

计算机系统导论

Introduction to Computer System

张磊

四川大学计算机软件学院

2023.10

Unit 3. 数字媒体

Digital Media

重点/Focus

- ❖ 数字音视频的获取
- ❖ 常用的声音、图形图像、视频文件格式
- ❖ 数字音视频格式转换

目录/Agenda

3.1 数字声音/Digital Sound

3.2 图像/Images

3.3 矢量和三维图形/Vector and 3-D Graphics

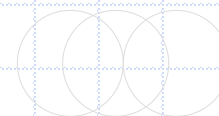
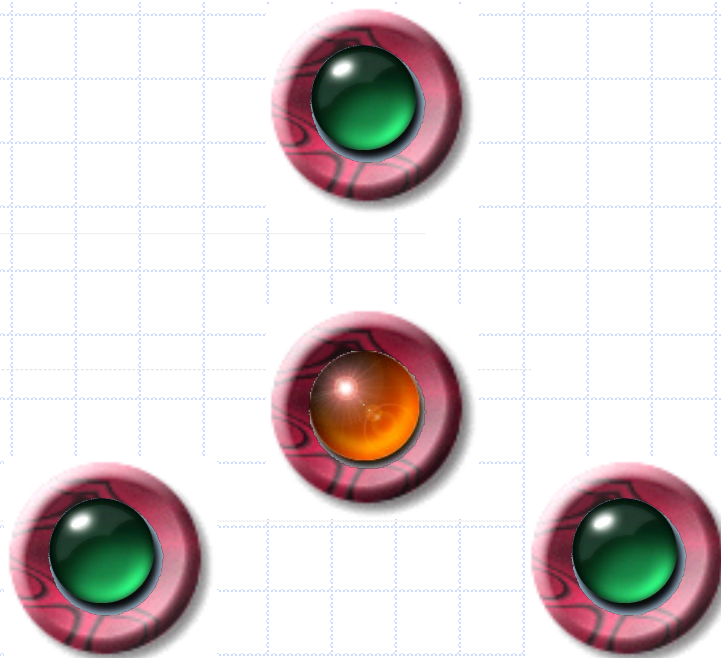
3.4 数字视频/Digital Video

3.1 数字声音

Digital Sound

- ❖ 数字声音的编辑
- ❖ 数字声音的格式转换

数字音频 (*Audio*)



