拍等。其声音效果不会因改变节拍而变调。

## MIDI不适合编制口语旁白的音频。

## 数字音频可从麦克风、录音带、CD、电视及其它来源获取。它把声音转换成存储器中的数字信息。

## 数字音频较为稳定,容易保持一致性,音频品质也较易获得保证。

## 缺点是记录非常详尽,数据量极大,数字音频文件较MIDI音频文件大出200倍以上。要修改数字音频细节非常困难,大大地增加了CPU的负担。

## 它可以适合任何一种音源,包括人的口语在内,故大多数节目仍采用这种音频。

# 3D音频

## 3D音效可分为以下几类：

## （1）扩展式立体声（使用声音延迟技术对传统的立体声（2D）进行处理，产生3D效果）。

## （2）环绕立体声（对多声道数据进行编码，如：AC3,DTS）。

## （3）交互式音效(音效随听者的位置移动发生变化)。

## 支持3D音频API种类:

## DirectX DirectSound 3D; Aureal 3D; EAX; Sensaura; Qsound; 杜比AC-3; 数字化影院系统DTS

# 2.2 数字图像编码

# 数字图像的种类：分辨率，1024x768

1、黑白图像（2）

2、灰度图像（256）

3、彩色图像（16位

24位，32位）

# 2.2.1 彩色空间（彩色图像）

## 常见模型**：**

### RGB彩色空间

### HSI彩色空间

### YUV彩色空间

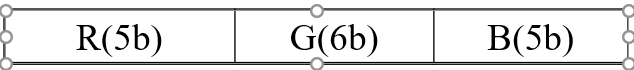
### YIQ彩色空间

# RGB彩色空间

## R、G、B是最基本的彩色表示模型,也是计算机系统中所使用的彩色模型。

### RGB5:6:5方式（16位增强色）

### 用2个字节表示一个象素，具体位分配。



## 因为人眼对绿色更敏感，因此，绿色用6个比特来存储

## RGB5:6:5方式

### RGB8:8:8方式（24位真彩色） ，32位真彩色

### R、G、B三个分量各占一个字节。

# HSI彩色空间

## 这种模型中, 用H(Hue,色调)、S(Saturation,饱和度)、I(Intensity,光强度)3个分量来表示一种颜色, 这种表示更适合人的视觉特性。

# YUV彩色空间 （广播电视领域广泛采用）

## Y为亮度信号,U、V是色差信号（基色信号减亮度信号即为色差信号）(B-Y,R-Y)，Cb，Cr。

## PAL制式彩色空间即为YUV。优点是亮度和色差信号分离,容易使彩色电视系统与黑白电视信号兼容。

## 国际无线电咨询委员会根据实验认为采用双倍度采样4:2:2（YUV各分量的采样率之比）方案效果较好, 提出CCIR601标准。

## 变换公式（YUV<--->RGB）

## Y = 0.299\*R + 0.587\*G+ 0.114\*B;

## U =-0.169\*R - 0.332\*G+ 0.500\*B;

## V = 0.500\*R + 0.419\*G - 0.081\*B

# YIQ彩色空间

## 广播电视系统另一种常用的亮度与色差分离的模型。NTSC制式彩色空间即为YIQ。这里Y是亮度, I和Q共同描述图象的色调和饱和度。

## 变换公式(YIQ<--->RGB)

## Y =0.299\*R+ 0.587\*G+ 0.114\*B;

## I =0.211\*R - 0.523\*G+ 0.312\*B;

## Q =0.596\*R - 0.275\*G - 0.322\*B

# 2.2.2 数字图象文件格式

## TIF

## PCX

## GIF、TGA、BMP、DVI、JPEG等

# 2.2.2 数字图象文件格式

## BMP文件结构

## BMP文件格式是Windows系统中广泛使用的图像文件格式，属于图像数据未经压缩的图像文件（即图像中直接保存图像中像素点阵中每个像素的颜色值）。

# 2.2.2 数字图象文件格式

## BMP文件的数据按照从文件头开始的先后顺序分为四个部分：

## 1）bmp文件头(bmp file header)：提供文件的格式、大小等信息

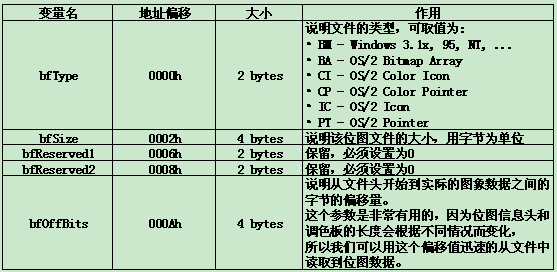
## 2）位图信息头(bitmap information)：提供图像数据的尺寸、位平面数、压缩方式、颜色索引等信息

## 3）调色板(color palette)：可选，如使用索引来表示图像，调色板就是索引与其对应的颜色的映射表

## 4）位图数据(bitmap data)：就是图像数据

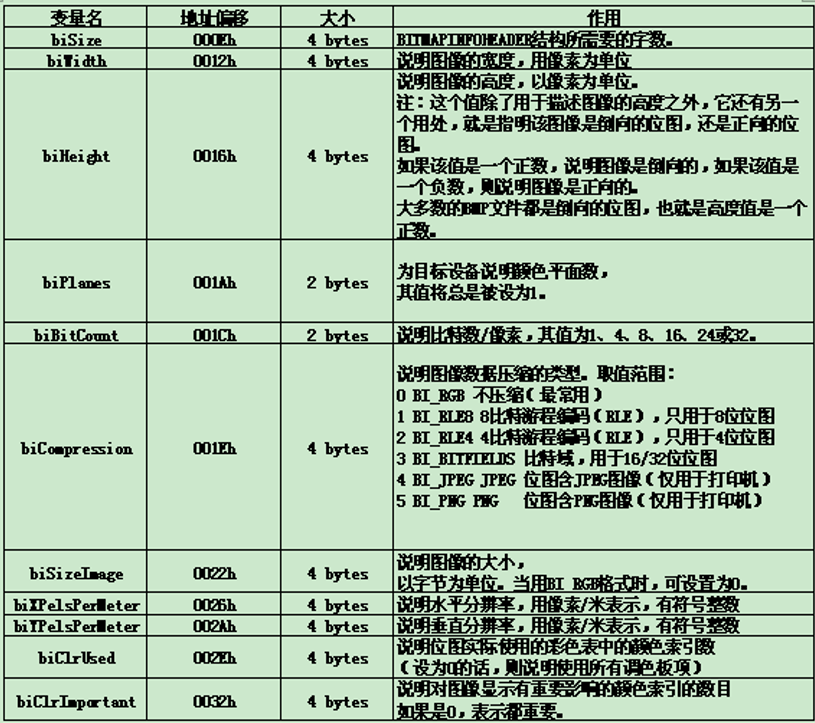
# 2.2.2 数字图象文件格式

## bmp文件头



# 2.2.2 数字图象文件格式

## bmp信息头



# 2.2.2 数字图象文件格式

## 调色板

## 调色板其实是一张映射表，标识颜色索引号与其代表的颜色的对应关系。

## 一般见到的图像以24位或32位图像，即R、G、B三种颜色各用8个bit来表示，这样的图像我们称为真彩色，这种情况下是不需要调色板的，也就是所位图信息头后面紧跟的就是位图数据了，而16位图像或256色图像是需要调色板的。

# TIF（tiff）文件格式

## 由美国Aldus Developer’s Desk和Microsoft制定

## 结构

### 文件头(8B)

### 参数指针表

### 参数数据表

### 图象数据

## 文件头

## 字节顺序(2B,表示存贮格式: II-Intel格式; MM—Motorola格式);

## 标记号(2B, 版本信息)，为固定值42,;

## 指向第一个参数指针表(4B)，相对于文件头的偏移量。

## 参数指针表

## 由每个长为12B参数块（描述图像的某一个属性）构成, 描述压缩种类、长宽、彩色数、扫描密度等参数。

## 较长参数(如调色板)只给出指针, 参数放在参数数据表中。其结构定义如下：

## *typedef struct {*

## *int tag-type（图像属性编号）;*

## *int number-size（该属性对应数据的类型）;*

## *long length（该属性对应数据的数量）;*

## *long offset（该属性数据相对于文件头的偏移量）;*

## *}TIF-FIELD;*

## 图象数据 按参数表中描述的形式按行排列（未压缩）

# PCX文件格式

## 由Z Soft公司最初制定

## 结构：

### 文件头(128字节)

### 数据部分(采用行程长度编码，压缩存储)

## 文件头结构定义

## ***typedef struct{***

## ***char manufacture; /\*always 0xa0\*/***

## ***char version；***

## ***char encoding; /\*always 1\*/***

## ***char bits-per-pixel; /\*color bits \*/***

## ***int Xmin, Ymin; /\* image origin \*/***

## ***int Xmax, Ymax; /\* image dimension \*/***

## ***int hres; /\* resolution values \*/***

## ***int vres;***

## ***char palette[48]; /\* color palette \*/***

## ***char reserved;***

## ***char color-planes; /\* color planes \*/***

## ***int bytes-per-line; /\* line buffer size \*/***

## ***int palette-type; /\* grey or color palette \*/***

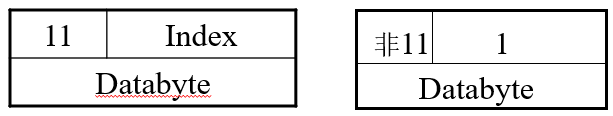
## ***char filler[58];***

## ***} PCXHEAD;***

## 其中Version若为5,文件内有个256色调色板（描述256中颜色的RGB分量值）,数据768字节,在文件最后。

## 文件体 对象素数据按行以字节为单位采用行程长度编码, 由包含Keybyte和Databyte的包组成。分2种情况:

## (1) 若Keybyte最高位为11, 则低6位为重复次数。但最多重复63次, 若再长重建一个包.



## PCX数据包的结构

## (2) 若Keybyte最高位不是11，那么该Databyte按原样写入图象文件。对一个字符的表示用长度为1的包。

# 2.3 **数字视频编码**

# 数字视频的结构

## 基本单位是帧

## 一帧即一幅图像，视频可以看做是按一定时间间隔连续播放的图像序列，这里没有考虑视频中包含的音频信息，实际上视频中还需包含音频信息，为讨论问题方便，我们这里只考虑视频中的图像数据。

## 若干同一场景的帧构成镜头（此处镜头为影视专业术语）

## 若干镜头构成情节

## 若干情节构成故事(节目)

# 国际视频标准（计算机领域）

## 计算机内能够直接处理的视频与数字电视系统中处理的视频其存储和处理方式是有区别的，原因在于，数字电视领域中一帧图像分为奇数场和偶数场显示，而计算机数字视频中的图像不需按此方式处理和显示，造成了计算机数字视频和广播电视领域数字视频的不同。

## 数字视频的标准制定机构有两个：ITU-T和ISO，其中 ITU-T发布的视频标准有H.261、H.262、H.263、H.263+、H.263++，其主要面向对象为数字广播电视领域，而ISO公布的MPEG系列标准有MPEG-1、MPEG-2、MPEG-4 和MPEG-7，并且计划公布MPEG-21，其主要应用领域为计算机。

# 国际视频标准（广播电视领域）

## **NTSC** 美国研制,是目前广泛使用的电视制式。它以525条横扫描线来组成一个屏幕帧,每秒30帧,其图象改变采用偶数线与奇数线相互交错更新的方式,造成视觉动态图象。

## **PAL** 中国、英国等国采用制式,W.Bruch1963年发明的,其基本原理类似于NTSC制式。以625条扫描线,每秒25帧,也是以奇偶数扫描线交错方式造成动态图象。

## **SECAM** 法国、俄罗斯等国采用制式。同样采用625条线和25帧, 但与NTSC和PAL相比, 其基础技术是采用频率调制, 传播方式也不同于以上两种。

# 2.4 常用数据压缩技术

# 2.4.1 数据压缩的基本原理

## 传统上用模拟方式表示声音和图象信息

## 易出故障，常产生噪音和信号丢失，且拷贝过程中噪音和误差逐步积累；

## 模拟信号不适合数字计算机加工处理。

## 数字化处理：巨大的数据量

## **采样定理:**

## 仅当采样频率≥2倍的原始信号频率时,才能保证采样后信号可被保真地恢复为原始信号。

## 采用8bit数字化（每个像素对应的比特数）,从而1秒钟电视信号的数据量约为99.2Mbits。

## 即约为100Mbps。650MB的CD-ROM仅能存约1分钟6秒钟的原始电视数据。若HDTV(1.2Gbps), 一张CD-ROM还存不下1秒的HDTV电视数据。

## 人说话的音频一般在20Hz到4KHz, 即人类语音的带宽为4KHz。依据采样定理, 设数字化精度为8b, 则1秒钟信号量为64Kbits。

## 因此, 人讲1分钟话的数据量为480KB。

## 数字化处理的关键问题－数据压缩

# 压缩的基础－数据冗余

## 空间冗余

## 时间冗余

## 信息熵冗余

## 结构冗余

## 知识冗余

## 视觉冗余

## 其它冗余

# 空间冗余

## 这是图象数据中经常存在的一种冗余。

## 在同一幅图象中,规则物体和规则背景的表面物理特性具有相关性,这些相关性的光成象结构在数字化图象中就表现为数据冗余。

# 时间冗余

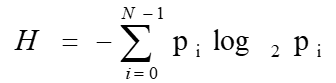
## 这是序列图象和语音数据中所经常包含的冗余。

## 图象序列中的两幅相邻的图象之间有较大的相关性,这反映为时间冗余。

## 人在说话时发音音频是一连续的渐变过程,而不是一个完全时间上独立的过程,因而存在时间冗余。

# 信息熵冗余

## 信息熵是指一组数据所携带的信息量,它定义为：



## *N*为数据类型数或码元个数, *Pi*为码元*yi*发生的概率.

## 为使信息编码单位数据量*d*接近于或等于*H*,应设：



## 其中*b(yi)*是分配给码元*yi*的比特数, 理论上应取*b(yi)=-log2Pi*. 实际一般取*b(y0)=b(y1)=…=b(yK-1)*.

## 例如, 英文字母编码码元长为7bit, 即*b(y0)=b(y1)= … =b(yK-1)=*7, 这样*d*必然大于*H*, 由此带来的冗余称为信息熵冗余或编码冗余。

# 结构冗余

## 有些图象从大域上看存着非常强的纹理结构，我们称它们在结构上存在有冗余.

## 例如布纹图象和草席图象

# 知识冗余

## 有许多图象的理解与某些基础知识有相当大的相关性。

## 例如,人脸的图象有固定的结构。比如说嘴的上方有鼻子, 鼻子的上方有眼睛, 鼻子位于正脸图象的中线上等等。

## 这类规律性的结构可由先验知识和背景知识得到, 我们称此类冗余为知识冗余。

# 视觉冗余

## 人类视觉系统对于图象场的任何变化,并不是都能感知的。

## 例如,对于图象的编码处理时,由于压缩或量化截断引入了噪声而使图象发生了一些变化,如果这些变化不能为视觉所感知,仍认为图象足够好。

## 事实上人类视觉系统一般分辨能力约为26灰度等级,而一般图象量化采用28灰度等级,这类冗余我们称为视觉冗余。

# 其他冗余

## 例如由图象的空间非定常特性所带来的冗余。

# 2.4 常用的数据压缩技术

## 根据解码后数据与原始数据是否完全一致，数据压缩方法划分为两类：

#### **可逆编码(无失真编码)** 解码图象与原始图象严格相同，压缩大约在2:1到5:1之间。如Huffman编码、算术编码、行程长度编码等。

#### **不可逆编码(有失真编码)** 还原图象与原始图象存在一定的误差，但视觉效果一般可以接受，压缩比可以从几倍到上百倍来调节。常用的有变换编码和预测编码

## 根据压缩的原理分类：

#### **预测编码** 利用空间中相邻数据的相关性来预测未来点的数据。差分脉冲编码调制(DPCM)和自适应差分脉冲编码调制(ADPCM)。

#### **变换编码** 将图象时域信号变换到频域空间处理。时域空间有强相关信号, 反映在频域上是某些特定区域内能量集中, 从而实现压缩. 正交变换如离散余弦变换, 离散付立叶变换和Walsh-Hadamard变换.

#### **量化与向量量化编码** 为了使整体量化失真最小, 就必须依照统计的概率分布设计最优的量化器。已知最优量化器是Max量化器。对象元点进行量化时, 也可以考虑一次量化多个点的向量量化。

#### **信息熵编码** 根据信息熵原理,让出现概率大的用短的码字表达,反之用长的码字表示。最常见的方法如Huffman编码、Shannon编码以及算术编码。

#### **子带编码** 将图象数据变换到频域后,按频域分带,然后用不同的量化器进行量化,从而达到最优的组合。或者分步渐近编码,随着解码数据的增加,图象逐渐清晰。离散小波变换

#### **模型编码** 编码时首先将图象中边界、轮廓、纹理等结构特征找出来,保存这些参数信息。解码时根据结构和参数信息进行合成,恢复出原图象。具体方法有轮廓编码、域分割编码、分析合成编码、识别合成编码、基于知识的编码、分形编码等。

# 2.4.1 预测编码

## 线性预测-DPCM

## 基本原理是基于图象中相邻象素之间具有较强的相关性。每个象素可根据已知的前几个象素来作预测。因此在预测编码中，编码和传输的并不是象素采样值本身，而是这个采样值的预测值与其实际值之间的差值

## 非线性预测（不讨论）

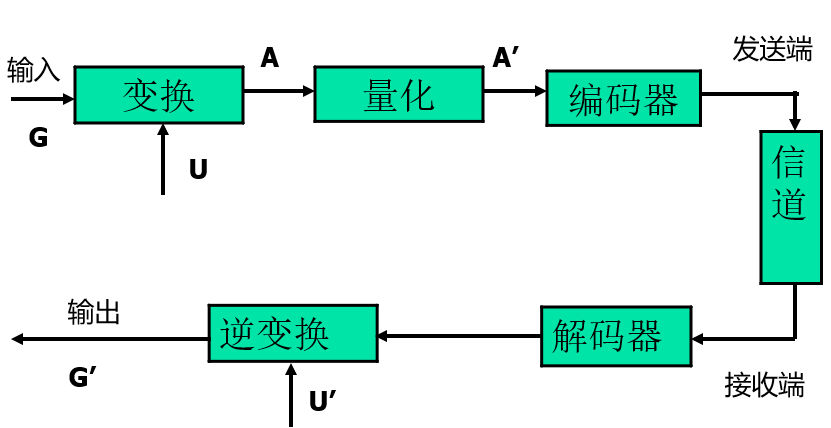
# 2.4.2 变换编码

## 输入图象G经正交变换U变换到频域空间,象素之间相关性下降,能量集中在变换域中少数变换系数上,已经达到了数据压缩的效果。

## 对变换系数A中那些幅度大元素予以保留,其它数量多的幅度小的变换系数,全部当作零不予编码,再辅以非线性量化,进一步压缩图象数据。

## 由于量化器存在,量化后变换系数A′和A间必然存在量化误差,从而引起输入图象G和输出图象G′间存在误差。图中U′是U的逆变换。

# 变换编码原理框图



## Karhunen-Loeve变换是**最佳变换**,但它的计算比较复杂。

## 实际应用中采用了一些准最佳变换如DCT,DFT和WHT等。

# 2.4.3 信息熵编码

## 又称为统计编码,它是根据信源符号出现概率的分布特性而进行的压缩编码。

## 基本思想: 在信源符号和码字之间建立明确的一一对应关系，以便在恢复时能准确地再现原信号,同时要使平均码长或码率尽量小。

## 如Huffman编码、算术编码。

# Huffman定理

## **定理** 在变长编码中,对出现概率大的信源符号赋于短码字,而对于出现概率小的信源符号赋于长码字。如果码字长度严格按照所对应符号出现概率大小逆序排列,则编码结果平均码字长度一定小于任何其它排列方式。

## Huffman定理是Huffman编码的理论基础

# 实现步骤

## (1) 将信源符号按概率递减顺序排列；

## (2) 把二个最小概率相加作为新符号的概率, 并按(1) 重排；

## (3) 重复(1)、(2), 直到概率为1；

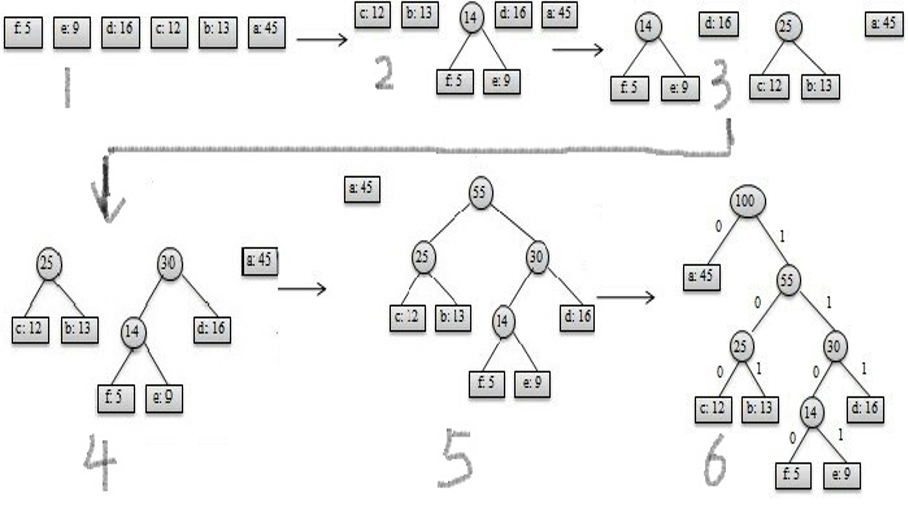
## (4) 在每次合并信源时, 将合并的信源分别赋“0”和“1”(如概率大的赋“0”,概率小的赋“1”)；

## (5) 寻找从每一信源符号到概率为1处的路径,记录下路径上的“1”和“0”；

## (6)写出每一符号的“1”、“0”序列(从树根到信源符号节点)。

# Huffman编码

# 已知一字符串中字符abcdef的出现次数如下，请确定该字符串的Huffman编码



## 上述编码的码字为:

## a:0

## b:101

## c:100

## d:111

## e:1101

## f:1100

# 算术编码

## 六十年代初,Elias提出了算术编码概念。

## 1976年, Rissanen和Pasco首次介绍了它的实用技术。其基本原理是将编码的信息表示成实数0和1之间的一个间隔(Interval),信息越长,编码表示它的间隔就越小,表示这一间隔所需的二进制位就越多。

# 算术编码举例

## 采用固定模式符号概率分配如下：

## 字符: a e i o u

## 概率: 0.2 0.3 0.1 0.2 0.2

## 范围:［0,0.2) ［0.2,0.5) ［0.5,0.6)［0.6,0.8)［0.8,1.0)

## 编码数据串为eai。令high为间隔的高端, low为低端，range为间隔的长度, rangelow为编码字符分配的间隔低端, rangehigh为编码字符分配的间隔高端。

## 初始high=1,low=0, range=high-low,

## 一个字符编码后新的low和high按下式计算：

## low=low+range×rangelow；

## high=low+range×rangehigh。

## (1) 在第一个字符e被编码时, e的rangelow=0.2, rangehigh=0.5, 因此:

## low=0+1×0.2=0.2

## high=0+1×0.5=0.5

## range=high-low=0.5-0.2=0.3

## 此时分配给e的范围为［0.2, 0.5)

## (2) 第二个字符a编码时使用新生成范围[0.2,0.5), a的rangelow=0, rangehigh=0.2, 因此:

## low=0.2+0.3×0=0.2

## high=0.2+0.3×0.2=0.26

## range=0.06

## 范围变成[0.2, 0.26)

## (3) 对下一个字符i编号, i的rangelow=0.5，rangehigh=0.6,range=0.06, 则：

## low=0.2+0.06×0.5=0.23

## high=0.2+0.06×0.6=0.236

## **结果**：用[0.23, 0.236)表示数据串eai,也就是从此区间中入一个数，作为编码进行传输，如果解码器知道最后范围是[0.23, 0.236),它马上可解得一个字符为e, 然后依次得到唯一解a、i, 最终得到eai。

## 解码过程：

## 解码的过程其实和编码的过程恰恰相反，解码是给定一个[0, 1)中的浮点数，通过解码操作之后就能完全获得原始的信号串。

## ·ÖÎöÉÏÊöÊµÀý£¬¼ÙÉèÊÜµ½Êý¾ÝÎª£º0.235则：

## 1）因0.235大于0.2，小于0.5,

## 故可判定第一个字符为e,然后计算(0.235-0.2)/(0.5-0.2)=0.117

## 2)因0.117大于0，小于0.2，

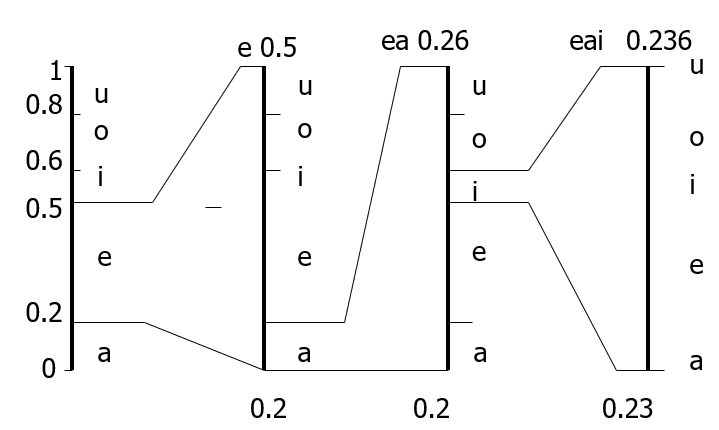
## 故可判定第二个字符为a,然后计算(0.117-0)/(0.2-0)=0.583

