

数字媒体技术基础

王晗

数字媒体教研室

wanghan@bjfu.edu.cn

<ftp://211.71.149.149/WangHan>

西配102

FTP课件下载及作业上传说明

- ▶ 在ftp://211.71.149.149 /WangHan文件夹下的teaching plan文件夹中,学生可以直接下载课件, 无需密码,
- ▶ 在homework文件夹,用于学生上传作业, 学生只能上传, 不能下载。
- ▶ 注意: 按照作业要求命名文件, 请勿重复上传!

课程简介

- ▶ 学时安排：48学时
 - ▶ 课堂教学：30学时
 - ▶ 上机实践：14学时，48以上课外学时
 - ▶ 研讨：4学时
- ▶ 课程验收

笔试50%+实验40%+平时成绩10%

课程内容

——分为八个专题

- ▶ 数字媒体技术基础概论
- ▶ 数字图像技术
- ▶ 数字视频技术
- ▶ 数字音频技术
- ▶ 数字媒体表示技术
- ▶ 数字媒体压缩技术
- ▶ 数字媒体传输技术
- ▶ 计算机视觉与人工智能

数字媒体技术基础

——专题一：数字媒体技术概论

王晗

wanghan@bjfu.edu.cn

<ftp://211.71.149.149/WangHan>

1. 媒体概念及特性

媒体是指交流、传播信息的工具，如报刊、广播、广告等。

--- 《现代汉语词典》（1998年修订本，商务印书馆）

媒体是数据记录的载体，包括磁带、光盘、软盘等。这两种解释说明媒体是一种工具，包括信息和信息载体两个基本要素。一张光盘不能称为媒体，只有记录了信息，并可进行信息传播时才称为媒体。

--- 《现代英汉词典》

1. 媒体概念及特性

► 1.1 传播范畴中的媒体含义

媒体的英文单词是medium，源于拉丁文的medius，其含义是中介、中间的意思，常用复数形式media。同时，媒体又是信息交流和传播的载体。

“现代大众传播学之父”施拉姆(Wilbur Schramm)认为，“媒介就是插入传播过程之中，用以扩大并延伸信息传送的工具。”

英国南安普顿大学的媒介教育学家A.哈特(Andrew Hart)把媒介分为三类：示现的媒介系统、再现的媒介系统、机器媒介系统。

1. 媒体概念及特性

► 1.1 传播范畴中的媒体含义

传播学研究领域最有影响的媒介研究学者、加拿大多伦多大学教授麦克卢汉认为，“媒介就是信息”。媒体包括两层含义：

(1) 传递信息的载体，称为媒介，是由人类发明创造的记录和表述信息的抽象载体，也称为逻辑载体，如文字、符号、图形、编码等。

(2) 存储信息的实体，称为媒质，如纸、磁盘、光盘、磁带、半导体存储器等。载体包括实物载体、或由人类发明创造的承载信息的实体，也称为物理媒体。

媒体概念及特性

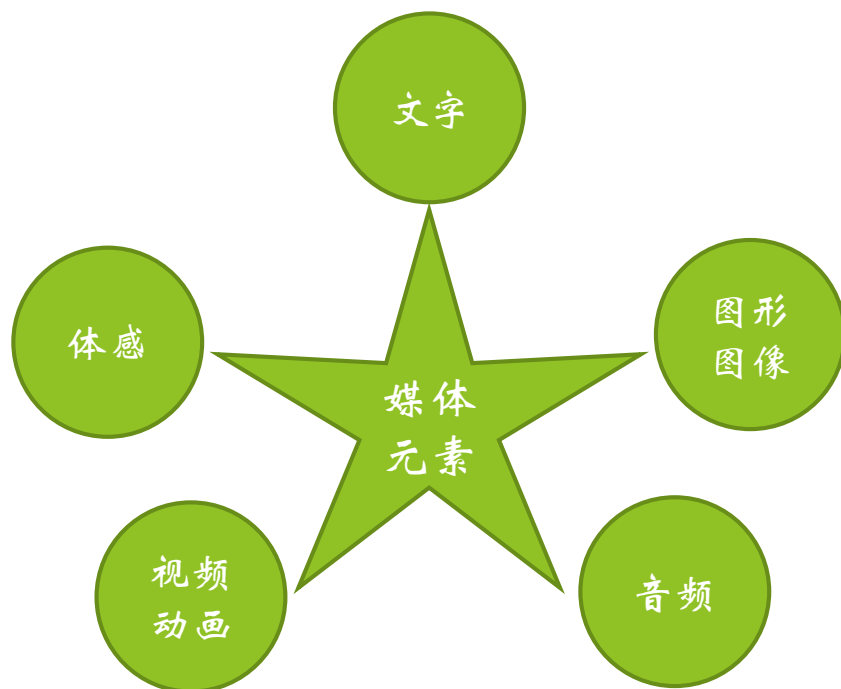
► 1.2 技术范畴中的媒体

《自然辩证法百科全书》中把技术定义为“人类为了满足社会需要依靠自然规律和自然界的物质、能量和信息来创造、控制、应用和改进人工自然系统的手段和方法”。

视频：人工智能与计算机视觉

1. 媒体的概念及特性

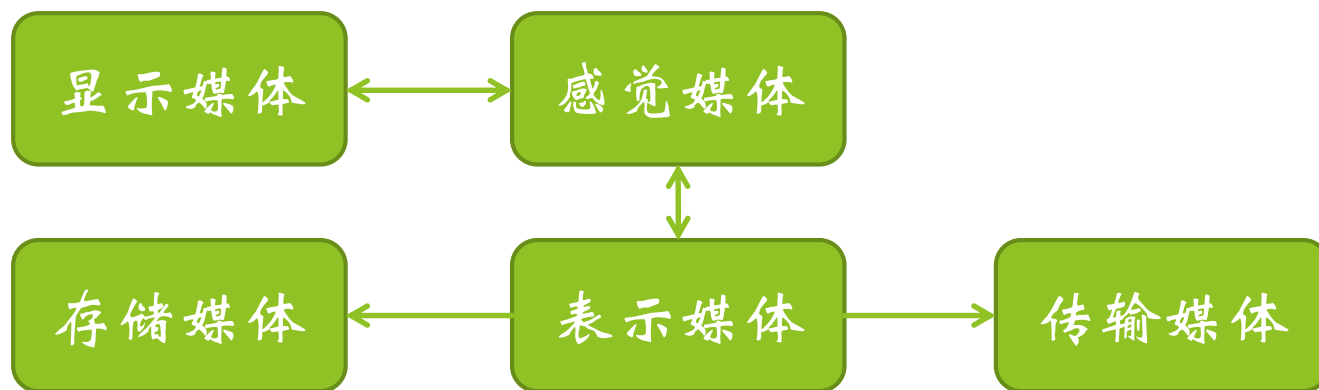
- ▶ 多媒体：使用计算机交互式综合技术和数字通信网络技术处理多种表示媒体，如文本、图形、图像和声音，并使多种信息建立逻辑连接，集成为一个交互式系统。（ITU）



媒体概念及特性

1.2 技术范畴中的媒体

- ▶ 媒体 (media)：人类信息的载体和表现形式，人类通过感官系统接收和发送这些信息。
- ▶ International Telecommunications Union, ITU将媒体定义为5种类型。



感觉媒体 (Perception) , 是指能够直接作用于人的感觉器官, 使人产生直接感觉(视、听、嗅、味、触觉)的媒体, 如语言、音乐、各种图像、图形、动画、文本等。

表示媒体 (Presentation) , 是指为了传送感觉媒体而人为研究出来的媒体, 借助这一媒体可以更加有效地存储感觉媒体, 或者是将感觉媒体从一个地方传送到远处另外一个地方的媒体, 如语言编码、电报码、条形码、语言编码, 静止和活动图象编码以及文本编码等。

显示媒体 (Display)，是显示感觉媒体的设备。显示媒体又分为两类，一类是输入显示媒体，如话筒，摄象机、光笔以及键盘等，另一种为输出显示媒体，如扬声器、显示器以及打印机等，指用于通信中，使电信号和感觉媒体间产生转换用的媒体。

存储媒体 (Storage)，用于存储表示媒体，也即存放感觉媒体数字化后的代码的媒体称为存储媒体。例如磁盘、光盘、磁带、纸张等。简而言之，是指用于存放某种媒体的载体。

传输媒体 (Transmission)，传输媒体是指传输信号的物理载体，例如同轴电缆、光纤、双绞线以及电磁波等都是传输媒体。

2. 数字媒体概念及特性

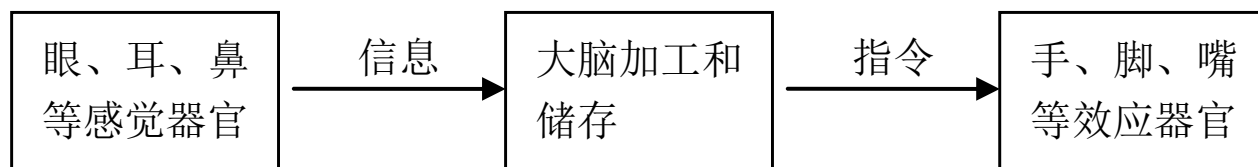
► 定义：

- 数字媒体 (Digital Media) 是指以二进制数的形式通过计算机产生、获取、记录、处理和传播的信息媒体，这些媒体包括感觉媒体、表示媒体、存储媒体、传输媒体等。
- 该定义主要包含两个方面：
 - 数字化的手段，即利用二进制进行各类信息的处理，这显然和计算机密切相关；
 - 媒体，也就是说和信息的传播和传达有关。

2. 数字媒体概念及特性

信息及其数字化

► 人们处理信息的全过程：



► 连续形式的模拟信息：

- 指在时间上和数值上都连续变化的信号
- 适合大脑处理
- 模拟信号 (analog signal)

► 离散形式的数字信息：

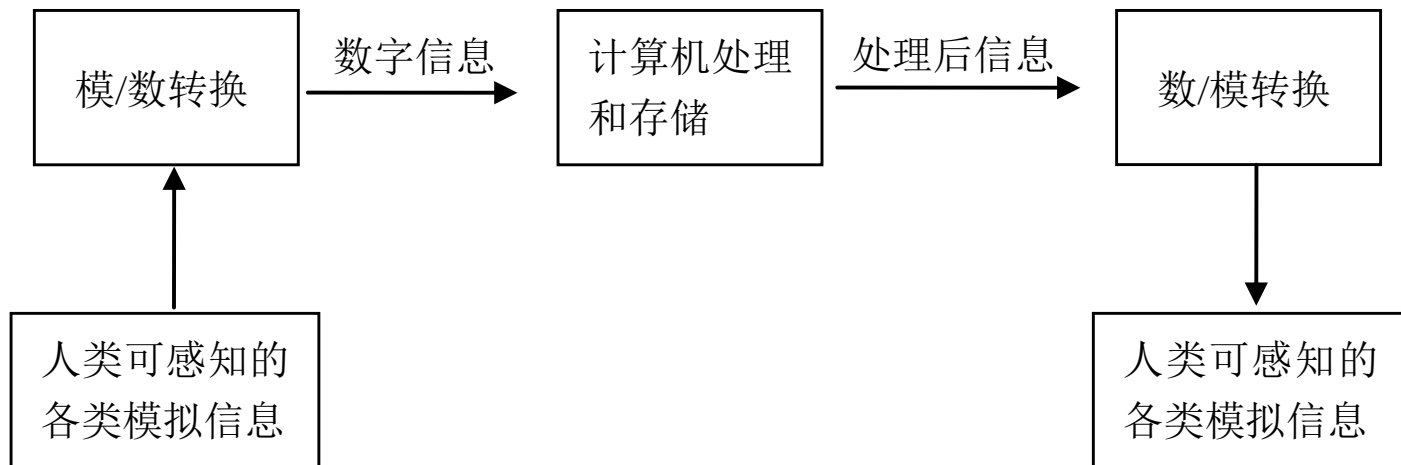
- 指在时间上和数值上都不连续变化的信号
- 计算机能够处理
- 数字信号 (digital signal)