数字音频技术

计算机音频技术主要包括声音的采集、数字化、压缩/解压缩以及声音的播放。

数字化主要包括采样和量化这两个方面。



(1) 数字声音采样/Sampling

❖ 模拟音频

➤ 声波信号

❖ 数字音频

➤ 数字信号

❖ 采样就是将模拟声音信号转化为数字声音信号的过程

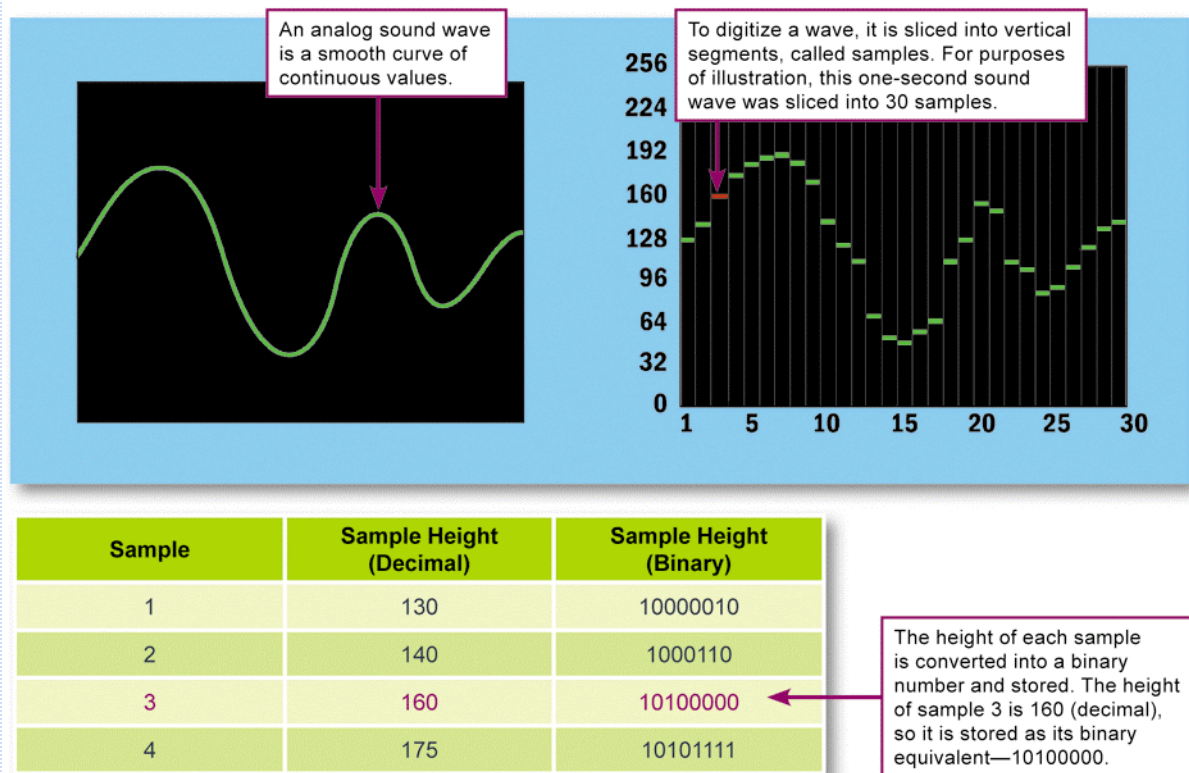
➤ 采样率：采样的频率

➤ 11kHz

➤ 44kHz

FIGURE 8-1

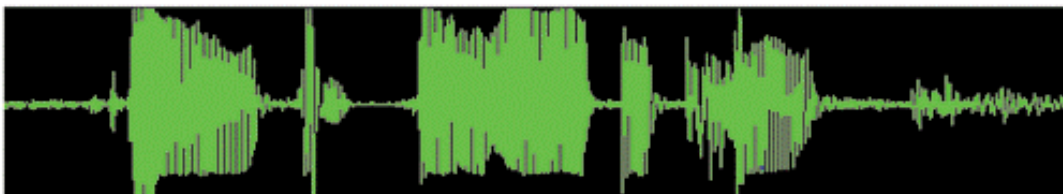
Sampling a Sound Wave



采样率/Sampling Rate

- ❖ 采样频率 (sampling rate) 是将模拟声音波形转换为数字时，每秒钟所抽取声波幅度样本的次数，单位是Hz（赫兹）
- ❖ 采样率越高，声音越逼真，但数据量增大
- ❖ 采样率提高一倍，单声道声音数据增大一倍

