**一，多媒体的基本概念**

媒体：承载信息之载体

分类：感觉、表示、显示、存储、传输

感觉媒体：用户接触信息的感觉形式|视觉、听觉、触觉等

表示媒体：信息的表示形式 |文字、图像、声音、视频、动画等

显示媒体：表现和获取信息的物理设备 |显示器、打印机、扬声器、键盘等

存储媒体：存储数据的物理设备 |磁带、磁盘、光盘等

传输媒体：传输数据的物理设备|光缆、电缆、电磁波、交换设备等

多媒体是融合两种或者两种以上媒体的一种人机 交互式信息交流和传播媒体，使用的媒体包括数字、文 字、图形、图像、声音、动画和视频等。

三个重要特性

媒体多样性

人机交互性

系统集成性

媒体多样性

文字（书籍/报纸/广告/短信）

声音（电话/磁带/声讯服务/语音短信）

静止图像（画像/照片/传真/ 手机彩信）

连续图像（电影/录像带/电视/电脑动画）

视觉（65%）：文字/图像/视频

听觉（20%）：声音

触觉（10%）：盲文/键盘/鼠标/游戏手柄

味觉（3%），嗅觉（2%）尚未用计算机进行处理

|  |  |  |  |
| --- | --- | --- | --- |
| 计算机发展的初期 | 1950’s~1970’s | 1980’s开始 | 2007年至今 |
| 用数值这种媒体承载信息 | 用文字作 为信息的载体 | 将声音、图形和图像作为新 的信息媒体 |  |
| 通过“0”和“1”表示信息 | 用文字（如英文）编程，输入计算机，处理结果也可用文字表示输 出 | 首次引进“位映射”的图 形机理，用户接口开始使用Mouse驱动的窗口技术和图符 | 集通话、拍照、录像、游戏、定位、网页浏览、支付、NPC等等于 一身 |
| 纸带机和卡片机是主要的I/O设备 | 打字机、键盘和显示终端 | 计算机的应用更为直 观、容易 | 智能手机广泛使用 |
| 机器语言时代 | 高级程序设计语言 |  |  |
| 极少数计算机专业人员 | 具有一般文化程度的科技人员 | 文化水平较低的公众，包括儿童在内 |  |

人机交互性

1.信息交流方式的巨变

2.交互性

传统媒体（电视、广播）：被动接受信息

多媒体：是一种人－机交互式媒体，向用户提供交互式 的使用、加工和控制信息的手段

简单的交互式应用：如视频点播（VOD）、从数据库中检 索出某人的照片等

高级的交互式应用：虚拟现实&增强现实（VR AR）

系统集成性

集成性

1.多种信息媒体的集成

2.处理这些媒体的设备与设施的集成

传统的大众传播媒体，是一对多

数字媒体传播的理想信道是具有足够带宽的、可以传输比 特流的高速网络信道。网络可能由电话线、光缆或卫星通 信构成。

数字媒体：多点之间的传播。

在数字多媒体传播模式中，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。

**二，媒体的种类**

视觉类媒体(70%-80%)

听觉类媒体(10%)

触觉、嗅觉和味觉类媒体(10%)

视觉是人类感知信息最重要的途径

文本(Text)：

最早的计算机媒体信息|英文文本，中文文本，WPS (.wps)、Word (.doc)、Notepad (.txt) …非格式化文本文件、格式化文本文件

图像(Image)、

一幅图像可定义为一个二维函数f(x,y)，(x,y)为空间坐标，f(x,y)为图像在点(x,y)处的强度；

数字图像：当x, y和f都是有限的、离散的

像素：构成数字图像的基本元素

图形(Graphic)、

图形一般是指计算机绘制的画面，如直线、圆、圆弧、矩 形、任意曲线和图表等

不直接描述数据的每一点，而是描述产生这些点的过程和 方法。

line (x1, y1, x2, y2, color) circle (x, y, r, color)

图形图像的区别

1.存储结构和 表示方法完全不同

2.图形是矢量结构的画面存储形式，抽象，数据量小， 但显示成本高

图像是栅格结构的画面形式，基本元素是像素，逼 真，数据量大

3.图形是更加抽象化的图像

视频(Video)、动画(Animation)

一幅幅有联系的静态图像的连续播放，利用了人眼 的视觉暂留性，即每秒连续播放20-30帧(frame)，产 生运动画面的效果

视频：单帧是采集的真实图像

动画：单帧是由计算机产生或人工画出的图像或 图形，动画包括：二维动画、三维动画、真实感 三维动画等

听觉类媒体

|  |  |  |
| --- | --- | --- |
| 波形声音（Wave） | 语音（Speech） | 音乐（Music） |
| 20~20kHz， | 300~3.4kHz， |  |
| 波形声音是对自然界声音进行采样和量化的结 果，是自然界所有声音的拷贝 | 语音（或称话音）是指人的说话声，包含语意和情感等信息 | 符号化了的声音，多媒体中专指 MIDI 音乐 |

多媒体技术主要的研究方向

1.数字多媒体表示与操作

2.数字多媒体压缩

3.数字多媒体存储与管理

4.数字多媒体传输

5.数字声音处理

6.数字图像处理

7.数字视频处理

8.数字动画设计

9.数字游戏设计

10.数字媒体压缩

11.数字媒体存储

12.数字媒体管理与保护

13.数字媒体传输技术

多媒体应用领域

1.教育培训

2.电子商务

3.信息发布

4.个人娱乐

5.电子出版

6.创意设计

7.虚拟现实

**三，数字音频处理技术**

记录声音：留声机，磁性录音，数字音频（光碟），

声音的基本特性

(声音的产生、频率、带宽、响度)

1.声音是通过空气传播的一种连续的波，叫声波

2.音频信号是时间依赖的连续媒体

3.模拟人耳的两个通道

4.语音处理包含情感和意向，涉及语言学、社会学和声学

声音的产生

1.声源：正在发出声音的振动物体

2.声波：物体振动或气流扰动而引起周围的空气或其它弹性介质发生 波动的现象

3.声场：声波所波及的空间范围

4.声音产生的三个条件：存在声源并振动，传播介质，听觉感受

声音的频率

|  |  |  |  |
| --- | --- | --- | --- |
| 人耳 | 亚音信号  （次音信号） | 音频信号 | 超音频信号（超声波） |
| 20～20000 Hz | 小于20 Hz | 20 Hz～20 kHz | 高于20 kHz |

人的发音器官：80～3400 Hz

人说话的信号频率：300～3000 Hz（话音信号）

频率响应曲线：用频率响应曲线表示音响特性

人体的固有频率

水平方向的固有频率约为3—6赫兹,竖直方向的固有频率约为48赫兹

工操作的各类振动机械的频率必须大于20赫兹。

次声波枪和次声波炸弹利用频率为16—17赫兹的次声波，与人体内的某些器官发生共振，

声音的带宽

复合信号：由许多频率不同的信号组成声音信号

分量信号：单一频率的信号

带宽是声音信号的一个重要参数，它用来描述组成复合信号的频率范围。

高保真音信号(high-fidelity audio)的频率范围为10 Hz～20000 Hz，它的带宽约为20 kHz（20000-10）

声音的响度:

声音的响度就是声音的强弱

听阈：当声音弱到人的耳朵刚刚可以听见时的声音强度

痛阈：声音强到使人耳感到疼痛时的声音强度。

(听阈和痛阈都是和频率相关的)

用分贝（dB）来描述声音的响度/强度



将某个声音的声压定义为“标准值”0分贝，任何一个声音的声压和标准值相除，然后取对数，得到该声音的分贝

15-60-100



模拟音频和数字音频

模拟磁性录音技术:

话筒模拟电压的幅度来表示声音强弱

直接记录音频信号的波形，唱针扫描槽纹或者用放音磁头

受电磁性能的影响较大

动态范围可达80dB

模拟信号:间和幅度上都是连续的信号.音频信号是模拟信号

数字音频

通过模数转换器（A/D Converter）将声波波形转换成一连串的二进制数据来再现原始声音

音频的数字化

样本:以每秒上万次的速率对声波进行采样，每一次采样都记录下原始模拟声波在某一时刻的状态，

样本精度/采样精度:每个样本分配一定的存储位（bit）来表达声波的振幅状态

采样：每隔一段时间就记录一次声音信号的幅度

量化：把采样得到的声音信号幅度转换为数字值

时间上的离散叫采样，幅度上的离散叫量化

采样频率:每秒钟所抽取声波幅度样本的次数,单位是kHz | 采样频率越高声音失真越小，但用于存储音频的数据量也越大。

奈奎斯特采样定律:采样频率不应低于声音信号最高频率的两倍

声音信号可以看成由许许多多正 弦波组成的，一个振幅为A、频率为f的正弦波至少需要两个采样 样本表示，因此，如果一个信号中的最高频率为Fm，采样频率 最低要选择2\*Fm。

Eg：电话话音的信号频率约为3.4 kHz，采样频率就选为8 kHz。

模拟音频中，声音的幅度通常用电压表示；

而在数字音频中， 电压通常用数字表示。

例如，0.5V电压用数字20表示

量化：把某一个幅度范围内的电压用一个数字表示。先将整个幅度划分成为有限个小幅度（量化阶 距），把落入某个阶距内的样值归为一类，并赋予相同的量化 值。

量化位数（Bit Depth）决定了每个采样点用多少二进制位来表示，量化位数 越多，音质越好，但数据量也越大。

音频数字化化过程：模拟信息 ——> 采样 ——> 量化 ——> 编码 ——> 数字信号

样本精度：样本精度是每个采样点能够表示的数据范围，常用的有8位、 12位和16位。8位量化级表示每个采样点可以表示256个（ 0~255）不同量化值。（2的8次方）

单声道：记录声音时，如果每次生成一个声波数据

双声道（立体声）：每次生成两个声波数据

数字音频文件大小的计算

数据量 （Byte） = 采样频率Hz ×（采样位数/8） ×声道数 ×时间s

声波信号->ADC（用于音频数字化）->声卡->操作系统->声音处理程序->内存/硬盘

声卡的功能主要包括以下几个方面

1.音频录放、编辑

2.音乐合成

3.文语转换

4. CD-ROM接口

5.MIDI接口

6.游戏接口

常见的音频文件

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CD | .wav | .mid  .rmi | .mp3 | .ra  .rm  .rmvb  .wma |
| 业界标准 | 用于存储 | 用于合成音乐 | 着重于压缩 | 用于网络传输 |
| 一种数字化声音，以16位量化级、44.1kHz采样频率的立体声存储，可以完全重现原始声音，每片CD唱片能记录约74分钟这种质量的音乐节目。 | 声音波形的直接记录。波形文件/波形音频，媒体计算机获得声音最直接、最简便的方式。模/数转换），然后以适当的格式存在硬盘上。 | 纯粹人造声音。MIDI文件存储的不是声音信号，而是各种乐器的发音命令，播放时系统根据这些命令合成乐曲。 | 存储压缩后的波形数据 |  |
| CD-ROM驱动器 | WAV是Windows所使用的标准数字音频文件。 | 非常小，可以满足长时间音乐的需要 |  |  |
|  | 缺点是产生的文件太用于控制音乐合成器，合成细节由合成器决定。大，不适合长时间记录。 | 于控制音乐合成器，合成细节由合成器决定。 |  |  |



电脑音乐：利用多媒体电脑及与音乐相关的软件为工具制作并播放出来的音乐。

示例：虚拟歌手初音未来

电脑音乐三个步骤：

编曲（确定旋律）-合成（不同的音乐片断合成等）-后期制作（加入人工录音等）

CakeWalk 、Vegas Audio、CoolEdit、AutoScore

声音的转换到其他媒体

示例：Siri

用声音表示各种信息：度告警装置用警示音表示温度超标

声音可视化：将声音进行小波转换后

模拟音频处理设备

|  |  |  |
| --- | --- | --- |
| 话筒 | 音箱 | 模拟调音台 |
| Microphone，麦克风 | Speaker，扬声器 |  |
| 声音的收集，完成声能向电能的转化 | 还原声音，将音频电流信号转化成声音信号 | 可以拾取多路信号，每路声音信号可单独进行处理。还可以进行各种声音的混合。拥有  多种输出。既能创作立体声、美化声音，又可抑制噪声、控制音量，是声音艺术处理必不可少的一种设备。 |

数字音频处理设备可以分为两类：

（1）专用数字音频设备

（2）非专为处理音频而设计的多媒体计算机

数字音频设备

|  |  |  |
| --- | --- | --- |
| 数字调音台 | 数字录音机 | 数字音频工作站 |
| 1.将每一路进行优化和调节  2.对多路声音进行混合输出。 | 采用数字记录方式来存储音频信号。用硬盘或者光盘记录方式 | 一台能够完成从录音、编辑、混合、压缩，一直到最后刻出母盘的全部音频节目制作过程的设备。  调音台、多轨录音机、编辑机、效果器等。 |
|  |  | 集成度高，免去录音连线的烦恼，且便于携带。 |

关键的硬件技术内核包括：

1）模数转换器（ADC）

2）数模转换器（DAC）

3）数字信号处理器（DSP）

如：模拟和产生声场，控制声效等。

数字音频编辑软件中几个关键的技术概念：

|  |  |  |
| --- | --- | --- |
| 声道：声道就是不同位置发出的声音 | 单声道 | 缺乏位置感 |
| 立体声 | 声音在录制过程中被分配到两个独立的声道，但所占空间比单声道多一倍。 |
| 四声道环绕 | 四个发生点：前左、前右、后左、后右，同时建议增加一个低音音箱， |
| 5.1声道 | 运用于各类传统影院和家庭影院中。4.1环绕+中置单元用于加强人声 |
| 7.1声道 | 5.1+中左和中右两个发声点 |
| 音轨：音轨就是在音频处理软件中看到的一条一条的平行“轨道” | 音序器软件中，一条音轨对应于音乐的一个声部或者对于一种乐器，它把MIDI 或者音频数据记录在特定的时间位置。 | 每一音轨对应一个原始音频素材文件或者前后对应多个音频文件。可以同时让多个音频素材同时播放，产生混音效果。 |
| 时序 | 所谓时序，其实也就是时间的顺序 | 人们在处理多个轨道多个音频素材时，这些素材的先后顺序如何去定义，这就是时序的思想。 |