► 信息的数字化包含：

► 模/数转换 (A/D)

► 数/模转换(D/A)

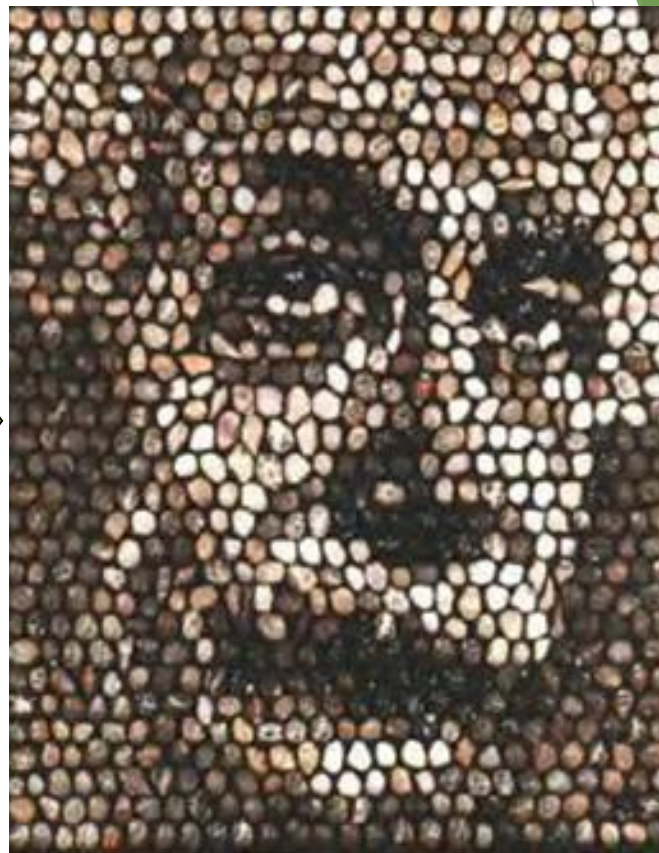
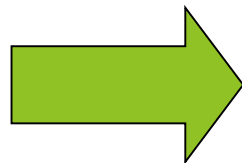
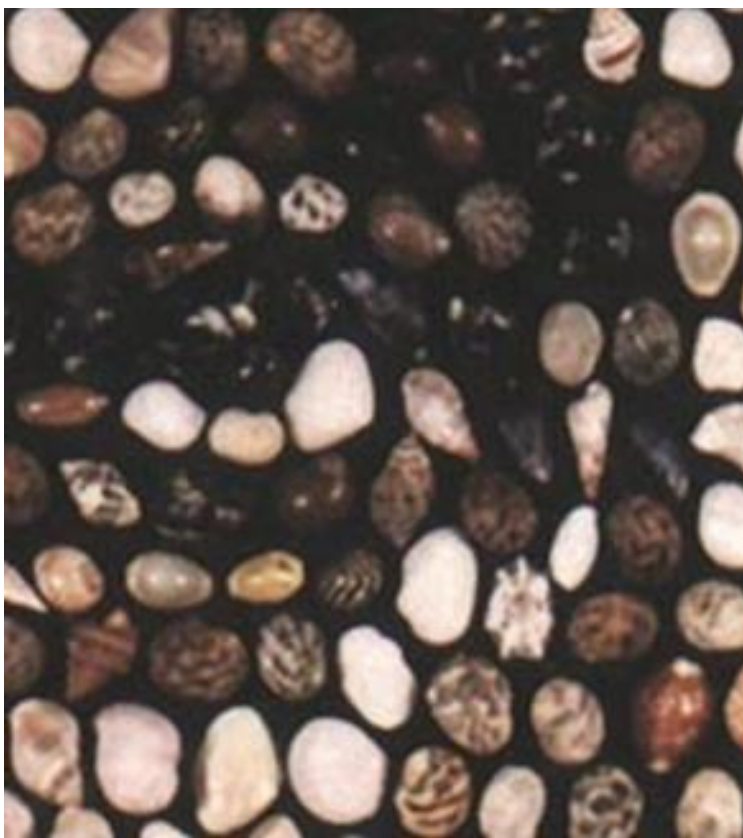
► 电子计算机处理信息的过程：



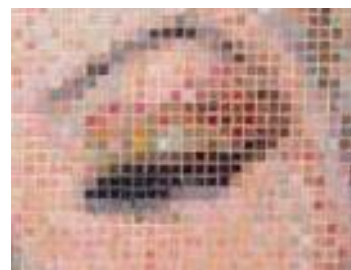
例 1-1 数字的含义

- ▶ 计算机中使用0和1两个数字来进行数值的运算和各类信息的表示。
- ▶ 由于计算机存储位数的限制，计算机只能表示有限位数的数值，致使计算机表示的数据都是不连续的。
- ▶ 为什么我们能够读懂计算机所表示的那些不连续数据呢？

例 1-1 数字的涵义



例 1-2 数字的涵义



2. 数字媒体概念及特性

▶ 2.2 数字媒体的分类

▶ 按时间

- ▶ 静止(still)媒体
- ▶ 连续(continues)媒体

▶ 按来源

- ▶ 自然 (natural) 媒体
- ▶ 合成 (synthetic) 媒体

▶ 按组成元素

- ▶ 单一媒体 (single media)
- ▶ 多媒体 (multimedia)

2. 数字媒体概念及特性

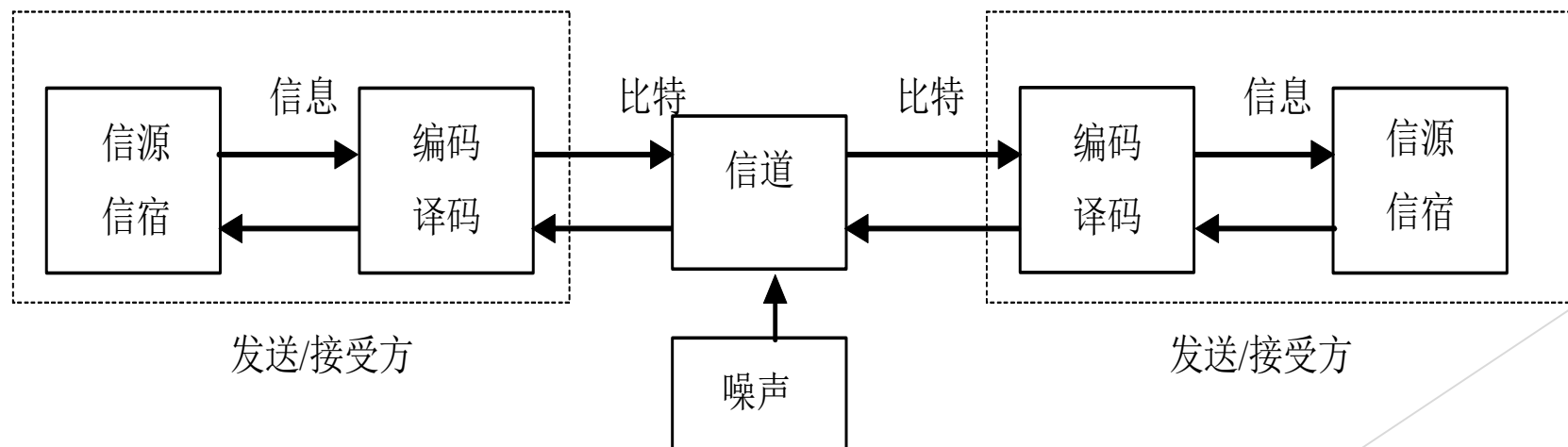
► 2.3 数字媒体特性

- (1) 数字化
- (2) 交互性
- (3) 趣味性
- (4) 集成性
- (5) 技术与艺术的融合

2. 数字媒体概念及特性

► 2.4 数字媒体传播模式

数字媒体系统完全遵循信息论的通信模式。从通信技术上看，它主要由计算机和网络构成，如下图所示。它在传播应用方面比传统的大众传播更有独特的优势。在数字媒体传播模式中，信源和信宿都是计算机。因此，信源和信宿的位置是可以随时互换的。这与传统的大众传播如报纸、广播电视等相比，发生了深刻的变化和革命。



3. 数字媒体技术的内涵

► 3.1 计算机技术与多媒体



3. 数字媒体技术的内涵

► 3.2 数字媒体艺术

数字媒体艺术是随着20世纪末数字技术与艺术设计相结合的趋势而形成的一个新的交叉学科和艺术创新领域，一般是指以“数字”作为媒介素材，通过运用数字技术来进行创作的，具有一定独立审美价值的艺术形式或艺术过程，是一种“在创作、承载、传播、鉴赏与批评等艺术行为方式上推陈出新，颠覆传统艺术的创作手段、承载媒介和传播途径，进而在艺术审美的感觉、体验和思维等方面产生深刻变革的新型艺术形态。数字媒体艺术是一种真正的技术类艺术，是建立在技术的基础上并以技术为核心的新艺术，以具有交互性和使用网络媒体为基本特征。