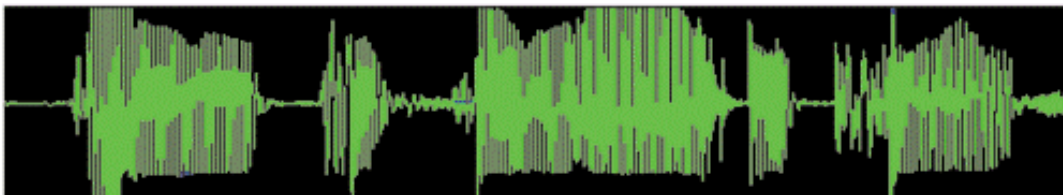
Low sampling
rate: File size =
66 KB



Medium sam-
pling rate: File
size = 124 KB

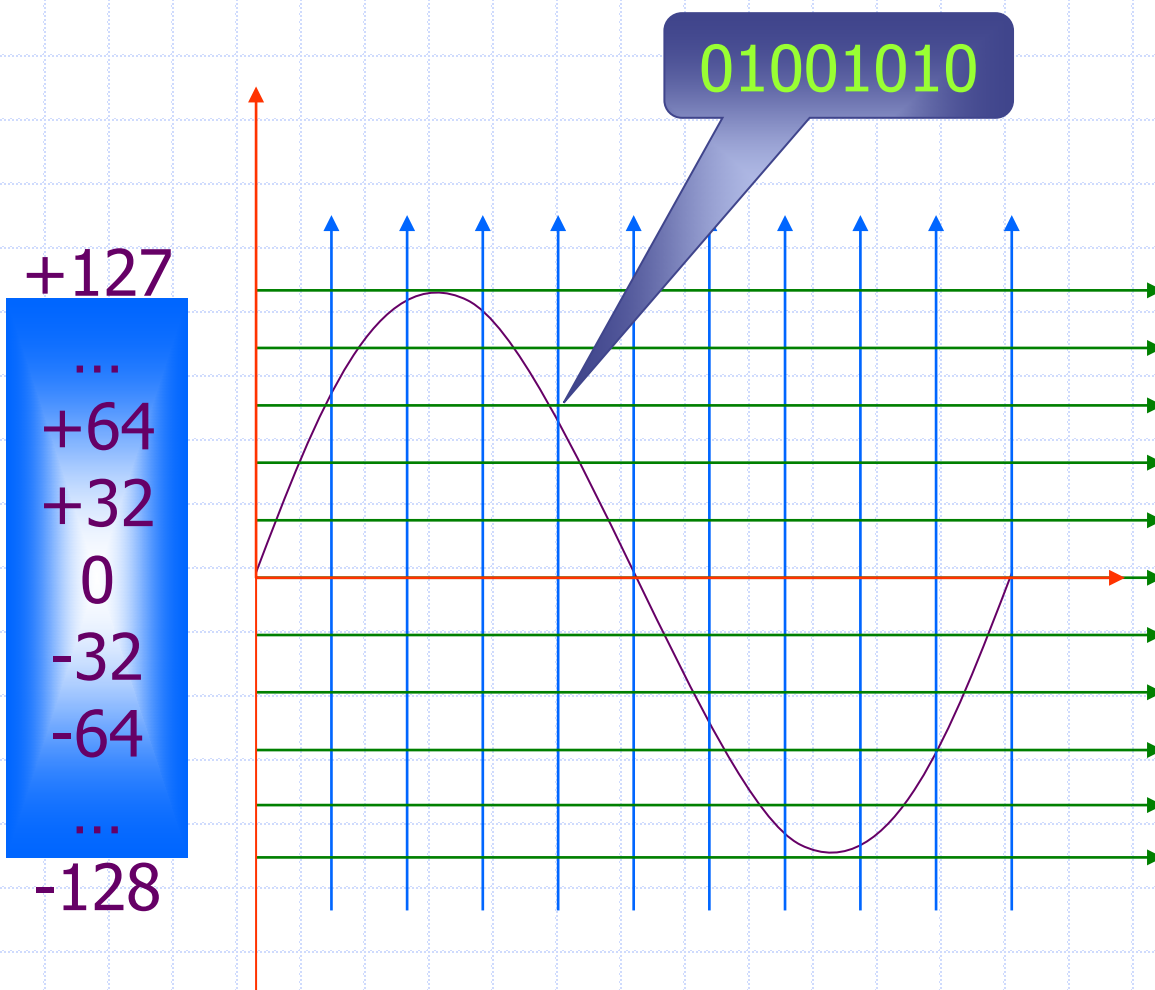


High sampling
rate: File size =
235 KB



© MediaTechnics

Comparison:

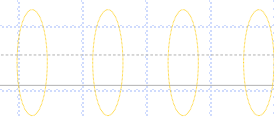


*Sampling
Resolution
8 bit*

*Sampling
Frequency
11 KHz*



Comparison:



0010101100011000

+32767

...