**四，数字图像处理技术 I**

视觉的常识

1.颜色是视觉系统对可见光的感知结果

2.可见光波长：（紫外线）380 nm～780 nm（红外线）

3.我们看到的大多数光是由许多不同 波长的光组合成的。

可见光的度量参数：波长和强度

光谱的功率分布（spectral power distribution, SPD）可以 完全表征

280nm－298nm是使皮肤变红最主要的紫外线波段

图像：物体在平面坐标上的直观再现，是其所表示物体 的信息的直接描述和概况表示。可以用一个二元函数来表示 f(x, y)，x, y分别表示横坐标和纵坐标，函数f的值表示亮度（intensity）或 者灰度值（gray level）

数字图像：像素组成的二维或三维排列，可以用矩阵表示。x, y和函数f的值都是有限、离散的

单色（灰度）图像：每个像素的亮度用一个0-255的数值来表示（0黑255白）

彩色图像：用红、绿、蓝三元组的二维矩阵来表示（RGB）

二值图像：灰度值只有0和1两个的图像。0代表黑色。1代表白色。

索引图像：包括存放图像数据的二维矩阵，还包括一个颜色索引矩阵 （MAP）。

RGB彩色图像（真彩色图像）：由R、G、B来表示每个像素的颜色。颜色值直 接存放在矩阵中，无需索引。即 M\*N\*3

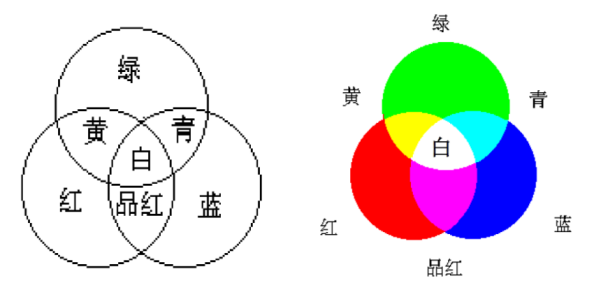
数字图像处理的目的

1.便于人类理解和分析。图像增强、图像复原、医学图像

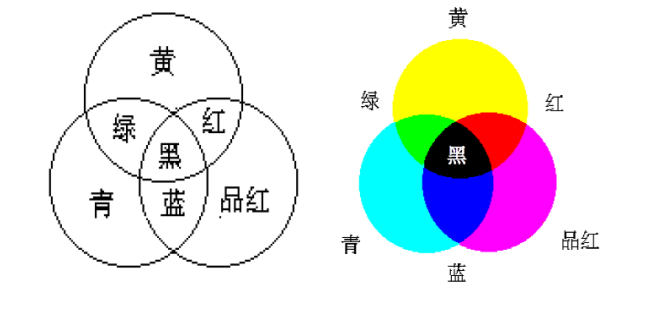
2.机器识别。自动字符识别、人脸识别、指纹识别，生物特征识别

颜色模型

RGB相加颜色模型：红色 + 绿色 + 蓝色



CMY相减颜色模型：青色(Cyan) + 品红(Magenta) + 黄色(Yellow)

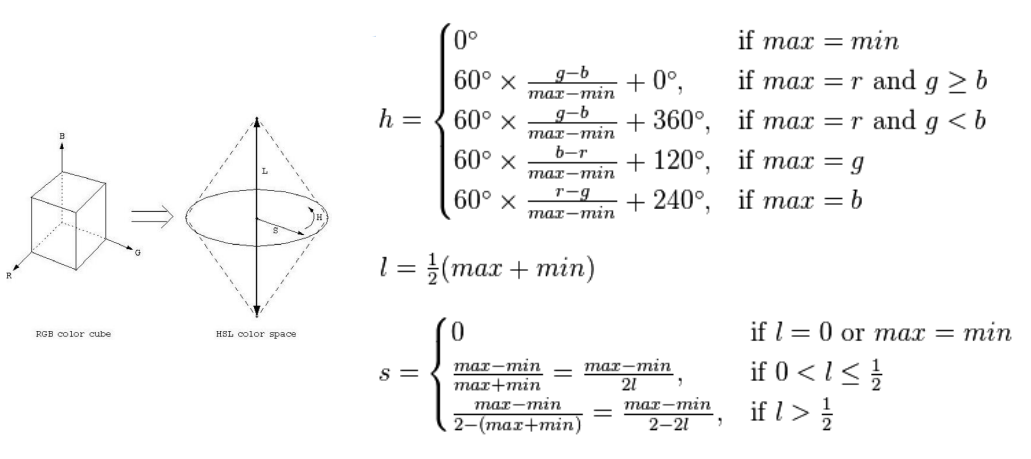


Rgb和CMY缺点：对一个像素的每一步操作都需要涉及所 有红、绿、蓝三个通道。效率较低

HSL颜色模型：色相 (hue)、饱和度 (saturation)、亮度 (luminance)

YUV颜色模型：Y表示亮度、U：色度，V：浓度

C = 255 - R M = 255 - G Y = 255 - B



真彩色 ：RGB每个基色分量直接决定显示设备的基色强度。24位色显示2^24=16800000种颜色

增强色：16位色显示2^16=65536种颜色

伪彩色：把像素值当作彩色查找表(color look-up table，CLUT)的 表项入口地址，去查找一个显示图像时使用的R，G，B强度值。8 位色显示2^8=256种颜色

为什么要使用伪彩色：65536色以下的显示方式中一般采用伪彩色。标准的调色板是在 256K色谱（即25万6千种色彩）中按色调均匀地选取16种或256种色 彩。 一般应用中，有的图像往往偏向于某一种或几种色调，此时如果采 用标准调色板，则色彩失真较多。因此，同一幅图像，采用不同的 调色板显示可能会出现不同的色彩效果。

矢量图：用一系列计算机指令来描述和记录 一幅图，这幅图可分解为一系列子图如点、线、面等。不受分辨率的影响。难以表现色 彩层次丰富的逼真图像效果。

位图：用像素点来描述或映射的图，也即位映 射图。位定义图像中每个像素点的颜色和亮度。缩放和旋转容易失真，同时文件容量较大。色彩丰富，逼真。

图形：指用计算机绘制工具绘制的画面，包括直线、曲线，圆/圆弧， 方框等成分。图形一般按各个成分的参数形式存储。

图像：由输入设备捕捉的实际场景或以数字化形式存储的任意画面。 图像可以用位图或矢量图形式存储。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 位图文件BMP | TIFF | GIF | JPEG | PNG |
| 最普及的图像文件格式，Windows采用的图像文件存储格 式 | 较为通用的图像文件格 式，在苹果公司的Mac计算机中广泛应用 | 数据是经过压缩的， | 连续色调静态图像的数字压缩和编码Digital  Compression and Coding of Continuous -tone Still Image）  。 | 可携式网络图像Portable Network Graphics |
| 特点是包含的图像信息较丰富，几乎不进行压缩，占用磁盘空间大 | 格式灵活易变，支持多种编码方法 | 支持256种色彩，可以存多幅彩色图像，可构成一种最简单的动画 | 采用JPEG压缩编码算法压缩的图像，其压缩比约为5:1至50:1，甚至更高。 | 新兴的网络图像格式，结合了GIF和JPEG格式的一些优点，允许连续读出和写入图像数据并使用无损压缩。 |
| 单机，PC机运行 Window时的墙纸 | 扫描仪 |  |  |  |

**五，数字图像处理技术 II**

数字图像质量很大程度上由 M, N, L决定number of samples (M, N)|gray levels (L)

缩放图像：

1. 创建新的像素位置

2. 灰度分配

像素空间

像素p的邻域

|  |  |  |
| --- | --- | --- |
| 4近邻（4-neighbors）：N4 (p) | 对角近邻（D-neighbors） ：ND (p) | 8近邻（8-neighbors） ：N8 (p) |
|  |  | N8 (p) = N4 (p) + ND (p) |

连通性：描述区域(region)和边界(boundary)

具有连通性的两个必要条件：

1.相邻

2.灰度级是否满足相似性准则（V是具有相似灰度的集合）

|  |  |  |
| --- | --- | --- |
| 4连通（4-adjacency） | 8连通（8-adjacency） | m连通（m-adjacency，混合连通） |
| p和q的像素值都属于集合V，q属于集合N4 (p) | p和q的像素值都属于集合V，q属于集合N8 (p) | q属于集合N4 (p)，或者q属于集合ND (p)，且N4 (p) ∩N4 (q)中没有像素值属于集 合V的像素 |
|  |  |  |

通路（path）：

一系列像素：(x0 , y0 ) (x1 , y1 ) ,…,(xn , yn )，其中(x0 , y0 )是像素p的坐标， (xn , yn )是像素q的坐标，像素(xi , yi ) 和像素(xi-1 , yi-1 ) 是 连通的，1<=i<=n。n称为通路的长度，如果(x0 , y0 )= (xn , yn )，则称通路是闭合的

距离度量：

对于像素 p, q 和 z，坐标分别为 (x, y)，(s, t) 和 (v, w)，如 果满足：

(1) D(p, q) >= 0 (D(p, q)=0 if p=q)

(2) D(p, q) = D(q, p)

(3) D(p, z) <= D(p, q) + D(q, z) 则称 D 是一个距离函数

|  |  |  |
| --- | --- | --- |
| 欧式距离 | D4距离  （city-block距离） | D8距离  （chessboard距离） |
|  |  |  |
|  |  |  |

一张图大小为400x500，共有多少相邻像素对？4近邻：399\*499\*2 + 399 + 499

几何失真（几何畸变）：图像中像素之间的空间关系会 发生变化

几何失真校正：

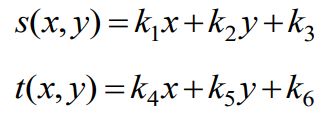
1.空间变换：对图像平面上的像素进行重新排列以恢复原空间 关系

2.灰度插值：对空间变换后的像素赋予相应的灰度值以恢复原 位置的灰度值

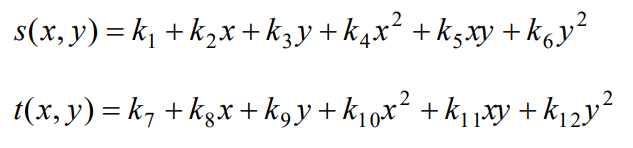
空间变换：

图像f (x, y)受几何形变的影响变成失真图像 g(x' , y' ) x' = s(x, y) y' = t(x, y)

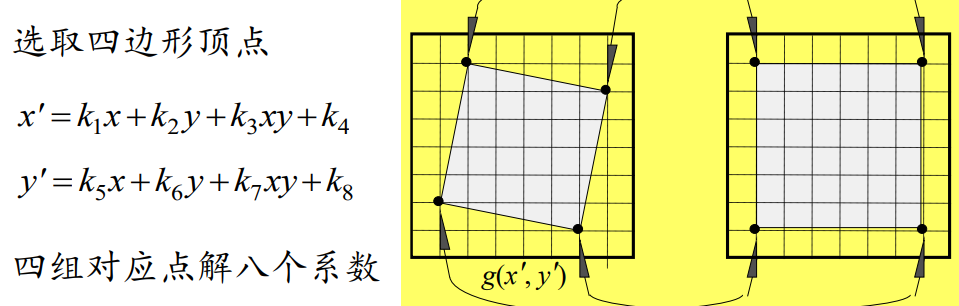
线性失真



（非线性）二次失真

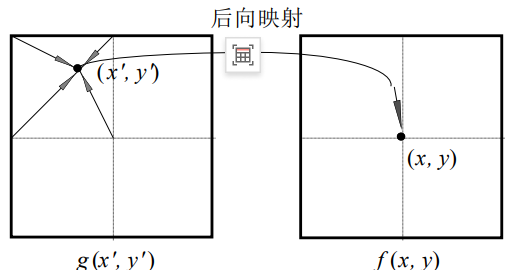


约束对应点方法：在输入图（失真图）和输出图（校正图）上找一些其位置 确切知道的点，然后利用这些点建立两幅图间其它点空间 位置的对应关系



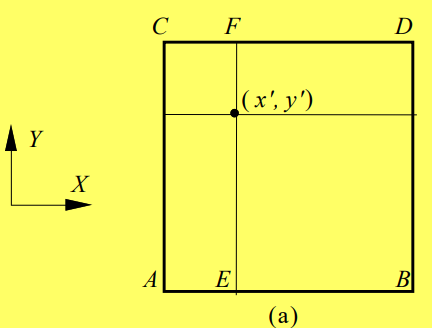
灰度插值、

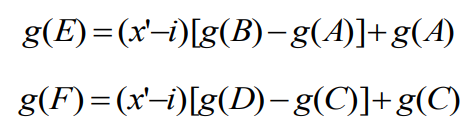
用整数处的像素值来计算在非整数处的像素值。(x, y)总是 整数，但(x’ , y’ )值可能不是整数