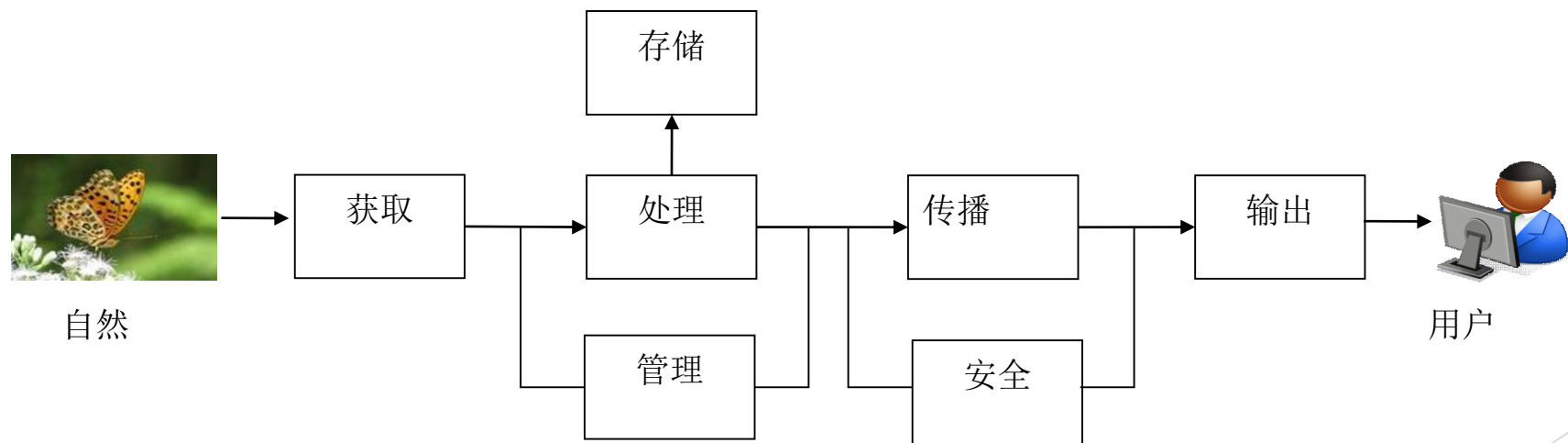
3. 数字媒体技术的内涵

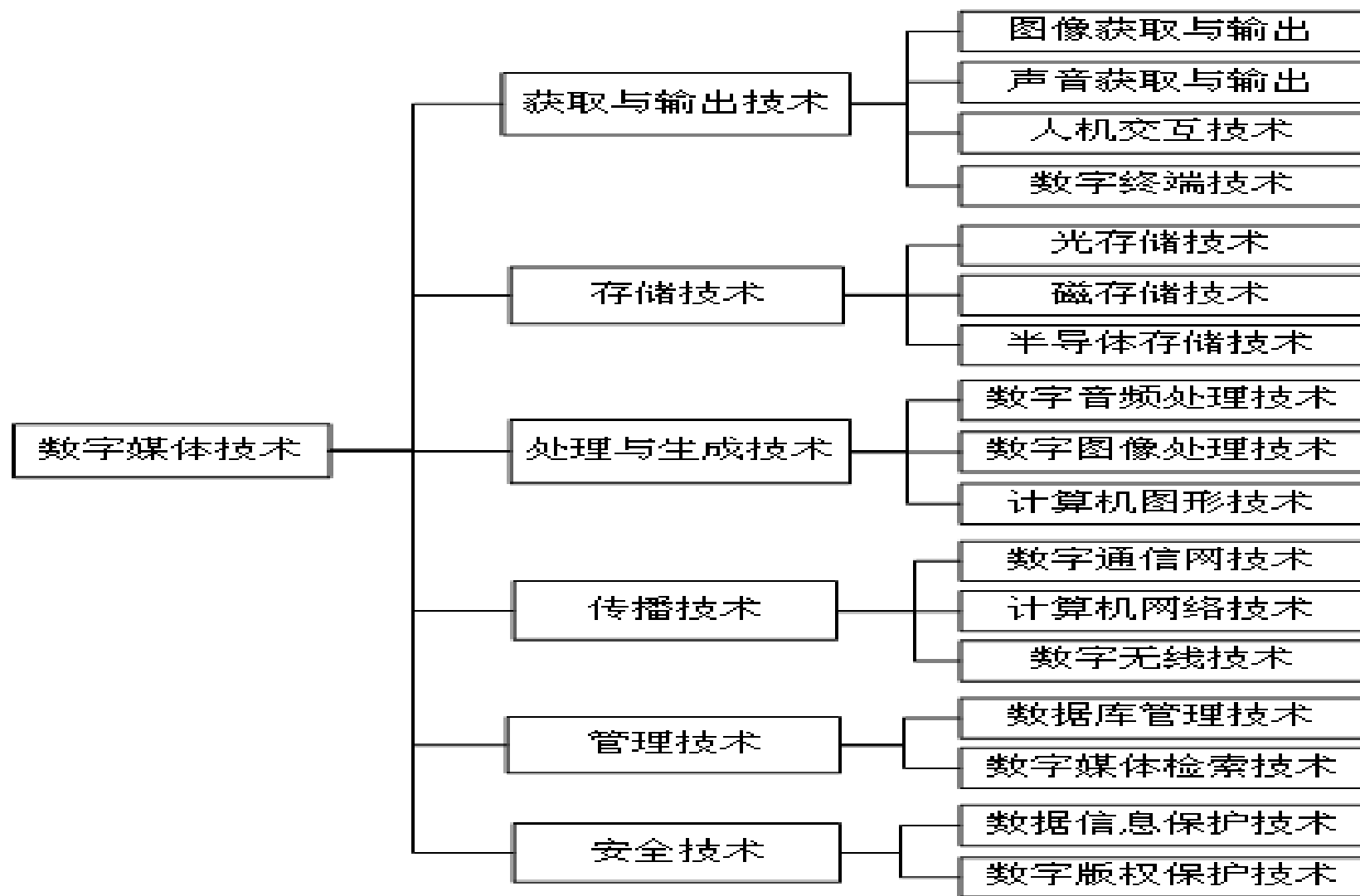
► 3.3 数字媒体技术

数字媒体涉及的技术范围很广，技术很新、研究内容很深，是多种学科和多种技术交叉的领域。主要技术范畴包括：

- (1) 数字媒体表示与操作，包括数字声音及处理、数字图像及处理、数字视频及处理、数字动画技术等。
- (2) 数字媒体压缩，包括通用压缩编码、专门压缩编码（声音、图像、视频）技术等。
- (3) 数字媒体存储与管理，包括光盘存储（CD技术、DVD技术等）、媒体数据管理、数字媒体版权保护等。
- (4) 数字媒体传输，包括流媒体技术、P2P技术等。

► 数字媒体的流通过程





3.3.1. 数字信息获取与输出技术

- ▶ 数字信息的获取设备，也称为Input设备。其技术基础是现代传感技术。
- ▶ 数字信息输出技术主要研究如何将计算机信息转化成更符合人类认知的信息表现形式的技术，主要包括显示技术、硬拷贝技术、声音系统、以及三维显示技术等。数字信息的输出设备也称为Output设备。

演示视频

丰富多彩的人机交互设备



(a) 视线跟踪设备



(b) 力反馈感应设备



(c) 三维扫描仪



(d) “真” 三维显示器



(e) 可穿戴计算机交互设备

3.3.2 数字信息存储技术

► 目前占据主流地位的存储技术：

► 磁存储技术

- 运用电磁效应原理将携带信息的电信号转换成具有相同变化规律的磁场，然后将磁性记录介质层磁化，并以介质层的剩磁的形式形成信息的物理标志长期保存。
- 主要形式有硬磁盘、磁带等。

► 光存储技术

- 采用光学方式来记录和读取二进制信息的技术。
- 主要形式有CD、VCD、DVD、BD等。

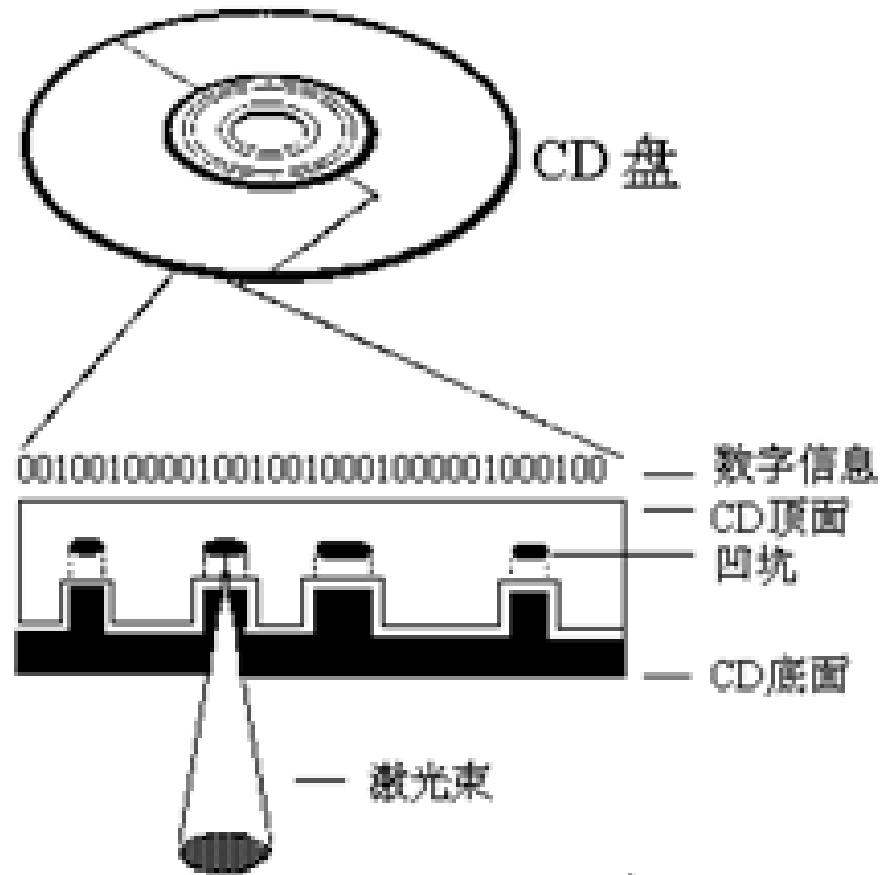
► 半导体存储技术

- 采用半导体材料来记录和读取二进制信息的技术。
- 主要形式是RAM和ROM。

蓝光存储技术

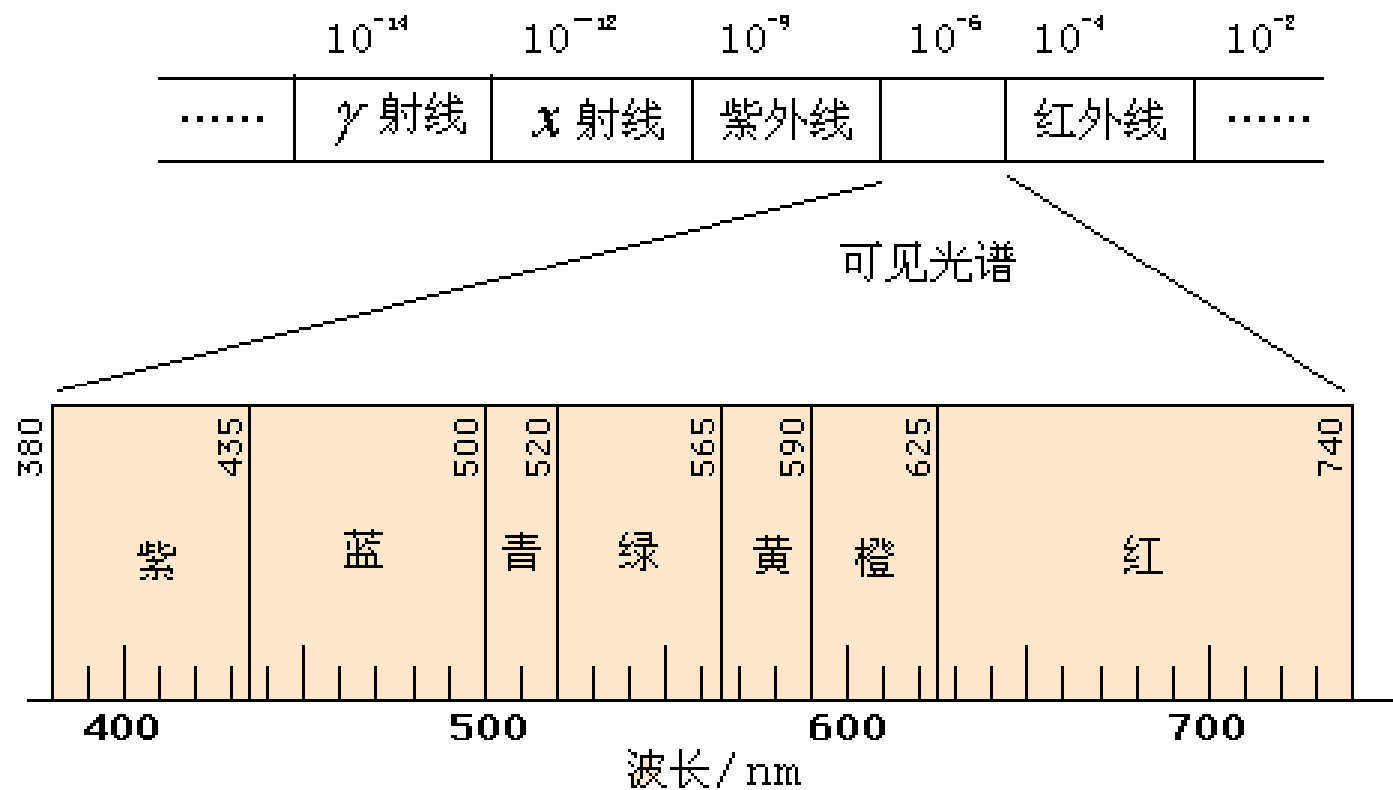
- ▶ 光的波长直接影响着读取光盘数据的尺度，通常来说波长越短的激光，能够在单位面积上记录或读取更多的信息。光盘的发展经历了VCD、DVD阶段后，就要进入BD（Blu-ray Disc，缩写为BD，即蓝光盘）阶段。VCD是采用波长为780nm的红光、DVD是采用波长为650nm的红光、BD则采用405nm的蓝色来读取或写入数据。
- ▶ BD光盘根据物镜NA大小的不同分为HD-DVD（High Definition DVD）和Blu-ray Disc两种。HD-DVD的单面单层容量为20G，Blu-ray Disc的单面单层有23.3GB、25GB及27GB三种盘片。物镜NA的变大，让光点能在高密度光盘中更准确读写资料，但缺点是容易造成光差增加。