+512

+128

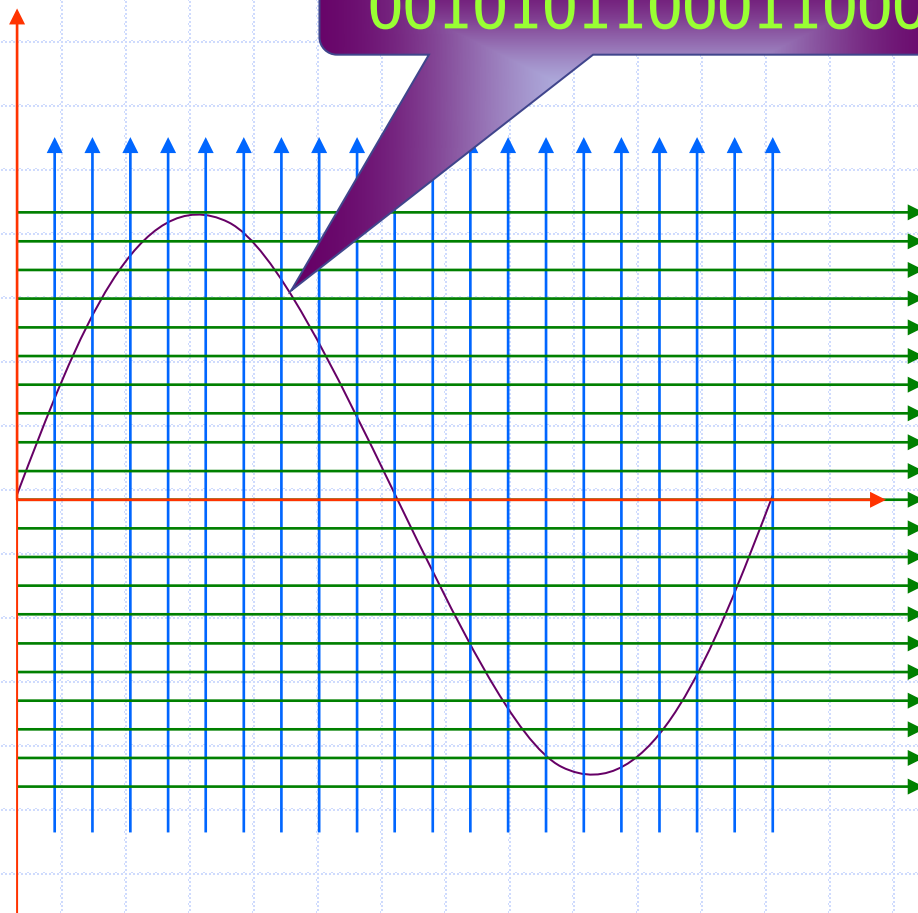
0

-128

-512

...

-32768



Sampling
Resolution
16 bit

Sampling
Frequency
22 KHz



量化/quantization

- ❖ **量化数据位数（也称量化级）**：每个采样点能够表示的数据范围，经常采用的有8 位、12 位和16 位。
- ❖ 8 位量化级表示每个采样点可以表示256 个不同量化
- ❖ 16位量化级则可以表示65536 个不同的量化值
 - 16位的声音数据比8位声音数据增大一倍
- ❖ **声道**
 - 每个声道单独采样和量化
 - 记录声音时，如果每次生成一个声道数据，称为**单声道**；
 - 每次生成两个声波数据，称为**立体声（双声道）**。

(2) 数字声音文件格式

- ❖ 常见的： WAV, WMA, AAC, MP3
- ❖ 音频转换时的参数设置
 - 码率：每秒钟的字节数Bps或每秒的位个数bps
 - ◆ $Bps = bps * 8$
 - ◆ VBR/CBR, 可变/不变码率
 - 采样率
 - ◆ 22kHz, 44kHz
 - 量化/精度
 - ◆ 8位/16位
 - 立体声
 - ◆ Mono/stereo

数字音频的存储量

存储量 = 采样频率 × 量化位数 ÷ 8

若使用双声道，存储量再增加一倍

例如，数字激光唱盘（CD-DA）的标准采样频率为44.1 kHz，量化位数为16位，立体声。一分钟 CD-DA 音乐所需的存储量为

$$44.1 \text{ K} \times 16 \times 2 \times 60 \div 8 = 10584 \text{ KB}$$



Audio Conversion

Sampling rate

- ☐ No change
☐ 8000Hz
☐ 11025Hz
☐ 16000Hz
☐ 22050Hz
☒ 44100Hz
☐ 48000Hz

☐ Custom

☐ Integral conversion

☐ High quality

Precision

- ☐ No change
☐ 8-bit
☒ 16-bit

Channels

- ☒ No change
☐ Mono
☐ Stereo
☐ Left
☐ Right

Bandwidth required: 87KB/s

OK

Cancel

Select audio compression

<No compression (PCM)>
 AC-3 ACM Codec
 AC-3 ACM Extensible
 CCITT A-Law
 CCITT u-Law
 GSM 6.10
 IMA ADPCM
 Microsoft ADPCM
 MPEG Layer-3
 MPEG Layer-3

320 kBit/s, 32.000 Hz, Stereo	40KB/s
320 kBit/s, 44.100 Hz, Stereo	40KB/s
320 kBit/s, 48.000 Hz, Stereo	40KB/s
256 kBit/s, 32.000 Hz, Stereo	32KB/s
256 kBit/s, 44.100 Hz, Stereo	32KB/s
256 kBit/s, 48.000 Hz, Stereo	32KB/s
224 kBit/s, 32.000 Hz, Stereo	28KB/s
224 kBit/s, 44.100 Hz, Stereo	28KB/s
224 kBit/s, 48.000 Hz, Stereo	28KB/s
192 kBit/s, 32.000 Hz, Stereo	24KB/s
192 kBit/s, 44.100 Hz, Stereo	24KB/s
192 kBit/s, 48.000 Hz, Stereo	24KB/s
160 kBit/s, 32.000 Hz, Stereo	20KB/s
160 kBit/s, 44.100 Hz, Stereo	20KB/s

☐ Show all formats

Format information

Format ID 0x0055
 Bytes per block 1 bytes
 Data rate 24000 bytes/sec
 Granularity 24000.0 blocks/sec

OK

Cancel

MP3 编码选项

编码质量

- ☒ 恒定码率 (CBR) 64 kbps
☐ 平均码率 (ABR) 64 kbps
☐ 变长码率 (VBR) Fast Standard

立体声

☒ 对于双声道采用联合立体声编码

确定

取消

MP3/mp3PRO?Encoder Options

Presets

Add

Delete

- ☒ CBR (Constant Bitrate) ☒ MP3
☐ VBR (Variable Bitrate) ☐ mp3PRO?