1.最近邻插值（零阶插值）：将离(x’ , y’ )点最近的像素的灰度值作为(x’ , y’ )点的灰度 值赋给原图(x, y)处像素

2.双线性插值

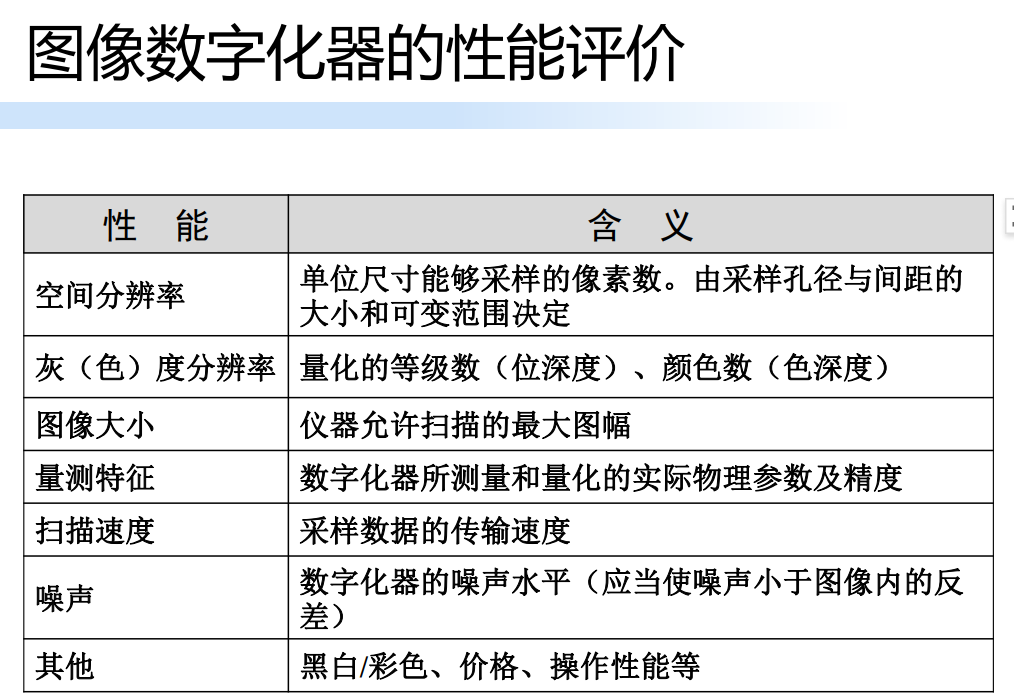






图像输入输出设备

|  |  |  |  |
| --- | --- | --- | --- |
| 图像输入设备 | | | 图像输出设备 |
| 数字化器是将模拟图像转换成数字图像的数字化输入装置 | | | 数字图像的显示是图像数字化的逆过程（D/A） |
| 数码电视摄像机 | CCD摄像机 | 数码相机  扫描仪 | 打印机、 胶片纪录仪、静电绘图仪等 |
| 光电转换器件主要有光电摄像管（vidicon）、光电倍增管、 视像管等。 | 体积小、重量轻、寿命长、成像速度快、 成本低、灵敏度高、图像无几何失真，已成为视频信号的 常用输入设备。 |  | 显示器是典型的暂时显示设备，而打印机等永久显示设备。 |
| 目前摄像管摄像机正逐渐被固体光电转换器件摄像机取代。 | CCD的数量可分为： 单片、三片式摄像机，三片式摄像机的质量最好。 |  |  |
|  | 灵敏度：最低环境照度  分辨率：成像后可以分辨的水平黑白线对的数目。  信噪比：主要取决于ADC的位数。 |  |  |



显示分辨率：显示屏上能够显示出的像素数目。

显示分辨率为640×480个显示屏就含有640\*480=307200个 显像点

图像分辨率：组成一幅图像的像素密度的度量方法

DIP：每英寸多少点

用300 DIP 来扫描一幅8″×10″的彩色图像，就得到一幅2400×3000 个像素的图像。

图像分辨率 是确定组成一幅图像的像素数目，而显示分辨率是确定显 示图像的区域大小。

电脑显示器参数

1.分辨率（640 ×480、1024 ×768）

2.刷新率（85Hz、110Hz）

3.带宽（200MHz、115MHz）

带宽= 最大分辨率×最大分辨率下的刷新率×1.3(or 1.5)

打印机的两个重要参数

分辨率：300DPI，600DPI（大多数），1200DPI

打印速度：6～12PPM（个人和小型办公室激光打 印机），20PPM以上（工作组网络激光打印机）

（PPM：每分钟可以输出的页数）

像素深度（图像深度）：

存储每个像素所用的位数，它也是用来度量图像的分辨率

R，G，B每个8位，一个像素用24位，像素深度为24，2^24=16 777 216种颜色。

**六，数字视频处理技术 I**

原理：以一种连续贴图的方式快速播放，再加上人类”视觉暂留”的因素， 产生动画。

视频是时间上连续的一系列图像的集合

数字视频是时间上连续的一系列数字图像的集合

动画和视频信息是连续渐变的静态图像或图形序 列，沿时间轴顺次更换显示，从而构成运动视觉 的媒体。

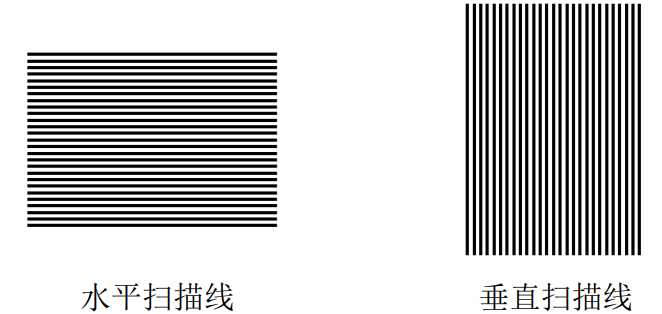
动画：序列中每帧图像是由人工或计算机产 生的图像

视频：序列中每帧图 像是通过实时摄取自然景象或活动对象

视频的分辨率：水平像素数 x 垂直像素数

VGA：640 x 480；SVGA：800 x 600

线：清晰度的单位是“电视行 （TVLine）”也称线。



电视的画面清晰度是以水平清晰度（垂直扫描线的数目 ）作为单位

一个图框的扫描线数是525条或625条。

美国：每幅画面525条扫描线，每秒钟必30幅画面

欧洲：每幅画面625条扫描线，每秒钟必25幅画面

常见动画文件格式：GIF文件（\* .gif）Flic文件（\* .FLI/\* .FLC ）FLASH文件（\* .fla）AVI 文 件 （ \* .avi ）QuickTime文件（\*.mov/\*.qt）MPEG文件（.mpeg/.mpg/.dat ）RealVideo文件（.rm/.rmvb）

电视工作原理

发送端：摄像机（摄像管）把景物转变成电信号（光-电转换）

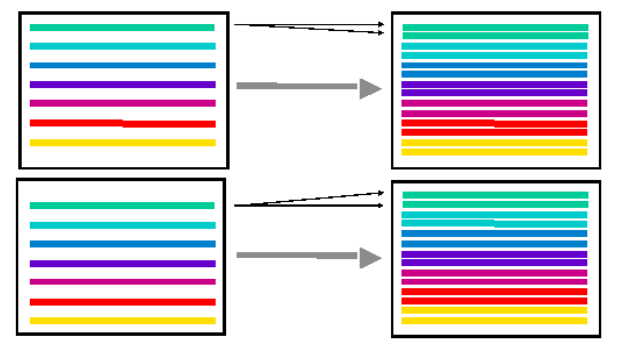
接收端：由显示设备（显像管）重现原景物（电-光转换）

扫描的机制：

1.逐行扫描（p）

2.隔行扫描（i）：两场扫描，先奇后偶

隔行扫描在传送信号带宽不够时有用，但稳定性更差，行间闪烁比较明显，



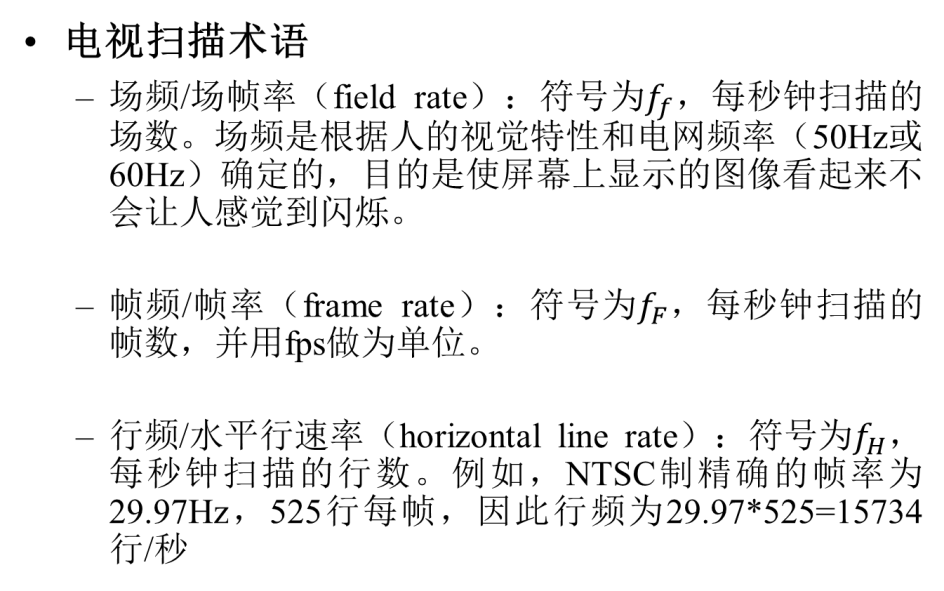
帧速率和场

视频中因为逐行扫描和隔行扫描的原因，在采用隔行扫描方式进行播放的设备中，每一帧画面都会被拆分开进行显示，而拆分后得到的残缺画面就称为 “场”

帧速率为30fps-每秒要播放60场画面-每秒钟会有30帧或30画

先有场后有帧

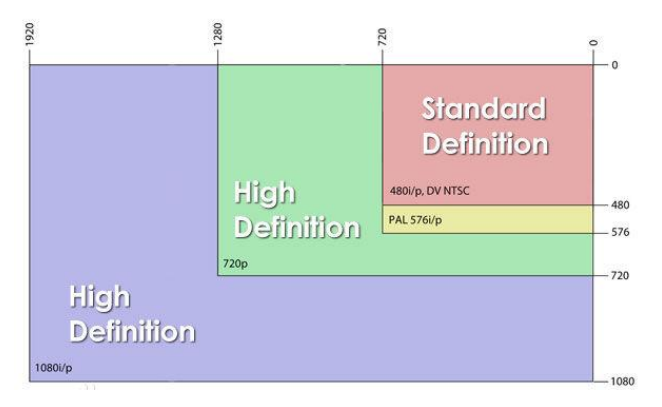
PAL制式：分别隔一场抽掉奇数行和偶数行



电视扫描格式

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 格式 | 垂直扫描线 | 可见垂直扫描线 |  | 分辨率 |  | 行频 |  |
| D1 | 480i | 525 | 483 | 4：3或16：9 |  | 隔行/60Hz | 15.75KHz | 和NTSC模拟电视清晰度相同 |
| D2 | 480p | 525 | 480 | 4：3或16：9 | 640×480 | 逐行/60Hz | 31.5KHz | 和逐行扫描DVD规格相同 |
| D3 | 1080i | 1125 | 1080 | 16:9 | 1920×1080 | 隔行/60Hz | 33.75KHz | 标准数字电视显示模式 |
| D4 | 720p | 750 | 720 | 16:9 | 1280×720 | 逐行/60Hz | 45KHz |
| D5 | 1080p | 1125 | 1080 | 16:9 | 1920×1080 | 辨率为1920×1080逐行扫描，专业格式。 | |

标清与高清



：

每场中25行为场回扫，575行是有效行

宽高比4:3

隔行扫描，2场/帧，312.5行/场

颜色模型：YUV

|  |  |
| --- | --- |
| PAL制式 | NSTC制式 |
| 625行(扫描线)/帧，25帧/秒(40 ms/帧) | 525行(扫描线)/帧, 30帧/秒 |
| 宽高比4:3 | 宽高比4:3 |
| 隔行扫描，2场/帧，312.5行/场。 | 隔行扫描，一帧分成2场(field)，262.5线/场 |
| 每场中25行为场回扫，575行是有效行 | 在每场的开始部分保留20扫描线作为控制信息，故只有485条线的可视数据，每行63.5微秒，水平回扫时间10微秒(包含5微秒的水平同步脉冲)，所以显示时间是53.5微秒 |
| 颜色模型：YUV | 颜色模型：YIQ |

黑白电视中只需传送一个亮度信号，彩色全电视信号=亮度信号+2个色差信号（与黑白电视兼容、而且不增加为黑白电视所规定的信道带宽）

主要的彩色电视制式有3 种：

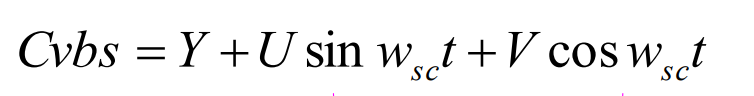
1.NTSC：正交平衡调幅制

2.PAL：逐行倒相正交平衡调幅制

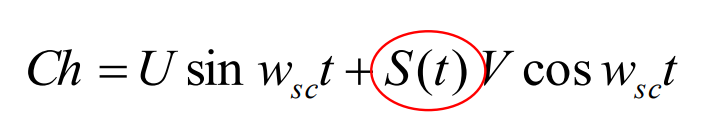
3.SECAM：顺序传送彩色与存储制

NTSC制

U和V（用副载波wsc）经平衡正交调制，加到亮 度Y上，形成彩色全电视信号,U信号调制在副载波的零相位上， 而V信号调制在90 o相位上，



PAL制式是为克服NTSC制式系统对信道微分相位敏感，导致图像色 调失真的缺点



S(t)称为PAL开关函数，它是双极性矩形脉冲，其重复 周期为行周期的两倍，幅度为+1和-1

视频信号类型

1.复合电视信号（CVBS）：包含亮度信号、色差信号和所有定时同步信号的单一电视信号， 或称全电视信号。易导致亮色串扰、清晰度降低。

2.分离电视信号（S-Video）：是亮度和色差分离的一种电视信号(Y/C)。图象质量优于复合 视频信号

3.分量电视信号：是指每个基色分量作为独立的电视信号，如：RGB或YUV。表示颜色的最好方法，但需要比较宽的带宽 和同步信号。

电视信号的数字化

好处

1.可直接进行随机存储，电视图像的检索方便

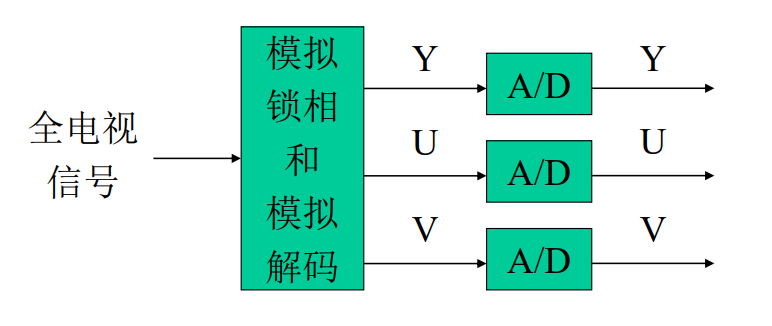
2.制数字电视图像和在网络上传输数字电视图 像都不会造成质量下降

3.很容易进行非线性电视编辑

方法

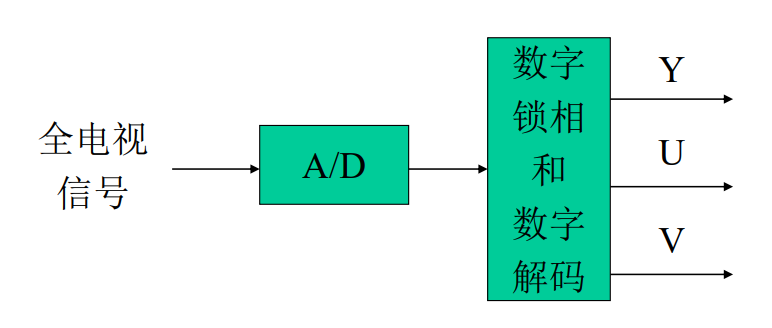
1.分量电视信号的数字化

先从复合彩色电视图像中分离出彩色分量，然后数 字化



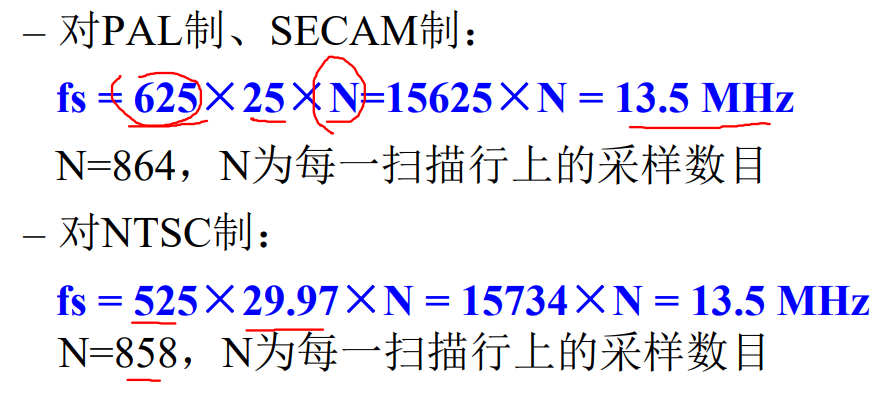
2.复合电视信号的数字化

对色度信号和亮度信号共频带所形成的复合电视信 号直接进行数字化



为了保证信号的同步，采样频率必须是电视信号 行频的倍数。

NTSC、PAL和SECAM制式，电视图像采样标准：fs＝13.5MHz



对于4 : 2 : 2的采样格式（ITU-R BT.601推荐使用），亮度信号Y用 fs 频率采样，两个 色差信号（Cr、Cb）分别用 fs/2 = 6.75MHz 的频率采样