蓝光存储技术







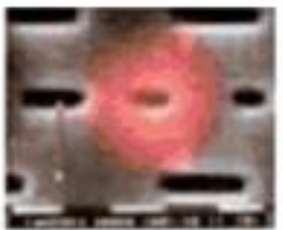

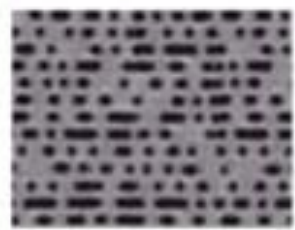

CD盘的读出原理图

蓝光存储技术



可见光光谱图

蓝光存储技术

CD	DVD	HD-DVD	Blu-ray Disc
 $\lambda=780\text{nm}$ $\text{NA}=0.45$	 $\lambda=650\text{nm}$ $\text{NA}=0.6$	 $\lambda=405\text{nm}$ $\text{NA}=0.65$	 $\lambda=405\text{nm}$ $\text{NA}=0.85$
			

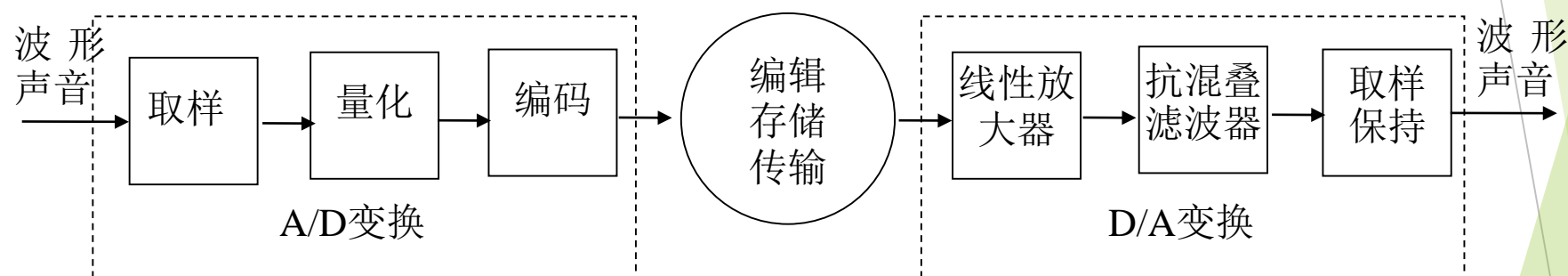
各类光盘参数比较图

3.3.3 数字信息处理与生成技术

- ▶ 数字媒体信息的获取途径有两种：
 - ▶ 一种是直接来源于现实生活，再经数字化处理，转化成数字媒体，例如音频、图像、视频等；
 - ▶ 另一种是间接来源于数字生活，由计算机直接合成得到，无需经历数字化过程，例如计算机合成音乐、图形、动画等。

a. 数字音频处理与生成技术

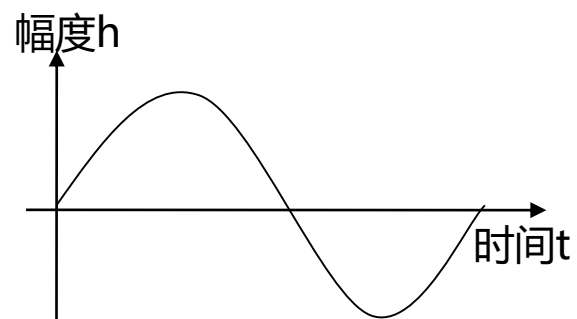
► 声音的数字化过程经过取样、量化和编码三步。



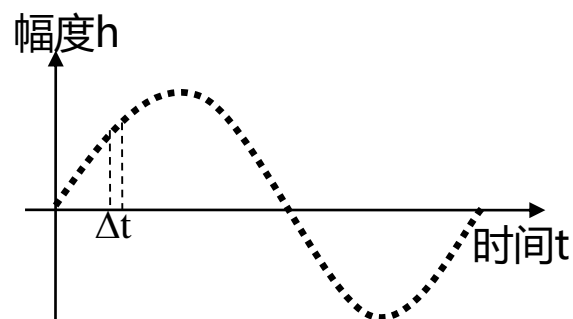
声音的数字化过程图

现实生活中的声音都是以波的形式存在的。人耳能够听到的声波的频率范围为20Hz~20kHz之间的物理波，其中人的说话声是频率为300~3400Hz的波，成为言语 (speech)

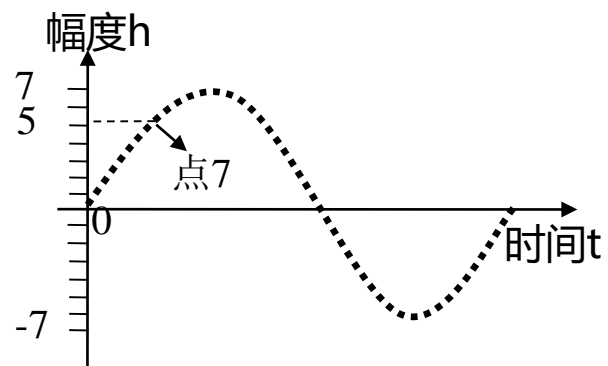
声音信息的数字化过程



(a) 原始声音信号



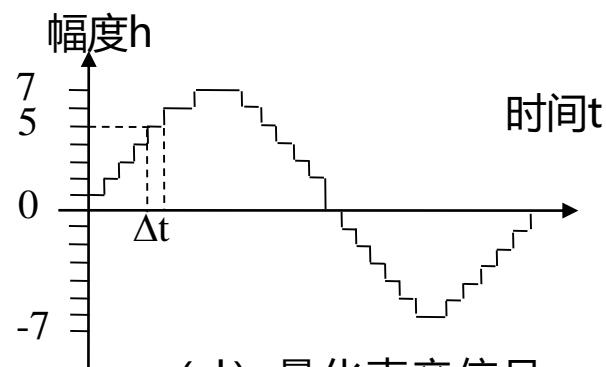
(b) 采样声音信号



(c) 量化声音信号中

0001 0010 0011 0100.....

(e) 声音信号编码



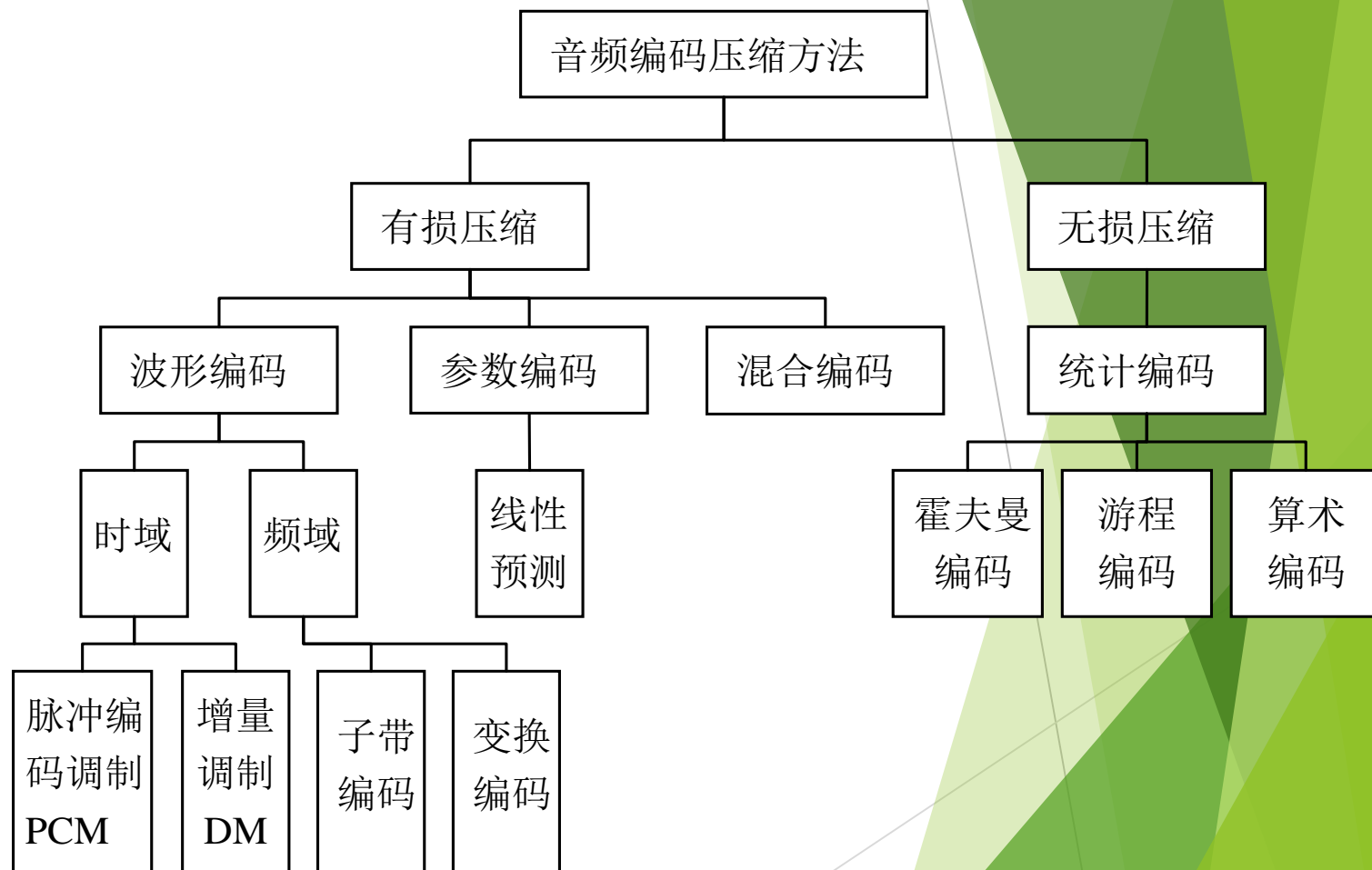
(d) 量化声音信号后

a. 数字音频处理与生成技术

- ▶ 经过数字化后的音频信号如果不进行压缩，则数据量会非常巨大。
- ▶ 例如，CD音质数字音频的采样频率为44.1kHz，量化精度为16bit，如果采用双声道立体声时，1分钟音频所需的存储空间为：
$$44.1\text{kHz} \times 16\text{b} \times 2 \times 60\text{s} = 84672\text{kb} \approx 10.3\text{MB}。$$

a. 数字音频处理与生成技术

- ▶ 根据压缩后音频信号是否能够完全重构出原始声音可以将音频压缩技术分为无损压缩和有损压缩两大类。
- ▶ 无损压缩一般是通过统计的方法，统计出数字信息中的冗余信息，
- ▶ 有损压缩则是通过损伤部分信号信息来达到较高的压缩比。