## 3)因0.583大于0.5，小于0.6，

## 故可判定第一个字符为i

## 实际上，此过程可以一直重复下去，所以解码端需知道发送的字符个数才可以正确解码。

# 算术编码过程表示



# 算术编码的特点

## 不必预先定义概率模型,自适应模式具有独特的优点;

## 信源符号概率接近时,建议使用算术编码,这种情况下其效率高于Huffman编码(约5%)。JPEG扩展系统采用。

# 2.5 多媒体数据转换

## 不同媒体表示不同的信息表示方式。研究媒体之间转换十分有意义。

## 有些媒体之间的转换是非常困难的事情,需要研究人类本身对各种媒体理解原理和解释过程。

## 有些媒体之间的转换则相对容易,几乎不用做什么工作。

# 部分媒体的转换关系



# 第三章 多媒体数据编码标准

## 静态图像编码标准JPEG

## 运动图像编码标准MPEG

## 视听通信编码解码标准H.26X

## AVS标准

## 声音压缩标准

# 3.1 JPEG标准

# 3.1.1 JPEG标准主要内容

## ISO/IEC 10918号标准“多灰度连续色调静态图象压缩编码”即JPEG标准, 选定ADCT作为静态图象压缩的标准化算法。

## 该标准为保证通用性,包含以下两种方式：

### 空间方式 可逆编码

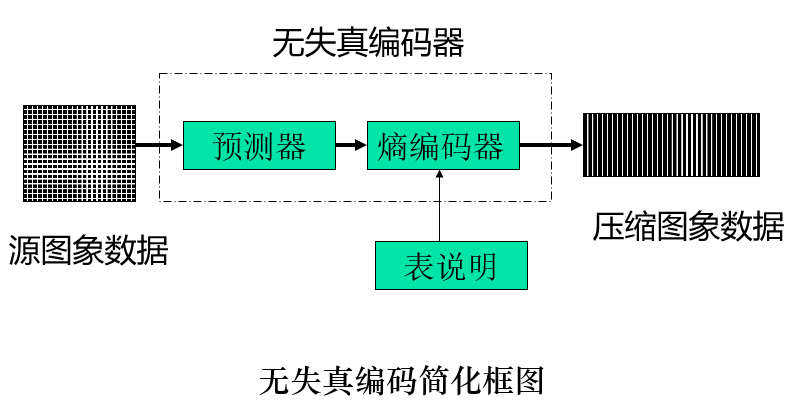
### 空间方式对于基本系统和扩展系统来说,被称为独立功能。

### DCT方式 非可逆编码, 包含基本系统(必须保证的功能)和扩展系统(扩充功能)

## 基本系统是实现DCT编码与解码所需的最小功能集, 大多数的应用系统只需要应用此标准, 就能满足要求（baseline）。

## 扩展系统是为了满足更广领域的应用要求而设置的。

# 1.基于DPCM的无失真编码（空域无失真编码）



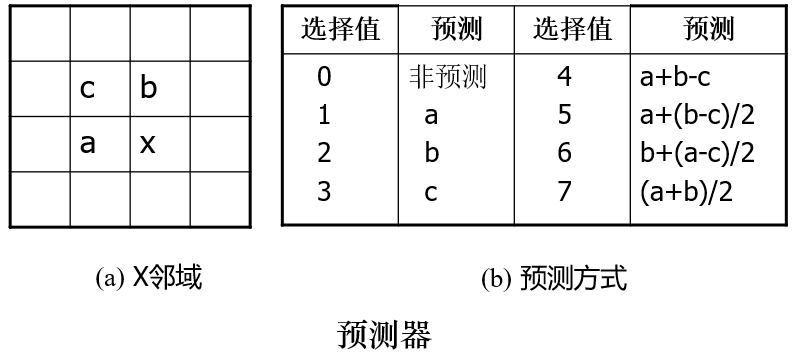
## 这里的预测器采用了基于DPCM（差分脉冲编码调制）的无失真编码技术，优点是硬件易实现，重建图象质量好。 缺点是压缩比太低, 大约为2:1。

## 熵编码器采用huffman编码或算术编码，同样也是无失真编码，在进行编码或解码时需要有Huffman码表或算术编码表。

## 经过这样的无失真编码后，总体压缩比可以达到2:1到5:1.得到的压缩图像可以无损恢复原图像 。

# DPCM无失真编码

工作原理是对X，不直接存储X，而是求X的预测值X’,将X-X’进行无失真熵编码，然后存储。对X’的求法见图给出的预测方式（共7种预测算子）。



# 2. 基于DCT的有失真压缩编码

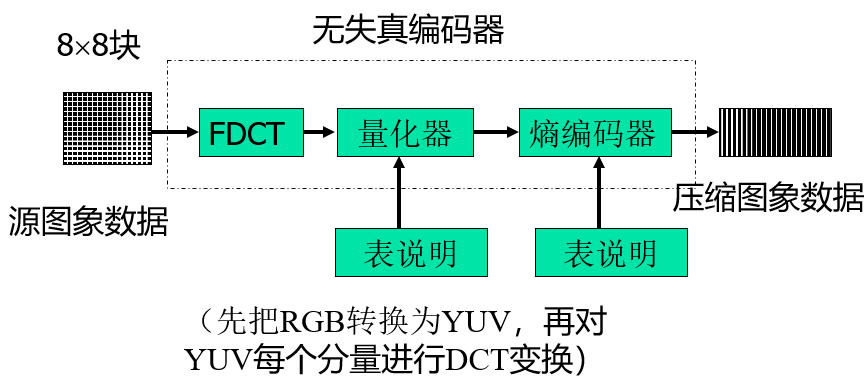
## 离散余弦变换

## 量化处理

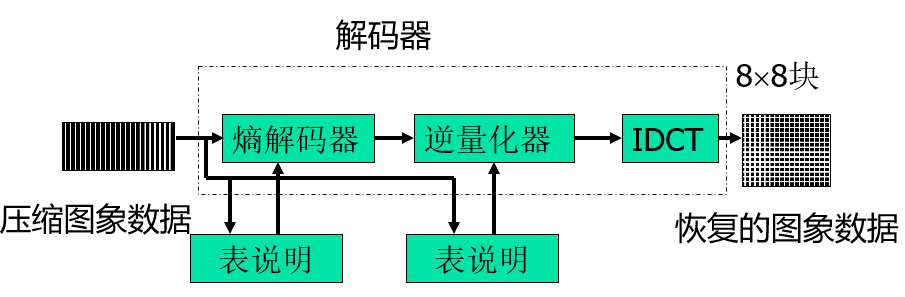
## DC系数的编码和AC系数的行程编码

## 熵编码

# 基于DCT编码过程解码过程



解码过程：



# 离散余弦变换

## (1)首先把原始图象顺序分割成8×8子块;

## (2)进行重采样，采样精度为P位(二进制), 把[0, 2P-1]（0-255）范围的无符号数变换成[-2P-1,2P-1]（-128-127）范围的有符号数, 作为离散余弦正变换(FDCT)的输入（减小绝对值波动）;

## (3)在输出端经离散余弦逆变换(IDCT)后又得到一系列8×8子块, 需将数值范围[-2P-1,2P-1]变换回[0, 2P-1]来重构图象。

## 这里用的8×8 FDCT的数学定义为：

## *F(u, v)*=(1/4)C(u) C(v)[∑x=07∑y=07f(x,y)·

## *cos*((2x+1)uπ/16)·*cos*((2y+1)vπ/16)]

## 8×8 IDCT的数学定义为：

## *f(x, y)=* (1/4)[∑u=07∑v=07 C(u)C(v)F(u,v)·

## *cos*((2x+1)uπ/16)·*cos*((2y+1)vπ/16)]

## 其中:C(u), C(v)= 1/√2当u, v=0

## C(u), C(v)=1其它

## 下面的编码针对FDCT输出的64个基信号的幅值(F(0,0),…,F(7,7)称作DCT系数)来进行

# 量化处理

## 量化是一个“多到一”的过程,是造成失真的原因。

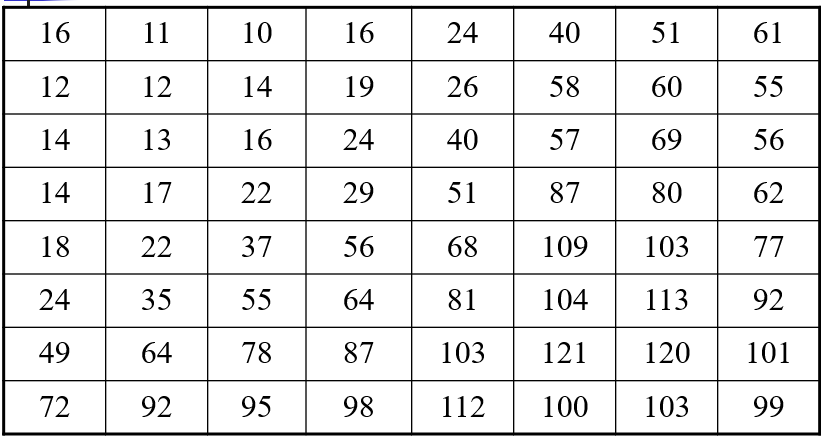
## 关键是找最小量化失真的量化器, JPEG采用线性均匀量化器,定义为对64个DCT系数除以量化步长, 然后四舍五入取整:

## *FQ(u, v)=Integer Round*[F(u, v)/Q(u, v)]

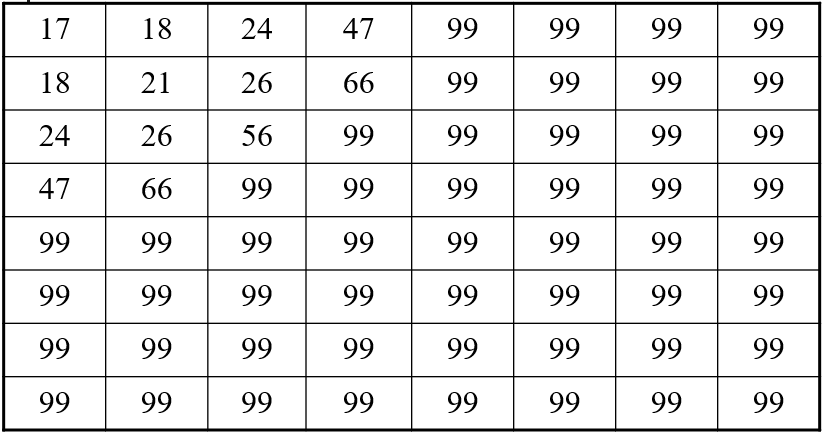
## *Q(u,v)*是量化器步长,它是量化表的元素。量化表元素随DCT系数的位置和彩色分量不同有不同的值,量化表尺寸为8×8与64个变换系数一一对应。

## 这个量化表应由用户规定(JPEG给出参考值-见表2.2,2.3), 并作为编码器的一个输入。

# 亮度量化表



# 色度量化表

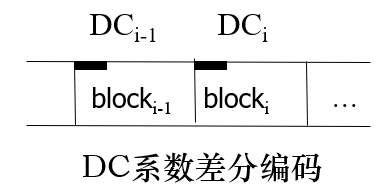


## 量化的作用: 在一定主观保真度图象质量前提下,丢掉那些对视觉影响不大的信息,通过量化可调节数据压缩比。

# DC系数的编码

## 64个变换系数经量化后, 坐标u=v=0的F(0, 0)称DC系数(直流分量), 它即64个空域图象采样值的平均值。

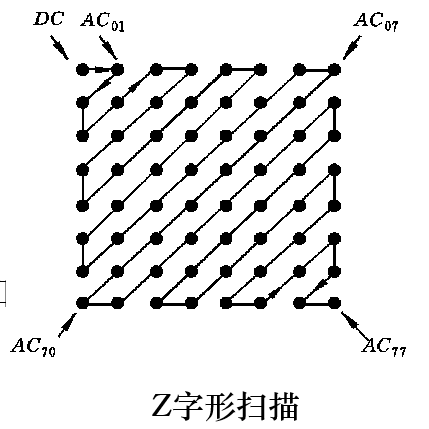
## 相邻8×8块之间DC系数有强相关性。JPEG对量化后的DC系数采用DPCM 编码, 即对DIFF= DCi-DCi-1编码。



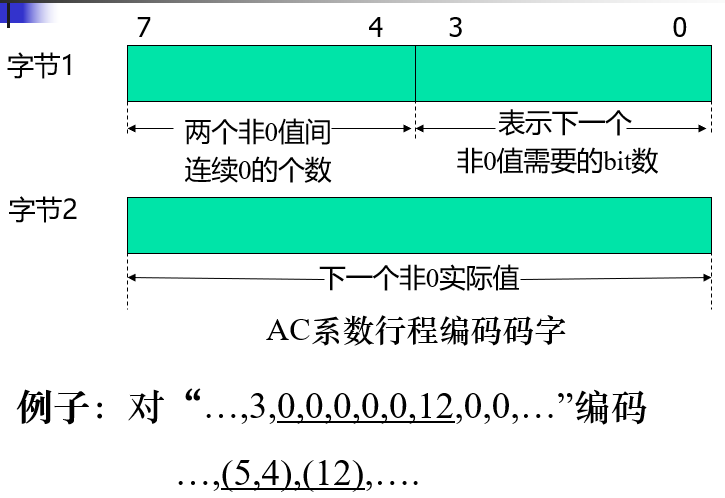
# AC系数的行程编码

## 量化后，AC系数出现较多的零（高频细节），其余63个交流系数(AC)采用行程编码。（zigzag）

## 从左上方AC0,1开始沿对角线方向“Z”字形扫描直到AC7,7扫描结束, 这样可增加行程中连续0的个数。



## AC系数编码的码字用两个字节表示,如图所示:



# 熵编码

## 为了进一步压缩数据,需对DC码和AC行程编码的码字再做基于统计特性的熵编码。

## JPEG建议的熵编码是Huffman编码和自适应二进制算术编码。

## 熵编码可分成两步进行:

### 把DC码和AC行程码转换为中间符号序列

### 给这些符号赋以**变长码字**

# AC系数熵编码的中间格式

## 需要进行熵编码的中间符号格式由两个符号组成:

### 符号1: (行程,尺寸)

### 符号2: (幅值)

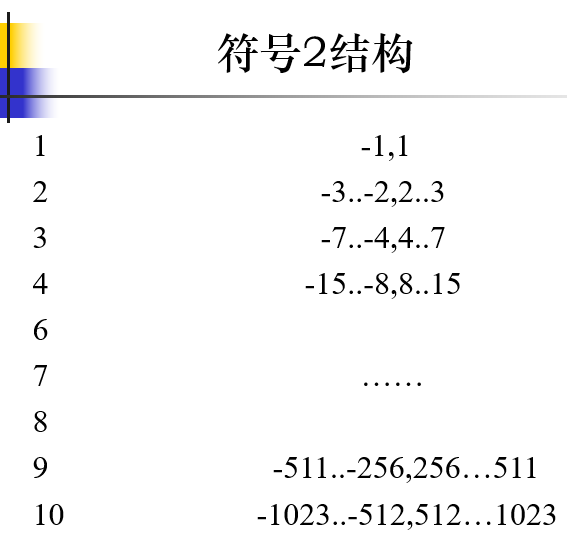
## 第一个信息参数“行程”表示前后两个非0的AC系数之间连续0的个数。

## 第二个信息参数“尺寸”是后一个非0的AC系数幅值编码所需比特数。

## 行程取值范围为1～15, 超过15时用扩展符号1 (15, 0)（前一个符号1）来扩充, 63个AC系数最多增加3个扩展符号1。编码结束时用(0,0)表示。

## “尺寸”取值范围为0～10。

## “幅值”用以表示非0的AC系数的值, 范围为［-210, 210-1］(最长10bit), 结构形式如表2-4所示。



# DC系数的熵编码

## 对于直流分量DC也有类似于AC系数的编码格式

### 符号1：(尺寸)

### 符号2：(幅值)

## “尺寸”表示DC差值的幅值编码所需的比特数, 而“幅值”表示DC差值的幅值, 范围为［-211,211-1**］**。可在表2.4中多加一级, 幅值尺寸以1到11比特表示。

## 将63个AC系数表示成为符号1和符号2序列,其中连续0的长度超过15时,有多个符号1; 块结束(EOB)时仅有一个符号1(0,0)。

## “…4,0,0,0,0,0,0,0,……0,3,0”. 4,3之间有31个0.

## (15,0),(15,0),(1,2),(3)

## 可变长度熵编码就是对上述序列进行变长编码。

## 对DC系数、AC系数中的符号1采用Huffman表中的变长码编码(VLC),这里Huffman变长码表必须作为JPEG编码器输入。

## 符号2用码字长度在表2.4中给出的变长整数VLI码编码。VLI是变长码,但不是Huffman码。VLI的长度存放在VLC中, JPEG提供VLI码字表供用户使用

## JPEG中为了更进一步节约空间，并不直接保存数据的具体数值，而是将数据按照位数分为16组，保存在表里面。这也就是所谓的变长整数编码VLI。即，第0组中保存的编码位数为0，其编码所代表的数字为0；第1组中保存的编码位数为1，编码所代表的数字为-1或者1......，如下面的表格所示，这里，暂且称其为VLI编码表：



## 对应DC系数中间码的符号1为存储DPCM差值对应需要多少位存储（也就是上表中的组号），符号2为对应编码。比如DC系数的DPCM差值为6，则对应的DC编码为（3）（6）

## JPEG提供2套Huffman码表: 亮度和色度。每套又有DC表和AC表各1个。共有4个表。

## 表定义（亮度DC系数码表）

## 16B说明码字长度：

## X’00 01 05 01 01 01 01 01 01 00 00 00 00 00 00 00’

## *第i个(1-16)元素值表示长度为i的Huffman码个数*。

## 紧跟一组值说明亮度表分类：

## X’00 01 02 03 04 05 06 07 08 09 0A 0B’

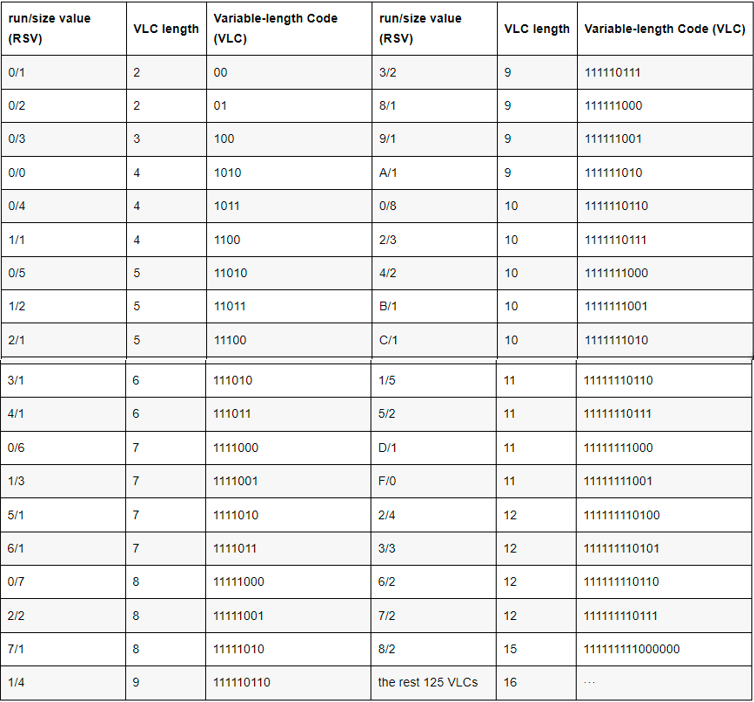
# 亮度DC系数表



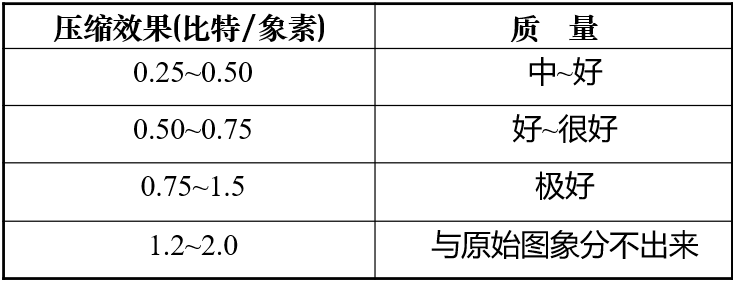
# 色度DC系数表



# 亮度AC系数表



# JPEG压缩效果评价



# 3. 基于DCT的累进操作方式编码

## 顺序方式: 每个图象分量的编码一次扫描完成的;

## 累进方式: 图象分量编码要经过多次扫描才完成。

## 累进方式

## 第一次扫描只进行一次粗糙图象的扫描压缩, 以相对于总的传输时间快得多的时间传输粗糙图象, 并重建一帧质量较低的可识别图象; 在随后的扫描中再对图象作较细的压缩, 这时只传递增加的信息,可重建一幅质量提高一些的图象。这样不断累进, 直到满意的图象为止。