4种采样格式

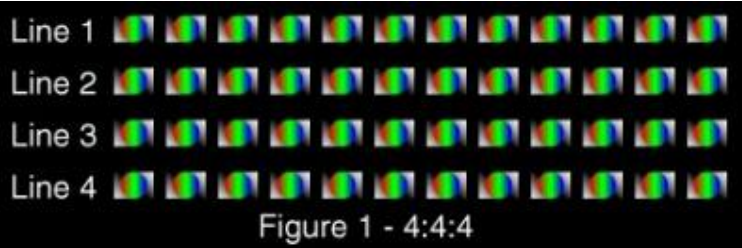
（亮度Y样本、红色差Cr样本，4蓝色差Cb 样本）

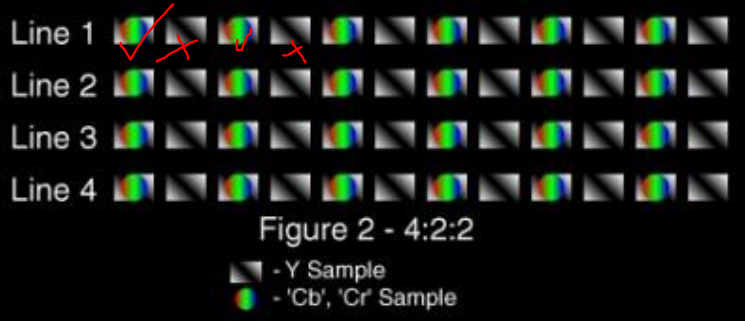
4:4:4 在每条扫描线上每4个连续采样点取4Y、4Cr，4Cb，每个像素3样本

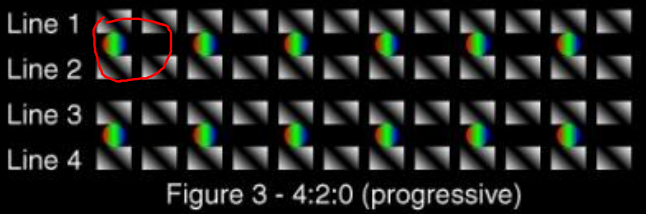
4:2:2 在每条扫描线上每4个连续采样点取4Y、2Cr，2Cb，每个像素2样本

4:1:1 在每条扫描线上每4个连续采样点取4Y、1Cr，1Cb，每个像素1.5样本

4:2:0 在在水平和垂直方向上每2个连续的采样点取2Y、1Cr，1Cb，每个像素1.5样本







**七，数字视频处理技术 II**

摄像机工作原理：把光学图像信号转变为电信号

摄像机主要组成部分：镜头系统、主机、寻像器和附件。

DVD数码摄像机最大的 特征是磁带上记录的信号为数码信号，而非模拟信号。

磁带录像机：利用磁记录原理把视频信号及其伴音信号记 录在磁带上的设备

视频采集卡是视频采集和压缩同步进行

视频编辑可以分为两种形式：

1.线性编辑：在编辑机上进行的编辑， 由一台放像机和一台录像机组成。

2.非线性编辑：是相对于传统上以时间 顺序进行线性编辑而言，非线性编辑 借助计算机来进行数字化制作。

前投就是观众和 投影机位于投影屏幕的同一侧

背投是观众和投影机位于投影屏 幕的两边

**八，多媒体压缩技术**

编码就是对应，对数字信息的编码表现为从一个比特流转换为 另一个比特流。

多媒体数据的编码

1.首先需要表示媒体信息

2.在最大程度上压缩比特流

3.适合于在网络上传输

两种表现形式

1.在等同的空间中容乃更多的信息

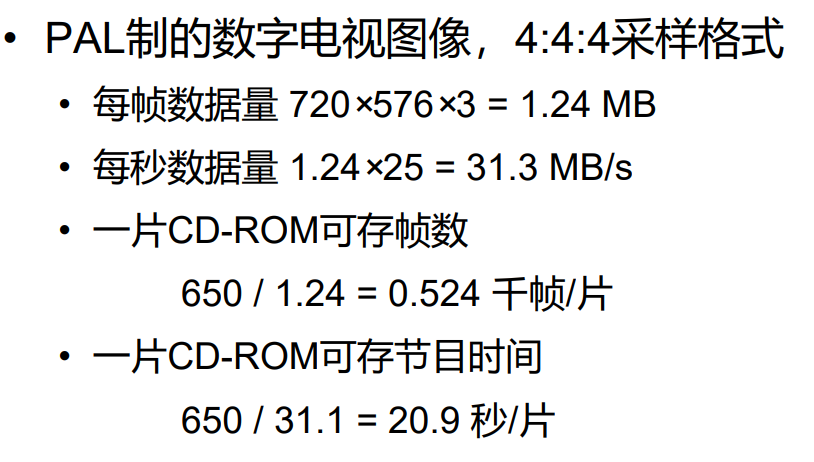
2.在等同的时间内表现更多的信息

比特流的压缩

1.用相同长度的比特流表示更多的信息

2.用尽量短的比特流表达尽量多的信息

650MB的CDROM





图像数据压缩技术就是研究如何利用图像数 据的冗余性来减少图像数据量的方法

空域冗余：图像在 空间上存在很大相 关性，如：相邻像 素值重复或非常接 近。 时域冗余：视

视觉冗余（感知冗余）：人眼感受不到图像中 的一些复杂细节信息，丢弃这些信息不会影响 视觉感受 频图 像在时间上存在很 大相关性。

多媒体数据压缩方法的分类

1.质量有无损失 ：有损失编码、无损失编码

2.作用域 ：空间方法、变换方法、混合方法

3.是否自适应 ：自适应性编码、非自适应性编码

无损压缩

1.使用压缩后的数据进行重构重构后的数据与原来的数 据完全相同；

2.用于要求重构的信号与 原始信号完全一致的场合。

3.磁盘文件的压缩。

4.霍 夫 曼 (Huffman) 算 法 和 LZW(Lenpel-Ziv &Welch)压缩算法

有损压缩

1.使用压缩后的数据进行重构，重 构后的数据与原来的数据有所不同，但不会让 人对原始资料表达的信息造成误解。

2.用于重构信号不一定非要和原始信号完全相 同的场合。

3.图像和声音的压缩

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 脉冲编码调制（PCM） | 预测编码 | 变换编码 | 统计编码 | 混合编码 |
| 实际上是连续模拟信号的数字采样表示。 | 编码器记录与传输的不是样本的真实值，而是它与 预测值的差 | 利用图像块内像素值之间的相关性，把 图像变换到一组新的基上，使得能量集中到少数几 个变换系数上，通过存储这些系数达到压缩的目的 | 对于出现频率大的符号用较少的位数表示，而对于 出现频率小的符号用较多的位数来表示 | 合并变换和预测技术 |
| 其它编码方法通常都是在多媒体数据模拟信号 经过PCM编码后再进行的 | 借助了空间的相关性 | 把原始图像分成许多个矩形区域子图像独立进行变 换 | 编码效率取决于信源的概率分布 | 对动态图像而言，空域的二维变换再加上时 间方向上的DPCM预测 |
|  | 差分脉冲编码调制（DPCM） | KLT、DCT、DFT | Huffman编码、算术编码、游程编码 |  |

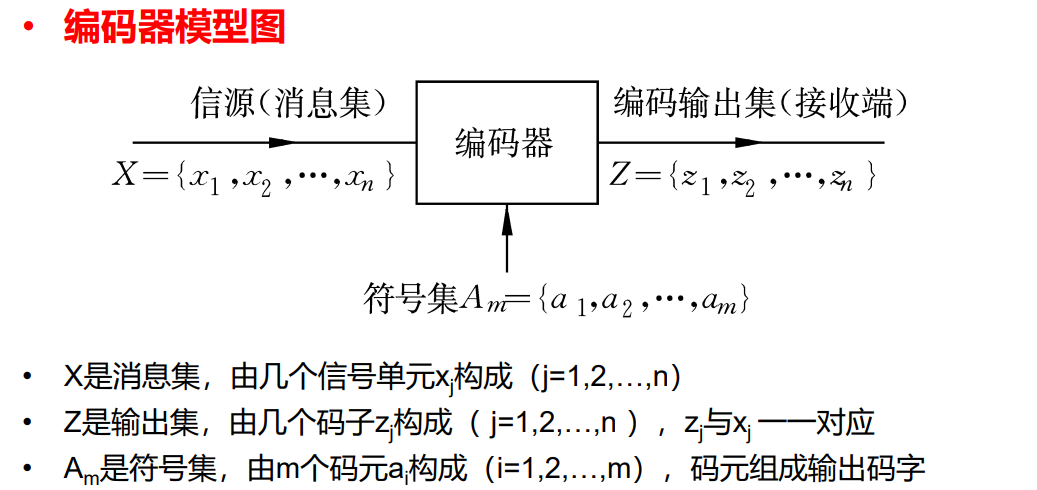
统计编码

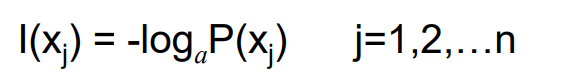
数据压缩的理论基础是信息论

数据压 缩的理论极限是信息熵

无失真编码（熵编码）:编码过程中不丢失信息，即要求保存信息熵

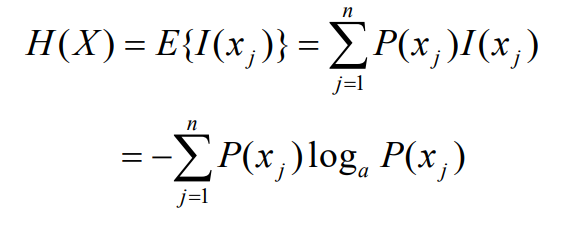
有失真编码:可允许信息部分损失以换取较高的压缩比



信息量

(P(xj )：信源X发出xj的先验概率 )

信息熵:自信息量的统计平均



信源的熵H(X)：信源X发出任意一个随机变量的平均信 息量

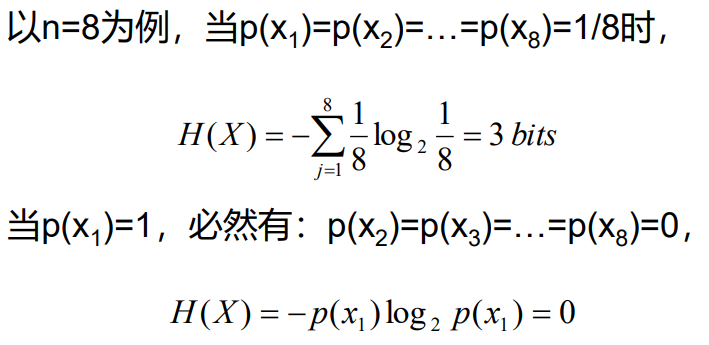
当a取2时，H(X)的单位为“比特”（bit）或者“香农”（Sh）

等概率事件的熵最大,极大值log2n

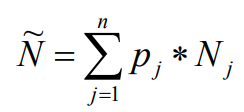
(n为信源中事件个数 )