案例6-6 霍夫曼编码

- ▶ 霍夫曼 (Huffman) 编码是一种无损压缩编码，属于统计编码。霍夫曼编码的码长是变化的，对于出现频率高的信息，编码的长度较短；而对于出现频率低的信息，编码长度较长。这样，处理全部信息的总码长一定小于实际信息的符号长度。
- ▶ 霍夫曼编码举例
 - ▶ 现有一个由5个不同符号组成的30个符号的字符串：
ABBECDABBDEDEDACDCBBDD
 - ▶ 计算
 - (1) 该字符串的霍夫曼码
 - (2) 该字符串的平均码长
 - (3) 编码前后的压缩比

案例6-6 霍夫曼编码

► 算法步骤：

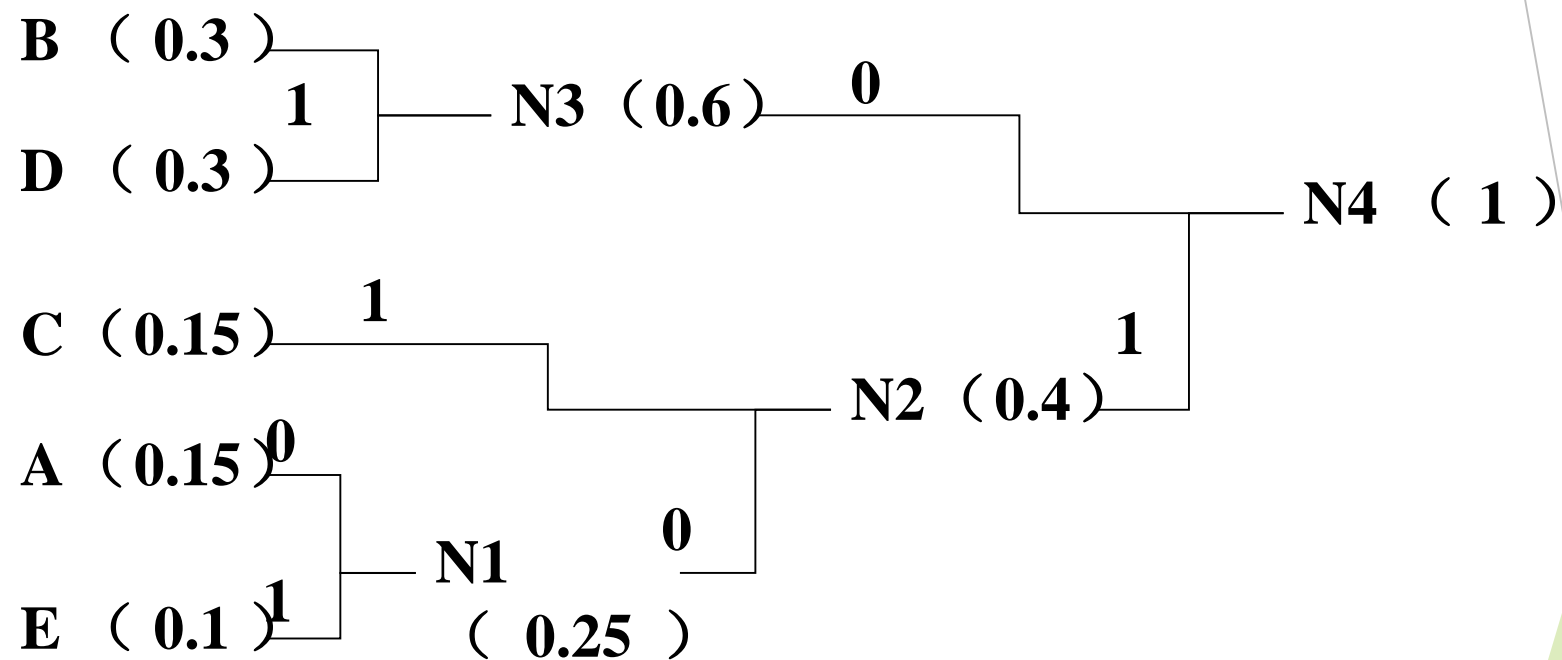
- (1) 初始化，将信号源的符号按照出现概率递减的顺序排列。
- (2) 计算，将两个最小出现概率进行合并相加，得到的结果作为新符号的出现概率。
- (3) 重复进行步骤1和2直到概率相加的结果等于1为止。
- (4) 分配码字，对所有出现的符号进行分配码字，概率大的符号用编码0表示，概率小的符号用编码1表示（当然也可以倒过来）。
- (5) 记录编码，记录下概率为1处到当前信号源符号之间的0，1序列，从而得到每个符号的编码。

案例6-6 霍夫曼编码

符号出现的概率

符号	出现的次数/概率	分配的代码	需要的位数
A	3/0.15	?	
B	6/0.3	?	
C	3/0.15	?	
D	6/0.3	?	
E	2/0.1	?	
合计	20		

案例6-6 霍夫曼编码



霍夫曼编码图

案例6-6 霍夫曼编码

符号	出现的次数/概率	分配的代码	需要的位数
A	3/0.15	100	3
B	6/0.3	00	2
C	3/0.15	11	2
D	6/0.3	01	2
E	2/0.1	101	3
合计	20		

利用霍夫曼编码最后得到的码长为：

$$3 \times 3 + 2 \times 6 + 2 \times 3 + 2 \times 6 + 3 \times 2 = 45,$$

比原来60个压缩了15个编码，压缩比为60：45=4：3

案例6-6 霍夫曼编码

▶ 霍夫曼编码的特点：

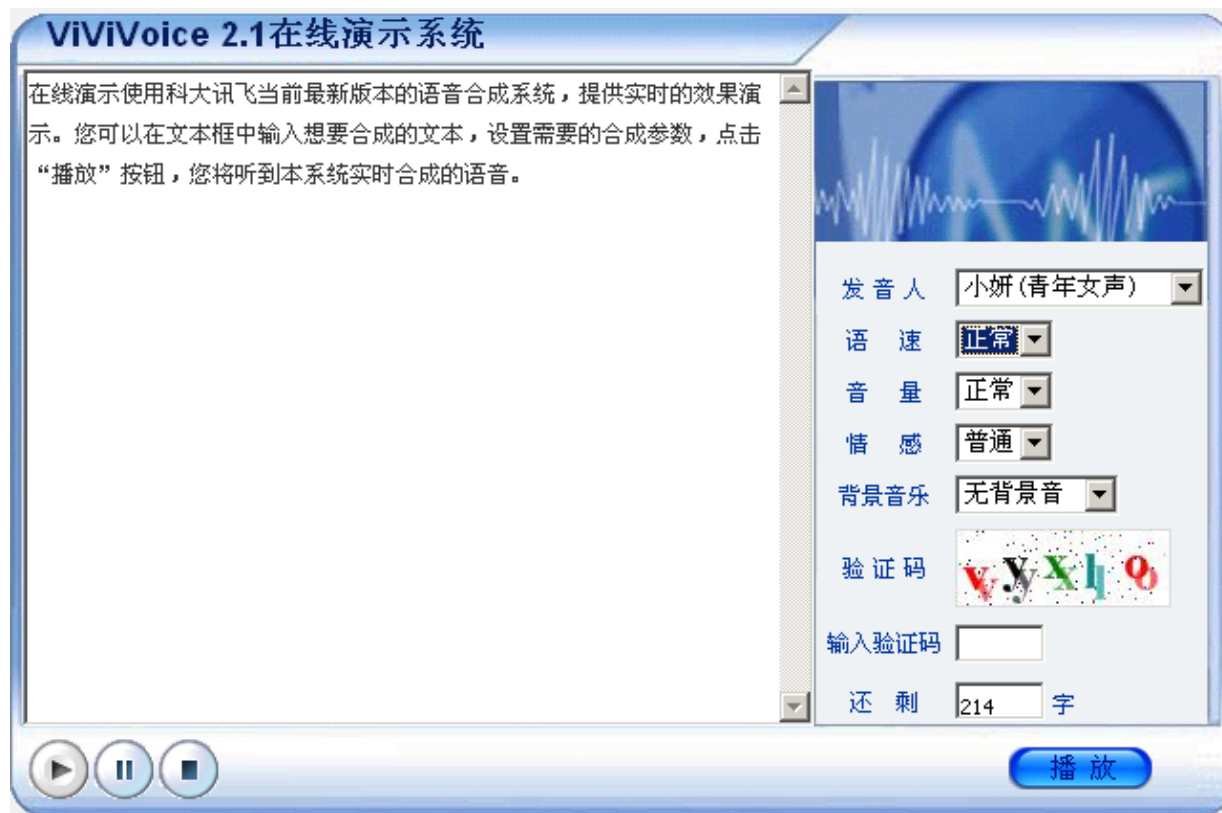
- ▶ 霍夫曼编码属于异字头码，保证了码的唯一可译性。
- ▶ 编码长度是可变的，因此译码时间较长，使得霍夫曼编码的压缩与还原相当费时。
- ▶ 由于编码长度不统一，硬件实现起来有难度。
- ▶ 对不同信号源的编码效率不同，当信号源的符号概率为2的负幂次方时，达到100%的编码效率；若信号源符号的概率相等，则编码效率最低。
- ▶ 由于"0"与"1"的指定是任意的，故由上述过程编出的最佳码不是唯一的，但其平均码长是一样的，故不影响编码效率与数据压缩性能。

a. 数字音频处理与生成技术

- ▶ 数字语音处理技术主要包括语音合成、语音识别和语音增强三部分。
 - ▶ 语音合成是指利用机器来模拟人类语言的技术，它不同于机器预先录制声音并回放的技术，而是指可以在任何时候将任意文本转换成具有高自然度的语音，从而真正实现让机器“像人一样开口说话”。
 - ▶ 语音识别就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的技术。
 - ▶ 语音增强就是从带噪声信号的原始语音中尽可能消除噪声，改进语音质量、提高语音的易懂度的技术。