## 需在量化器的输出与熵编码的输入之间,增加一个足以存储量化后DCT系数的缓冲区,对缓冲区中存储的DCT系数多次扫描, 分批编码。

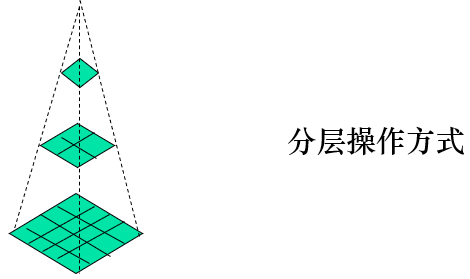
## 两种累进方式：

### 频谱选择法 扫描中只对64个DCT变换系数中某些频带的系数进行编码、传送, 随后对其它频带编码、传送, 直到全部系数传送完毕为止。

### 按位逼近法 沿着DCT量化系数有效位方向分段累进编码。如第1次扫描只取最高有效位的n位编码、传送, 然后对其余位进行编码、传送。

# 4. 基于DCT的分层操作方式

## 分层方式是对一幅原始图象的空间分辨率,分成多个分辨率进行“锥形”的编码方法,水平(垂直)方向分辨率的下降以2的倍数因子改变。



# 分层操作方式的过程

## (1)把原始图象空间分辨率降低。

## (2)对已降低分辨率的图象采用基于DCT的顺序方式、累进方式或无失真预测编码中的任何一种编码方法进行编码。

## (3)对低分辨率的图象解码,重建图象,使用插值滤波器,对它插值,恢复图象的水平和垂直分辨率。

## (4)把分辨率已升高的图象作为原始图象的预测值,对它们的差值采用基于DCT的顺序方式、累进方式或用无失真方式进行编码。

## (5)重复(3)、(4)直到图象达到完整的分辨率编码。

# 3.1.2 JPEG2000简介

## 基于Internet的多媒体应用给图像编码提出了新的要求. 2000年12月公布JPEG 2000标准(ISO 15444), 其目标是在高压缩率情况下保证图像传输质量。

## JPEG中采用DCT变换考察整个时域过程的频域特征或整个频域过程的时域特征。JPEG2000采用以小波变换为主的多分辨率编码方式。

## JPEG2000统一了面向静态图像和二值图像的编码方式, 是既支持低比率压缩又支持高比率压缩的通用编码方式。

## 该算法主要特点如下：

## (1)高压缩率。与JPEG相比,可修复约30％的速率失真特性。JPEG和JPEG2000在压缩率相同时, JPEG2000的信噪比将提高30％左右;

## (2)无损压缩。预测编码作为对图像进行无损编码的成熟方法被集成在JPEG2000中;

## (3)渐进传输。JPEG2000可实现以空间清晰度和信噪比为首的各种可调节性,从而实现渐进传输,即具有“渐现”特性.

## (4)感兴趣区域压缩。JPEG2000 支持所谓的“感兴趣区域”。

# 3.2 运动图像压缩标准MPEG

# 3.2.1 MPEG标准简介

## MPEG标准是面向运动图象压缩的一个系列标准。

## 最初MPEG专家组工作项目是3个,即在1.5Mbps, 10Mbps,40Mbps传输速率下图象编码, 分别命名为MPEG-1,MPEG-2, MPEG-3。MPEG-3后被取消.

## 为了满足不同的应用要求, MPEG又将陆续增加其它一些标准MPEG-4,MPEG-7,MPEG-21。

## MPEG算法编码过程和解码过程是一种非镜象对称算法(不对称), 解码过程要比编码过程相对简单些。

## MPEG-1和MPEG-2只规定了解码的方案, 重点将解码算法标准化。因而用硬件实现MPEG算法时, 人们首先实现MPEG的解码器,如C-Cube公司CL450解码器系列。

## 随着MPC性能提高,软件解压功能得到支持。

# 3.2.2 MPEG-1系统

## “用于数字存储媒体运动图象及其伴音速率为1.5Mbps的压缩编码”简称MPEG-1, 作为ISO/IEC 11172号建议于1992年通过。

## 主要用于在CD-ROM存储运动视频图像, 它针对标准分辨率(NTSC制为352×240; PAL制为352×288)的图像进行压缩, 每秒30帧画面, 具备CD音质。

## 它还用于数字电话网络上的视频传输, 如非对称数字用户线路(ADSL)、视频点播、教育网络等。

## 使用MPEG-1的压缩算法, 可将一部120分钟长的电影压缩到1.2GB左右。因此, 它被广泛地应用于VCD制作。

# MPEG-1分为5个部分

## (1)MPEG系统(11172-1), 定义音频、视频及有关数据的同步;

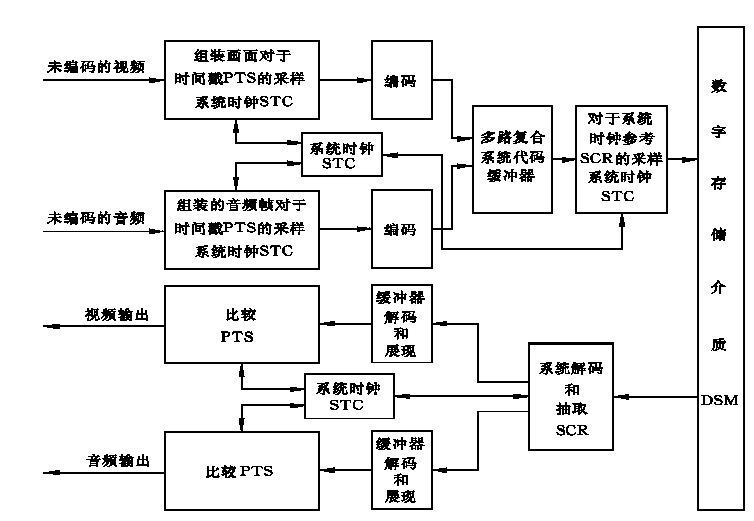
## (2)MPEG视频(11172-2), 定义视频数据的编码和重建图象所需的解码过程,亮度信号分辨率为360×240, 色度信号分辨率为180×120;

## (3)MPEG音频(11172-3), 定义音频数据的编码和解码;

## (4)一致性测试(11172-4);

## (5) 软件模拟(11172-5)。

# MPEG-1解码器原型



## MPEG-1编解码器原型:

## 多路复合而成的码流假设以介质特定格式存储在数字存储介质(DSM)或网络上, 标准不规定介质特定格式。

## 系统解码器从输入多路复合流中抽取定时信息,并对输入流进行分流处理, 输出两个基本流分别给视频和音频解码器。

## 视频和音频解码器分别解码输出视频和声音信号。

## 系统、视频、音频和介质4个解码器之间用定时信息进行同步。

## 多路复合流构造为2层: 系统层和压缩层。系统解码输入的是系统层; 而视频、音频解码器输入的是压缩层。

## 系统解码器执行两类操作:

### 一类是作用在整个多路复合流上的操作,称为复合流操作;

### 另一类是作用在单个基本流上的操作,称为特定流操作。

## 系统层分为两个子层:

### 一个子层称为包(pack),是复合流操作对象;

### 另一个子层称为组(packet),它用于特定流操作。

## MPEG音频标准的特点:

## 音频信号采样率可以是32KHz,44.1KHz或48KHz。

## 压缩后的比特流可以按以下4种模式之一支持单声道或双声道:

### 提供给单音频通道的单声道模式；

### 提供给两个独立的单音频通道的双-单声道模式；

### 提供给立体声通道的立体声模式；

### 联合立体声模式,利用立体声通道之间的关联或通道之间相位差的无关性,或者对两者同时利用。

## MPEG音频标准提供3个独立的压缩层次,用户可在复杂性和压缩质量之间权衡选择。

### 层1最简单,使用比特率384Kbps,主要用于DCC;

### 层2的复杂度中等,使用比特率192Kbps左右, 主要应用于数字广播的音频编码、CD-ROM上的音频信号以及CD-I和VCD。

### 层3最为复杂,使用比特率64Kbps,尤其适用于ISDN上的音频传输,有损压缩但音质保持逼真效果。

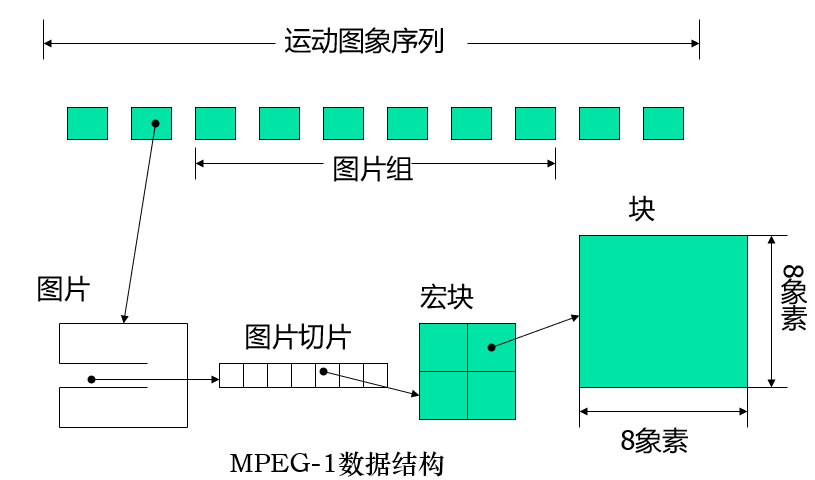
### MP3音乐 是利用 MPEG Audio Layer 3 的技术, 声音采用 1:10 甚至 1:12 的压缩率

## 压缩后的比特流具有预定义的比特率之一。MPEG音频标准也支持用户使用预定义的比特率之外的比特率。

## 编码后的比特流支持循环冗余校验(CRC)。

## MPEG音频标准还支持在比特流中载带附加信息。

# 3.2.3 MPEG视频数据流的结构



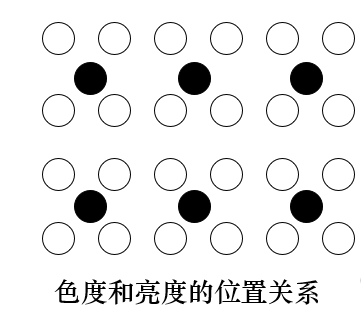
## 运动序列

## 图象组

## 图象信号分3个部分:

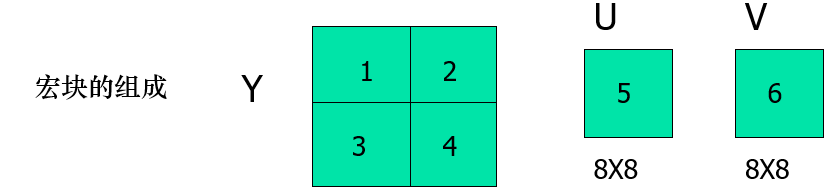
## 一个亮度信号Y和两个色度信号U、V。

亮度信号Y由偶数个行和偶数个列组成, 色度信号U、V分别取Y信号在水平、垂直方向的1/2。如图所示, 黑点代表色度U、V位置,亮度Y位置用白圈表示。



## 块: 一个块由一个8×8的亮度信息或色度信息组成。

## 宏块一个宏块由一个16×16的亮度信息和两个8×8色度信息构成，如图所示。



## 图象切片 由一个或多个连续的宏块构成。

# 3.2.4 MPEG-1视频编码技术

## 主要问题:一方面无法达到很高的压缩比，另一方面用单一的静止帧内编码方法能最好地满足随机存取的要求。

## 解决方法：对这两个方面做了折衷考虑。即为了减少时间上冗余性的基于块的运动补偿技术和基于DCT变换的减少空间上冗余性的ADCT技术

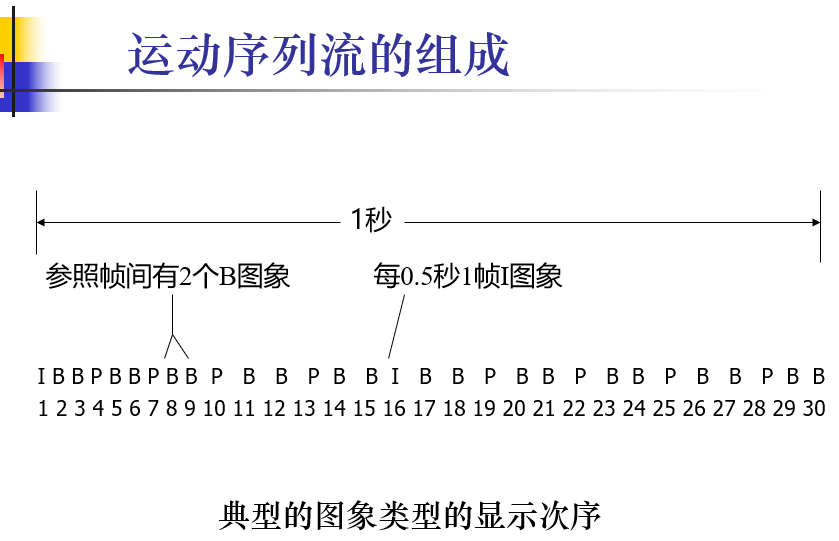
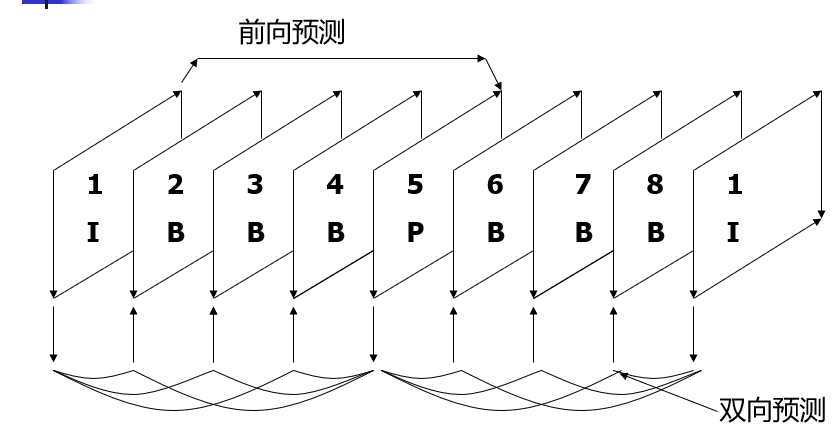
## 在MPEG中将图象分为3种类型：

### I图象 利用图象自身的相关性压缩, 提供压缩数据流中的随机存取的点.

### P图象 用最近的前一个I图象(或P图象)预测编码得到(前向预测)。

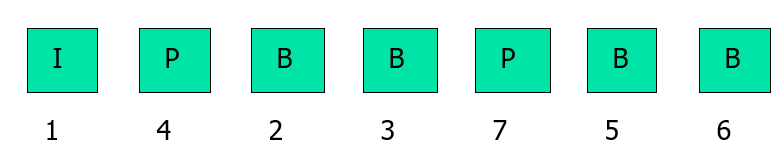
### B图象 B图象在预测时, 既可使用了前一个图象作参照, 也可使用下一个图象做参照或同时使用前后两个图象作为参照图象(双向预测)。

# 帧间预测



# 传输顺序

## MPEG编码器需对上述图象重新排序, 以便解码器高效工作, 因为参照图象必须先于B图象恢复之前恢复。上述1～7帧图象重排后图象组次序为：



# 运动补偿技术

## 运动补偿技术: 用于消除P图象和B图象在时间上的冗余性提高压缩效率。

## MPEG中,运动补偿技术工作在宏块一级。

## B图象宏块有4种类型

## 帧内宏块, 简称I块；

## 前向预测宏块, 简称F块；

## 后向预测宏块, 简称B块；

## 平均宏块, 简称A块。

## 对于P图象, 其宏块只有I块和F块两种。

## 无论B图象和P图象,I块处理技术都与I图象中采用技术一致即ADCT技术。

## 对于F块、B块和A块,MPEG都采用基于块的运动补偿技术。

### F块预测时其参照为前一个I图象或P图象

### B块预测时其参照为后一个I图象或P图象

### 对于A块预测其参照为前后两个I图象或P图象

# 基于块的运动补偿技术

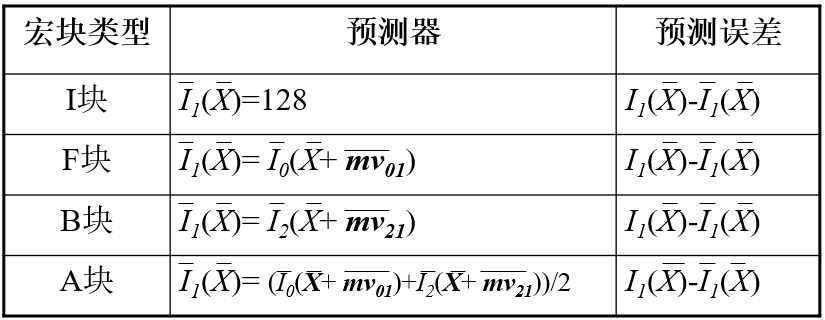
## 基于块的运动补偿技术,就是在其参照帧中寻找符合一定条件,当前被预测块的最佳匹配块。

## 找到匹配块后,有两种处理方法：

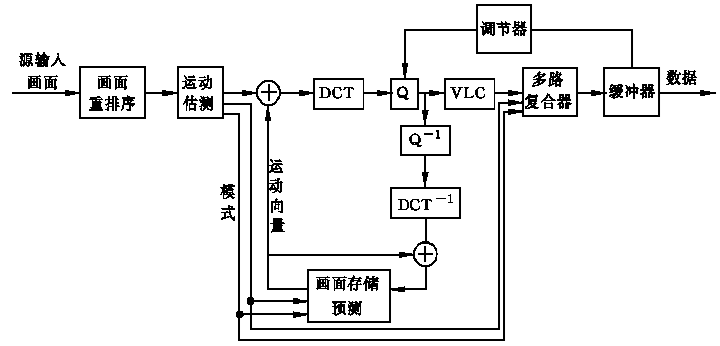
### 一是在恢复被预测块时,用匹配块代替；

### 二是对预测的误差采用ADCT技术编码,在恢复被预测块时,用匹配块加上预测误差。

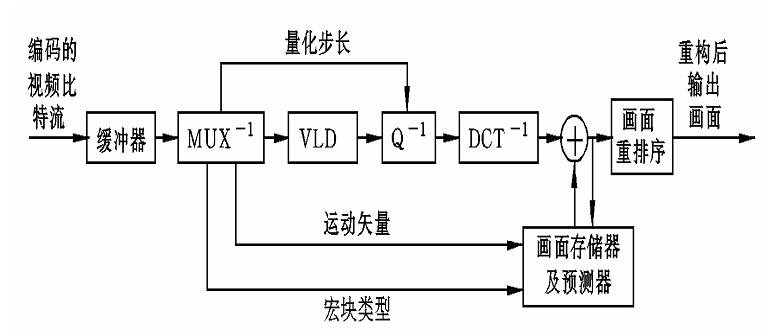
# 宏块的预测方式



# MPEG-1视频系统



## 简化的视频编码框图



## 基本的视频解码器框图

# 3.2.5 MPEG-2标准

## MPEG-2(ISO/IEC 13818)标准制定于1994年。它利用网络3～100Mbps的数据传输率来支持具有更高分辨率图象的压缩和更高的图象质量。

## MPEG-2可支持交迭图象序列,支持可调节性编码,多种运动估计方式,提供一个较广的范围改变压缩比,以适应不同画面质量、存储容量和带宽的要求。

## 它与MPEG-1兼容基础上实现了低码率和多声道扩展：MPEG-2可以将一部120分钟长的电影压缩到4～8GB(DVD质量),其音频编码可提供左右中及两个环绕声道、一个加重低音声道和多达7个伴音声道。

## 除了作为DVD指定标准外，MPEG-2还可用于为广播、有线电视网、电缆网络等提供广播级数字视频.

## MPEG-2 Video 定义了不同的功能档次(Profiles)

## 每个档次又分为几个等级(Levels),一个等级为N的解码器能够对最高为该等级的数码流解码。

# 5个档次

## • 简单型（Simple）

## • 基本型（Main）

## • 信噪比可调型（SNR Scalable）

## • 空间可调型（Spatial Scalable）

## • 增强型（High）

# 4个等级

## • 低级（Low） 352×288×30, 它面向VCR并与MPEG-1兼容；

## • 基本级（Main） 720×460×30或720×576×25，它面向视频广播信号；

## • 高1440级（High-1440） 1440×1080×30或1440×1152×25，它面向HDTV；

## • 高级（High） 1920×1080×30或1920×1152×25，它面向HDTV。

# 11种规范

## 高级的基本型 MP@HL

## 高级的增强型 HP@HL

## 高-1440级的基本型 MP@H1440

## 高-1440级的空间可调型 SSP@H1440

## 高-1440级的的增强型 HP@H1440

## 基本级的简单型 SP@ML

## 基本级基本型 MP@ML

## 基本级的信噪比可调型 SNP@ML

## 基本级的增强型 HP@ML

## 低级的基本型 MP@LL

## 低级的信噪比可调型 SNP@LL

# MPEG-2音频

## 基本特性之一是向后与MPEG-1音频兼容。

## 可以是5.1也可以是7.1通道的环绕立体声。

### 5.1也称为“3/2-立体声加LFE”,其含义是播音现场前面可有3个喇叭通道(左、中、右),后面可有2个环绕声喇叭通道。LFE是低频音效加强通道。

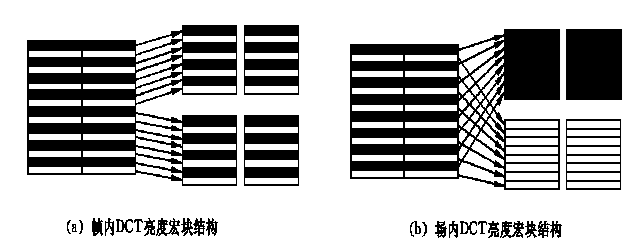
### 7.1通道环绕立体声与5.1类似,它另有中左、中右2个喇叭通道。

# MPEG-2编码方法

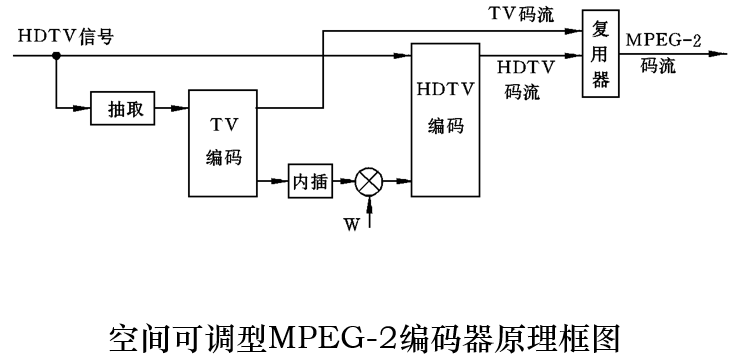
## MPEG-2的编码方法和MPEG-1区别主要是在隔行扫描制式下, DCT变换是在场内还是在帧内进行由用户自行选择, 亦可自适应选择。

## 一般情况下, 对细节多、运动部分少的图象在帧内进行DCT, 而细节少、运动分量多的图象在场内进行DCT。

## MPEG-2采用可调型和非可调型两种编码结构。还可以使用一个基本层加上多个增强型的多层编码结构,这由用户按质量和压缩比要求选择使用.



# MPEG-2亮度宏块结构



# 3.2.6 MPEG-4标准

## MPEG-4“甚低速率视听编码”1998年11月公布, 它针对低速率视音频编码,更加注重系统交互性和灵活性。

## MPEG-4引入了AV对象(AVO), 使得更多的交互操作成为可能：“AV对象”可以是一个孤立的人,也可以是这个人的语音或一段背景音乐等。

## MPEG-4对AV对象的操作主要有：

## 采用AV对象来表示听觉、视觉或者视听组合内容；

## 组合已有AV对象来生成复合的AV对象，并生成AV场景；

## 对AV对象的数据灵活地多路合成与同步，以便选择合适的网络来传输这些AV对象数据；

## 允许接收端用户在AV场景中对AV对象进行交互操作等。

## MPEG-4**标准主要构成部分**：

## (1)**传输多媒体集成框架(DMIF)**。主要用于解决交互网络中、广播环境下以及光盘应用中多媒体应用操作问题。它是MPEG-4制订会话协议,用来管理多媒体数据流。通过传输多路合成比特信息来建立客户端和服务器端的连接与传输。

## (2)**场景描述**。场景声音视频对象间的关系的描述体现在两个层次：BIFS描述场景中对象的空间时间安排,观察者可以有与这些对象交互的可能性；在较低的层次上,对象描述子定义针对每个对象的基本流的关系，并提供诸如访问基本流需要的URL地址、译码器的特性、知识产权等其它信息。