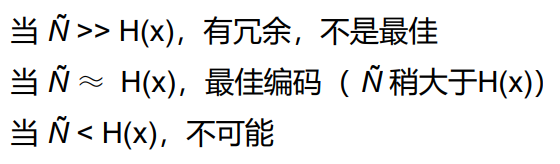
熵的范围为：0≤H(X)≤log2N



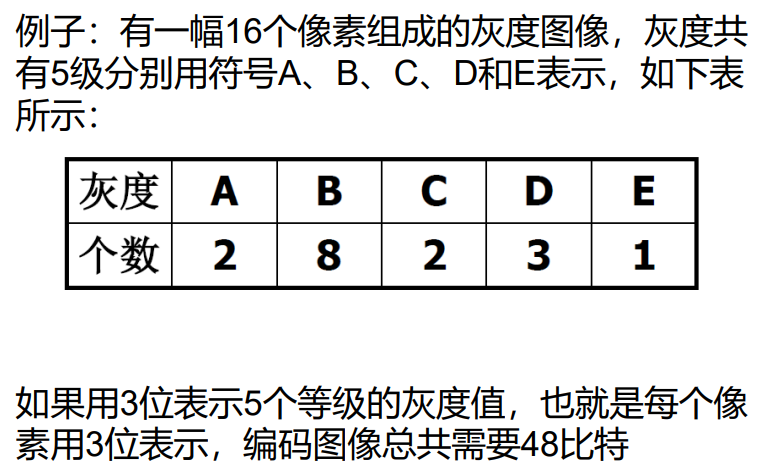
在编码中用熵衡量是否为最佳编码。

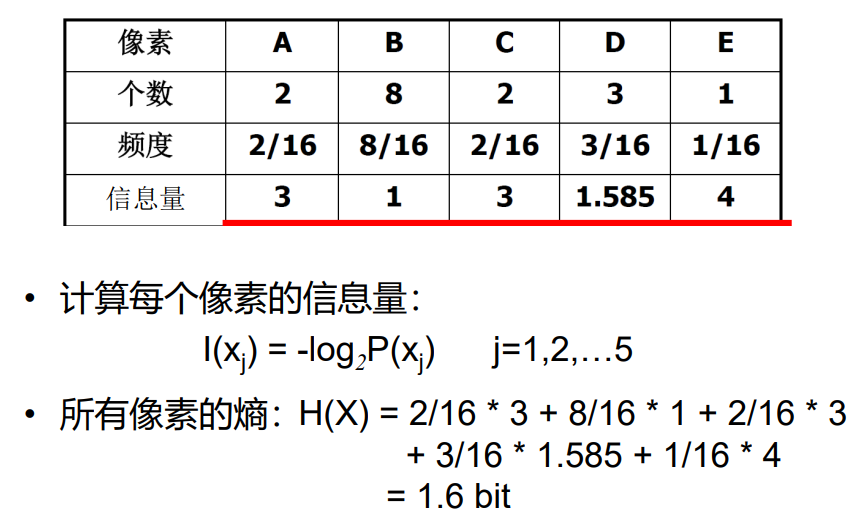
编码器输出码子的平均码长：

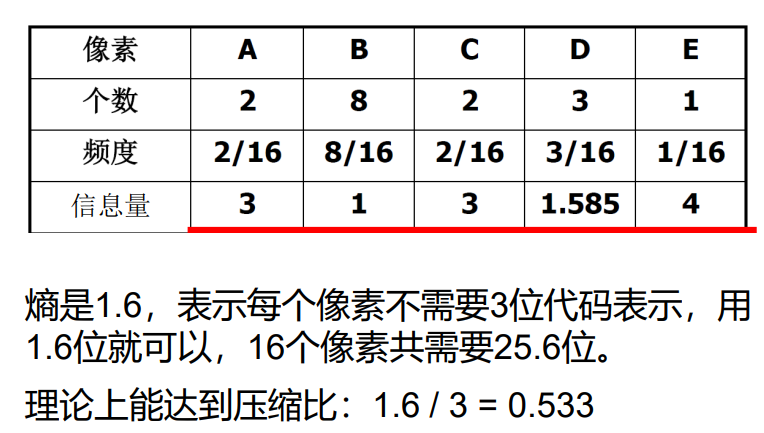
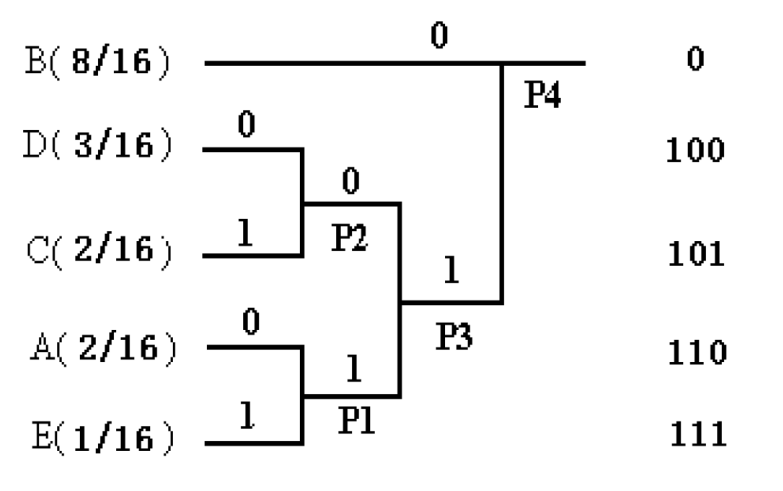
熵值是平均码长的下限：当分配给xj的比特数Nj = I(xj ) = -log2p(xj )时，平均码长取极小值，

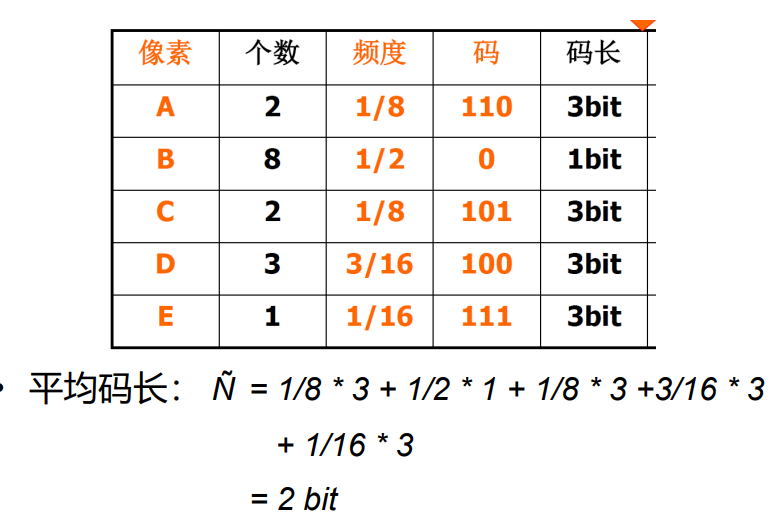


Huffman编码





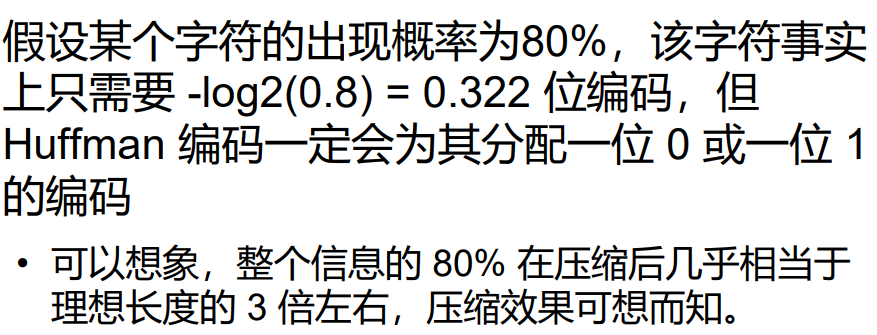
 



3 bits（等长码需要的比特数）>平均码长 Ñ > H（熵）

Huffman编码的局限

某个字符概率非常大时不适用

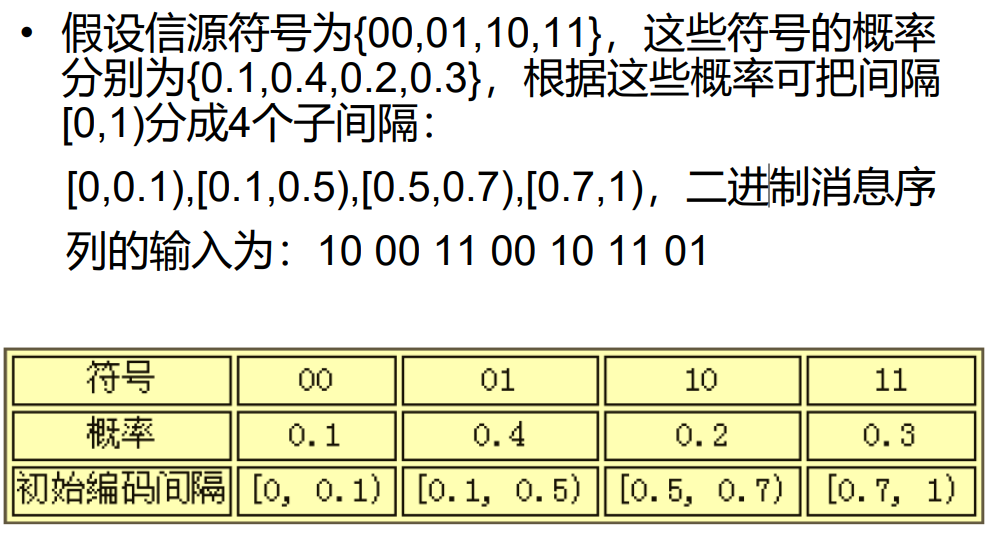


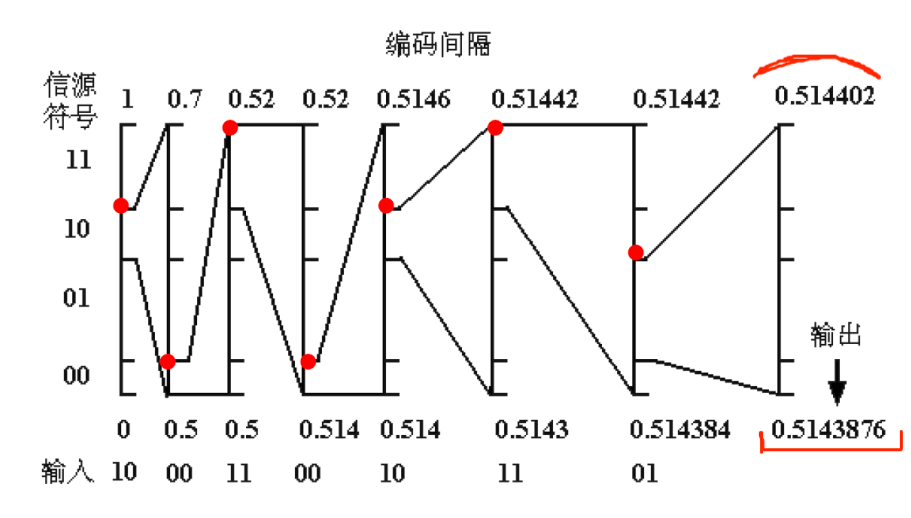
算术编码

1.基本思想：算术编码不是将单个信源符号映射 成一个码字，而是把整个信源表示为实数线上 的0到1之间的一个区间，其长度等于该序列的 概率，再在该区间内选择一个代表性的小数， 转化为二进制作为实际的编码输出