科大讯飞语音合成技术

- ▶ 安徽科大讯飞信息科技股份有限公司 (<http://www.iflytek.com/>)，简称科大讯飞。

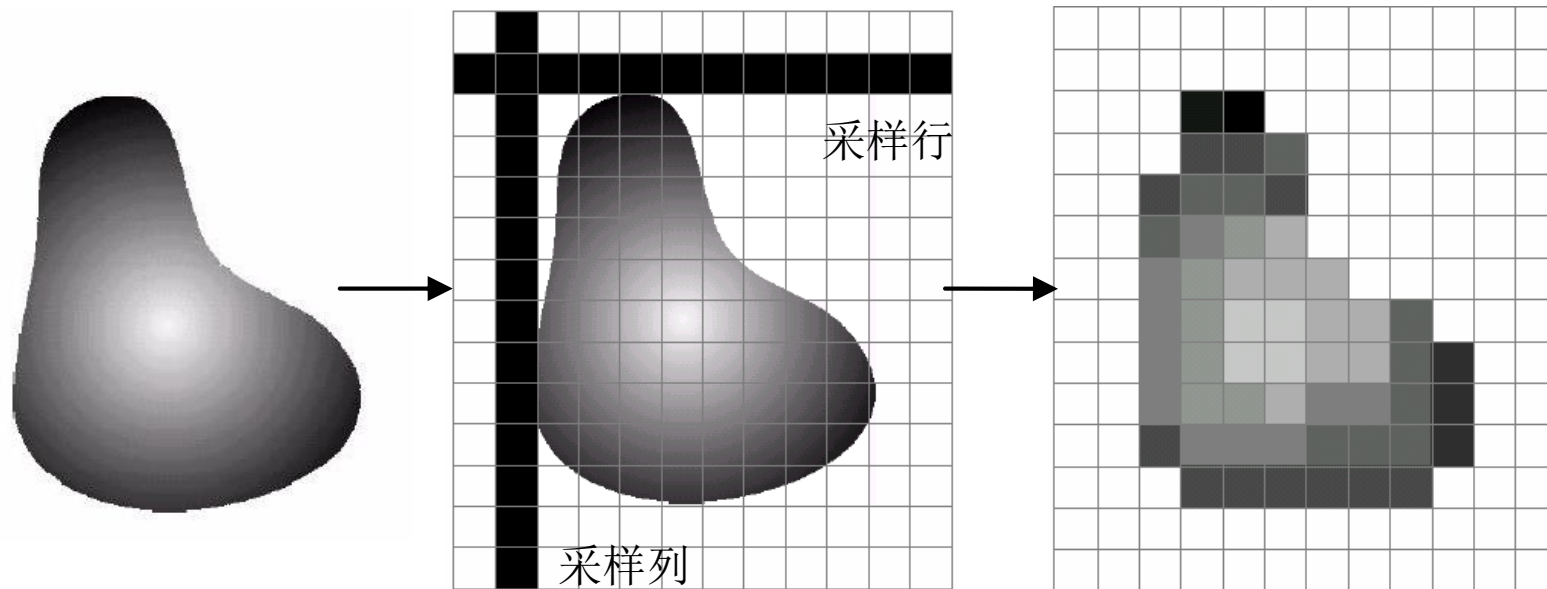


ViViVoice 2.1在线演示系统的界面

b. 数字图像处理技术

- 现实生活中的图像是在空间上和色彩上均连续的二维景象，所以图像的数字化不仅要在空间上将二维图像离散化，而且在色彩上也要将其离散化。

原始图像(景物) \longrightarrow 采样 \longrightarrow 量化



图像的数字化

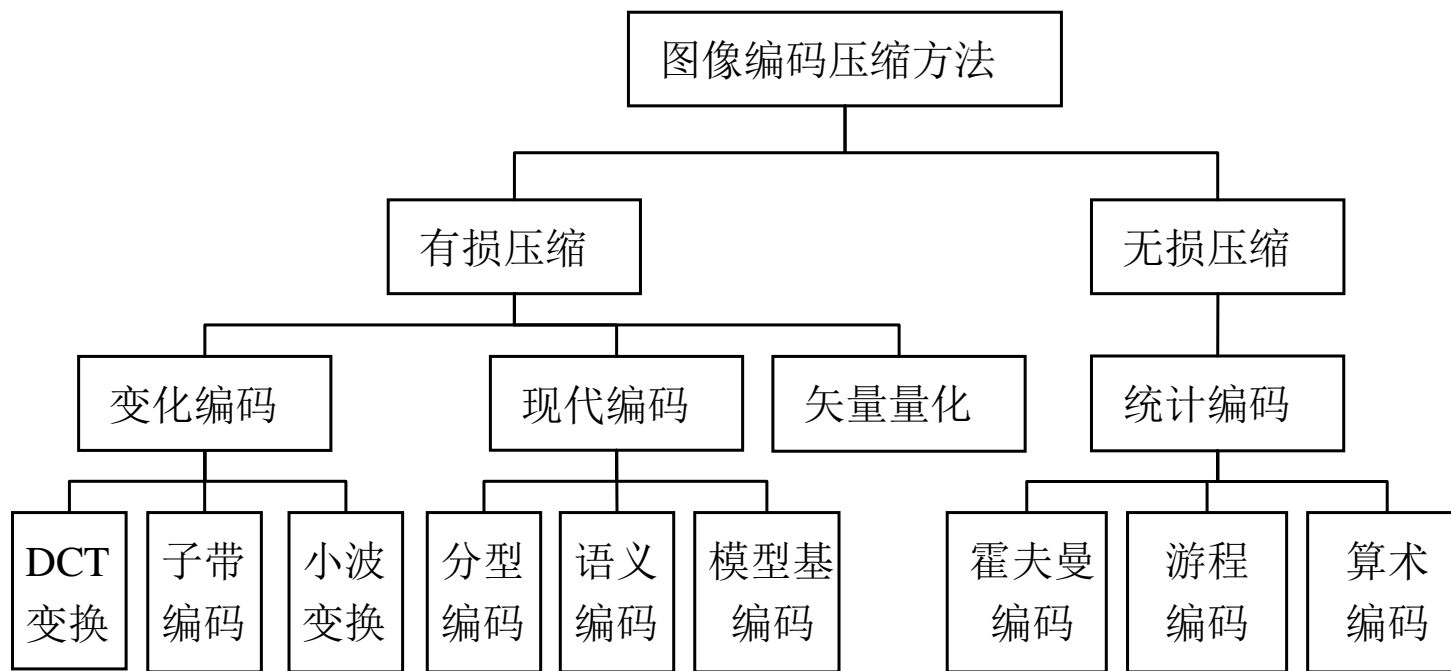
► 数字化过程：

1. 采样：图像是一种二维信号，需要变为一维信号后采样，先沿垂直方向，按一定间隔从上往下顺序地沿水平方向直线扫描的方式，取出各水平行上的一维扫描线，再对该一维扫描线信号按一定间隔采样得到离散信号。经过采样后，一副图像的像素数目也称为图像的分辨率，图像分辨率一般用水平方向的像素个数M乘以垂直方向的像素个数N来表示，即 $M \times N$ 。
2. 量化：经过采样，模拟图像已在空间上离散化为像素，但所得的像素值（即颜色值或灰度值）仍是连续量，把取样后所得的这些连续量表示的像素值离散化为整数值的过程叫量化。量化时所确定的离散取值个数称为量化级数，表示量化的亮度值（或色彩值）所需的二进制位数称为量化字长，也称图像深度。图6-11所采用的量化级数为16，量化深度为4。量化字长越长，就越能真实反映图像的原有效果。
3. 编码：把离散的像素矩阵按一定方式编成二进制编码组，所得到的图像数据按某种图像格式记录在图像文件中称为图像的编码。

b. 数字图像处理技术

- ▶ 影响图像质量的两个重要参数就是图像分辨率和颜色深度，图像分辨率越高，颜色深度越深，则数字化后的图像效果就越逼真，图像数据量就越大。对于一幅图像，其分辨率为 $M \times N$ ，其颜色深度为 D ，则图像的数据量可利用以下公式来计算：
- ▶ 图像数据量 $=M \times N \times D / 8$ (Byte)
- ▶ 例如，一幅 1024×768 的32位彩色图像，其文件大小的计算过程如下：
 $1024 \times 768 \times 32 / 8 = 3145728\text{B} = 3\text{MB}$ 。

b. 数字图像处理技术



JPEG和JPEG 2000图像压缩标准

- ▶ JPEG (Joint Photographic Experts Group, 联合图像专家小组) 是从事静态图像压缩标准制定的委员会。
- ▶ JPEG允许用户根据自己的需要选择不同的压缩比, 是一种很灵活的文件格式, 一般选择的压缩比率在10: 1到40: 1之间, 压缩比越大, 品质就越低; 相反地, 压缩比越小, 品质就越好。

压缩比	质 量
3%--6%	中--好, 满足某些应用
6%--10%	好--很好, 满足多数应用
10%--20%	极好, 满足绝大多数应用
20%--25%	与原始图像几乎分辨不出来

JPEG和JPEG 2000图像压缩标准

- ▶ JPEG 2000与JPEG最大的不同：
 - ▶ 以小波转换(Wavelet transform) 为主的多解析编码方式代替以离散余弦转换为主的区块编码方式
 - ▶ 小波转换的主要目的是要将图像的频率成分抽取出来。
- ▶ JPEG2000相对于JPEG：
 - ▶ 高压缩率和低误码率，其压缩率比 JPEG 高约 30% 左右。
 - ▶ 同时支持有损和无损压缩，而 JPEG 只能支持有损压缩，因此更适合保存重要图片。
 - ▶ 能实现渐进传输（如同GIF格式图像的“渐现”特性）。JPEG2000在传输过程中，先传输图像的轮廓，然后逐步传输数据，不断提高图像质量，让图像由朦胧到清晰显示，而不是JPEG的由上到下的显示模式。
 - ▶ 支持“感兴趣区域”特性，用户可以任意指定图像上感兴趣区域的压缩质量，还可以选择指定部分先解压缩。

b. 数字图像处理技术

- ▶ 图像识别技术是数字图像处理技术的一个重要组成部分。
- ▶ 图像识别是运用模式识别的原理对图像对象进行分类的学问。
- ▶ 图像识别问题简单过程为：
 - ▶ 首先，对图像进行特殊的预处理，
 - ▶ 再经过分割和描述提取图像中有效的特征，
 - ▶ 进而加以判决分类。
- ▶ 目前，图像识别技术广泛应用于工业、交通、军事等领域，而且也为人机交互、身份鉴别、数字媒体信息检索和管理等提供了更人性化、更安全、更有效地手段，比如汉字识别、手写输入、指纹识别、人脸识别以及图像检索等。

生物识别技术

- 指通过人类生物特征进行身份认证的一种技术，这里的生物特征通常具有唯一的（与他人不同）、可以测量或可自动识别和验证、遗传性或终身不变等特点。

生物特征	普遍性	独特性	稳定性	可采集性	性能	接受程度	防欺骗性
人脸	高	低	中	高	低	高	低
指纹	中	高	高	中	高	中	高
手型	中	中	中	高	中	中	中
虹膜	高	高	高	中	高	低	高
视网膜	高	高	中	低	高	低	高
签名	低	低	低	高	低	高	低
声音	中	低	低	中	低	高	低

c. 数字视频处理技术

- ▶ 视频 (Video) 是多幅静止图像 (图像帧) 与连续的音频信息在时间轴上同步运动的混合媒体, 多帧图像随时间变化而产生运动感, 视频也被称为运动图像, 所以视频的处理技术与图像处理技术有很多类似之处。
- ▶ 由于视频具有相邻图像间相似度高、数据量特别巨大等特点, 其处理技术又与图像处理技术有所差别。
- ▶ 视频的播放速度一般为25~30帧/s, 即每秒种播放25~30张图像。

3. 数字视频处理技术

- ▶ 视频可以看做是二维信号的图像再增加时间轴而形成的三维信号的图像，所以视频的数字化是一个跟时间有关的连续过程。
- ▶ 视频的数字化过程同样经历了采样、量化和编码。
- ▶ 由于视频数据特别巨大，视频的数字化过程和视频压缩往往是同步进行的。

3. 数字视频处理技术

- ▶ 例如，一幅全屏的、分辨率为 640×480 像素的256色图像需要有307200字节的存储容量，这样一秒钟（30帧）的数字化视频如不压缩需要大约9M的存储空间，两个小时的电影所需的存储空间将超过66G。

视频数字化



- ▶ 数码摄像机工作原理为：通过感光元件将外部景物的光信号转变成电流，再将模拟电信号转变成数字信号，由专门的芯片进行处理和过滤后，存放于摄像机的数字存储介质中。
- ▶ 视频采集卡工作流程：一端连接录像机、摄像机和其他视频信号源，接收来自视频输入端的模拟视频信号，对该信号进行采集、量化成数字信号，然后压缩编码成数字视频序列，存放入硬盘中。

3. 数字视频处理技术

► 视频的压缩不仅体现在空间方向上的对图像的压缩，也体现在时间轴上对相邻图像间的压缩，所以，视频的压缩往往是综合采用多种压缩技术的组合。

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

MPEG视频压缩编码技术

- ▶ MPEG (Moving Pictures Experts Group/Motin Pictures Experts Group, 动态图像专家组) 跟JPEG一样, 也是一个国际标准组织(ISO)下的委员会, 专门从事音频和视频压缩编码的制定。
- ▶ 已经制定的标准主要有以下五个: MPEG-1、MPEG-2、MPEG-4、MPEG-7及MPEG-21, 其中涉及压缩编码的是前3个标准。
- ▶ MPEG标准的视频压缩编码技术主要利用了具有运动补偿的帧间压缩编码技术以减小时间冗余度, 利用DCT技术以减小图像的空间冗余度, 利用熵编码减小在信息表示方面的统计冗余度。通过这几种技术的综合运用, 大大增强了压缩性能。