128 Kbps, 44100 Hz, Stereo (11.0:1)

☐ Convert to Mono


mp3PRO? audio coding technology licensed from Coding Technologies, Fraunhofer IIS and Thomson multimedia.

Constant Bitrate (CBR) encoding will use the same bitrate throughout the entire file. This is the most common method and is most predictable for bandwidth and file size.

OK

Cancel

Help

Advanced >>

©2000-2002 Syntrillium Software Corporation

Exercise 3.1

❖ 安装K-Lite_Codec_Pack_1155_Mega, 千千静听, Cool Edit Pro

1 用千千静听自带的格式转换, 将提供的mp3歌曲, 转换为wav格式

- 转换前的字节数, 转换后的歌曲字节数
- 查看mp3和Wav两个文件的MedioInfo信息, 读取信息:
 - ◆ Bit rate 数据率(bps)
 - ◆ Chanel 通道/立体声
 - ◆ Sampling rate 采样率(Hz)
 - ◆ Bit depth(量化/深度/高度)

2 用Cool Edit转换为160kbps/44100Hz/Stereo格式

- 查看转换前后的上述文件信息

(3) MIDI Music

- ❖ MIDI (Musical Instrument Digital Interface) 乐器数字接口，为电声乐器存储音乐数据制定的标准格式
 - 保持音符和控制参数，它指示MIDI设备如何演奏哪个音符以及多大音量等。
 - 类似于电子曲谱。
- ❖ MIDI音乐被解码为MIDI序列
 - Wavetable/波表/预先录制的声音
- ❖ 特点：
 - 声音文件小
 - 无法产生高质量的声音，无法产生真实回音
- ❖ 早期的手机，只支持midi铃声

(4) 语音合成与语音识别

- ❖ 语音合成/Text To Speech/TTS
 - 按字、词顺序播放事先录制好的音频，形成语音
- ❖ 语音识别
 - 将声音数据按人工智能方法，识别出成单词、文本、命令
 - 美国的电话声讯系统，基本都有自动语音识别功能
- ❖ IBM的语音识别系统
- ❖ 识别引擎
 - 微软 [Microsoft Speech SDK](#)
 - Google
 - 科大讯飞

目录/Agenda

3.1 数字声音/Digital Sound

3.2 图像/Images

3.3 矢量和三维图形/Vector and 3-D Graphics

3.4 数字视频/Digital Video

3.2 图像

Images

- ❖ 图像的数字化
- ❖ 数字图像的格式和压缩

图像

图像（*Image*）是指由输入设备捕捉的实际场景画面，或以数字化形式存储的任意画面。

静止的图像是一个矩阵，阵列中的各项数字用来描述构成图像的各个点（称为像素点 *pixel*）的强度与颜色等信息。这种图像也称为位图（*bit-map*）或者点阵图。



(1) 图像的数字化

- ❖ 图像 (Image) 与图形 (Graphics) 的区别
 - 图像是真实场景的反映，用栅格方式表示
 - 图形是人工绘制的，用抽象的数学模型表示
- ❖ 象与像
 - 我们认为，image应该翻译为“图象”

用于生成和编辑位图图像的软件通常称为“*paint*”程序。

图像文件在计算机中的存储格式有多种，如 BMP、PCX、TIF、TGA、GIF、JPG 等，一般数据量都较大。

图像处理时要考虑三个因素：

分辨率

图像深度与显示深度

图像文件大小



图像大小与像素

❖ 像素Pixel: Picture Element

- 图像的最小单位
- 每个像素一般用RGB三原色表示

❖ 分辨率: 数字化图像的大小, 即该图像的水平与垂直方向的像素个数。

- $1920 \times 1080 = 2,073,600$, 2百万像素
- $3264 \times 2448 = 7,990,272$, 8百万像素
- $5184 \times 3456 = 17,915,904$, 1千8百万像素