2.消息序列中元素越多，所得到的区间就 越小，表示这个区间所需的位数就更多

3.采用算术编码每个符号的平均编码长度可以 为小数，因此可以更加接近无损压缩的熵极 限

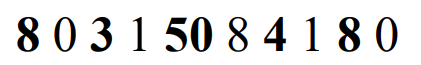




游程编码

游程编码和词典编码核心思想：利用重复结构信息

游程编码：相同符号序列->一个计数字段再 加上一个重复字符标志





RLE通常需要和其他的压缩编码技术联合应用

词典编码

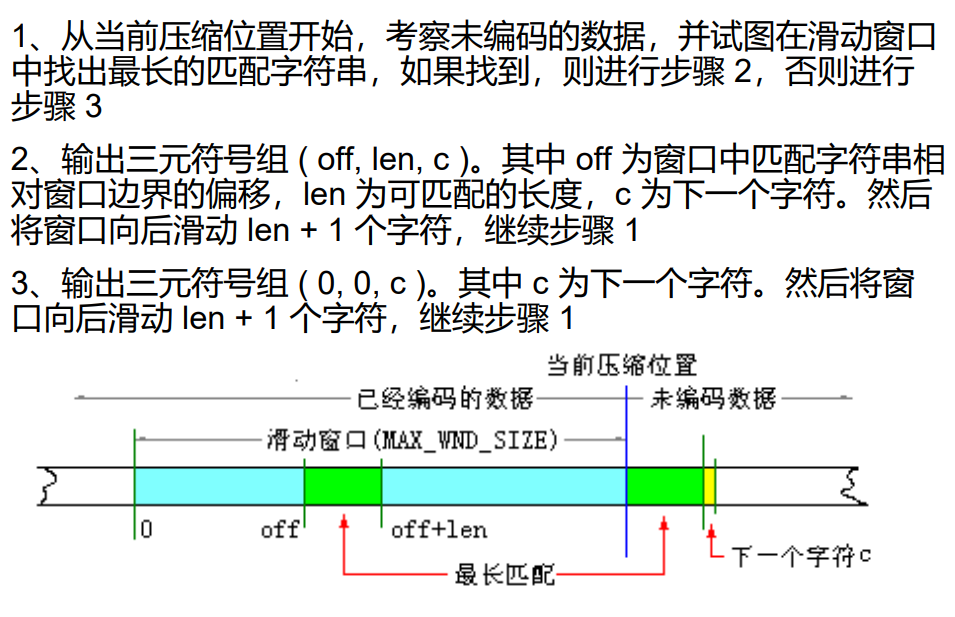
1.重复的部分用指向早期出现过的字符串的“指针”输出

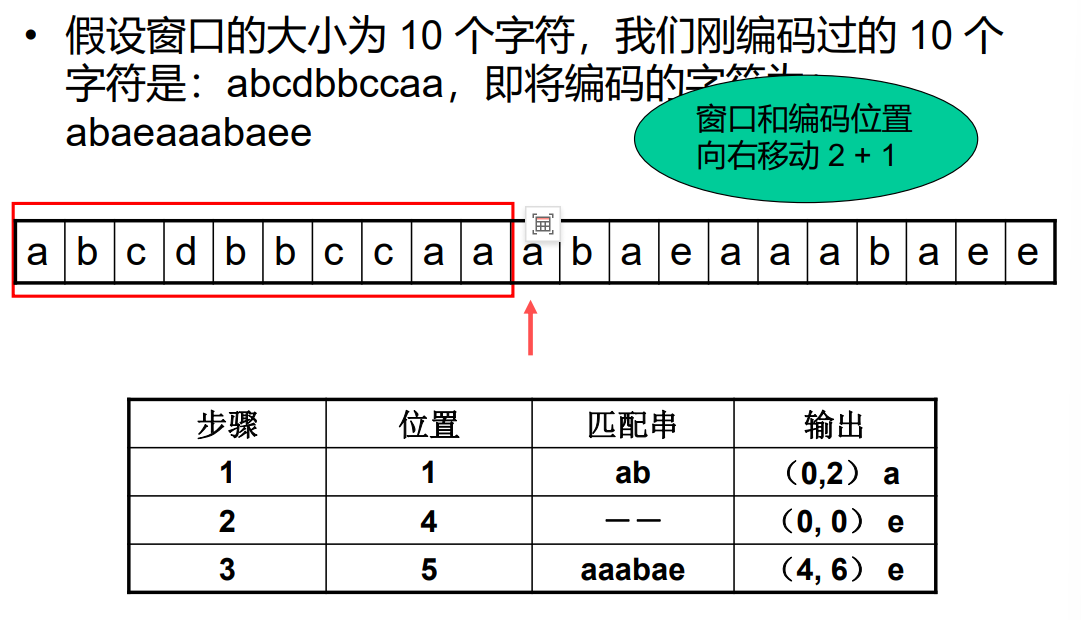
LZ77，LZSS

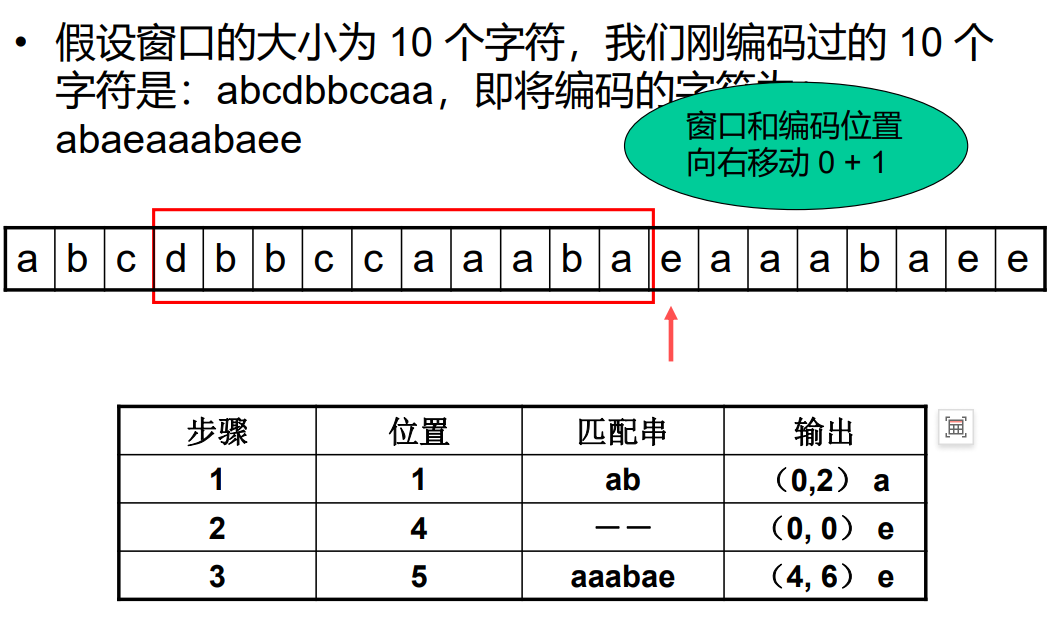
2.创建一个“短语词典”，遇到已经在词典中出现的“短语”时，编码 器就输出这个词典中的短语的“索引号” ，

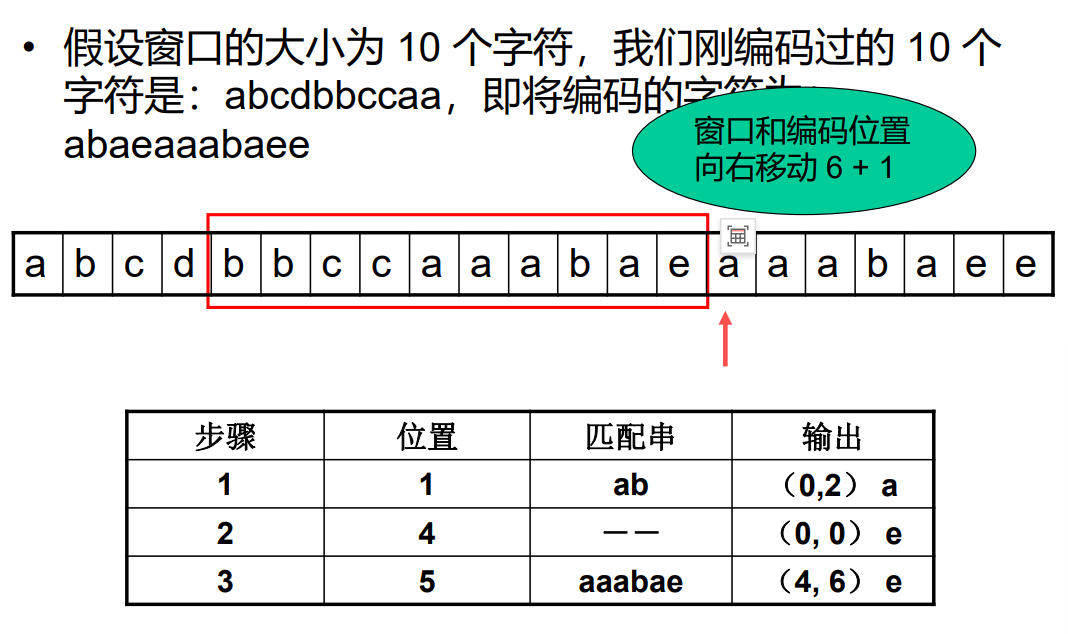
LZ78，LZW

LZ77算法









LZSS



预测编码

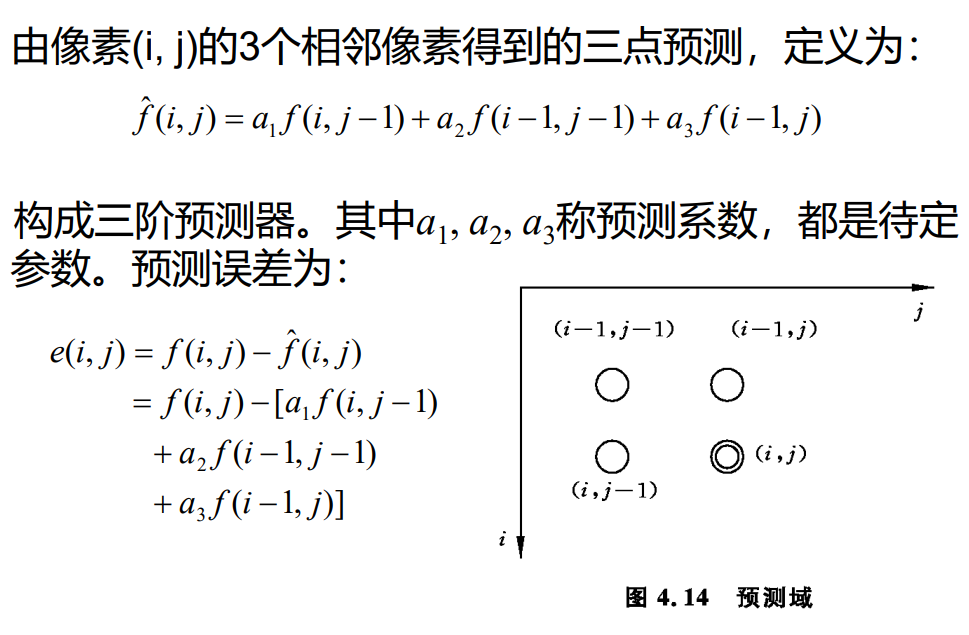
基本思想

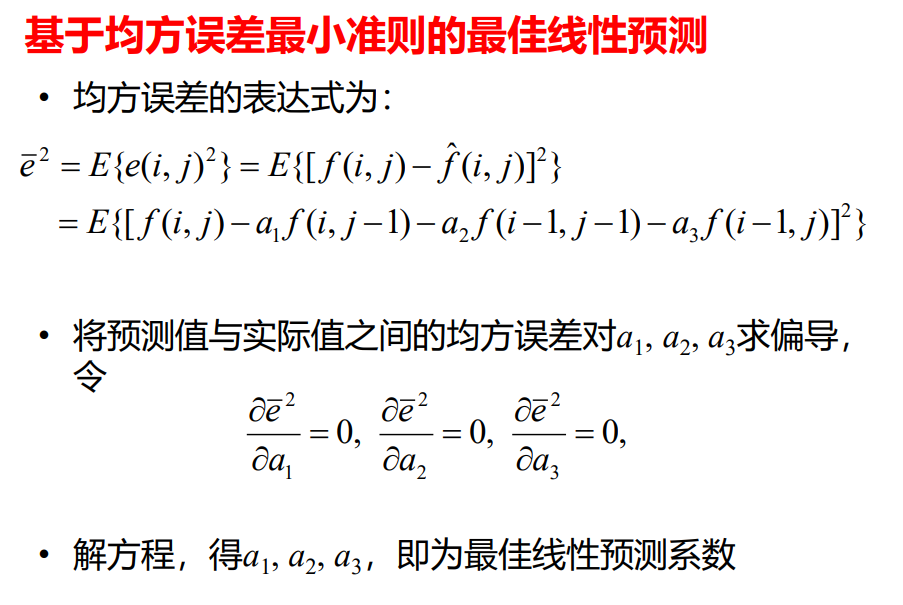
1.根据某一模型利用以往的样本值对于新的样本值进 行预测，然后将样本的实际值与其预测值相减得到 一个误差值，对这一误差值进行编码

2.如果模型足够好且样本序列在时（空）间上的相关 性较强，误差信号的幅度将远远小于原始信号，因 此可使用较少的位数对其量化

3.对预测误差e进行量化、编码、发送。（f：实际灰度，f'：预测灰度）

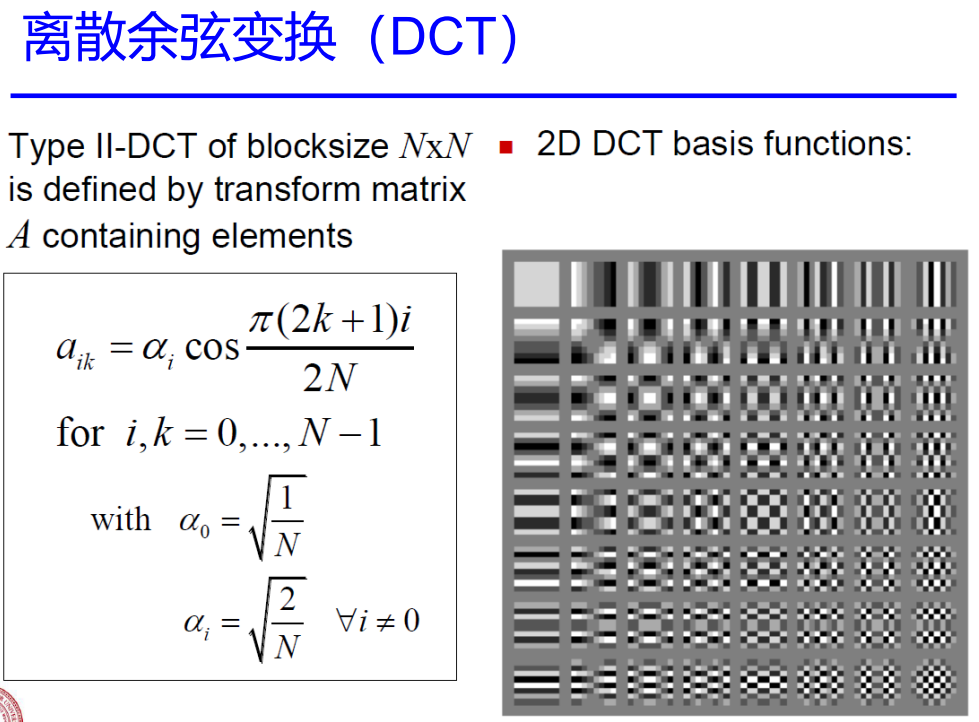






变换编码

变换编码不是直接对空域图像信号编码，而是首 先将空域图像信号映射变换到另一个正交矢量空 间（变换域或频域），产生一批变换系数，然后 对这些变换系数进行编码处理



九，JPEG压缩编码

JPEG算法具有以下4种操作方式

1.顺序编码(Sequential) – 从左到右，从上到下完成一次扫描

2.累进编码(Progressive) – 图像编码在多次扫描中完成

3.无失真编码(lossless)

4.分层编码(Hierarchical) – 图像在多个空间分辨率进行编码

JPEG有损顺序编码算法步骤：

1.彩色图像分量：将源图像分成几个颜色平面

2.FDCT：分成8×8数据块进行正向离散余弦变换

3.量化

4.Z字形排列量化结果

5.使用 DPCM对直流系数(DC)进行编码，再做熵编码

6.使用行程编码(RLE)对交流系数(AC)进行编码 ，再 做熵编码

JPEG采用的是8×8大小的子块的二维离散余弦 变换（DCT）

把原始图像分割成一系列8×8的子块

设原始图像为P位，输入时将[0, 2 P -1] 范围的无 符号整数变成[-2 P-1 , 2 P-1 -1]范围的有符号整数， 以此作为正向离散余弦变换（FDCT）的输入

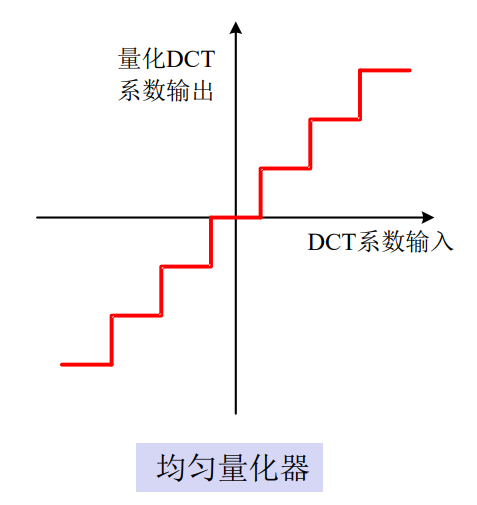
对于解码器输出端的离散余弦逆变换（IDCT） 的结果，需将其数值范围由[-2 P-1 , 2 P-1 -1]再变回 到[0, 2 P -1] ，来获得重构图像