# 

## MPEG-4具备与Web3D X3D和W3C SMIL的互操作性。XMT格式可在SMIL播放器、VRML 和 MPEG-4 播放器间互换。

## (3)**音频编码。**MPEG-4不仅支持自然声音,而且支持合成声音。MPEG-4的音频部分将音频的合成编码和自然声音的编码相结合,并支持音频的对象特征。支持MIDI和TTS.

## (4)**视频编码**。MPEG-4也支持对自然和合成的视觉对象的编码。合成的视觉对象包括2D、3D动画和人面部表情动画等。

## (5)**缓冲区管理和实时解码**。MPEG-4定义了一个系统解码模式(SDM), 该解码模式描述了一种理想的处理比特流句法语义的解码装置,它要求特殊的缓冲区和实时模式。通过有效地管理,可以更好地利用有限的缓冲区空间。

## 与MPEG-1和2相比,MPEG-4更适于交互AV服务以及远程监控,其设计目标使它具有更广适应性和可扩展性：传输速率可在4.8-64kbps之间,分辨率为176×144, 可以利用很窄带宽通过帧重建技术压缩和传输数据。

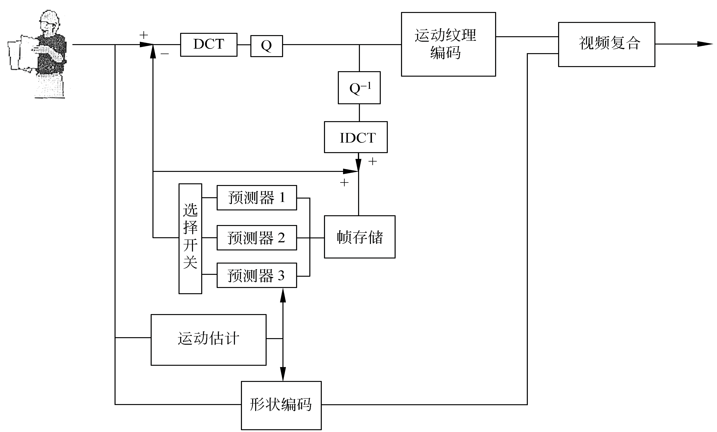
## 它将应用在数字电视、动态图像、互联网、实时多媒体监控、移动多媒体通信、Internet/Intranet上的视频流与可视游戏、DVD上的交互多媒体等方面。

## MPEG-4用MPEG-4压缩算法的ASF可以将120分钟的电影压缩为300MB左右的视频流；采用MPEG-4压缩算法的DIVX编码技术可以将120分钟的电影压缩600MB左右，也可以将一部 DVD影片压缩到 2 张 CD-ROM上.

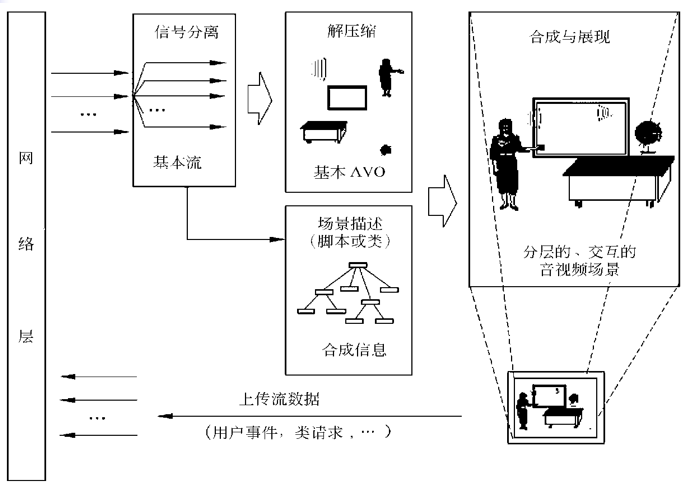
## MPEG-4属于一种高比率有损压缩算法,其图像质量始终无法和DVD的MPEG-2相比, 毕竟DVD的存储容量较大。

## 要想保证高速运动的图像画面不失真, 必须有足够的码率,目前MPEG-4的码率虽然可以调到和DVD差不多,但总体效果还有不小的差距。因此, 对图像质量要求较高的专业视频领域暂时还不能采用。

# MPEG-4视频编码器的算法方框图



# MPEG-4终端的构成(接收端)



## 背景全景图+视频对象(VO)=合成图象

# 3.3 视听通信编码解码标准H.26X

# 3.3.1 H.261

## ITU推荐H.261方案标题“64Kbps视声服务用视象编码方式”, 又称为**P×64Kbps视频编码标准。**

## P取值范围为1-30。P=1或2时,仅能支持QCIF(176×144)分辨率格式, 每秒帧数较低的可视电话;当P≥6时,则可支持图象分辨率格式为CIF(352×288)的电视会议。

## P×64Kbps压缩算法采用基于DCT的变换编码和带有运动预测的DPCM预测编码的混合方法。

## P×64Kbps标准的压缩算法与MPEG-1标准有许多共同之处, 只是传输速率P×64Kbps覆盖较宽的信道频带, 而MPEG-1是基于较窄的频带上传输。

# 3.3.2 视频层次数据结构

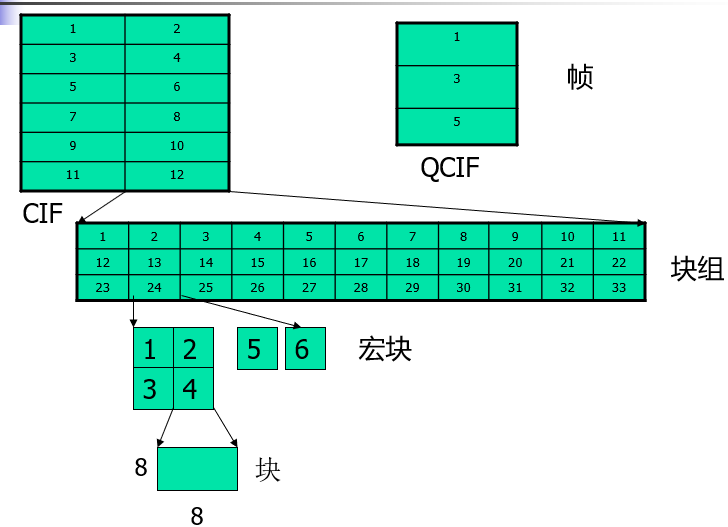
## P×64Kbps标准采用层次块的视频数据结构形式, 使高压缩视频编码算法得以实现。

## P×64Kbps标准的视频编码定义一个视频数据结构CIF保证解码器对接收到的比特流进行没有二义性的正确解码。

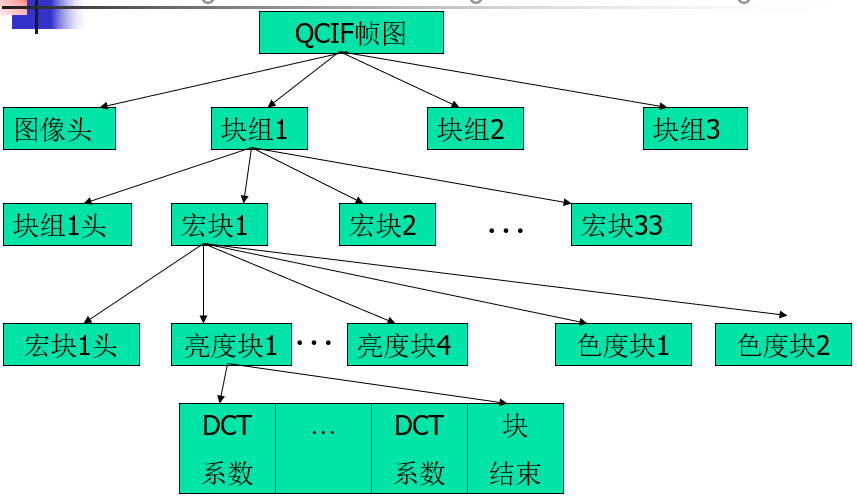
## 利用CIF格式, 可使不同制式的各国电视信号变换为统一的中间格式, 然后输入给编码器, 从而使编码器本身不必意识信号是来自哪种制式的。

## H.261标准适合各种各样实时视觉应用,如位率不同(P不同), 运动效果和图象质量不同,位率提高、画面质量改善。

# 图象数据层次结构



# 视频数据结构图



## 为了适应B-ISDN ATM传输需要, ITU与MPEG联合发布ISO/IEC 13818, 分别称为H.262和MPEG-2标准, 它与H.261和MPEG-1兼容。

## H.263是ITU-T制定的适合于低速视频信号压缩标准。大多数用户最方便的是公用电话线, 以V.34为标准的调制解调器支持在电话线中传输速率可达28.8kbps或33.6kbps, 甚至56kbps.

## 与MPEG-4基于对象编码不同,H.263采用基于帧编码.

## H.263是在H.261基础上扩展形成的, 支持的图象格式包括Sub-QCIF(128×96), QCIF, CIF, 4CIF, 16CIF (1408×1152)等。其中主要采用的改进技术有：

# 

## **(1)半象素精度的运动补偿** 在H.261中,运动矢量的精度为1个象素, H.263运动矢量的估值精度达到半个象素。精度的提高使运动补偿后的帧间误差减少,从而降低了码率。

## **(2)不受限的运动矢量** 当运动跨越图象边界时, 由运动矢量所确定的宏块位置可能有一部分落在边界之外, 此时可以用边界上的象素值表示界外的象素值,从而降低预测误差。

## **(3) 用基于句法的算术编码代替Huffman编码**(可选项) 这是一种效率较高的自适应算术编码。

# 

## **(4) 先进的预测模式(可选项)** 对宏块中的4个亮度块分别进行运动估值获得4个运动矢量。虽然此时传输运动矢量的比特数增加一些,但由于预测误差的大幅度降低,仍然使总码率降低。

## **(5) PB帧模式(可选项)** 虽然使用双向预测B帧可以降低码率, 但却要引入附加的编码延时和解码延时.为降低延时, H.263采用了P帧和B帧作为一个单元来处理的方式, 即将P帧和由该帧与上一个P帧所共同预测的B帧一起进行编码。

# 2.5.4 H264/AVC标准

## ITU-T和ISO/IEC联合开发组共同开发的最新标准，同等质量下压缩效率比提高了2倍以上。

## H.264最大的技术优势体现在4个方面：

## 1）将每个视频帧分离成由像素组成的块，因此视频帧的编码处理的过程可以达到块的级别。

## 2）采用空间冗余方法，对视频帧一些原始块进行空间预测、转换、优化和熵编码。

## 3）对连续帧的不同块采用临时存放的方法，这样只需对连续帧中有改变的部分进行编码。

## 4）采用剩余空间冗余技术,对视频帧里残留块进行编码。对源块和相应预测块不同,再次采用转换、优化和熵编码。

## H.264是保留了以往压缩技术的长处又具有其它压缩技术无法比拟的许多优点。

## 1）低码流：采用H.264技术压缩后的数据量只有MPEG-2的1/8, MPEG-4的1/3。

## 2）高质量的图像：H.264能提供连续、流畅的高质量图像（DVD质量）。

## 3）容错能力强：H.264提供了解决在不稳定网络环境下容易发生的丢包等错误的必要工具。

## 4）网络适应性强：H.264提供了网络适配层 , 使得H.264的文件能容易地在不同网络上传输(例如互联网, CDMA, GPRS, WCDMA, CDMA2000等)。

# 3.4 AVS标准

## AVS标准是《信息技术-先进音视频编码》系列标准的简称，它包括9个部分：系统(第1部分)、视频(第2部分)、音频(第3部分)、数字版权管理(第6部分)技术标准、一致性测试(第4部分)、参考软件(第5部分)、移动视频(第7部分)、系统知识产权IP(第8部分)、文件格式(第9部分)等支撑标准。

## 2006年2月22日，国家标准化管理委员会颁布通知：《信息技术-先进音视频编码》第2部分视频（GB/T 20090.2）于2006年3月1日起开始实施。标准其他部分将继续开展工作, 陆续进入标准报批和审核程序。

## 我国具备自主知识产权的第二代信源编码标准，它是数字音视频产业的共性基础标准。

## AVS标准具备先进性、自主性、开放性。

## AVS是基于我国创新技术和部分公开技术的自主标准，技术方案简洁,芯片实现复杂度低,达到第2代标准最高水平；

## AVS通过简洁的一站式许可政策，解决了AVC专利许可问题死结，制定过程开放、国际化，是开放式制订的国家、国际标准，易于推广；

## AVC仅是一个视频编码标准,而AVS是一套包含系统、视频、音频、媒体版权管理在内的完整标准体系，为数字音视频产业提供更全面的解决方案。

# 2.5 声音压缩标准

# 3.5.1 声音编码

## 声音包括语音和音乐, 是多媒体系统中两类重要数据。

## 声音数据表征是一个一维时变系统,特别对于语音数据,人们已经找到了较合理的声道模型,因此声音数据的压缩要比图象数据的压缩容易。

## 统计表明,语音过程是一个近似的短时平稳随机过程. 短时是指在10～30ms的范围。

## 由于语音信号的这一性质,使得我们有可能将语音信号划分为一帧一帧进行处理,每一帧内的信号近似地满足同一模型—这是本方法假设的基本前提。

## 在实用中,一般一帧的宽度为20ms。

## 语音的基本参数包括

### 基音周期

### 共振峰

### 语音谱

### 声强。

# **语音生成机构的模型**

## 语音生成机构的模型由3部分组成

### 声源

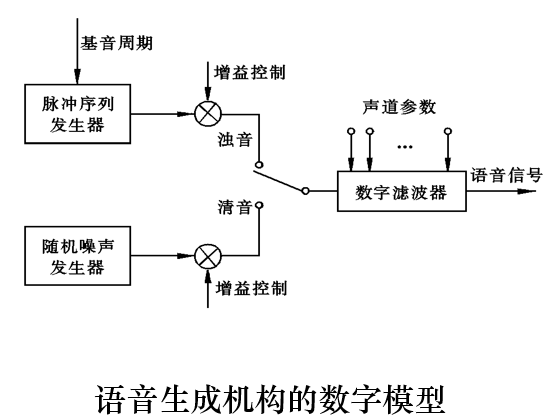
### 共鸣机构

### 放射机构

## 声源共有3类: 元音、摩擦音、爆破音。

## 共鸣机构 也称声道,由鼻腔、口腔与舌头组成。

## 放射机构 由嘴唇和鼻孔组成,是功能是发出声音并传播出去。



# 模型的描述

## 与此语音生成机构模型相对应的声源由基音周期参数描述, 声道由共振峰参数描述, 放射机构则由语音谱和声强描述。

## DPCM,ADPCM等波形预测技术是音乐和实时语音数据压缩技术的主要方法。

## 虽然该方法与基于语音识别的方法和基于参数分析合成的方法相比有压缩能力差的缺点, 但算法简单,容易实现, 以及能够较好地保持原有声音的特点等优点, 因而在语音数据压缩的标准化推荐方案中最先被考虑

## 参数编码的压缩率很大, 但计算量大, 保真度不高, 适合语音信号的编码。

## 混合编码介于波形编码和参数编码之间, 集中了两者优点。

# 2.5.2 ITU语音标准化方案

## 16Kbps ITU语音标准化方案G.728

## 32Kbps ITU语音标准化方案G.721

## 64Kbps ITU语音标准化方案G.722

# 1. 16Kbps语音标准化方案G.728

## 使用领域统一在包括可视电话、数字移动通信、无绳电话、卫星通信、DCME、ISDN等范围内。

## 对于以上所提到的应用范围，约束条件是语音质量在32Kbps ADPCM的同等或以上，且编码延迟时间在5ms以下。

## 1992年,CCITT制定了基于短延时码本激励线性预测编码(LD-CELP)的G.728标准, 这是一种基于AbS原理并考虑了听觉特性的编码方法, 它具有以下特征:

### 以块为单位的后向自适应高次线性预测：

### 后向自适应型增益量化；

### 以向量为单位的激励信号量化。

## G.728用在64Kbps的ISDN线路的可视电话,带宽分配为语音16Kbps,图象48Kbps。语音的多重化传送装置和个人计算机用的编码也是有希望的应用领域。

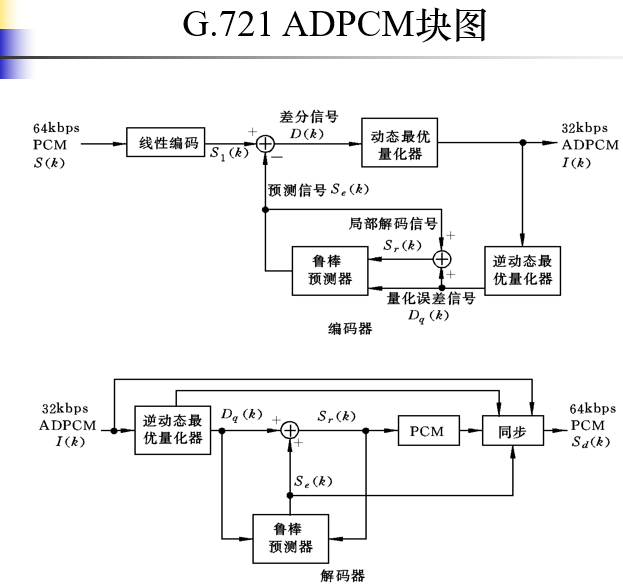
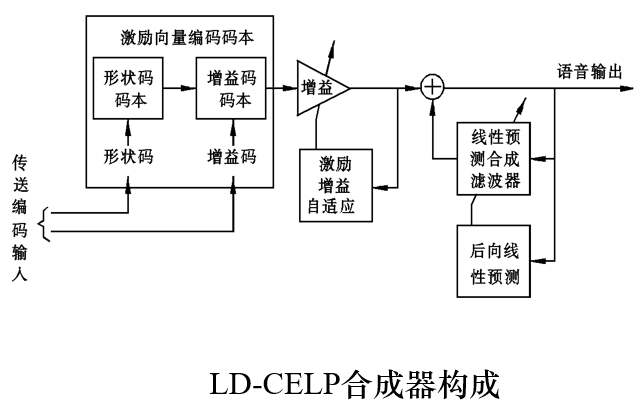
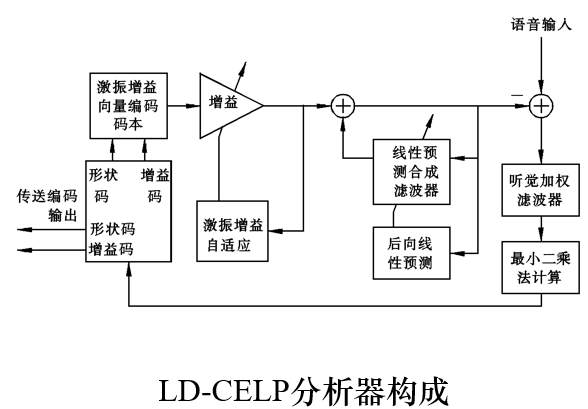
# 2. 32Kbps标准化方案G.721