图像的量化和色彩深度

- ❖ **图像深度**（也称图像灰度、颜色深度）：表示数字位图图像中每个像素上用于表示颜色的二进制数字位数。
- ❖ **量化**：每种颜色的数值表示
 - 单色8位，三色/通道24位，四色32位
 - 单色12位，三色/通道36位，四色48位
- ❖ **灰度图像和彩色图像**
 - 灰度图像为单通道图像
 - 彩色图像为3/4通道图像，真彩色

颜色深度与显示的颜色数目

颜色深度	颜色总数	图像名称
1	2	单色图像
4	16	索引16 色图像
8	256	索引256 色图像
16	65536	HI—Color 图像
24	16672216	True Color 图像



图像文件大小

用字节表示图像文件大小时，一幅未经压缩的数字图像的数据量大小计算如下：

图像数据量大小 = 像素总数 × 图像深度 ÷ 8

例如：一幅 640×480 的 256 色图像为
 $640 \times 480 \times 8 / 8 = 307200$ 字节

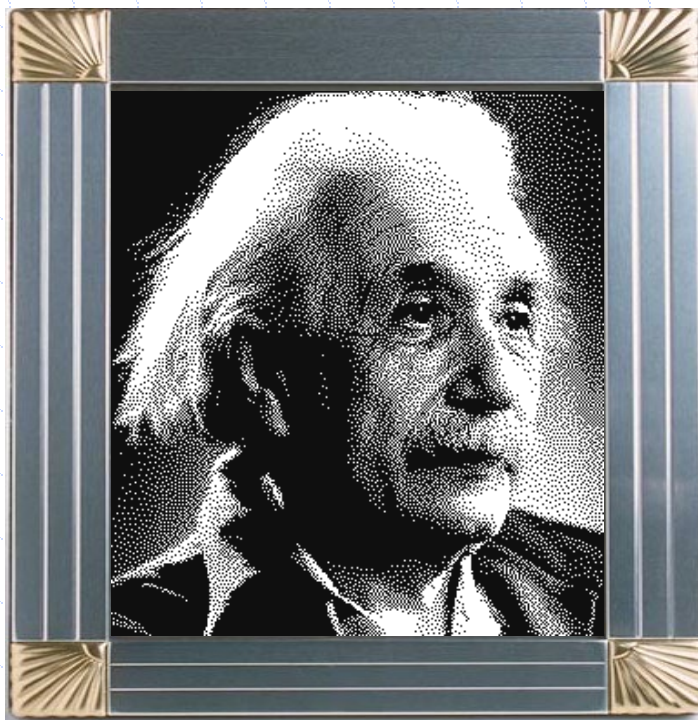


图像文件大小

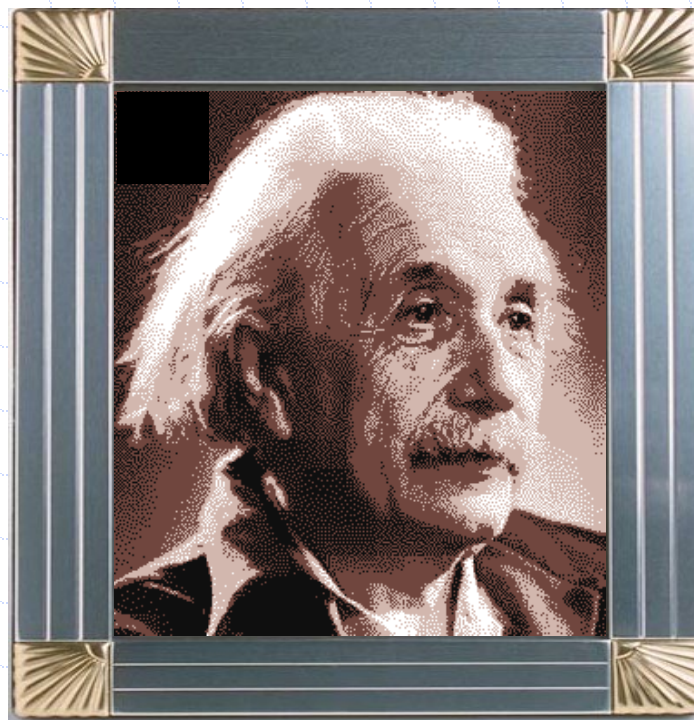
❖ 三色的普通位图存储所需空间 (不压缩)

- $1920 \times 1080 = 2,073,600$, 2百万像素
 - ◆ $24/8 \times 1920 \times 1080 = 6\text{百万字节} = 6\text{MBytes}$
- $5184 \times 3456 = 17,915,904$, 1千8百万像素
 - ◆ $24/8 \times 5184 \times 3456 = 5\text{千}4\text{百万字节} = 54\text{MBytes}$