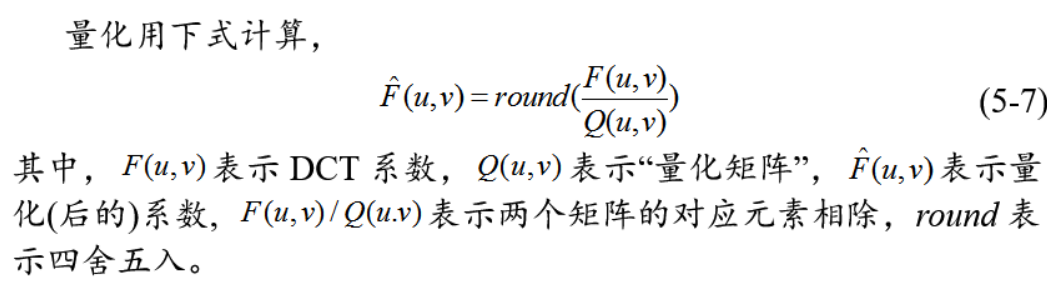


量化

目的：降低非0系数的幅度，增加0值系数的数目

量化是造成图像质量下降的最主要原因

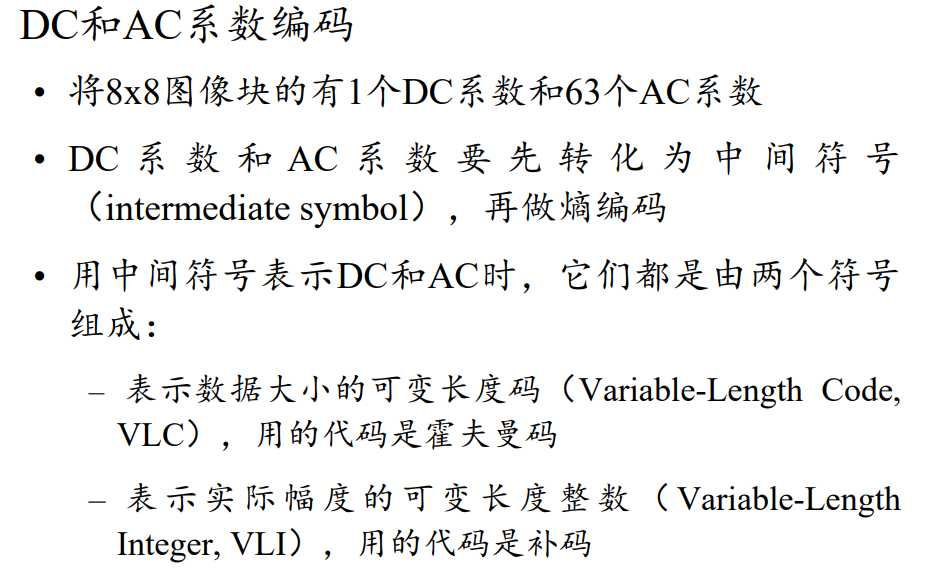




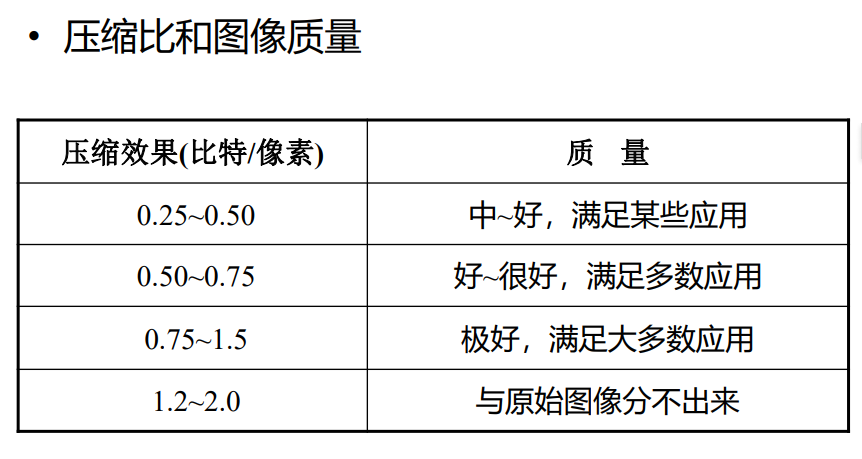
量化表的尺寸为8×8，与64个变换系数一一 对应

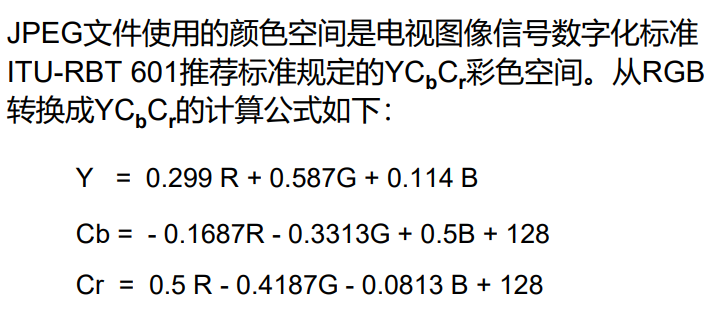
Z字形编排：重新编排量化后的系数，增加连续的0值系数数目

DC和AC系数编码



|  |  |
| --- | --- |
| DC | AC |
|  |  |
| 位数+实际值（差值） | 0的个数/位数+实际值 |
|  |  |
| 查表+vli | 查表+vli |
|  |  |
|  |  |





**十，语音编码**

语音(speech)：频率为300~3400 Hz的信号

声音(audio)：频率为20~ 20 000Hz的信号

语音是声音，但声音不一定是语音

采样精度

度量声音波形幅度的精确程度，通常用存储每个声音样本 的位数(n)来表示

例如，每个声音样本用16位(2字节)表示，测得的声音样本值 是在[0～65535]范围里的数，采样精度是1/65536

量化阶

精度是度量模拟信号的最小单位，称为量化阶(quantization step size)

如将0～1 V的电压用n=256个数表示，其量化阶等于1/256 V。

量化阶距=（最大值-最小值）/分组数

语音编码

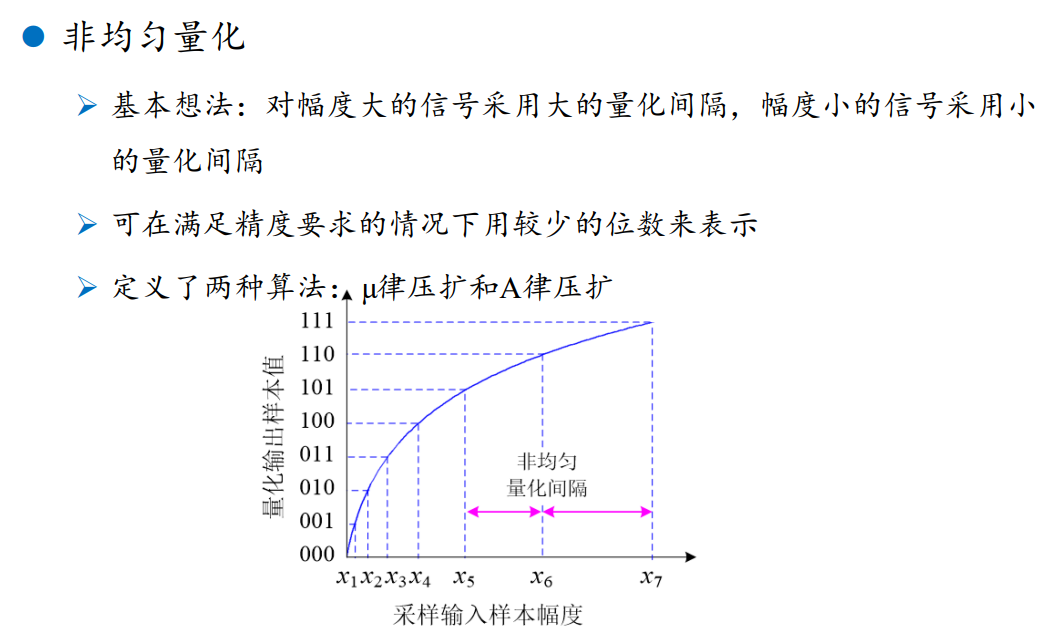
|  |  |  |
| --- | --- | --- |
| 波形编码 | 参数编码 | 混合编码 |
| 用数字形式精确地表示模拟信号波形的编码方法，不考虑 语音产生和感知特性 | 利用发音器官生成语音信号的模型，对从语音信号中抽出 的语音特征参数(如发音模型、有声/无声、音量大小、音 调)进行编码的方法 | 综合使用波形编码和参数编码技术，组合波形特性和语音 特征参数的编码方法，用在语音质量要求高的移动通信系统 |
| 输出的数据速率为9.6~64 kbps或更高 | 2~4.8 kbps | 4.0～16 kbps |
| PCM  APCM  DPCM  ADPCM | LPC  信道编码  相位编码  共振峰编码 | APC  RELP  CELP  SBC |
| 语音质量高 | 数据速率 低 | 既有波形编码语音质量高的优点，又有参数编码数据速率 低的优点，复杂性高，计算量大 |

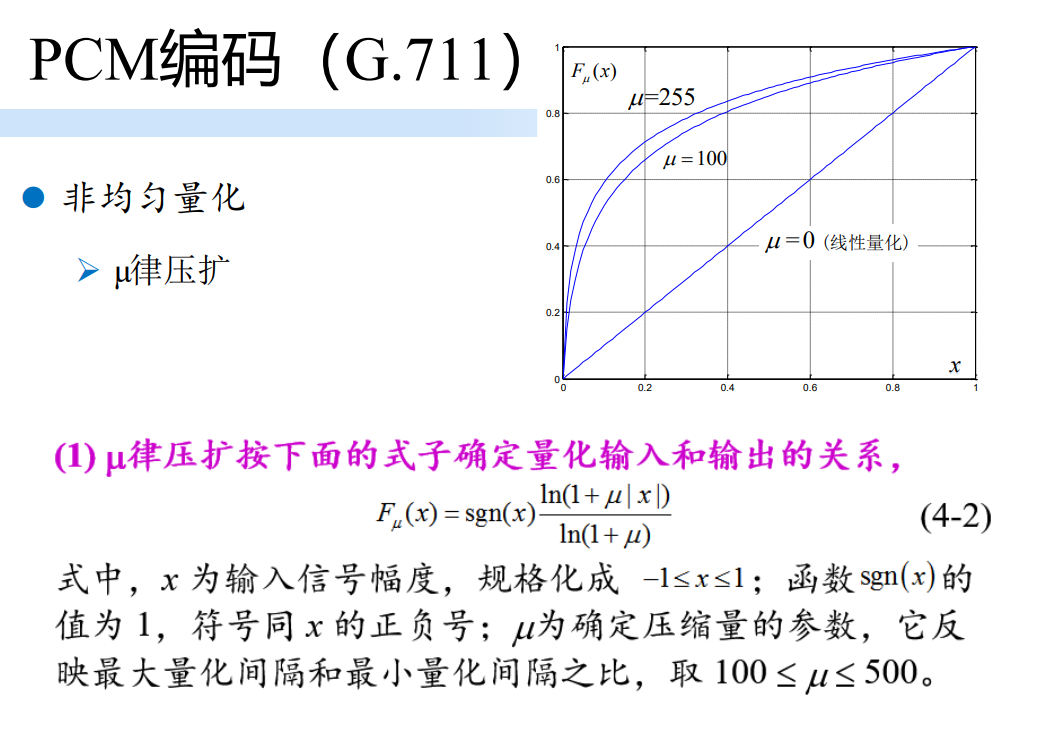
波形编码

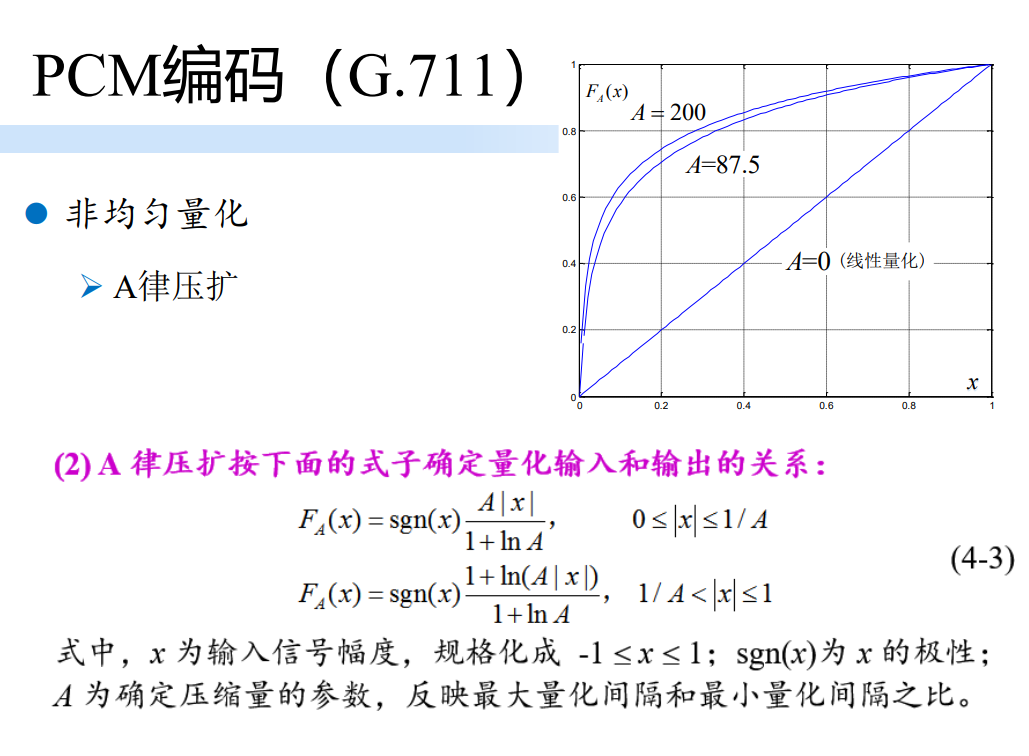
PCM

概念最简单、理论最完善，最早研制成功、使用最为广泛，数据量最大的编码系统

均匀量化(线性量化)