## 1984年10月CCITT公布了使用ADPCM的标准G.721,速率为32Kbps.目的是最终取代PCM电路传送方式.

## 作为对象的信号包括在电话线中流通的所有的信号, 如语音、个人计算机通信的调制解码信号, 按键电话的信号等等。

## 本方案针对PCM(8KHz采样,8比特样点)规定G.721用PCM的一半速率(8KHz采样,每样点4比特)完成。

## G.721方案采用算法是编码符号延迟为0的且对传送通道的误码率要求不高的ADPCM方式。



## ADPCM方式在算法上的特征

## (1)为了提高预测精度(特别对于性质相差很大的语音信号和调制解调器对的信号),采用了动态对数量化器。

## (2)ADPCM本身采用了按每个采样点进行自适应控制的鲁棒自适应预测器。

## (3)追加了PCM和ADPCM间不论进行多少次转换都不会引起特性降低的同步功能。

## G.721方案的应用 最初是面向卫星通信,长距离通信以及信道价格很高的语音传输。

## 目前的应用领域除了最初的目标外, 还被使用在包括电视会议的语音编码, 为提高线路利用率的多媒体多路复用装置, 数字录音电话的数字记录部件, 以及高质量的语音合成器等等。

# 3. 64Kbps标准化方案G.722

## 1983年开始,CCITT开始了7KHz带宽以语音和音乐为对象的标准化音响编码方案,1988年公布为G.722标准。

## G.722方案编码方法是使用在64Kbps位速率以内工作的SB-ADPCM方法音频编码,它将50Hz到7KHz间频带从4KHz处分割为高频区和低频区,分别利用ADPCM算法编码。

## 算法分为3种基本工作模式,即64Kbps,56Kbps和48Kbps模式。56Kbps和48Kbps两工作模式可分别在总体64Kbps位速率中设置一个8Kbps或14Kbps的数据通道.

## 编码器的构成

## (1) 将输入信号以16KHz的速率采样后,每个样点量化为14比特,然后到进行发送的功能模块。

## (2) 编码速率为64Kbps的SB-ADPCM的编码器。

## 解码器的构成

## (1) 与编码器操作完全相反的(逆)SB-ADPCM解码处理。解码器的速率依工作模式的选择可以在64Kbps, 56Kbps和48Kbps之间变化。

## (2) 生成在16KHz采样速度下与14比特的线性量化精度的数字信号相对应的模拟信号的接收功能模块。

## G.722的应用

## 主要应用对象是电视会议系统, 这是多媒体通信一个子领域。要解决的主要问题是高质量的语音传送。语音通信会议一般涉及3个或3个以上的不同地点, 因此64Kbps音频编码标准必须支持多地点间的会议系统。

# **本章小结**

## 首先介绍媒体数据的表示方法,重点是表示图象数据的彩色空间和文件格式, 并介绍了常用的数据压缩方法。

## 本章重点讨论多媒体系统中几个重要的图象压缩标准, 其中JPEG标准、MPEG标准以及H.26X标准是最重要的得到广泛应用的图象压缩标准。

## 最后简要介绍了多媒体系统的几个重要的声音压缩标准。

# 第四章 多媒体计算机系统组成

## 多媒体存储技术

## 多媒体功能卡

## 多媒体信息获取与显示设备

## 多媒体个人计算机

## CD-I交互式多媒体系统

## DVI多媒体计算机系统

## VCD与DVD播放系统

## 多媒体工作站

# 4.1 多媒体存储技术

# 4.1.1 多媒体信息存储的特点

## 多媒体信息存在和表现的形式

### 正文 包括文字和数据

### 向量图形 图元组成的图形

### 位图图象

### 数字化声音和高保真音响

### 数字化视频

# 4.1.2 光盘存储原理

## 光存储技术

### 光存储技术的产品化形式是由光盘驱动器和光盘片组成的光盘驱动系统。

### 驱动器读写头是用半导体激光器和光路系统组成的光头,记录介质采用磁光材料。

## **光存储技术原理** 改变一个存储单元的性质，使其性质的变化反映出被存储的数据;识别这种性质的变化, 就可以读出存储数据。

## 光存储单元的性质, 例如反射率、反射光极化方向等均可以改变, 它们对应着存储二进制数据0(不变)、1(改变), 光电检测器能够通过检测出光强和光极性的变化来识别信息。

## 高能量激光束可以聚焦成约1微米的光斑, 因此光存储技术比其它存储技术有更高的容量。

## 光盘系统的特点

## 与硬盘相比,具有可拆卸性; 驱动器较贵,但盘片便宜; 读写速度慢。

## 与磁带相比, 具有容量大、随机存取性强的优点。

## 激光头与介质无接触, 不受环境影响而退磁，信息保存时间长, 可达30年以上。

# 

## 光盘系统技术指标

## 容量 光盘盘片的容量

## 平均存取时间 在光盘上找到需要读写信息的位置所需时间

## 数据传输率

## 接口标准及格式规范等

## 存储容量

## 指它所能读写的光盘盘片的容量。

## 光盘容量又分为格式化容量和用户容量，采用不同的格式和不同驱动器, 光盘格式化后容量不同。

## 一般用户容量比格式化容量要少,因为光盘还需要存放有关控制、校验等信息。

## 平均存取时间

## 是在光盘上找到需要读写的信息的位置所需要的时间。

## 指从计算机向光盘驱动器发出命令, 到光盘驱动器可以接受读写命令为止的时间。

## 一般取光头沿半径移动全程1/3长度所需要的时间为平均寻道时间,盘片旋转一周的一半时间为平均等待时间,两者加上读写光头稳定时间就是平均存取时间。

## 数据传输率有多种定义方式。

## 一种是指从光盘驱动器送出的数据率,可以定义为单位时间内光盘的光道上传送的数据比特数, 这与光盘转速、存储密度有关。

## 另一种定义是指控制器与主机间的传输率, 它与接口规范、控制器内的缓冲器大小有关。

# 光盘的分类

## CD-ROM只读光盘

## WORM一次写多次读光盘

## Rewritable可重写光盘

## CD-ROM

## 第一代光盘系统，直径约12cm，因为它容量大，约650MB，价格便宜，市场上颇受用户的欢迎。

## CD-ROM光盘是由母盘压模制成的，一旦复制成形，永久不变，用户只能读出信息。

## CD-ROM工作特点采用激光调制方式记录信息，将信息以凹坑和凸区形式记录在螺旋形光道上。

## WORM一次写多次读光盘

## WORM光盘在使用前首先要进行格式化, 形成格式化信息区和逻辑目录区, 利用激光照射介质, 使介质变异, 利用激光不同的变化, 使其产生一连串排列的“点”, 从而完成写的过程。

## WORM光盘引入文件分配表的概念,在光盘的根目录下面是用户定义的逻辑目录, 逻辑目录对应文件管理区。

## 在逻辑目录建立同时, 用户可以根据需要, 对其中重要数据进行加密。

## 一旦写入就不能再更改。

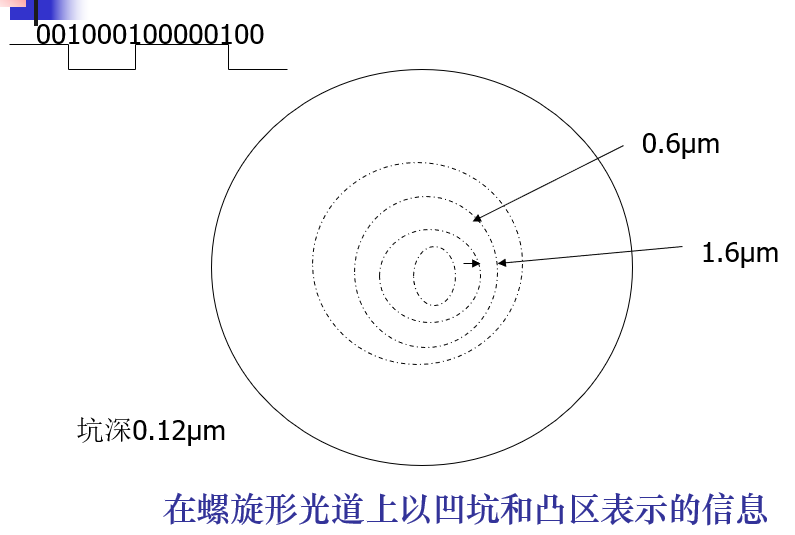
## Rewritable可重写光盘

## 可重写光盘或称可擦写光盘是最理想的光盘类型，也是最有应用前途的光盘类型。它像硬盘一样可读写，利用浮动磁光头在磁光盘上进行磁场调制，可进行高速重写磁光记录。

# 只读光盘读原理

## 只读光盘上的信息是沿着盘面螺旋形状的信息轨道以凹坑和凸区的形式记录的。

## 它既可以记录模拟信息(如Laser Vision系统)，也可以记录数字信号(如CD-DA)。



## 在光盘上记录模拟信息

## 模拟信号先进行频率调制(FM)，声音信号加在经过频率调制的视频信号上，所得到的综合信号经过双向限幅，再转换成光盘上长度不等的凹坑和凸区，边缘之间的长度反映了视频信号频率的高低和声音信号的频率和幅度。

## 在光盘上写/读数字信息

## 光道上凹坑或凸区的长度是0.3微米的整数倍。凹凸交界的正负跳变沿均代表数字“1”，两个边缘之间代表数字“0”,“0”的个数是边缘之间的长度决定的。

## 通过光学探测仪器产生光电检测信号，从而读出“0”、“1”数据。

## 数字信号记录的优点是抗干扰能力强，由于盘片损坏或变脏而造成的读出错误也容易得到纠正。

## EFM编码

## 为了提高读出数据可靠性,减少误读率,存储数据采用EFM(Eight to Fourteen Modulation)编码,即将1字节的8位编码为14位的光轨道位, 并在每14位之间插入3位“合并位” 以确保“1”码间至少有2个“0”码, 但最多有10个“0”码。

# 可重写光盘的擦写原理

## 光盘记录方式可分为两大类：

### 磁光式

### 相变式

# 4.1.3 光盘标准

# 光盘发展历史

## 1972年9月5日Philips公司向国际新闻界展示了长时间播放电视节目的光盘系统，在光盘上记录的是模拟电视信号(Laser Vision)。

## 1978年, SONY生产的影碟机正式投放市场, 光盘的直径为30厘米, 一片双面盘的播放时间可达2小时。

## 1979年, Philips公司发表了激光唱机(Compact Disk Player)。

# 光盘的规范和标准

## CD-DA

## CD-ROM

## CD-V(Video)

## 可录CD

## CD-I

## CD-ROM XA(Extended Architecture)

## Photo-CD

## Video CD

# 

## CD-DA

## 1981年制定红皮书(Red Book)，即CD-DA (Digital Audio)激光数字音频光盘的规范。这个标准是CD的最基本标准。

# 

## CD-ROM

## 1985年制定黄皮书(Yellow Book), 经修订, 1988年正式作为国际标准ISO9660, 1991年又推出了ISO 9660Ⅱ。

# 

## CD-V(Video)

## 从红皮书发展而来, 在影碟机上使用, 视频信息可以输出到电视机。

# 

## 可录CD

## 可录CD(Recordable Compact Disk)盘的橙皮书(Orange Book)标准。可录CD分为两类, 即CD-MO和CD-WO。CD-MO称为磁光盘,可重写；CD-WO又称CD-R, 这种盘一旦用户写入数据就不能抹掉。

# 

## CD-I

## 1987年制定绿皮书(Green Book)规范用于交互式多媒体CD-I系统中。

## 1992年推出第二代CD-I，可播放交互式视频图象。

# 

## CD-ROM XA(Extended Architecture)

## 1988年, Philips、SONY及 Microsoft制定CD-ROM扩展结构, 1991年又制定CD-ROM XAⅡ规范, 对应于ISO9660Ⅱ。

# 

## Photo-CD

## 像片光盘, 1991年Philips和KODAK对外发布Photo-CD, 1992年制定规范。用于存放数字化的静态照片。

# 

## Video CD

## 1993年制定的白皮书(White Book)规范,采用MPEG压缩算法压缩动态图象。它能使Video CD节目能够在CD-I、CD-ROM/XA和Video CD播放机上播放。

# CD-ROM的性能指标

## (1)容量 约为650MB。

## (2)数据传送速率 最初推出为150KB/S, 称为单速, 后又推出倍速(300KB/S),四速(600 KB/S), …, 48X等光驱。

## (3)存储缓冲器 早期为64KB，目前常用的为128KB或256KB。

# 

## (4)存取时间 200～400ms

## (5)误码率 1/1012～1/1016，采用复杂的纠错编码技术降低了误码率。

## (6)体积：光盘驱动器的大小一般为41mm(H)× 146mm(W)×206mm(D)

## (7)接口 采用SCSI接口、IDE接口和SATA总线接口。接口可以集成在音频板、视频板或主板上，也可以是一块单独的板。

# 

## (8)MTBF(Mean Time Between Failures) 平均无故障时间约为25000小时左右。

## (9)兼容性 支持Photo-CD和CD-ROM XA。

# 光盘的规范及格式

## CD-DA规范及格式

## CD-ROM规范及格式

## CD-I光盘的数据格式

## 激光视盘

# CD-DA规范及格式

## CD-DA即激光唱盘, 光盘的物理规格为直径12cm,内径1.5cm,厚度0.12cm, 重量14克。。

## 其螺旋线光道上等长分段, 每段称为一个扇区。每个扇区都存放一定量数据块, 并以一个特定的地址标记,其单位为“分”、“秒”、“扇区”,即1分=60秒,1秒=75扇区。

## 光道总长度为74分, 即可存放74分钟高音质非压缩的音频信号。

# 

## CD-DA每个扇区的音频数据分为许多称为帧的单元, 每帧共有33个字节。

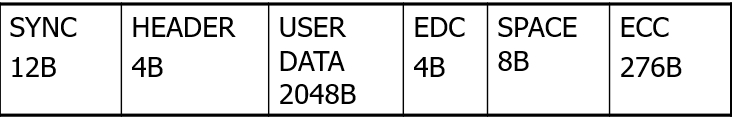
## 一帧中每个通道有6个音频数据，有左右2个通道，每个通道的样本值是16位的数据，共24个字节。一帧中有8个校验字节和1个“控制与显示(C&D)”字节。

## 错误的检测和校正采用的是CIRC(Cross Interleave Reed-Solomon Code)码。

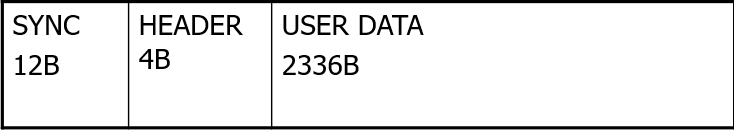
## CD-ROM同样是把光轨道分为等长的扇区,使用分、秒、扇区的数据编址方法, 采用常线速伺服方式。它与CD-DA的不同主要在每个扇区中数据格式的不同。

## CD-ROM光盘有两种格式: Mode1和Mode2。

## Mode1格式



## Mode2格式



## Mode1和Mode2格式相同之处

## 这两种方式的扇区首部都是12字节的同步码(SYNC), 其前后为“00H”而中间10个字节存放“FFH”数据。紧接着的4个字节为地址字段, 或称扇区头(HEADER),它采用分、秒、扇区号的制式确定地址标号, 地址字段中设置了MODE字节, 指明该扇区是哪种格式

# 

## Mode1和Mode2格式不同之处

## 用户数据量不同。Model1为2048个字节, Mode2为2336个字节。

## 存贮数据的类型不同。Mode1用于存放对错误极为敏感的数据, 如计算机程序等; 而Mode2用于存放对错误不太敏感的数据, 如声音、图象、图形等。

## Mode2的数据经过CIRC检验后的误码率为1/109, 对声音、图象类的数据可以不必做进一步校验; 而要满足计算机数据误码率小于1/1012的要求, 则应对Mode1的数据作进一步校验。

# 

## Mode1中的容错机制

## 用了4个字节作为错误检测码(EDC),采用的循环冗余校验码CRC,只能检测是否有错

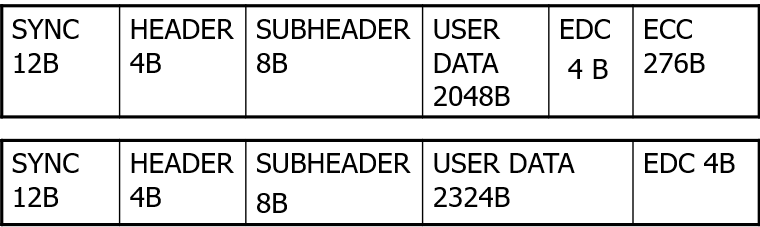
## 用276个字节作为错误校正码(ECC),可以校正扇区中多个字节错误。

## 通过两级校验，Mode1中数据误码率可以降到1/1012。

# 扩展结构CD-ROM XA

## 它所定义格式包括与CD-I格式相同的FORM1和FORM2格式，解决了普通CD-ROM驱动器不能读CD-I格式光盘的问题。

## CD-ROM XA的数据格式(FORM1,FORM2)



## CD-ROM XA在HEADER后面增加了8个字节信息来进一步说明扇区中用户数据,其中存放有数据类型(音频、视频、数据等)格式形式, 触发位(记录开始、文件结束、实时性等), 数据编码信息(ADPCM、CLUT、DYUV等)。

## 这样CD-ROM XA驱动器可通过对子头信息的识别, 读出数据区中多种媒体的信息, 特别地能正确读出CD-I中采用ADPCM自适应差分脉冲编码调制压缩的音频数据。

# 

## CD-ROM卷和文件结构

## 卷和文件结构由逻辑块(512×2n字节)和逻辑扇区(2048×2n字节)、记录、文件、卷、卷集等多级结构定义。

## 由于CD-ROM驱动器的平均寻道时间较长，为了能高速检索CD-ROM光盘上文件，ISO为CD-ROM光盘的文件目录结构规定了路径表

# CD-I光盘的数据格式

## CD-I光盘的数据格式是从CD-DA和CD-ROM光盘格式演变而来的

## 其扇区格式与CD-ROM XA相同, 它有三个区：导入区(Lead-in Area),节目区(Program Area)和导出区(Lead-out Area)。

## CD盘上的信息均采用EFM记录方式进行记录。由于光盘原始误码率较高, 所以都采用能纠突发错误的CIRC码。

# 

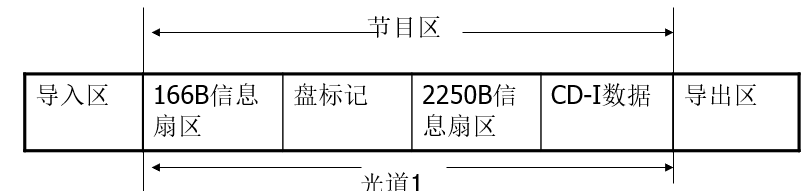
## CD-I光盘的导入区 是由若干个空扇区组成的,这样做的目的是使识别节目区变得容易。

## CD-I光盘可以有多达99条光道, 编号为1～99。

## CD-I光盘的导出区 或者是空扇区(最后一条光道是CD-I光道时)或者是无声的帧(最后一条光道是CD-DA光道)

# CD-I光盘的结构

## 只有CD-I光道的CD-I盘结构



## 有CD-I和CD-DA光道的CD-I盘结构



## 含有CD-DA光道的CD-I光盘, 第一条光道必须是CD-I光道; 且任一条CD-I光盘上的CD-DA光道必须在CD-I光道之后。

## 一片CD-I光盘上的CD-DA光道可以有一条或多条CD-DA光道,但最多不超过98条。而一条光道的长度可以是300个扇区(相当于4秒)和325000个扇区(相当于最长的超级HiFi播放72分钟)之间的数。

## CD-I光盘的导出区或者是空扇区(最后一条光道是CD-I光道时)或者是无声的帧(最后一条光道是CD-DA光道)

## CD-I盘上的所有数据都以文件形式存放，任何一个文件都可以通过盘上的路径表取出。

## 每个文件都有文件描述符记录，存放于文件目录中。文件描述符记录包含有文件名、文件号、文件大小、地址、拥有者、属性、交叉存取因子、读取许可权。

## 文件分为目录文件、实时文件和标准文件。

## CD-I的数据以两种专门的数据格式FORM1和FORM2记录。如光盘标号上的数据是用FORM1记录，因为它有EDC和ECC码，CD-I系统利用它们可以获得误码率小于1/1013数据。

# **激光视盘**

## 激光视盘也是一种只读光盘, 家用激光视盘播放机又称为影碟机, 是独立的视频播放设备, 与音响设备和电视机(监视器)相连就可以播放视盘。

## 计算机可以通过外设接口与视盘播放机相连, 视盘在多媒体应用形式主要是“交互式视盘”,由计算机来控制视盘的播放、视频帧的寻址和显示。

# 

## 视盘与前述的几种光盘原理结构一样

## 视盘的型号 大-12英寸，小-12CM

## 视盘有两种信息记录格式

## 常线速(CLV)

## 常角速(CAV)

# 

## CLV型视盘

## 扇区长度为常数,以紧凑形式存放信息,每盘可以存放60分钟的视频信号。

## 当驱动器从内圈到外圈读盘时,由于内圈和外圈存放的信息量不同,所以转速也不同。在内圈可达到每分钟1800转,在外圈轨道每分钟约600转。

# 

## CAV型视盘

## 以类似于磁盘的方式划分扇区,扇区长度从内圈到外圈逐渐增加,每盘仅可存放30分钟的视频信号。

## 优点是驱动器读盘时,从内圈到外圈,转速一致, 对NTSC制式,CAV视盘转速为每分钟1800转,对PAL制式, CAV视盘转速为每分钟1500转。

## 这种信息存放格式有利于单帧访问、搜索、帧序列的随机访问等功能,适合于多媒体平台

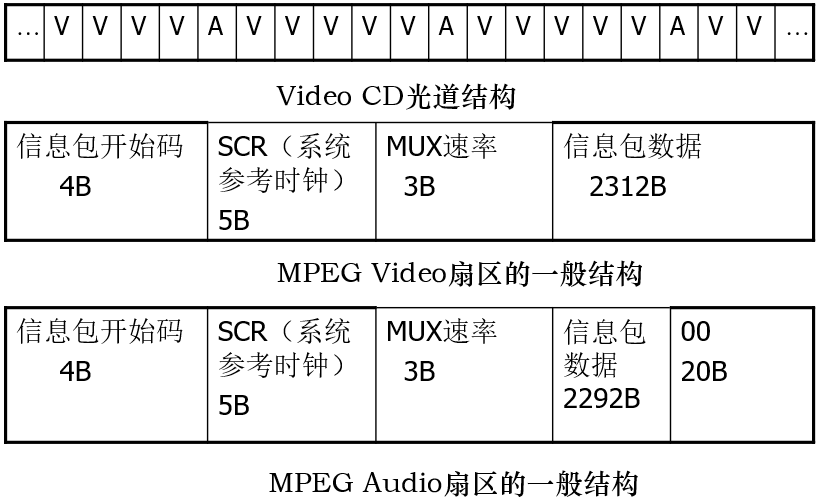
# **Video CD标准**

## Video CD标准是90年代流行的视频光盘标准, 它描述了一个使用CD格式和MPEG-1标准的数字视频系统。

## Video CD定义了MPEG光道的结构, 它由MPEG-Video扇区和MPEG-Audio扇区组成。光道上的Video(电视图象)和Audio(声音)是按MPEG-1的规定进行编码。

## MPEG-Video扇区和MPEG-Audio扇区是交错存放在光道上

# Video CD结构



# CD-ROM光盘制作过程

## (1) 数据准备

## (2) 主盘制作

## (3) 复制光盘

# 4.1.4 DVD光盘技术

# DVD光盘

## MPEG-2的视频质量是广播级质量,为解决其MPEG-2数据存储问题，研制了DVD并于1995年完成标准化方案。

## DVD盘片尺寸与CD相同,容量最高的双层双面盘可达17GB。单面单层DVD盘能够存储4.7GB的数据,存储133分钟的MPEG-2视频，其分辨率与现在电视相同, 并配备Dolby AC-3/MPEG-2音频质量的声音和不同语言的字幕。

## DVD系列标准与CD系列标准对应关系

## **DVD系列** **CD系列**

## Book A：DVD-ROM CD-ROM

## Book B：DVD-Video Video CD

## Book C：DVD-Audio CD-Audio

## Book D：DVD-Recordable CD-R

## Book E：DVD-RAM CD-MO

# DVD光盘

## DVD-Video的规格

## 数据传输率 可变速率,平均速率为4.69Mbps,最大速率10.7Mbps

## 图象压缩标准 MPEG-2标准

## 声音标准

## NTSC: Dolby AC-3或LPCM,可选用MPEG-2Audio

## PAL/SECAM：MPEG MUSICAM 5.1或LPCM,

## 可选用Dolby AC-3

## 通道数 多达8个声音通道和32个字幕通道

# DVD光盘

## 从外观和尺寸上看,DVD盘和CD-ROM盘没什么差别, 直径均为120mm,厚度为1.2mm; 新的DVD播放机能够播放已有的CD激光唱片和VCD。

## 不同点:

## DVD光道之间的间距由原来的1.6μm缩小到0.74μm,而记录信息的最小凹坑凸区长度由原来的0.83μm缩小到0.4μm，这是DVD盘存储容量提高到4.7GB的主要原因。

## DVD信号的调制方式和错误校正方法也做了相应的修正以适合高密度的需要,它采用效率较高的8比特到16比特+(EFM PLUS)调制方式,DVD校验系统采用更可靠的RS-PC(Reed Solomon Product Code)。

## DVD播放机也采用波长更短（由780nm减小至635/650nm）的激光源来提高聚焦激光束的精度。

# 4.2 多媒体功能卡

## 包括视频信号捕捉、压缩、处理、播放的视频卡,音频卡,VGA与TV的转换卡,图形加速卡, SCSI接口卡, 光盘接口卡等。

## 通过这些功能卡将计算机与各种外部设备相连,构成一个制作和播出多媒体系统的工作环境。

## 具有代表性的多媒体功能卡：声音卡、视频卡

# 4.2.1 声音卡

## 声音卡或音频卡(Audio Card)是处理音频信号的计算机插件,它是普通计算机向MPC升级一种重要部件。目前作为微机必备功能集成在主板上.

## MPC所用声音卡由专用DSP芯片管理声音的输入输出和MIDI操作。音频数据是8位或16位的PCM数据或压缩格式ADPCM数据

# 4.2.2 视频卡

## MPC视频处理芯片可分为两类：

## 一是专用固定功能的芯片，这类芯片主要围绕数据压缩标准JPEG，MPEG等开发的；

## 二是可编程的多媒体处理器，如Intel 750系列，TI公司的TMS320系列高效可编程多媒体处理器以及Philips和Sony共同开发的CD-I等产品。

# C-Cube公司的视频卡

## 第一个把JPEG算法集成在一块芯片上。

## CL550 芯片

## CL450 芯片

# CL550简介

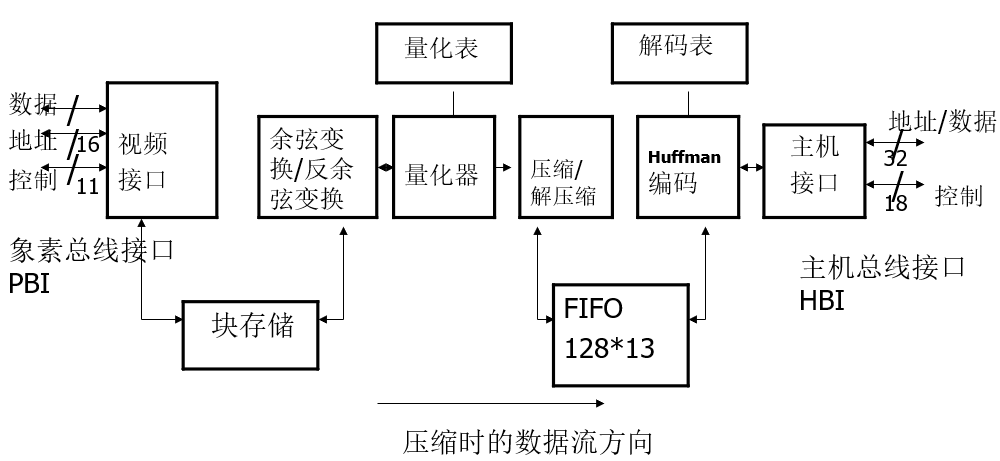
## CL550用了40多万只晶体管集成了JPEG压缩编码所需的DCT/逆向DCT单元、量化器、可变长编码器等单元。