Comparison:



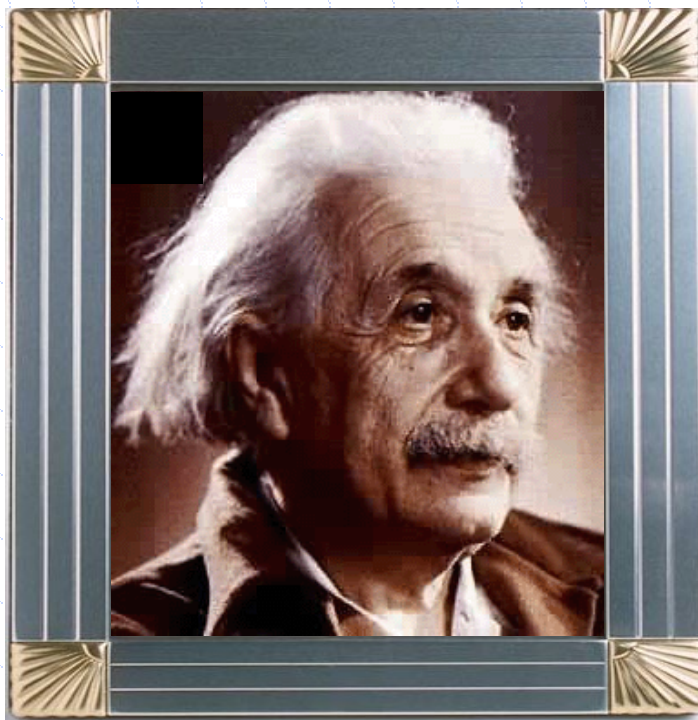
宽度：271
高度：300
颜色：2
大小：9.9 KB



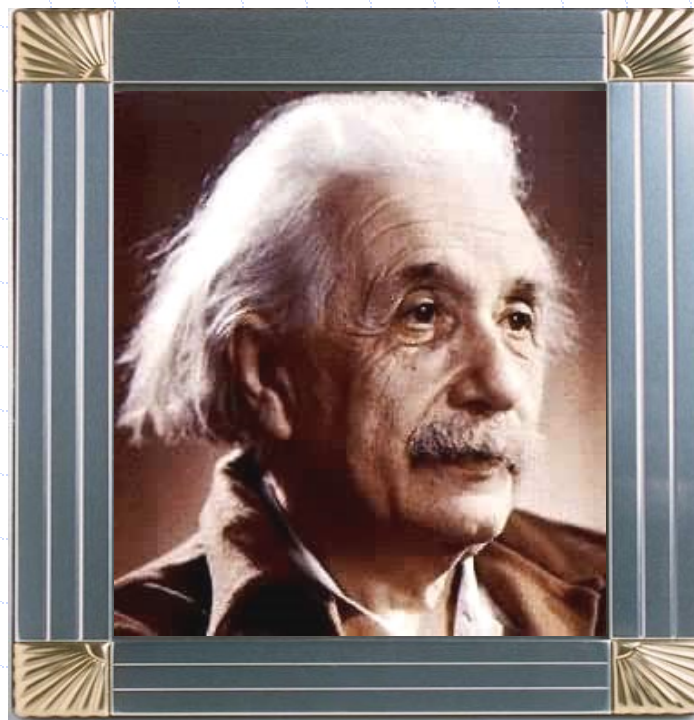
宽度：271
高度：300
颜色：4
大小：19.8 KB



Comparison:



宽度：271
高度：300
颜色：256
大小：79.4 KB



宽度：271
高度：300
颜色：真彩色
大小：238.2 KB



(2) 图像的压缩

❖ 无损压缩

- 去除冗余的信息
- 如图像中有一串序列是 0000 0000 0000 0000
- 则可以表示为：0x04 0x00 (4个0x00)

❖ 有损压缩

- 图像像素之间存在强关联，因为图像是连续变化的
 - 从图像的频谱特征来说，高频分量表示剧烈变化的部分，如边缘和细节，低频分量表示平缓变化的部分，如背景
 - ◆ 因此，可以对高频部分和低频部分，实行不同的量化策略
 - 本质上说，是欺骗人类视觉系统
- ❖ 普通文件的压缩，如zip, rar, 都必须是无损压缩
 - ❖ 图像视频文件的压缩，可以有损，也可以无损