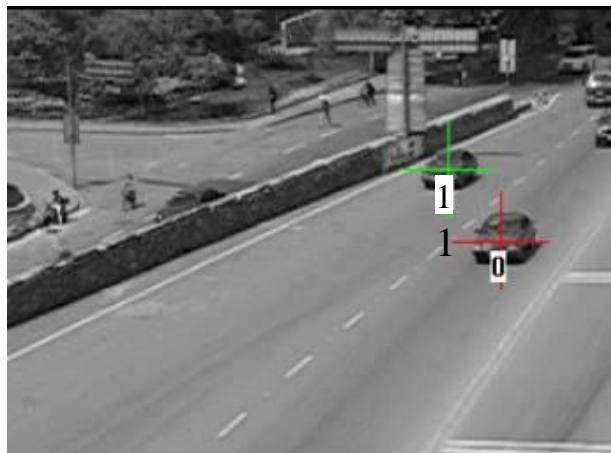
3. 数字视频处理技术

- ▶ 视频移动目标识别和跟踪技术是目前视频特征提取和识别技术研究的热点。
- ▶ 研究目的是使计算机能够模拟人类视觉运动感知功能，并赋予机器辨识序列图像中运动目标的能力，为视频分析和理解提供重要的数据依据。
- ▶ 目前已广泛应用与交通车辆识别和跟踪、人脸识别和跟踪、导弹识别与跟踪等领域。

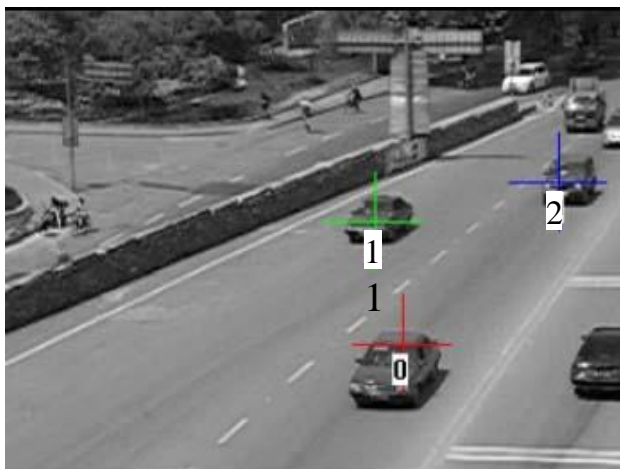
车辆识别与跟踪



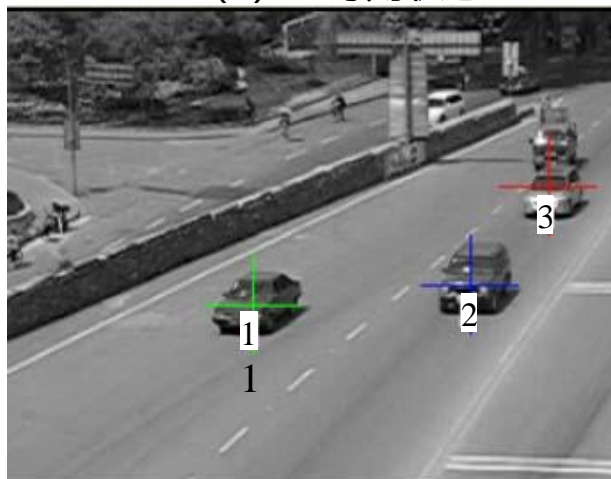
(a) t_1 时刻状态



(b) t_2 时刻状态



(c) t_3 时刻状态



(d) t_4 时刻状态

视频处理demo

- ▶ 人群状态估计1
- ▶ 人群状态估计2
- ▶ Kinect
- ▶ 视频摘要生成

d. 计算机合成音乐

- ▶ MIDI (musical instrument digital interface, 乐器数字接口) 是数字音乐国际标准。
- ▶ MIDI是乐器和计算机使用的标准语言，是一套指令（即命令的约定），它指示乐器（即MIDI设备）要做什么、怎么做，如奏出音符、加大音量、生成音效等。这里强调一下，MIDI不是声音信号，而是发给MIDI设备或其他装置让它产生声音或执行某个动作的指令。
- ▶ MIDI声音将电子乐器键盘的弹奏信息记录下来，包括键名、力度、时值长短等，这些信息称为MIDI信息。

d. 计算机合成音乐

- ▶ MIDI文件相对于普通声音文件有两大优点：
 - ▶ 一是所需存储容量小，例如CD-DA格式的波形声音，如播放1小时的立体声音乐，需要600MB的存储容量，而播放同时时间的MIDI音乐仅需要400KB左右的存储容量，两者相差1000倍以上；
 - ▶ 二是编辑修改十分灵活，例如可任意修改曲子的速度、音调，也可改换不同的乐器等。

计算机上MIDI的产生过程

- MIDI是将电子乐器的弹奏过程记录下来，当需要播放某首曲目时，根据记录的乐谱指令，通过音乐合成器生成音乐声波，经放大后由扬声器播出。

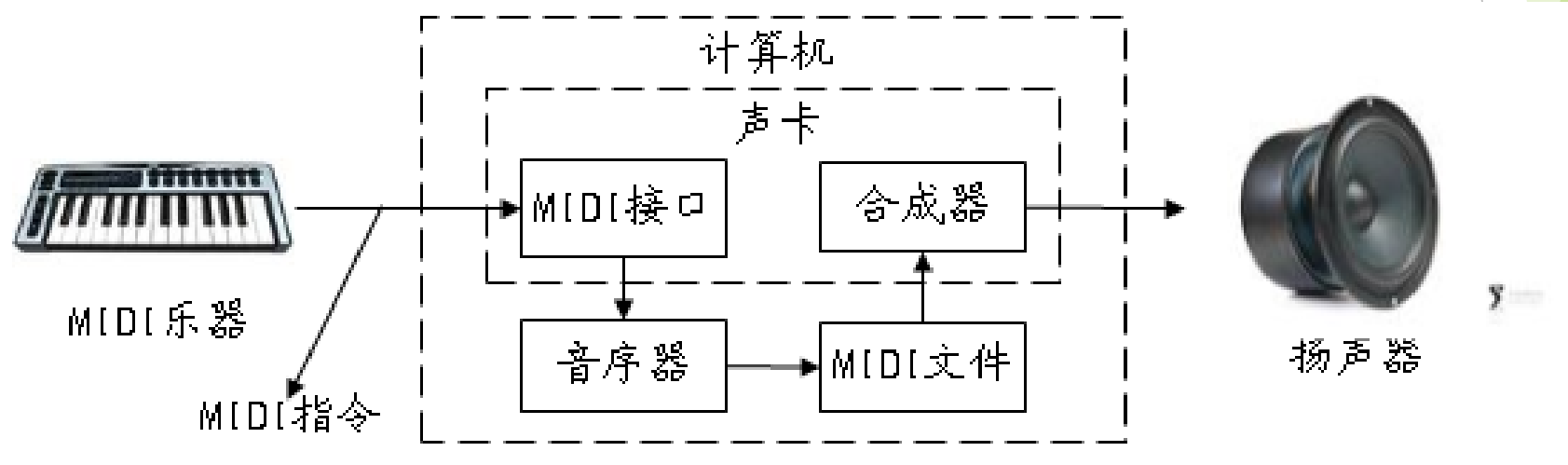


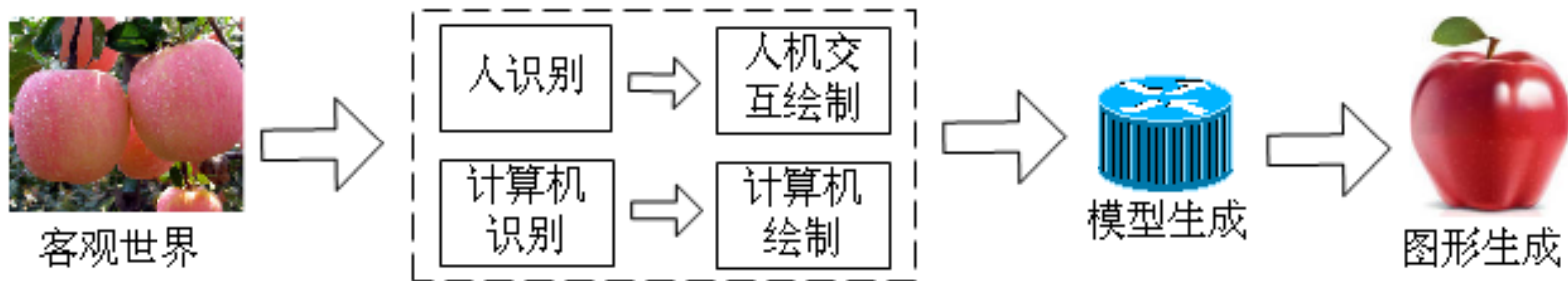
图6-21 MIDI音乐的产生过程示意图

c. 图形处理技术

- ▶ 构成图形的要素有几何要素（刻画对象的轮廓形状的点、线、面、体等）和非几何要素（刻画对象表面属性或材质的颜色、灰度等）。
- ▶ 图形可用形状参数和属性参数来表示，即参数表示法；也可用带有灰度或色彩的点阵图来表示，简称为像素图、图像（数字图像），即点阵表示法。

c. 图形处理技术

- ▶ 图形是计算机图形学(Computer Graphics, 简称CG)研究的对象。
- ▶ ISO给计算机图形学的定义为：研究用计算机进行数据和图形之间相互转换的方法和技术。



计算机图形生成模型图

案例6-16 真实感绘制技术和非真实感绘制技术

- ▶ 真实感图形绘制是借助数学、物理、计算机等学科的知识使用计算机生成三维场景的真实逼真图形、图像的过程。
- ▶ 非真实感绘制是利用计算机生成不具有照片般真实感,而具有手绘风格的图形的技术,其目标不在于图形的真实性,而主要在于表现图形的艺术特质、模拟艺术作品(甚至包括作品中的缺陷)或作为真实感图形的有效补充。



e. 动画处理技术

- ▶ 计算机动画是采用连续播放静态图像的方法产生景物运动的效果，也就是使用计算机产生图形、图像运动的技术。
- ▶ 一般而言，计算机动画中的运动包括景物位置、方向、大小和形状的变换，虚拟摄像机的运动、景物表面纹理、色彩的变化。
- ▶ 根据视觉空间的不同，计算机动画可分为：
 - ▶ 二维动画，也称平面动画，它的每帧画面是平面的展示动画内容。
 - ▶ 三维动画，也称立体动画，它包含了组成物体模型完整的三维信息，根据物体的三维信息在计算机内生成影像的模型、轨迹、动作等，可以从各个角度表现角色，具有真实的立体感。

6. 动画处理技术

硬件平台	主机	由于动画系统对数据运算与图形图像处理能力要求较高，所以主机一般为图形工作站。	
	输入输出设备	2D/3D鼠标	2D鼠标就是只能进行二维平面操作的鼠标，平时用的鼠标基本属于此类。3D鼠标就是指可以仅凭鼠标就完成X、Y、Z三个空间轴上进行像素移动的鼠标。
		图形输入板	提供类似于传统绘画模式的专业图形输入设备
		图形扫描仪	提供纹理、贴图等素材的输入
		三维扫描仪	提供三维物体的扫描输入
		刻录机、编辑录像机	动画视频输出设备
软件平台	系统软件	包括操作系统、高级语言、诊断程序、开发工具、网络通信软件等。	
	应用软件	图形设计	Photoshop、Illustrator等。
		2D	Animator Studio、Flash等。
		3D	3DS MAX、Maya等。
		特效与合成	Combustion、Maya Fusion、Shake等。