## 压缩率可以通过修改量化表和VLC表的内容来改变。当执行JPEG的有损压缩算法时,可按不同的图象质量、存储器容量、带宽等应用环境来设置不同的压缩比。

## 压缩比可以从8:1到100:1之间任意选择。

## CL550专用芯片上还提供有数字视频接口和直接与系统总线相连的接口, 视频接口支持8位灰度、RGB、CMYK及YUV数字信号的输入和输出。

# CL550框图



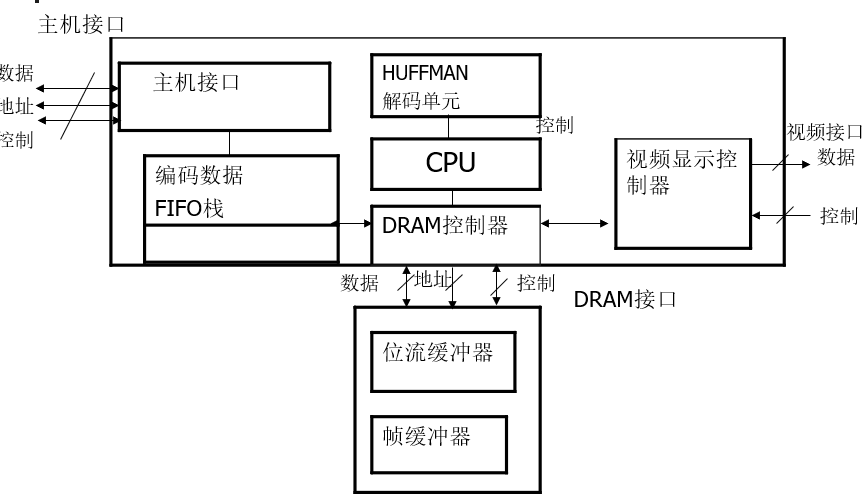
# CL450

## C-Cube公司推出CL450系列产品及CL680解码器。

## CL450包括RISC处理器、Huffman解码器、DRAM控制器、视频显示控制器等单元, 三条总线是主机总线、DRAM总线、象素总线。

## CL450完全遵从MPEG标准, 能实现RGB和YUV格式的相互转换, 支持NTSC和PAL制式, 能完成CIF分辨率(352×240, 30Hz或352×288, 25Hz)的实时解码, 并支持视频和音频的同步, 可全部或部分显示解码后图象

# CL450解码器框图



# 4.3 多媒体信息获取与显示设备

## 多媒体计算机必须配置必要的外部设备来完成多媒体信息获取和显示的功能

## 常见有鼠标、光笔、扫描仪、摄像机、触摸屏、彩色显示器、打印机等设备。

# 4.3.1 图像获取设备

## 数字化图像获取设备常见的有扫描仪(Scanner)、数字照相机(Digital Camera)等静态图像获取设备和摄像机等视频图像获取设备。

# 图象数字化

## 自然景物成象后的图象信息以照片或视频记录介质形成保存,这些图象必须数字化成计算机能处理的数字化信息, 才能被多媒体计算机处理。

## 对于照片和视频图象来说, 不管是从图象信息的空间分布和亮度(颜色)分布都是连续的, 这些连续的信息量必须离散化, 离散化的过程即数字化处理过程, 它应该包括空间位置的离散和亮度电平的离散化。

## 步骤：采样, 量化, 模数变换

## 空间采样

## 一幅图象在二维方向上分成M×N个网格,每个网格用一个亮度值来表示该区域亮度,这样一幅图象就离散化为M×N个亮度值来表示。

## 这个过程称为图象的采样, 其中M×N称为采样的分辨率, 网格的亮度值即为采样值

## 空间采样的分类

## 一维采样用扫描方式(如传真、扫描仪等)把二维图象转化为一维随时间变化的信号。这样得到的一维行扫描信号经采样实现图象数字化。

## 二维采样是固体摄象器件采用的通用方法,它把光电转换和采样功能结合起来。固体摄象器件由M×N个光敏元件构成, 每个光敏元件对应一个采样点, M×N个光敏元件构成M×N个采样点。

# 

## 量化

## 所谓量化就是把连续的亮度值分为K个区间,每个区间上对应着一个亮度I。落于区间i中的任何亮度值都以亮度值Ii表示,共有K个不同亮度值。

## 按照量化区间划分方法, 量化可分为均匀量化和非均匀量化。

# 

## 模数变换

## 实现上述量化过程称为模数变换, 这个过程一般采用PCM量化器来实现,PCM量化是均匀量化。

## 非均匀量化一方面可利用PCM量化的结果,根据信号特性处理为非均匀量化的数据;另一方面也可以利用专门的非均匀量化器来实现。

## 另外要考虑图象数据采样过程中产生的失真和噪声,包括叠加噪声、孔径效应及插入噪声等。

# 图象扫描仪

## 基本原理是将反映图象特征的光信号转换成计算机可接受的电信号 。

## 工作过程 – CCD,光电转换

## 图象扫描仪是最常用的静态图象输入设备

## 它往往配置大量的管理和控制扫描过程软件及文字识别、排版、图文数据库等软件,使扫描仪提供很强图文信息获取能力。

## 另外,扫描仪还提供设置扫描区域、分辨率、亮度、图象深度等参数,扫描后图象还可进一步处理

## 图象扫描仪的种类

## 平板式

## 手持式

## 滚动式

# 摄像机

## 摄象机由摄象镜头管、同步信号发生电路、偏转电路、放大电路、电源等部分组成。

## 来自被摄物体的光通过光学系统在摄象管的靶上形成光学图象，这个光学图象经摄象管转换成电信号，以视频信号方式输出被摄图象。

## 彩色图象摄取重要的是分离出三基色信号，利用滤色片、分色镜或棱镜等把光分解成三基色

## 最新产品不用电子管作光电转换，用电荷耦合器件CCD等固态摄象器件，这种器件具有体积小、重量轻、省电、寿命长、可靠性高等优点。

# 4.3.2 显示设备

## 多媒体计算机系统中信息显示的设备主要有PC机的显示系统。包括两部分:显示器,显示适配器。

## 显示适配器由寄存器、视频存储器和控制电路三部分组成，其中视频存储器包括显示帧缓存RAM和存放BIOS的ROM。显示适配器可以是一块插入PC总线的扩展卡或集成在主板上。

## 显示器与显示适配器相配。

## 从扫描频率角度来分，显示器主要有固定扫描频率与可变扫描频率两种。

# 4.3.3 触摸屏

## 触摸屏最早出现于70年代，90年代随着多媒体应用得到成熟和推广

## 触摸屏应包括三个部分：传感器；控制部件；驱动程序

# 

## 触摸屏的分类（按工作原理）

## 红外线触摸屏

## 电阻式触摸屏

## 电容式触摸屏

## 表面声波技术

## 底座式矢量压力测力触摸屏

## 等。

# 红外线触摸屏

## 工作原理

## 红外触摸屏有内置式和外挂式两种

# 电阻式触摸屏

## 电阻式触摸屏感应器是一块覆盖电阻性栅格的玻璃，再在上面蒙上一层涂有导电涂层并有特殊模压凸缘的聚脂薄膜。

## 凸缘避免其表面的涂层与玻璃的涂层接触。控制器向玻璃的四个角加有稳定的5伏电压,并读取导电层的电压值。

## 当屏幕被触摸时,压力使聚脂薄膜凹陷而碰到玻璃,导电层接触。控制器向玻璃的两个邻角加电压,并把对面两个角接地,于是电阻栅格使玻璃片上形成从矩形的一边到另一边线性变化的电压阶梯,控制器从两个方向测出触摸点的电压值,从而计算出触摸的精确位置.

# 电容式触摸屏

## 这种触摸屏由一个模拟感应器和一个智能双向控制器组成。

## 感应器是块透明的玻璃，表面有导电涂层，其上覆盖一层保护性玻璃外层。

## 它工作时在感应器边缘的电极产生分布的电压场，用**手指或其它导电体触摸**导电涂层时，电容改变，电压场变化，控制器检测这些变化，从而确定触摸的位置。控制器把数字化的位置数据传到主机，以实现人机的交互。

## 电容式触摸屏的感应器安装在监视器内部，外部与普通监视器一样，可靠性较高。

# 表面声波触摸屏

## 表面声波是应变能仅集中在物体表面传播的弹性波。

## 触摸屏在一片玻璃的每个角上装有两个发射器和两个接收器，一系列的声波反射器被嵌进玻璃中，沿着两面从顶至底穿过玻璃。发射器朝一个方向发射5MHz的短脉冲。当脉冲离开一角后，就会不断地被每个反射器反射回来一部分声波。

## 当触摸玻璃的某点就阻碍了脉冲能量通过那点反射到达接收机，于是从接收的脉冲信号中就见到一段缺口。脉冲起点至下跌点间的时间长度就确定了触摸点的坐标。控制器通过互换两对发射器和接收器，就可测出触摸在X及Y方向的坐标

# 矢量压力测力触摸屏

## 这种触摸屏的原理是在CRT外面盖上一块四角装有应力计的平板玻璃。

## 当玻璃受到压力时, 应力计就会出现电压或电阻等电气特性的变化。压力越重,变化值就越大。每个角记录这些变化。

## 控制器读取每个角的记录值, 并计算触压位置。这种触摸屏分辨率较低 。

# 4.4 交互式多媒体系统

# 4.4.1 概述

## Microsoft公司与IBM等数十家软硬件公司于1990年成立MPMC；

## 1990年10月提出了MPC技术规范1.0,1993年MPMC发布了MPC技术规范2.0

## 1995年6月, IBM、COMPAQ、APPLE、DELL及Microsoft等著名公司都宣布支持MPC技术规范3.0, 其技术指标具体为PENTIUM 75MHz, 8M RAM,540M HD,四倍速CD-ROM驱动,MPEG-1硬件或软件回放,波表合成声卡等配置。

# 

## MPC配置特点

## (1) 一个功能强大，速度快的中央处理器(CPU)；

## (2) 大容量的存储器空间；

## (3) 高分辨率显示接口与设备；

## (4) 可处理音频的接口与设备；

## (5) 可处理图象的接口与设备；

## (6) 可存放大量数据的配置等。

## MPC配置（扩充）

## (1) 光盘驱动器

## (2) 音频卡

## (3) 视频卡

## (4) 打印机接口

## (5) 交互控制接口

## (6) 网络接口

## (7) 图形加速卡

# MPC

## Macintosh多媒体计算机

## Amiga多媒体个人计算机

## Windows+Intel

# 交互式多媒体系统结构



# 4.4.2 CD-I交互式多媒体系统

## CD-I系统是家用交互式多媒体系统, 它是Philips公司和Sony公司于1986年4月联合推出的一种电视计算机或称Smart TV系统。

## 该系统把各种多媒体信息存放在容量为650MB的只读光盘上,用户可通过CD-I系统读取光盘的内容来进行演播,光盘的数据使用CD-I格式存放。

## CD-I的正式商品于1991年面市, 用户可以交互式地把家用电视机和计算机相连, 通过鼠标器、操纵杆、遥控器等装置选择人们感兴趣的视听节目进行播放, 是一种较好多媒体系统产品

# CD-I基本系统结构

## CD-I基本系统主要有5部分构成

## (1)音频处理子系统

## (2)视频处理子系统

## (3)多任务的操作系统

## (4)CD播放机

## (5)微处理器、存储器、键盘、定位装置和CSD字体模块

# CD-I音频子系统

## CD-I基本系统有4种标准音质的运行方式和一种非实时的语音音质运行方式。

## CD-I除继承CD-DA超级高保真音质运行方式外，还有A、B、C三个音质等级的运行方式。A级相当于Laser Vision音质, B级相当于FM调频广播的音质, C级相当于AM调幅广播的音质。这4种音质的语音为实时的语音。

# CD-I音频子系统

## 非实时语音音质是文本到语音编码转换而成的音质。CD-I有两种接口用来辅助编码这种音频信息,它们是上层接口和下层接口。上层接口是处理器默认的字符集,下层接口是对8位PCM数据进行实时解码。这两种接口之间转换由微处理器控制。

## 声音数据的解码和控制是由CD-I音频处理器来完成的.

# 

## CD-I音频处理器的组成

## (1) 解码器ADPCM。

## (2) 音频处理单元。

## (3) 特技处理器及声音输出。

## (4) CD-I接口单元。

## (5) 音频信号存储器。

## (6) 控制器。

# CD-I视频子系统

## CD-I视频处理子系统的功能是把CD-I光盘上的数字化视频信号通过存储和控制进行实时解码、颜色切换、重叠控制, 经过混合处理而产生RGB信号输出。

# 视频压缩与解码原理

## (1) 一维的DYUV编码 自然图象

## (2) RGB 5:6:5编码 高质量图形

## (3) CLUT（彩色查找表）编码 动画

## (4) 一维行程编码（RL） 动画

# 4.4.3 DVI多媒体计算机系统

## DVI技术最早是由美国David Sanaoff研究中心研究开发的交互式数字视频装置,这项技术研究成功后被GE公司购买, 后又被Intel买到手。

## Intel和IBM公司联合开发,于1989年在美国计算机博览会(Comdex/Fall’89)推出第一代产品Action Media750

## 1991年又推出第二代产品Action MediaⅡ, 在Comdex一举获得了最佳多媒体产品奖和最佳展示奖。Intel公司已将成功的DVI多媒体系统做到一个母板上,并要把DVI技术集成在一个芯片上。

## DVI多媒体计算机系统特点：

## (1) 提供一种全数字化的方法。

## (2) 先进的视频压缩技术

## (3) 声音压缩技术

## (4) 合成图形

# DVI系统结构及其工作原理

## **DVI系统结构**

## 第一代DVI系统(DVI Ⅰ)由3块插板组成: DVI视频板, DVI音频板及DVI多功能板。

## 1991年推出的第二代DVI系统(DVI Ⅱ)将上述3块板集成在一个板上,视频、音频的获取部分也都装在上面,仅占一个IBM PC标准插槽,为用户提供了方便。

## Intel将系统外围逻辑电路集成为3个门阵列电路即82750H主机接口门阵列, 82750LV VRAM/SCSI/ Capture门阵列, 82750LA音频子系统接口门阵列。

# 

## 其它设备包括1~16MBVRAM视频处理器、音频信号处理器、D/A转换器及模拟滤波器和DVI总线。

## DVIⅠ的核心部件是视频象素处理器82750PA和视频显示处理器82750DA, DVIⅡ将这两个芯片升级为82750PB和82750DB, 使运算速度提高了一倍。

# 

## 以DVIⅡ为代表DVI多媒体硬件系统特点：

## ①采用了高速专用视频处理器i750B, 具有实时处理视频功能；

## ②DVI总线保证了高速传输；

## ③外围逻辑集成到三个门阵列, Action MediaⅡ体积缩小；

## ④外围接口设计方便了用户。

# DVI软件开发环境

## 1989年推出的第一代DVI系统软件是基于DOS环境,采用了层次结构模型,具有模块化特点, 其核心是AVSS(Audio Video Support System)。

## Intel和IBM开发第二代DVI系统中采用基于Windows的DVI系统软件, 其核心为音频/视频内核AVK(Audio/Video Kernel),AVK能在不同的操作系统环境下工作, 而且为了实时响应,能够最少地依赖主机CPU。

# 音频视频子系统AVSS

## 应用层

## 高层次接口DVI标准

## 高层次模块低层次接口DVI标准

## 驱动器接口模块

## 系统RAM接口模块

## 硬件

# DVIⅠ型系统软件层次结构

## 最下层是DVI系统硬件。

## 硬件之上和硬件直接打交道的软件是驱动程序,驱动程序模块包括视频驱动程序、音频驱动程序以及多功能板驱动程序。

## 驱动程序模块层之上是驱动程序接口模块层。DVI系统中共有4个驱动程序接口模块：①微码接口模块;②视频接口模块;③多功能接口模块;④音频接口模块

## 驱动程序接口模块层之上是应用支持层, 它主要包括两个高层次的软件包即一个图形软件包［Gr］,一个音频视频支持软件AVSS。

## 最高层是应用层, 它可以提供大量的应用程序。对其支持的DVI高层接口提供了多媒体编辑制作工具及创作语言, 方便了应用软件的开发。

# 基于窗口系统环境的AVK

## AVK的概念模型：

## 多媒体系统软件的核心是AVK, 其概念模型是“数字视频制作演播器”。

## 数字式制作演播器主要的组成部分是：模拟设备接口,显示管理器,采样器,效果处理器以及音频/视频混合器等 。

## DVIⅡ系统概念模型是“数字视频制作演播器”。这种模型要求多媒体技术模拟现代电视制作演播室,由特定硬件完成各项功能,并使应用开发者具有同实际演播室一样的创作自由度。

## 一个典型的制作演播器应包括混合器,磁带,监视系统,特技处理器以及为了记录、修改和播放视频和音频信息联在一起的其它设备。

## 数字式制作演播器主要组成部分：

## (1) 模拟接口

## (2) 显示系统

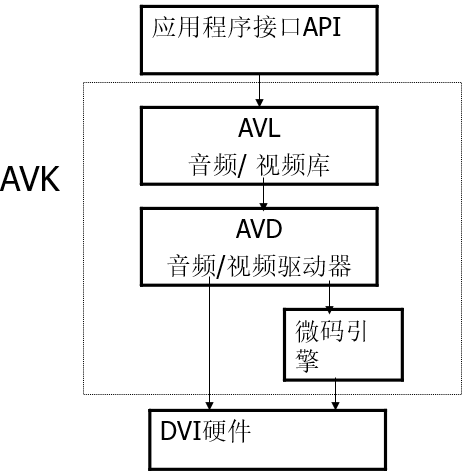
## (3) 采样器

## (4) 数据流控制器

## (5) 效果处理器

## (6) 混合器

# AVK系统的结构



# DVI图象格式及压缩算法

## 静态图象处理：静态图象压缩的算法有JPEG和行程算法。

## 视频图象压缩：DVI技术采用两种视频图象压缩算法，即PLV和RTV算法。

# 4.4.4 VCD与DVD播放系统

## VCD播放系统

## DVD播放系统

# VCD播放系统

## VCD播放机是基于MPEG-1标准的交互视频播放系统，它有两种形式：

## 一种是使用PC机构成的播放系统，它是在PC机加上MPEG解压卡或解压软件升级而成；

## 另一种是VCD播放机加上电视机构成。

# VCD播放机的基本结构

## VCD播放机由3个核心部件组成：

### (1)   CD驱动器，或称CD加载器。

### (2)   MPEG解码器。

### (3)   微控制器。

# VCD播放机的基本功能

## (1) 支持VCD2.0标准的播放控制功能。

## (2) 可把NTSC制电视转换成PAL制电视

## (3)播放不太清洁或者缺陷不大的VCD盘时不会产生断续图象，C-Cube称之为ClearView技术。

## (4) 支持单盘和多盘加载器。

## (5)支持下列CD盘格式：VCD2.0,VCD1.1,CD-DA,卡拉OK-CD1.0,CD-I。

## (6)支持的播放特性有：1/2,1/4,1/8和1/16的播放速度；快速向前播放；按时间搜索。

## (7) 卡拉OK功能。

# DVD播放系统

## 与VCD播放系统相差不大

## 主要有下列部分组成

### (1) DVD盘读出机构

### (2) DVD-DSP。

### (3) 数字声音/视频解码器MPEG-2

### (4) 微处理器。

# **本章小结**

## 首先介绍了多媒体光盘存储技术，叙述光盘的读写原理、类型、规格、格式及制作方法。

## 多媒体功能卡是构成多媒体系统的关键部件, 本章以典型的产品为例讨论视频卡、音频卡和专用多媒体功能芯片的原理和功能。

## 多媒体的输入输出设备是系统的重要配置, 本章也简要介绍了扫描仪、显示器、触摸屏等设备的工作原理。

## 交互式多媒体计算机系统,本章介绍了它们的代表性系统或产品。

# 第五章 多媒体软件开发环境

# 5.1 多媒体数据获取

## 多媒体应用软件的开发过程

## 明确使用对象, 了解用户需求;

## 选择开发方法;

## 准备多媒体数据;

## 完成系统集成。

## 多媒体节目特点是以内容情节为导向的, 而其它的软件如文本处理软件、数据库管理系统等一般是工具性的软件。

## 用户需求往往是针对多媒体技术从内容和设备配置方面具体要求, 如

## 用户是否有不使用鼠标和键盘而直接通过触摸屏幕来获取信息的要求；

## 系统中是否需要语音和音乐；

## 数据类型中有无图象、视频、动画、字幕的要求；

## 等等。

# 

## 有两种方法可供选择：

## (1)由开发人员编码来实现多媒体节目；

## (2)利用多媒体开发工具来制作多媒体节目。

## 前者的优点是不需较大的投资, 但需编制大量的程序，维护也不方便; 后者需要一定的投资，但开发周期短，维护问题少，关键是要选择一种功能较强价格合理的工具软件。

# 

## 当开发方法确定后,就进入了具体实施阶段。实施阶段的基本工作是多媒体数据的准备。

## 多媒体节目里包括音频、视频、动画、静态图象、文字、图形等多种媒体素材, 这些素材在系统集成之前必须准备好。

# 

## 制作者通过所选择的开发方法将节目情节具体化程序化, 并将准备好的多媒体素材按照需要进行编辑加工, 最终集成为一个由程序和数据组成的软件产品。

## 开发的节目往往又记录在某种介质(如CD-ROM)中，便于销售和使用。

# 多媒体数据获取方法

## 多媒体数据的准备又包括数据的获取、数据的整理、数据的编辑加工等阶段。

## **数据的整理**主要是对采集的媒体数据按照指定的方法进行登记与分类，便于后面步骤的使用。

## **数据的编辑加工**指的是根据情节的要求对媒体数据进行剪辑、修改、格式转换等处理，一般通过专门的软件工具来完成.

# 图象数据获取方法

## 购买数字化的图象或图片；

## 自己动手使用特定软件创作电子图象；

## 用扫描仪将照片、图片做数字化处理；

## 用摄像机或帧捕捉器捕捉视频画面并进行数字化处理；

## 向有关单位无偿或有偿交换拷贝或通过网络获取公开的图象文件。

# 音频数据获取方法

## 最常用的方法是利用录制设备录制音源,然后数字化处理并存入计算机。

## 另一种方法就是购买或和有关部门交换音频数据文件,如大型图书馆、电台等部门存放有珍贵的名人讲话原始录音,制作者可通过合适途径获取。

# 文本和数据文件的获取方法

## 人工输入。这是最自然、最易掌握的普遍采用的方法。英文输入方法比较简单，中文输入方法较为复杂，流行的有几十种方法。

## OCR扫描识别。通过扫描原稿输入文字数据图表技术已经成熟，市面上有多种产品问世。目前印刷体文本的识别率可达99%以上，手写体文本识别率一般也能达到75%以上。

## 购买和交换文本

# 图形和动画的获取方法

## 图形和动画的视觉效果是由人类创作设计的虚幻的或仿真的画面所表现的，它表达了人类对空间和物体的主观认识。

## 图形和动画的获取以用户制作为主要途径，有时也可通过购买或交换得到 。

# 音频编辑软件

## 若要在多媒体节目中加入完整的音频,则必须有赖于编辑声音的软件及PC上加装音频卡。

## 音频编辑软件一般包括下列各种工具或功能：

### (1)菜单条

### (2)控制板

### (3)显示板

### (4)剪辑板

# 数字音频编辑工具

## Microsoft提供的“录音机”, 使用它可以录制、播放、编辑Microsoft的波形数字音频

## 两种。

# 图象处理软件

## Photoshop可完成简单的亮度和色彩校正以及复杂的图象编辑,例如图象扫描、校色、滤色、润色、混色、特殊效果及颜色分离;

## 支持大量的流行的图象文件格式, 如TIFF、TGA、PCX、GIF、BMP等格式, 并很容易文件格式转换；

## 支持RGB、CMYK、HLS多种彩色模式及灰度图象。

# PhotoShop的功能

## (1) 扫描图象

## (2) 选择和蒙片

## (3) 图象编辑、转换和特殊效果

## (4) 图象合成

## (5) 图象校正和加强

## (6) 绘图和润色工具

## (7) 图象处理过滤器

## (8) 颜色分离打印

# 5.2 图形和动画的制作

## 图形数据

## 计算机图形学是研究怎样用计算机生成、处理和显示图形的一门学科。它着重讨论怎样将数据和几何模型变成可视的图象; 或对自然界已存在对象通过获取相关数据建立几何模型生成的图象, 以便于分析处理。

## 图形的显示形式也称为图象, 但和一般意义的图象处理技术不同, 后者侧重于将客观世界中原来存在的物体映象处理成新的数字化图象, 关心问题是如何滤掉噪声、压缩数据、提取特征、三维重建等内容。

## 在多媒体数据中,无法从客观世界直接摄取的可视信息,就可用图形技术来制作,这些数据主要包括文字、图形、动画。

## 图形包括二维空间及三维空间图形, 其中二维图形仅能表现图形中各部分简单的位置关系

## 三维图形经真实感处理, 将使图形能表达出空间、位置、材质、明暗等接近自然的真实感效果。

## 动画是图形对象赋予运动属性后制作的连续画面效果,需要专门的软件工具制作。

# 图形数据

## 图形文件的格式通过图形原语和它们的属性来描述。图形本身决定了哪些原语被支持,诸如线型、线宽、颜色之类的属性影响着图形画面的输出。

## 这种表示需要在图形显示时被转换成低级图象的点阵。高级原语的优点是减少了存储每幅图象需要的数据,容易操纵图形图象,缺点是需要更多的步骤来将图形原语及其属性转换成它的图象象素。

## 对图形软件开发产生广泛影响的标准有PHIGS，GKS和OPEN GL等 。

# 计算机动画

## 计算机动画(Computer Animation)是用计算机生成一系列可供实时演播的连续画面的技术,它可把人们的视觉引向一些客观不存在或做不到东西,并从中得到享受。

## 计算机动画是使用计算机作为工具来产生动画的技术,计算机在动画制作过程中起着大量的不同的作用,表现在画面创建、着色、录制、特技剪辑、后期制作等各个环节。

# 历史与发展

## 1831年, 法国人Joseph Antoine Plateau利用视觉滞留原理发明了一种称之为“诡盘”的机器创造了运动画面的幻觉。第一部动画片是1906年由美国人J. Steward Blackton 制作的, 名字叫“Humorous Phases of Funny Face” 。