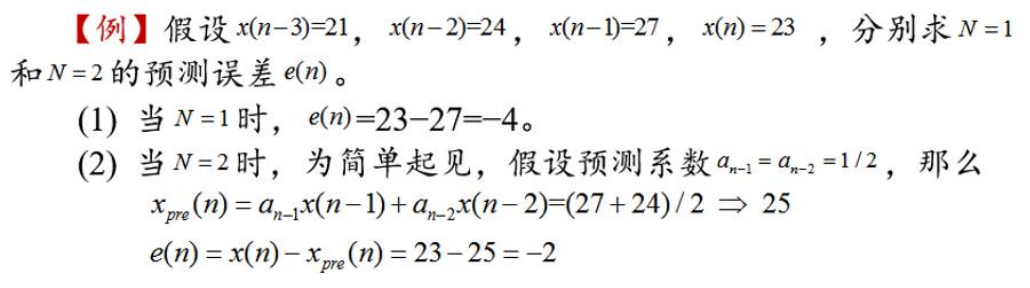


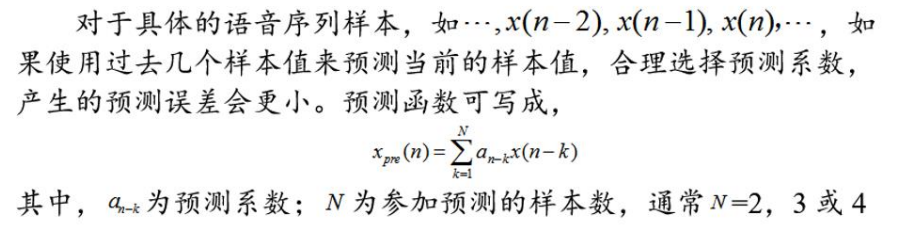


APCM(自适应脉冲编码调制)

前向自适应(forward adaptation)：根据未量化的样本 值的均方根估算输入信号的幅度，以此来确定量化阶 大小，并对其幅度进行编码作为边信息

后向自适应(backward adaptation)：从量化器刚输出的 过去样本中来提取量化阶信息，不需要传送边信息

预测编码



子带编码

1.用一组带通滤波器(band-pass filter，BPF)把输入声音信号的频带分 成若干个连续的频段，每个频段称为子带

2.对每个子带的声音信号采用单独(ADPCM，APCM或PCM)的编码方 案编码

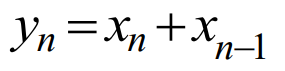
参数编码：

对语音的特征参数进行编码，也称声源编码

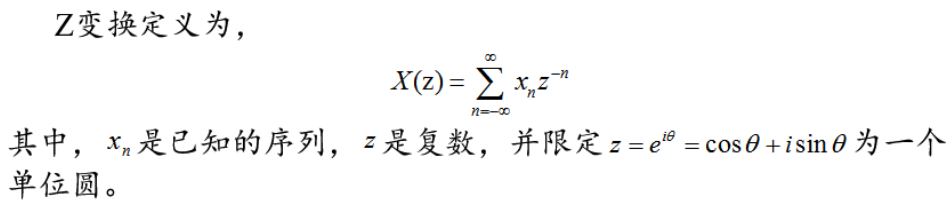
参数编码

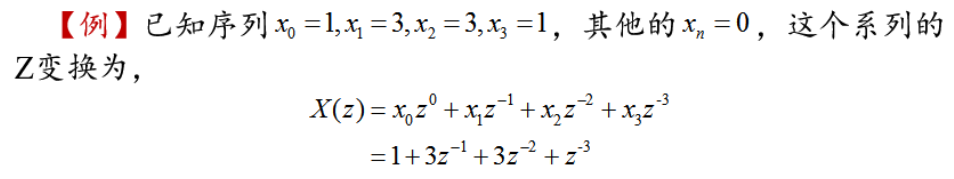
数字滤波器：

可对数字信号执行滤波功能的软硬件

最简单的滤波器 

Z变换





**十一，MPEG压缩编码**

MPEG视频是用MPEG视频标准压缩和解压缩的电视图像。现有 的MPEG视频标准包括MPEG-1 Video，MPEG-2 Video，MPEG4 Visual，H.264/MPEG-4 AVC和H.265/ MPEG-H HEVC

视频压缩原理

视频数据中存在大量 的冗余数据，包括时间冗余、空间冗余、结构冗余、视觉 冗余、知识冗余和数据冗余

MPEG视频标准利用的冗余：时间冗余、空间冗余、视觉 冗余和数据冗余

|  |  |
| --- | --- |
| 时间冗余1 | 在某个时间间隔上出现场景相同或基本相 同的连续帧 |
| 空间冗余1 | 在单帧图像中，相邻像素的值常有相 同或变化不大的情况 |
| 结构冗余 | 从宏观上看一帧图像，有些图像存在 相同或类似的结构，如用地板图案构成的图像 |
| 视觉冗余1 | 对图像的亮度变化敏感而对颜色变化 不敏感，对剧烈变化区域敏感而对缓慢变化区域不敏感， |
| 知识冗余 | 在单帧图像中含有为人熟知的知识，称为 先验知识。例如，正面人头像有相对固定的结构 |
| 数据冗余1 | 视频数据本身存的冗余 |

ISO/IEC制定-MPEG-x-广播电视、DVD和视频流媒体

ITU-T制定-H.26x-实时视频通信

ISO领导的

1.JPEG的目标：静止图像的压缩

2.MPEG的目标：运动图像的数据压缩

MPEG主要包括MPEG视频、MPEG音频和MPEG系统(视音频 同步)三个部分

MPEG-1和MPEG-2是MPEG组织制定的第一代视音频压缩标准

MPEG-4是基于第二代视音频编码技术制定的压缩标准

MPEG-7是多媒体内容描述标准，支持对多媒体资源的组织管理 、搜索、过滤、检索

MPEG-21的重点是建立统一的多媒体框架，

MPEG-7/21不涉及视频的压缩技术

MPEG-1压缩率约26:1

基本方法

1.空间方向上，与JPEG类似的算法

2.时间方向上，移动补偿算法

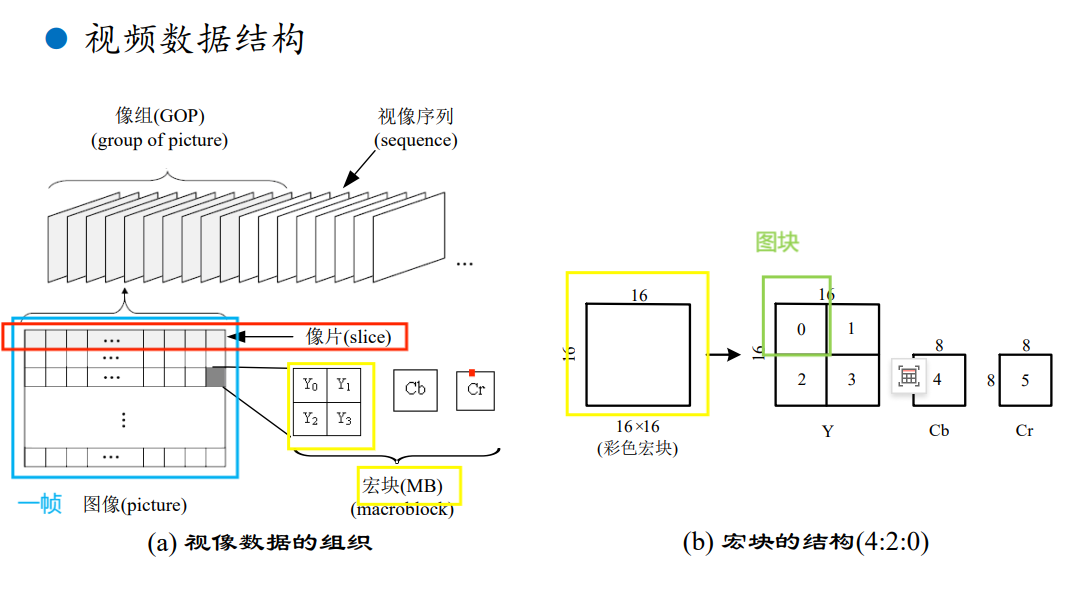
1.视频序列分成许多像组

2.像组中每一帧图像分成许多像片，每个像片16行

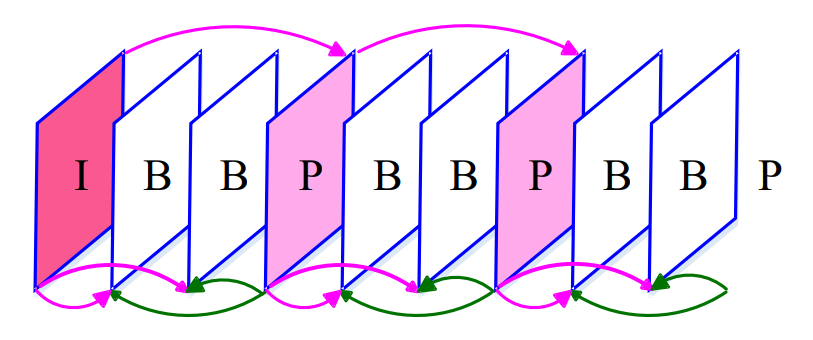
3.像片分成16行×16像素/行的宏块

4.宏块分成8行×8像素/行的图块

5.一个宏块由4个亮度(Y)图块和2 个色度图块(Cb和Cr)组成



|  |  |  |  |
| --- | --- | --- | --- |
| 帧内图像I（I图像或I帧）（关键帧） | 包含内容完整的图像，用于为其他帧图像的编码和 解码作参考 | 不参照过去的帧和将来的帧，采用与JPEG类似的压缩算法 以减少空间的冗余数据 |  |
| 预测图像P（P图像或P帧） | 以在它之前出现的帧内图像I作参考的图像，对预测 图像P进行编码就是对它们之间的差值进行编码 | (1) 求解差值的方法 |  |
|  |  | (2) 求解移动矢量的方法 |  |
| 双向预测图像B（B图像或B帧） | 以在它之前和之后的帧图像(I和P)作参考的图像，对 B进行编码就是对帧内图像I和预测图像P的差值分别 进行编码 |  |  |



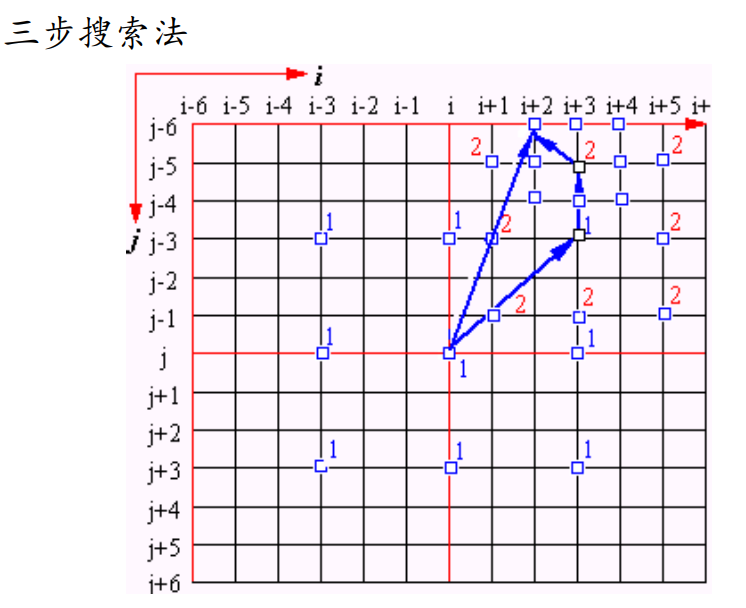
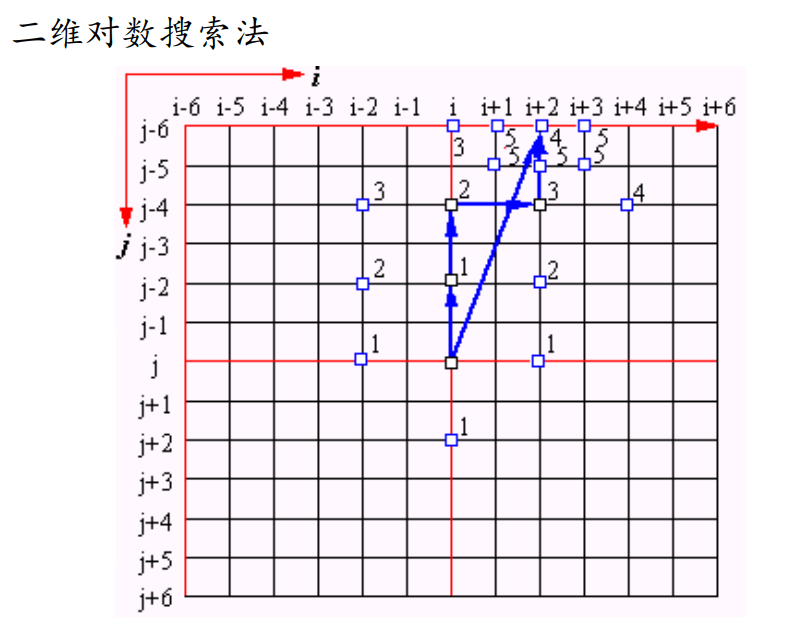
有必要每 隔一段时间传送1帧只采用帧内编码的图像作为参考图像

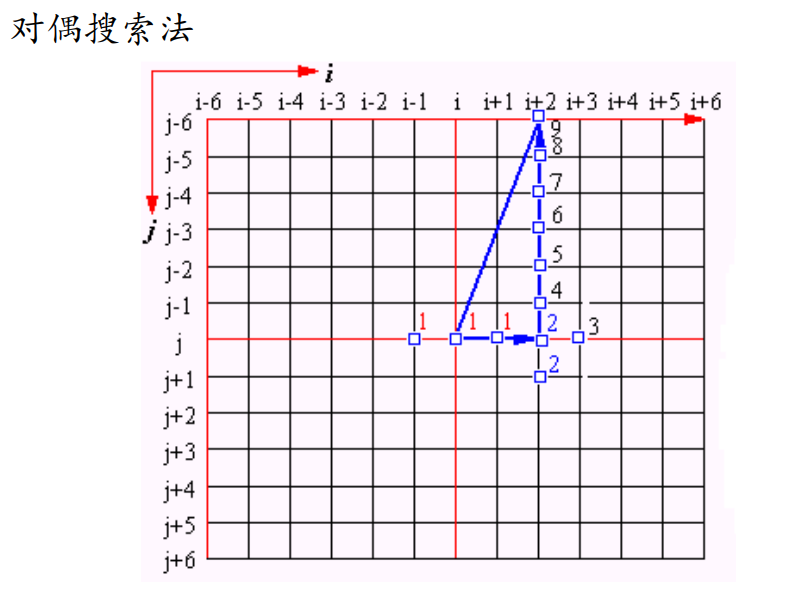
GOP：两个I帧之间的图像构成一个组称为GOP





预测图像P最佳匹配搜索算法





帧内图像I的数据量最大，而双向预 测帧图像B的数据量最小



对快速运动的图像，帧内图像I的频率可以选择高一些，双 向预测图像B的数目可以选择少一些