运动捕捉技术在人体动画中的应用

- 完整的运动捕捉系统一般由以下几个部分组成:传感器、信号捕捉设备、数据传输设备、数据处理设备。



(a) 人脸表情捕捉



(b) 舞蹈动作捕捉

图6-24 运动捕捉系统演示图

3.3.4. 数字传播技术

- ▶ 一种在线媒体传输技术，它运用可变带宽技术，以“流”（Stream）的形式进行数字媒体的传送，使人们可以一边下载一边观看高质量音频和视频节目的技术。
- ▶ 较好的解决了尽力而为的互联网络不能保证提供数字媒体信息业务的QoS和文件下载时间过长的的问题。
- ▶ 流媒体系统要比下载播放系统复杂得多。

3.3.5. 数字信息管理

- ▶ 信息管理是离不开数据库技术的。
- ▶ 早期数据多为文本、数字数据，采用关系数据库技术能够将数据以统一的格式存放在数据库中，非常方便查找和管理。
- ▶ 随着多媒体信息的急剧增长，图形、图像、声音、视频等类型的数据信息属于非格式化数据，利用关系数据库技术已经无法进行有效的管理了，这就需要一种更有效的方式进行存储和管理，多媒体数据库技术就应运而生了。

1. 数字媒体数据库技术

► 扩充关系数据库的方式

- 一种是关系数据库中记录多媒体文件的存放路径信息，而多媒体文件则通过文件的方式进行管理；
- 另外一种是在关系数据库中引入新的数据类型用以描述和管理多媒体数据，最常见的是引入一种大二进制对象的数据类型。

► 面向对象数据库的方式。

- 采用对象、方法、属性、消息、类的层次结构和继承特点描述多媒体数据模型。
- 优点：更好地处理复杂的多媒体对象的结构语义；便于支持新的数据类型及其操作；采用的方法使多媒体对象是独立的，且良好地封装了多媒体数据之间在于系统模式方面的巨大差异，并且很容易实现并发处理，也便于系统模式的扩充和修改。
- 是主要方向，但实现难度很大。

2. 数字媒体信息检索技术

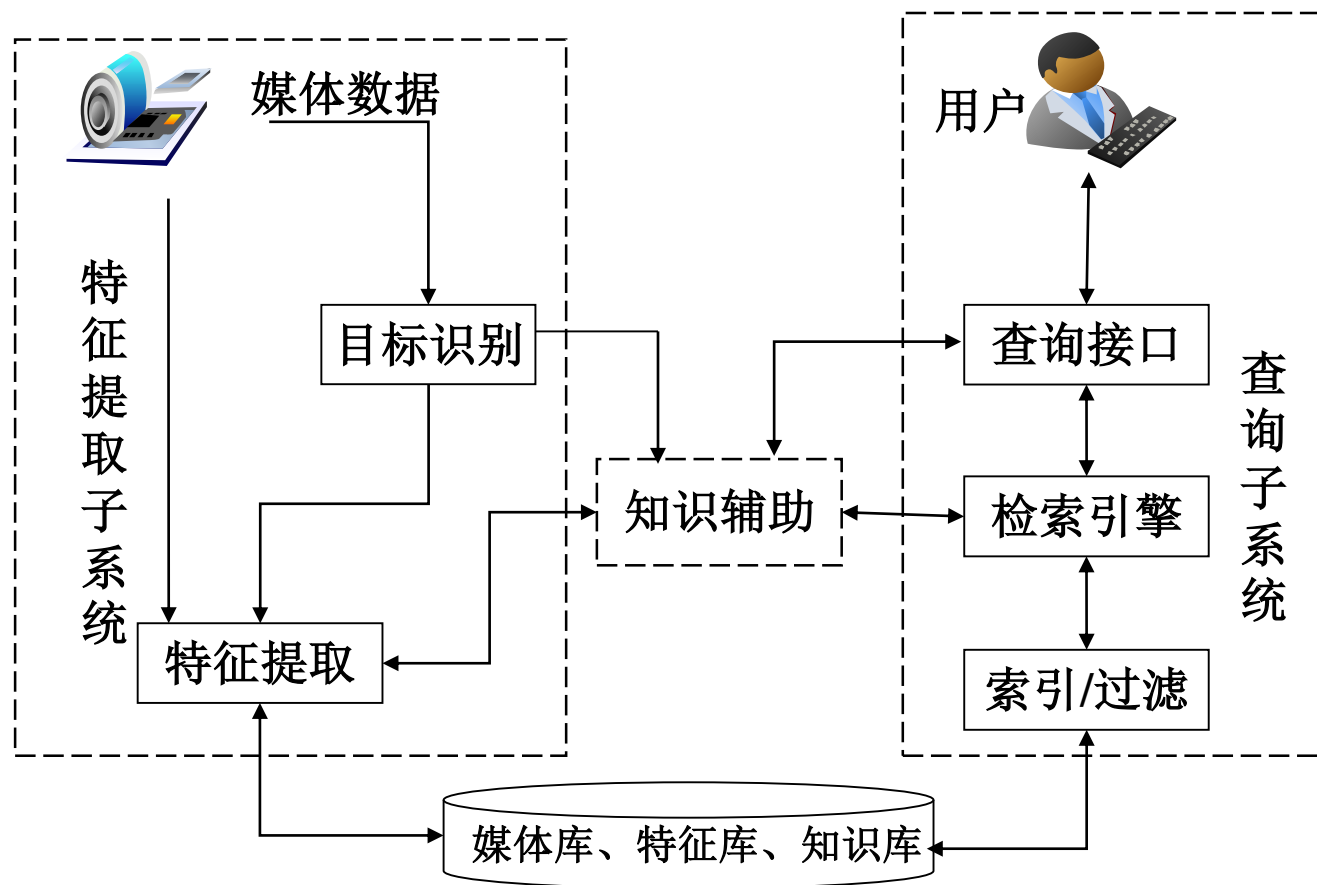
► 基于文本的检索方法

- 利用文本描述的方式来描述数字媒体的特征
- 缺点：无法解释和表达数字媒体信息的实质内容和语义关系，并带有很大的主观性。

► 基于内容的检索

- 就是根据媒体信息的内容来检索，包含信息内容和检索两方面内容
- 直接对图像、视频、音频内容进行分析，抽取特征和语义，利用这些内容特征建立索引并进行检索
- 这些内容特征主要包括图像中的颜色、纹理、形状、相对位置，视频中的镜头、场景、镜头的运动，声音中的音调、响度、音色等。
- 数字媒体信息检索技术的发展方向。

2. 数字媒体信息检索技术



基于内容检索的体系结构

基于内容的图像检索系统

- ▶ 基于内容的图像检索（Content Based Image Retrieval, CBIR）是从图像数据库中找出与检索内容相似的图像的检索技术。
- ▶ 利用从图像中自动抽取出来的各种特征，进行计算和比较，检索出符合用户需求的结果图像集。
- ▶ 图像的特征分为底层特征、中层特征和高层特征，其中底层特征有颜色、纹理、轮廓和形状等，中层特征有图像中的对象、图像背景、不同对象间的空间关系等，高层特征为语义特征，主要有场景、事件、情感等。

基于内容的图像检索系统

Selected sample queries for all classes:

Please click on an image to start a query.



Query Image



Weights: Perceptual Grouping = 0.2, Color = 0.4, Texture = 0.4, L, A, B channels.

Retrieved Images



(a) 示例图片输入 (b) 选择第一张示例图片检索出的部分结果
基于示例图片输入的图像检索图

基于内容的图像检索系统

Manmade: Buildings

Please select one of the following images:

man_bld_sony_D6_mvc-020f.jpg

View Image

Submit Query

Reset Form



Weights (should sum to 1):

Perceptual grouping: 0.33 Color: 0.33 Texture: 0.33

Texture:

☒ L, A and B channels

☐ L channel only (~Grayscale texture)

(a) 参数设定

Query Image



Relevance feedback type: Cluster. Weights: Perceptual Grouping = 0.33, Color = 0.33, Texture = 0.33
For relevance feedback, please select the check boxes below each image, and then select the feedback

Retrieved Images



☐ Yes ☐ NS ☒ No ☐ Yes ☐ NS ☒ No ☐ Yes ☐ NS ☒ No

(b) 检索结果

人工参与下的参数设定法检索图

数字信息安全

- ▶ 随着网络和无线通信的普及，许多传统媒体内容都向数字化转变，如MP3的网上销售，数字影院的大力推行，网上图片、电子书籍销售等等，在无线领域，随着移动网络由第二代到第三代的演变，移动用户将能方便快速的访问因特网上的数字媒体内容。但是数字媒体内容的安全问题成了一个瓶颈问题，主要包括
 - ▶ 安全传递
 - ▶ 访问控制
 - ▶ 版权保护

3. 数字媒体技术的内涵

► 3.4 数字媒体技术的研究领域

(1) 数字声音处理

包括音频及其传统技术（记录、编辑技术）、音频的数字化技术（采样、量化、编码）、数字音频的编辑技术、话音编码技术（如PCM、DA、ADM）。数字音频技术可应用于个人娱乐、专业制作、数字广播等。

(2) 数字图像处理

包括数字图像的计算机表示方法（位图、矢量图等）、数字图像的获取技术、图像的编辑与创意设计。常用的图像处理软件有Photoshop等。数字图像处理技术可应用于家庭娱乐、数字排版、工业设计、企业徽标设计、漫画创作、动画原形设计、数字绘画创作。

3. 数字媒体技术的内涵

► 3.4 数字媒体技术的研究领域

(1) 数字声音处理

包括音频及其传统技术（记录、编辑技术）、音频的数字化技术（采样、量化、编码）、数字音频的编辑技术、话音编码技术（如PCM、DA、ADM）。数字音频技术可应用于个人娱乐、专业制作、数字广播等。

(2) 数字图像处理

包括数字图像的计算机表示方法（位图、矢量图等）、数字图像的获取技术、图像的编辑与创意设计。常用的图像处理软件有Photoshop等。数字图像处理技术可应用于家庭娱乐、数字排版、工业设计、企业徽标设计、漫画创作、动画原形设计、数字绘画创作。

(3) 数字视频处理

包括数字视频及其基本编辑技术、后期特效处理技术。常用的视频处理软件有Premiere等。数字视频处理技术可应用于个人、家庭影像记录、电视节目制作、网络新闻。

(4) 数字动画设计

包括动画的基本原理、动画设计基础（包括环节：构思、剧本、情节链图片、模板与角色、背景、配乐）、数字二维动画技术、数字三维动画技术、数字动画的设计与创意。常用的动画设计软件有3DMAX、Flash等。数字动画可应用于少儿电视节目制作、动画电影制作、电视节目后期特效包装、建筑和装潢设计、工业计算机辅助设计、教学课件制作等。

(5) 数字媒体压缩

包括数字媒体压缩技术及分类、通用的数据压缩技术（行程编码、字典编码、熵编码等）、数字媒体压缩标准，如用于声音的MP3、MP4、用于图像的JPEG、用于运动图像的MPEG。

(6) 数字媒体存储

包括内存存储器、外存储器、光盘存储器等。

(7) 数字媒体管理与保护

包括数字媒体的数据管理、媒体存储模型及应用、数字媒体版权保护概念及框架、数字版权保护技术，如加密技术、数字水印技术、权利描述语言等。

(8) 数字媒体传输技术

包括流媒体传输技术、P2P技术、IPTV技术等。等等。