## 1909年美国制作第一部卡通片; 1915年引入Cel动画技术; 1928-1938,Disney时代,米老鼠、唐老鸭风靡一时;

## 1963年,BELL实验室制作第一部动画片;

## 1960年代, 2维计算机动画发展时期。

## 1970年代, 动画片屡屡获奖; 三维动画系统是研究重心。

## 1980年代后期, 随着以SGI为代表的高性能图形工作站出现以及图形学理论的飞速发展，推出了一批可生成具有高度真实感的实用化、商品化的三维动画系统。

## 1990年代后,计算机动画在影视制作方面取得辉煌成就。

# 

## 我国的计算机动画的研究与应用虽然起步较晚，举办1990年亚运会时,开始有所起步;

## 90年代发展很快,一些大学和科研 单位相继开展计算机动画的研究工作,缩短了与国际上的差距，有些领域的工作在国际上产生一定的影响。

## 在应用方面,以电视片头、卡通片、动画广告制作为重点,在社会上产生了动画热。

## 目前我国的动画研究水平逐步缩短了与国际差距,越来越多的学者活跃于国际学术界.

# 计算机动画研究内容

## 运动控制方法

## 动画描述模型与动画语言

## 中间画面的生成技术

## 三维动画中的物体造型技术

## 动画绘制技术

# 运动控制方法

## 计算机动画中用于控制动画物体随时间而运动或变化的运动控制模型主要有

#### 运动学方法；(起点,终点,时间)

#### 物理推导方法；(受力分析)

#### 随机方法；(分形技术,粒子系统)

#### 行为规则方法；(角色分类,行为规则)

#### 自动运动控制方法等。(人工智能,机器人)

# 运动控制方法

## 粒子系统动画

## 分形动画

# 动画描述模型与动画语言

## 用户和动画制作系统的交互方式是评价动画制作系统的重要因素之一,这种交互方式的抽象层次和自然语言化程度主要依赖于动画描述模型的影响。

## 对动画描述较有影响的描述模型有面向对象方法,角色理论,记号系统,时间轴描述,基于时序算子的描述,基于知识的描述等.

## 基于动画描述模型开发的动画描述语言主要有3类: 记号语言(动作编码);通用语言(扩展的流行语言);图形语言(专门设计)

# 中间画面的生成技术

## 动画的中间画面的生成主要有3种途径：

### 关键帧方法；

### 算法生成；

### 基于物理的动画生成。

# 三维动画中的物体造型技术

## 动画中物体表示可分为3个层次

## 线框 物体由一系列线框表示。

## 表面 物体由一系列面素表示。

## 体 物体看作一系列体素组成或看作3D空间包围部分。

## 曲面造型 代数曲面;拟合曲面(Bezier, B样条, NURBS曲面)

## 实体造型 边界表示,CSG,推移,八叉树,单元分解

## 人体造型 骨架-肌肉-皮肤

## 人体动画 造型和运动控制十分困难,计算机动画技术的最大挑战

# 动画绘制技术

## 真实感图形绘制技术是计算机图形学研究的一个重要内容,人们已经提出了许多光照模型和绘制算法, 其中有代表性的常用的光照模型有Phong模型, Cook-Torrance模型, Whitted模型等。

## 绘制技术有扫描线算法,Phong明暗处理算法, 光线跟踪技术,辐射度技术等。动画绘制除了使用上述方法外, 还可根据其目标是生成一系列连续画面图象的特点, 利用相关性来加速绘制过程。

# 计算机动画的应用

## **面向影视制作的应用**不强调画面的真实性,只追求观赏性和趣味性,其中角色的运动可以有些虚幻,但绘制技术要求较高,能模拟出各种真实感效果。

## **面向模拟的应用**着眼于各种真实问题的仿真研究,它追求数据的正确性和结果的可信性以及能使各种以前仅能得到大批数据的科学试验可视化,这类动画对绘制效果没有前者要求的高。

# 计算机动画的应用

## (1) 影视制作

## (2) 广告制作

## (3) 教育领域中的辅助教学

## (4) 科研领域

## (5) 工业领域

## (6) 视觉模拟

## (7) 娱乐工业

# 三维动画制作软件3DS max

## 3DS max是美国Autodesk公司所推出的在PC机上运行的实体设计及动画制作软件, 提供给专业绘图人员建立高质量图象或制作动画所需功能。

## 3DS max提供用户组合各种色彩的设置、各种透明度的控制、表面纹理以及各种反射的特性, 以帮助产生任何可以想象出来的材质。

## 3DS max由5个模块组成: 二维造型, 三维放样, 三维编辑, 材质编辑, 关键帧编辑。

# 动画制作过程

## 用二维造型(2D Shaper)绘制各种平面几何图形

## 在三维放样(3D Lofter)中将平面几何沿着给定路径放样成三维立体形体, 并转入三维编辑(3D Editor)

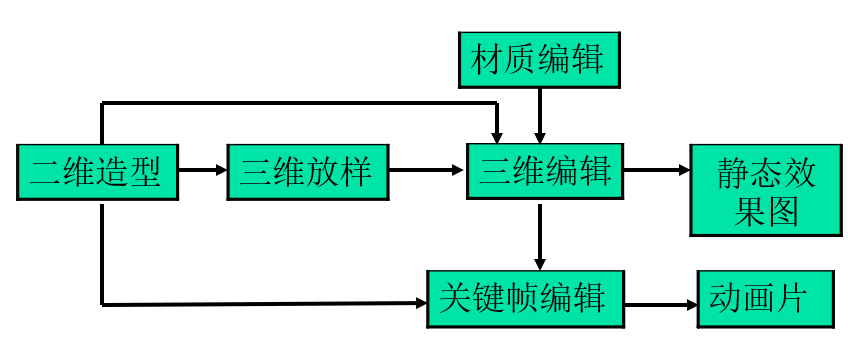
## 再在三维编辑器中, 对物体所在的场景进行各种设置和调整, 包括设置场景中的光源和摄象机。

## 材质编辑(Material Editor)的作用是制作物体所需的各种表面材质和纹理质感。

## 在关键帧编辑(Keyframer)中, 可以对三维编辑器中制作的三维场景设置关键帧, 以定义各种物体的运动轨迹。

## 真正的动画效果还需要绘制过程才能体现

# 3DS max各模块之间关系



## 动画创意

## 动画是科学与艺术相结合的产物,又称之为计算机艺术,好的动画作品,需要完美的艺术构思和创造性设计,这就是动画的创意。

## 动画创意至少需要下列3方面知识：

### 设计者必须具备一定的艺术修养和创造能力；

### 设计者必须知道计算机能否实现它,实现的复杂程度如何；

### 设计者必须具有大量的计算机动画制作经验。

# 

## 二维造型

## 创作各种二维的线条和图形, 这种平面造型可由三维放样模块沿着一条路径扩展长成三维空间的对象,然后这些对象可由三维编辑器编辑。

## 二维造型可用来设计某个对象运动的路径,交关键帧编辑器从而形成立体对象的动画。

## 在进行三维放样前,需保证二维图样是闭合且无重叠交叉,其合法性可用Shape/check来检查。

# 

## 三维放样

## 工作步骤如下：

### ①读入二维造型所建立的造型；

### ②赋予读入造型一条在二维空间里的放样路径；

### ③将二维造型沿着路径增加厚度放样成一个三维的对象。

### 所产生的对象会自动由三维编辑器处理成三维的网状对象。造形和路径共同组成一个所谓的模型。在三维编辑里用户可以直接编辑这个对象,在对象表面上赋予材质,或指定光源, 将对象着色。

## 三维编辑

## 用户使用这个模块创造、排列、着色一个三维空间的场景。也可以建立及调整灯光作特殊效果, 并使用材质编辑创造的材质以增加真实感。

## 三维编辑和三维放样模块都提供6种观察方式(Front, Back, Top, Bottom, Left, Right)来观察正在建立的三维对象; 此外, 它还提供了摄象机观察方式, 用户可以通过此方式来进行对象材质的赋予及着色等动作。

# 

## 关键帧编辑

## 使用关键帧编辑器来制作动画的过程为：

## 先定义整个动画是包含那几个关键帧;

## 然后指定这些帧中每个元素位置与状态,该编辑器会计算出画面之间移动的过程以及状态的变化，以自动产生关键帧间的画面,使动画过程平滑和缓。

## 材质编辑器

## 主要给用户提供不同方式的贴图功能,所谓贴图是指可以赋予对象表面以图象效果。

## 用户可以建立以下种类贴图：

### ①纹理贴图，其效果就象将图象直接画到对象表面一样。

### ②透明度贴图，用来设置对象表面的透明程度，贴面里颜色越深的地方透明程度越高，越浅则越不透明。

### ③反射贴图,用来制造出图象从对象表面反射出来的效果。这些可通过材质编辑器调整物体材质的属性参数来实现。

# 5.3 多媒体编著工具

## 多媒体创作工具(或称编著工具)是一种高级的软件程序或命令集合。

## 创作工具旨在提供给设计者一个自动产生多媒体节目的综合环境, 使设计者可将不同的内容与各种功能结合在一起,形成一个结构完整的节目。

## 多媒体创作工具通常应包括制作、编辑、输入输出各种形式的数据,以及将各种数据组合成为一个连续性序列的基本工作环境 。

# 

## 多媒体创作工具分类

## (1) 以卡或页为基础的创作工具

## (2) 以图符为基础,基于事件的创作工具

## (3) 以时间为基础的创作工具

## (4) 以传统程序语言为基础的创作工具

## (5) 其它专用的创作工具

# 5.3.2 以卡或页为基础的多媒体创作工具

## 大多数以卡或页为基础的创作工具提供一种可以将对象连接于卡或页的环境。

## 一页或一张卡便是数据结构中的一个节点，它类似于书的一页或数据袋里一张卡片。

## 这种页或卡片上的数据比书上的一页或数据袋里一张卡片的数据更多样化。

## 在卡或页上的图符很容易理解和使用

## 以Asymetrix公司开发的ToolBook为例介绍这类软件。

# ToolBook简介

## ToolBook是一个面向对象开发环境

## 提供一种面向对象的程序设计语言OPENSCRIPT。

## 从使用观点来看,ToolBook分两个层次：读者和作者。在读者层次用户可以执行该书,阅览其内容。在作者层次上,设计者可以使用命令来编写新的书。

## ToolBook采用Windows用户接口。也采用下拉菜单、图符驱动、放弃和剪贴板功能、综合帮助功能,设置操作约定。

## ToolBook支持大量流行的文件格式,方便数据交换。

# 

## **设计编程制作一体化环境**

## ToolBook具有把图形、文字、数字视频图象、声音及动画集成为一个交互式节目的能力。

## 它提供了高级脚本语言OPENSCRIPT,配置了许多命令去播放各种类型的媒体,管理各种数据以便改变对象的性质。

## 擅长于制作把其它Windows应用软件集成在一起的多媒体节目,还可在媒体单元之间建立链接关系。

## ToolBook的书形隐喻符很容易使人理解,因此它缩短了用户开发节目之前学习创作工具的时间。

# 节目设计思想

## ToolBook按书的结构组织应用程序。

## ToolBook电子书的每屏被描述为一页,每页内可有多级的对象,它们分为背景和前景,其中背景的设置是满足用户要将生成的一系列页共享一些通用元素的要求,如一幅图象或象NEXT、QUIT这样的命令按钮。

## 开发电子书的过程是：

## 在屏幕上画出各种各样的对象, 然后生成潜在的“脚本”, 它在一给定对象以某种方式被选中或触发时, 引发一个或多个结果。这些脚本事实上是用OPENSCRIPT语言写的小段程序 。

# 

## 脚本特点综述如下:

### (1)脚本是一系列OPENSCRIPT语句或指令,它们告诉对象要做些什么。

### (2)脚本可分为一些处理单元，它们描述特定文件出现时,如读者触发按钮或按某个键,将会发生的事件。

### (3)脚本可以控制对象也可以控制信息。

### (4)页面上对象如字段、按钮以及图形的脚本往往对该页面或者同一本书中的一个页面发生影响。

### (5)书、页面以及背景和页面上的对象一样也可以有脚本。

# 5.3.3 基于图符和事件的创作工具

## 基于图符的创作工具提供可视化的程序设计环境。

## 设计之初须先用其他软件来制作各种元素;

## 然后在此系统中建立一个流程图,在流程图当中可以包括起始事件、分支、处理及结束等各种图符;

## 设计者可依流程图将适当的对象从所谓的图符库按下拉至工作区内。这些图符可以包括菜单条的选项、图形、图象、声音及运算等;

## 这个流程图也是事先安排的次序, 同时也表示整个节目的逻辑蓝图;

## 这类创作工具最典型是Authorware。

# Authorware功能特点

## Authorware是一个交互式多媒体节目创作工具, 它使用图符设计流程图,无需编程,非常方便使用。

## Authorware可以用流程图来当作导航图,设计者只要将图符用鼠标按下拉至流程图的某个位置上, 便可以使每一个环节相互连接。

## 变量可以互相传输参数。Authorware提供了200个以上的系统变量及功能来决定属性、数据抓取、对象处理及显示等工作, 甚至控制作业流程的分支, 跳画面及循环等效果。

## Authorware 最大特点是使用15个图符组成的界面 。

# 5.3.4 以时间为基础的创作工具

## 常见的一种多媒体编辑系统,常用于制作电影与卡通片的节目。

## 大多是以时间轴来决定事件的顺序与对象显示上演的时段。

## 这种时间关系可以许多频道形式出现,以便安排多种对象同时呈现。

## 这类系统中都会有一个控制播出的控制面板,它很象录音机、录放像机的控制板, 含有倒带、倒退、停止、演出及快进等按钮。

## 如Action!

# Action!

## 由Macro Media公司所发行,可在Windows与Macintosh下执行的多媒体编辑创作工具。

## 结合了动作、声音、文字、图形、动画多媒体显示环境,使用时间轴来组织其元素。

## 使用时间轴及控制面板来组织一个场景。

## 有一个内容表可以显示出一节目的全部场景,以及每一个场景当中的全部对象。

## 另有一个场景排序器可显示出节目中某一场景的全貌、场景各称及其连接的模板,也显示出每一个场景最后的状态及场景之间的声响等。

# 5.3.5 传统程序语言为基础的创作工具

## 精通编程的程序员对于多媒体编辑创作系统的限制及依赖工具箱产生对象的方式较不容易接受

## 因此,一方面保留传统语言的特性,另一方面改进其程序设计环境成为可视化的操作系统。

## 这样程序员既可以用传统的语言来编写程序又可方便地使用媒体开发工具箱，使这些工具箱内的编码可以直接被采用成为重用的编码 。

## Visual BASIC

## Visual C++

# Visual BASIC

## Microsoft Windows环境程序语言。

## VB提供各式的图形界面。

## VB是基于事件的语言,程序的行为附着于对象,等到对象被调用或被用户引发时才被执行。

## VB提供给鼠标与键盘双重的输入管道。同时也可摄取剪辑板,动态数据交换及对象连接与嵌入等设备,并通过MCI使音响、影片、动画等均可融入其中。

## 它还可将数据文件引进来使用.在完成一个多媒体产品后,可以将它制作成为一个可以直接执行的EXE文件而成为单独的一个应用程序。

# Visual C++

## Visual C++是Microsoft所推出多媒体程序设计软件, 它与Visual BASIC很相似 。

## Visual C++的工具包括有Visual Workbench、AppStudio、AppWizard、ClassWizard等模块。

## 设计VC程序的方法是先利用Visual Workbench及AppStudio来产生或编辑新的资源, 接着利用ClassWizard来产生类, 最后将这些资源在AppWizard中组织起来通过Build来完成构造一套新的应用节目或多媒体的节目。

# 5.4 Windows多媒体开发环境

# 5.4.1 媒体控制接口(MCI)

## MCI在控制音频、视频等设备方面,提供了与设备无关的API接口。

## 用户应用程序可使用MCI控制标准多媒体设备

## 不同设备其驱动控制方式不同:

## 一些MCI设备驱动程序(影碟机)直接控制目标设备;

## 一些MCI设备驱动程序(MIDI函数)可使用MMSYSTEM函数间接控制目标设备;

## 还有一些MCI设备驱动程序(影片演播器)则提供了与其它Windows DLL的高层接口。

## 应用程序通过设备的类型来区分设备 。

## 如果要通过MCI去控制设备,必须将相应的MCI驱动程序和设备的驱动程序,DLL(如果需要)装入。

## MCI驱动程序的安装可通过Windows中的控制面板来完成。在Windows中SYSTEM.INI文件中的〔mci〕部分包括了一个已安装了的设备类型表 。

# MCI接口分类

## Windows采用两种MCI接口:

## 一是使用命令消息接口函数,直接控制MCI设备;

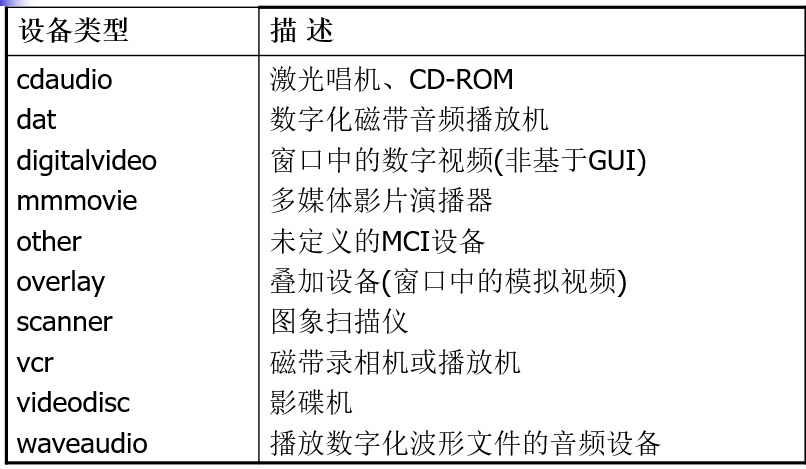
## 二是使用命令字符串接口函数,基于文本接口或命令脚本来控制MCI设备。

## 不同之处在于它们基本命令结构及其发送消息到设备的原理不同。

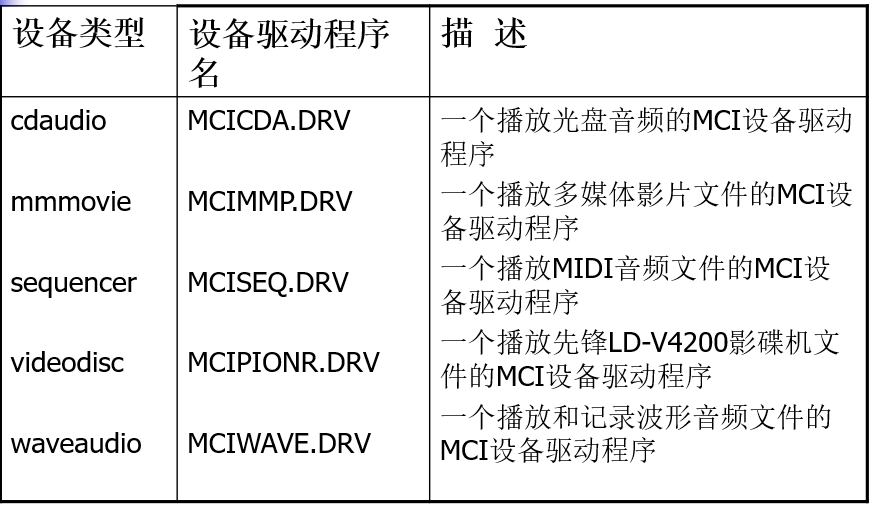
## 命令消息接口使用消息控制MCI设备；

## 命令字符串接口使用文本命令控制MCI设备。

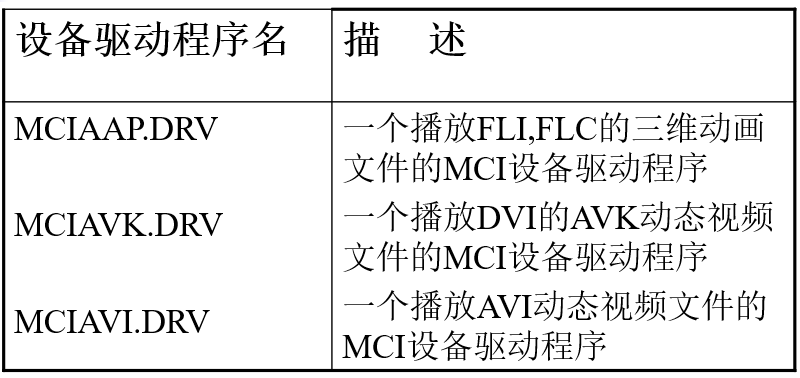
# MCI设备类型



# Windows提供的MCI设备驱动程序



# Microsoft其它MCI设备驱动程序



# 5.4.2 命令消息接口

## 使用命令消息接口发送MCI命令3个函数:

### MciSendCommand 发送一个命令消息到一个MCI设备;

### MciGetDeviceID 当打开一个设备时,返回这个设备的ID号;

### MciGetErrorString 返回对应于一个错误代码的字符串。

# 

## 发送命令消息 MciSendCommand函数定义:

## DWORD MciSendCommand(WORD DeviceID,WORD Message,DWORD Param1,DWORD Param2)

## 其中: DeviceID标识一个MCI设备;

## Message 标识要发出的消息,如MCI-OPEN等;

## Param1 为消息指定标志;

## Param2 为指定一个指向消息数据结构的指针。

## 该函数调用如果成功返回0,否则返回一个错误代码, MciGetErrorString 可获得对这个错误的文本描述。

## MCI命令消息分类:

## 直接由MCI解释的命令;

## 由所有的MCI设备所支持的命令;

## 基本命令;

## 扩展命令。

## 对于不同的设备类型, MCI使用一组不同的扩展命令控制此类设备特殊性能:

## 

## 第一组是MCI元素文件操作扩展命令组, 包括MCI\_COPY,MCI\_CUT,MCI\_DELETE, MCI\_PASTE,一般具有编辑MCI数据能力的设备支持;

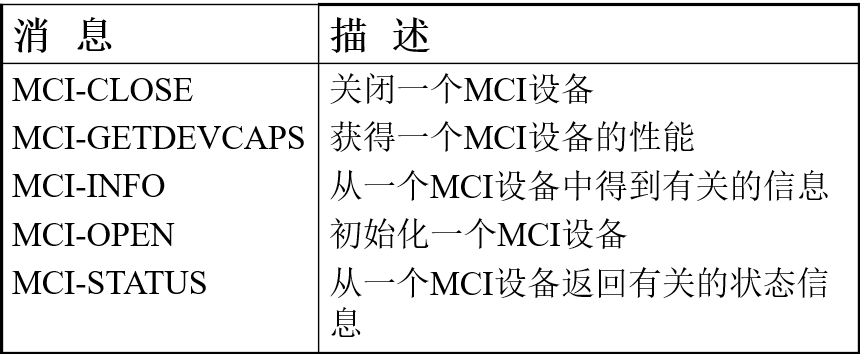
## 第二组是MCI设备操作及定位扩展命令组, 包括MCI\_CUE,MCI\_ESCAPE,MCI\_SEEK,MCI\_STEP;

## 第三组是窗口或视频设备的扩展命令组,包括MCI\_FREEZE, MCI\_PUT,MCI\_REALIZE, MCI\_UNFREEZE, MCI\_UPDATE, MCI\_WHERE, MCI\_WINDOW。

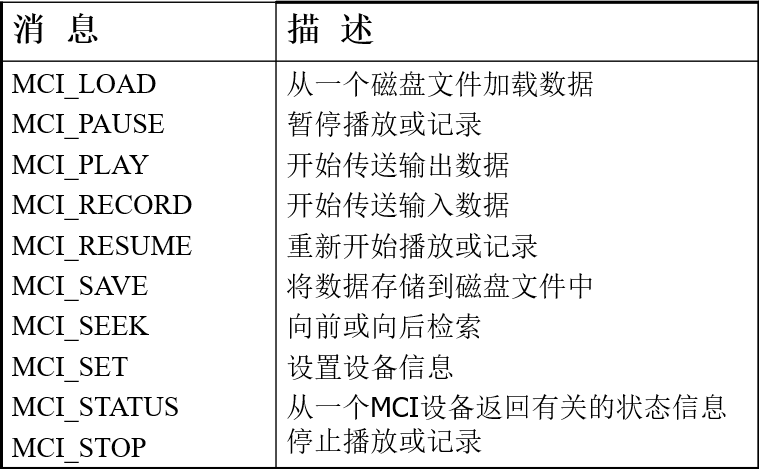
# 直接由MCI解释的命令



# 所有的MCI设备支持的命令消息



# 基本命令消息



# 打开一个设备

## 使用设备之前,必须使用MCI\_OPEN命令消息来初始化该设备

## 打开MCI设备的方法有以下几种:

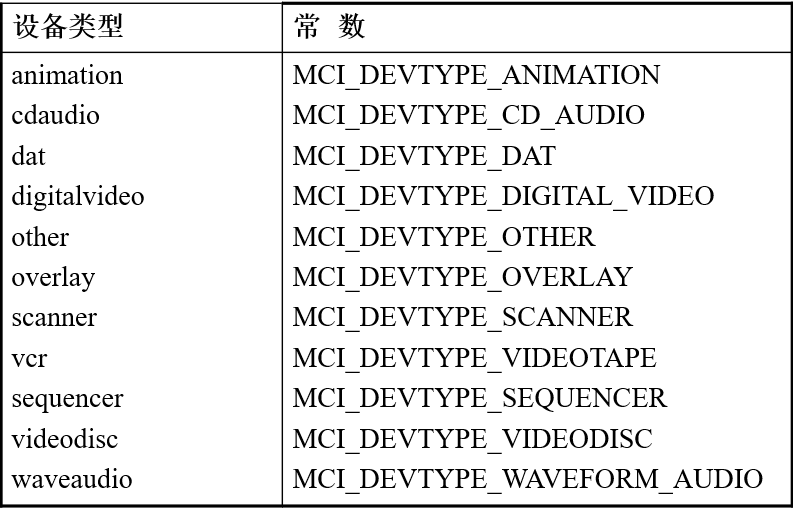
### (1)用设备类型字符串指定待打开的MCI设备

### (2)用MCI设备驱动程序名指定待打开的MCI设备

### (3)用设备类型常数(见下表)指定待打开的MCI设备

### (4)仅用设备元素指定打开的复合设备

# 设备类型及常数



# 关闭一个设备

## MCI\_CLOSE命令消息取消对一个设备或者设备元素的访问,它类似于一个文件的关闭操作。

## 为了有助MCI管理设备,应用程序在它使用完这个设备之后,应该明确地关闭它所使用过的每一个设备或者设备元素。

# 5.4.3 命令字符串接口

## 使用命令字符串接口3个函数:

### MciSendString向一个MCI设备驱动程序发送一个命令字符串。这个函数同时也具有对于回调函数和返回字符串的参数。

### MciGetErrorString返回一个同错误代码相对应的错误字符串。

### MciExecute向一个MCI设备驱动程序发送一个命令字符串。

# 

## 发送命令字符串 MciSendString函数的语法定义如下:

## WORD FAR PASCAL MciSendString(LpstrCommand, LpstrRtnstring, WORD Rtnlength, hcallBack)

## 指针LpstrCommand 指向一个以NULL结尾的MCI控制命令的字符串。这个字符串的形式如:

## *Command device\_name arguments*

## *如：* open 难忘的一天.mp3 alias mysong

## 指针LpstrRtnstring指向一个由应用程序提供的返回字符串缓冲区。 Rtnlength是缓冲区大小.

## 句柄hcallBack用来指定接收并处理MCI向应用程序发出的MM\_MCINOTIFY消息的窗口句柄。

# 

## 使用MciExecute发送命令字符串

## MciExcute函数是MciSendString的简化形式。其语法定义如下:

## BOOL MciExecute(Lpstr Command)

## LpstrCommand是一个指向以NULL结束的控制命令的字符串,字符格式同MciSendString。若函数调用成功返回TRUE,否则返回FALSE。

# 

## MciGetErrorString函数返回一个MCI错误代码的文本描述字符串,其语法如下:

## WORD MciGetErrorString(DWORD Error, LpstrBuffer, WORD Length)

## Error是错误代码,是上一次MciSendCommand或MciSendString函数调用的返回值;

## LpstrBuffer指向一个缓冲区指针,用来接收系统返回的文本描述;

## Length指定LpstrBuffer的长度。函数调用成功返回TRUE,否则表示查询的错误代码未知。

# Windows DirectX9.0家族成员

## DirectX Graphics：集成DirectDraw（视频输入输出）和Direct3D（3D图形API）

## DirectInput（有关鼠标、键盘、游戏杆和其他游戏控制设备，以及力回馈设备的一组API ）：支持输入输出

## DirectPlay（为开发者提供了开发诸如多人游戏或聊天程序的工具，完成了与用户连接相关的所有复杂工作）：网络游戏的通信组织

## DirectSetup：自动安装DirectX组件

# Windows DirectX9.0家族成员

## DirectMusic：MIDI音频播放

## DirectSound：实现.wav格式的波形声音数据的播放控制

## DirectShow：为在Windows平台上处理各种格式的媒体文件的回放、音视频采集等高性能要求的多媒体应用提供了完整的解决方案

## DirectX Media Objects：DirectShow Filter简化模型

# **5.4.5 DirectShow技术**

## 保证大量多媒体数据处理的高效性

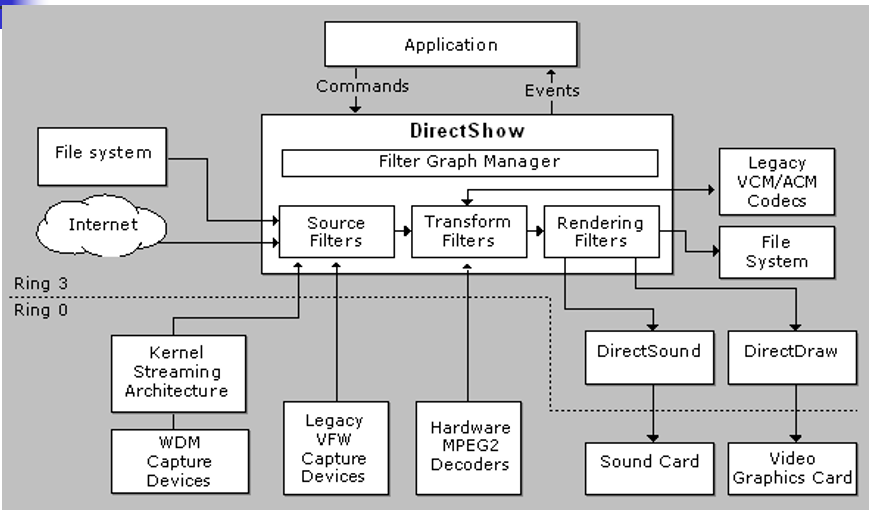
## 音视频同步

## 简单方法处理复杂媒体源问题：文件、网络、广播电视等

## 处理多种媒体格式的问题

## 支持不可预知的硬件

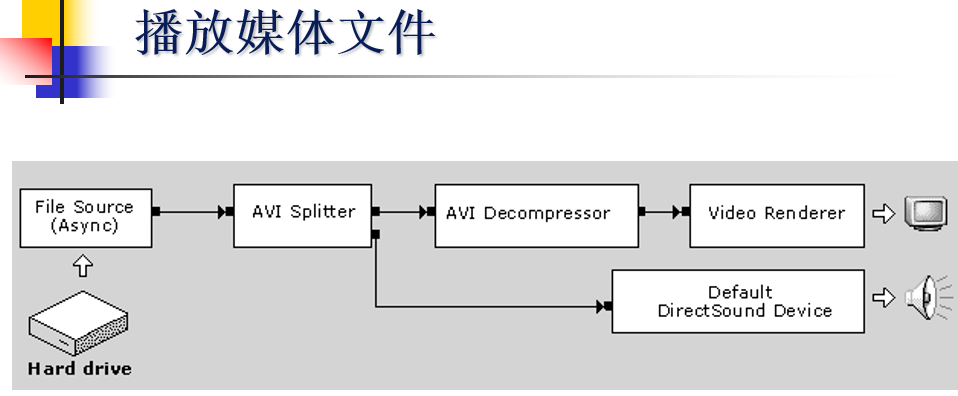
# DirectShow系统



# Filter Graph

## DirectShow的基本模块是Filter软组件,Filter通常实现一个对多媒体流的功能操作,比如读文件、从视频捕捉设备上获取视频、解码特殊的流格式、将数据传给图形卡或声卡等。

## 一系列连接的Filters被称为Filter Graph,通过Filters Graph Manager控制Graph中的数据流,它可以传递事件通知给应用程序, 以便程序能对事件作出反应。



# Filter概述

## Filter是DirectShow中最基本的概念

## Filter Graph是Filter的“容器”，Filter是Filter Graph中最小的功能模块

## Filter由一个或多个Pin组成，Filter之间通过Pin连接，构成一条顺序链路

### 仅含输出Pin，没有输入Pin为Source Filter

### 既有输入Pin，又有输出Pin为Transform Filter

### 仅有输入Pin，没有输出Pin为Rendering Filter

# Pin

## Filters的连接点被称为pin，每一个pin是独特的COM对象，提供IPin接口

## 每一个pin都有一个input或output的方向，必须与相反方向的pin进行连接

## 连接过程的步骤大致如下：

### Filter Graph Manager在输出Pin上调用IPin::Connect（带输入Pin的指针作为参数）

### 如果输出Pin接受连接，则调用输入Pin上的IPin::ReceiveConnection

### 如果输入Pin也接收这次连接，则双方连接成功

# 数据传输

* IPin接口主要是用于Pin连接, 用于数据传送一般是输入Pin上实现的IMemInputPin接口(方法IMemInputPin::Receive)
* 连接着双方Pin拥有同一个Allocator
* 数据传送时, 上一级Filter从输出Pin的Allocator中(调用IMemAllocator::GetBuff)得到一个空闲Sample的数据内存地址, 将数据放入其中
* 将这个Sample传送给下一级Filter的输入Pin

# 数据传送模式

## 数据传送主要有两种模式：

## 推模式（Push Model）

## 拉模式（Pull Model）

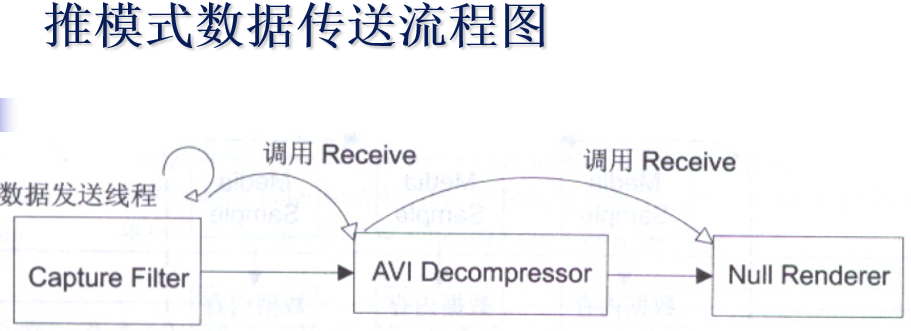
# 推模式

## 推模式最典型情况在Live Source(实时源)中。这种源能自己产生数据，并使用专门的线程将这些数据“推”下去

## 数据从Capture Pin出来，调用Transform Filter的输入Pin上的IMemInputPin::Receive函数，实现数据从Capture Filter到Transform Filter的传送

## 在Transform Filter内部，Filter将这块数据进行转换处理后将数据放到输出Pin的Sample中，调用Rendering Filter输入Pin上的IMemInputPin::Receive函数，而实现数据从Transform Filter到Rendering Filter的传送

## Rendering Filter接收到数据，进行必要的处理后就返回，数据传送的一个轮回完成



# 拉模式

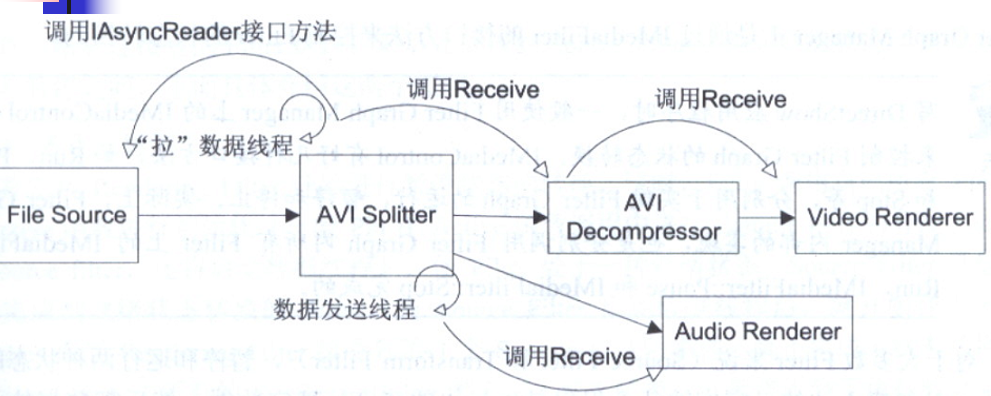
## 拉模式最典型发生在File Source(文件源)中。这种源管理数据, 但它没有把数据“推”下去的能力，而要靠后面的Filter来“拉”

## Source Filter输出Pin上实现IAsyncReader接口

## Source次级Filter的输入Pin上有一个“拉”数据的线程, 不断调用Source Filter输出Pin上的IAsyncReader接口方法来取得数据

## 在Transform Filter内部,将从Source Filter中取得的数据进行处理，然后通过各个输出Pin发送出去，往下的数据传送方式，与推模式相同

# 拉模式数据传送流程图



## DirectShow应用程序至少包含两个线程

## 应用程序主线程：**状态的改变**

## 数据传送子线程：**样本的传递**

# Filter项目的功能分析

## 功能单一化

### 需求过多，适度分解

### 需要在输入Pin上进行缓存的Filter,不需将缓存功能独立出去

## 选择一种Filter模型

### 大多数采用推模式，上级为被动源则用拉模式

### 区分Source Filter、Transform Filter或Rendering Filter

## 定义输入输出

### 输入、输出Pin的数量

### Pin上支持的媒体类型

## 接口定义

### Filter中哪些属性需要被外部知道和定制

## 其它一些特殊需求

### Pin上的数据缓存

### 给输出Sample打时间戳

### Filter内部需要其它线程及其同步问题

# Filter的设计

## 选择一个合适的父类

### 推模式Source Filter用CSource

### Filter接收输入，处理后输出用CTransformFilter（就地处理用CTransInPlaceFilter）

### Pin需要拉数据用CBaseFilter,并且Pin使用工具类CPullPin

### Filter不再需要输出用CBaseFilter（播放有标准Filter）

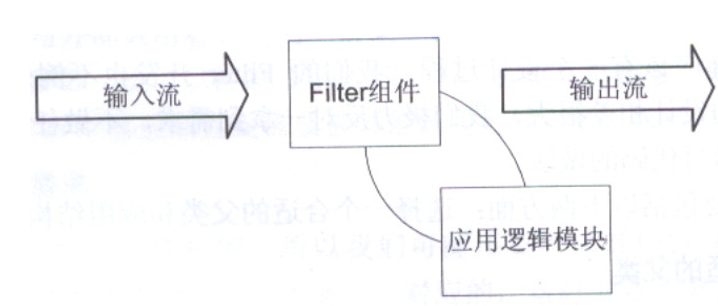
## 应用结构设计

### Filter是应用框架，应和应用逻辑分离

### Filter框架负责输入输出

### 应用逻辑负责数据处理控制策略，包括算法实现、效率优化等

# Filter框架与应用逻辑分离结构



# Source Filter Example：VOD Source Filter

## 流媒体接收Filter是一个典型的Source Filter

# VOD Source Filter功能分析

## 进行数据缓存,为下级Filter提供数据源

## 采用拉模式：MPEG-1 Stream Splitter只工作在拉模式下

## 一个输出Pin,支持媒体类型为MEDIASUBTYPE\_MPEG1VideoCD / MEDIASUBTYPE\_MPEG1System / MEDIASUBTYPE\_Avi

## 接口定义

### **设置流数据读取位置**

### **读取流中指定数量的数据**

### **获得数据总长度和当前可读取长度**

### **数据对齐方式**

### **数据读取同步对象的锁定/解锁**

## 特殊需求

### **网络传输、数据缓存、多线程问题**

# VOD Source Filter的设计

## Asynbase ：静态库项目，实现异步读写基类

### CAsyncStream：纯虚类，标识数据流源

### CAsyncRequest：表示输入输出请求

### CAsyncIo：实现了数据输入输出控制,包括同步/异步模式

### CAsyncOutputPin：输出Pin，从CBasePin派生

### IAsyncReader：拉模式的基本接口

## Async：使用asynbase库,实现拉模式Source Filter

### CMemStream：实现CAsyncStream，指定数据源内存区

### CMemReader：继承IAsyncReader，添加注册功能

# 媒体定位的实现

## 通过Filter Graph Manager上IMediaSeeking接口实现随机定位，其真正的实现在Filter上。

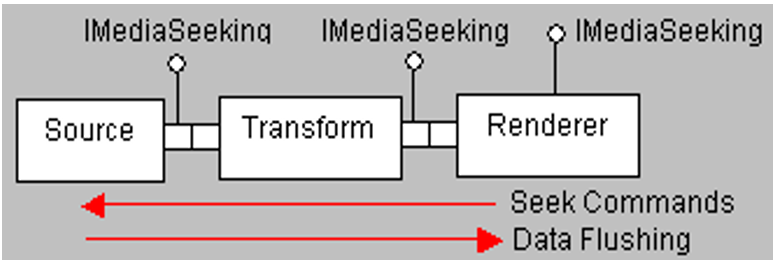
## Filter Graph Manager从Rendering Filter提出请求。

## 一般Rendering Filter不能真正执行定位操作而向上级Transform Filter输出Pin请求。

## 如果该级不能实现则向前继续请求,否则执行并返回

## 不成功的请求一直上溯到Source Filter

# IMediaSeeking接口的获取过程



# Flushing

## 丢弃采样的处理过程称为Flushing

## 当事件改变正常的数据流时,它使得Graph作出相应的反应

## 有BeginFlush和EndFlush两个方法

### BeginFlush：开始丢弃数据

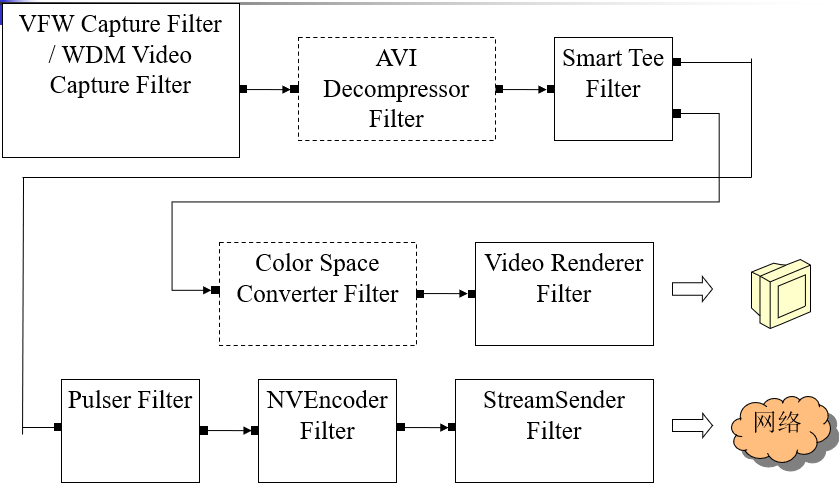
### EndFlush：结束丢弃数据

# Transform Filter Example：Pulser

## 视频媒体采样从用于捕捉的Filter产生后就被送到Rendering Filter进行回放然后再编码发送，可见视频回放的速率与捕捉的速率是一样的，如果没有速率转换Filter(Pulser)的话，编码发送的速率也将等同于捕捉的速率。

## 但是视频流发送速率一般为预定的值(使用前设定),在视频会议系统的客户端中默认为10帧每秒(出于带宽考虑)。但视频捕捉设备的捕捉速率一般为30帧每秒,而且大多数捕捉设备不能对速率进行调节, 因此必须有一种机制能够将媒体采样产生的速率降低，速率转换Filter正是起到这个作用

# 视频会议系统发送端Filter Graph



# Pulser Filter的设计

## 速率转换Filter是一种Transform Filter

## 用CPulserFilter类实现，从CTransInPlaceFilter基类派生而来

## Pulser Filter实现IPulser接口,用于调节速率

## Pulser Filter的输入Pin用CPulserInputPin类来实现，派生自CTransInPlaceInputPin基类

# 本章小结

## 首先介绍了媒体数据的获取方法，图形和动画的制作

## 对常用的多媒体创作工具进行了分类, 并以典型的软件为例介绍了它们各自的功能特点

## 最后介绍了Windows多媒体开发环境,着重介绍MCI的概念和使用方法, 以及DirectShow技术。

## 希望通过本章的介绍对读者开展多媒体应用有所帮助。