(3) 文件格式

- ❖ **Bmp** (Bit map), windows的位图格式, 一般为不压缩, 即使压缩也是无损压缩
- ❖ **Png** (Portable Network Graphic Format), 新格式的位图, 无损压缩
- ❖ **Raw**, 数码相机的未处理格式, 不压缩, 单通道12位
- ❖ **Jpg/jpeg**, Jpeg格式, 有损压缩
- ❖ **Tiff**, 有损压缩
- ❖ **Gif**, 有损压缩, 可以由多幅图像构成简单的动画

1. BMP格式

BMP 是标准的WINDOWS 和OS/2 的图形和图像的基本位图格式，有压缩（RLE）和非压缩之分。BMP 支持黑白图像、16色和256色的伪彩色图像以及RGB 真彩色图像。

2. PCX格式

是使用游程长编码（RLE）方法进行压缩的图像文件格式文件。支持黑白图像、16色和256色的伪彩色图像、灰度图像以及RGB 真彩色图像，



3. GIF格式

GIF 是压缩图像存储格式，它使用LZW 压缩方法，压缩比较高，文件长度较小。支持黑白图像、16色和256色的彩色图像。

4. TIF格式

TIF 格式是工业标准格式，支持所有图像类型。文件分成压缩和非压缩两大类。



5. JPG和PIC格式

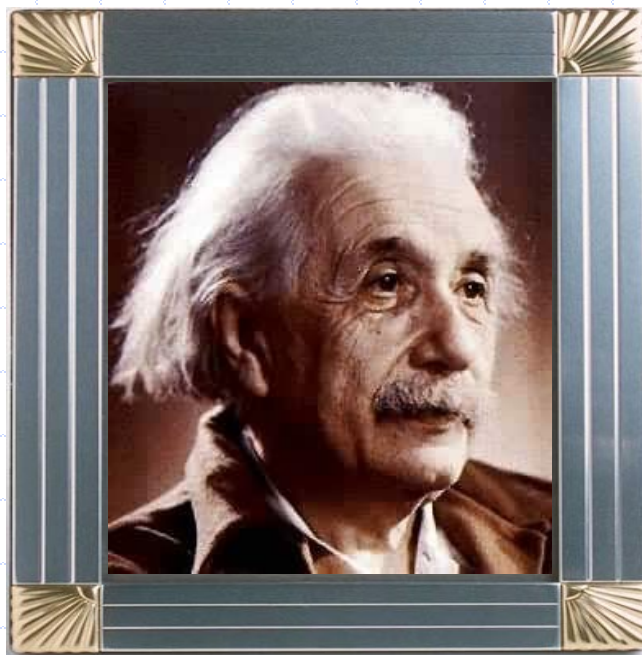
JPG 和PIC 都使用JPEG 方法进行图像数据压缩。这两种格式的最大特点是文件非常小。它是一种有损压缩的静态图像文件存储格式。支持灰度图像、RGB 真彩色图像和CMYK 真彩色图像。

6. PCD格式

PCD格式是Photo-CD 的专用存储格式，文件中含有从专业摄影照片到普通显示用的多种分辨率的图像，所以数据量都非常大。



Comparison:

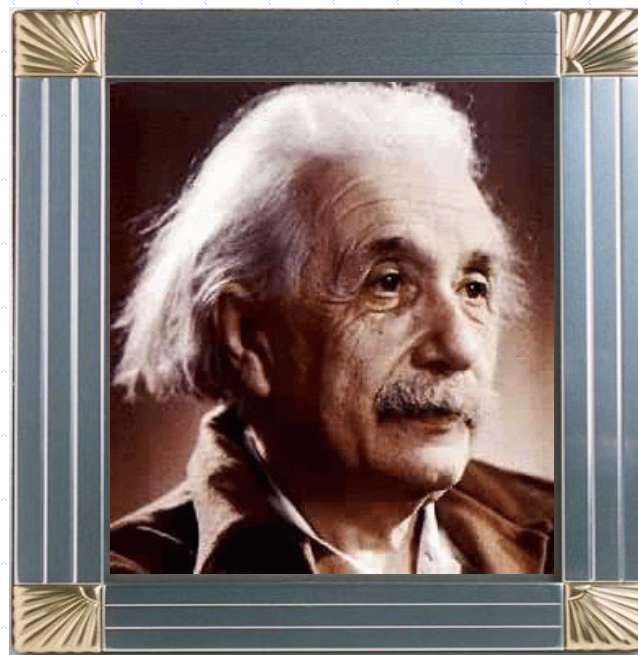


Einstein.bmp

颜色：真彩色

大小：239.1 KB

压缩比：1.0



Einstein.gif

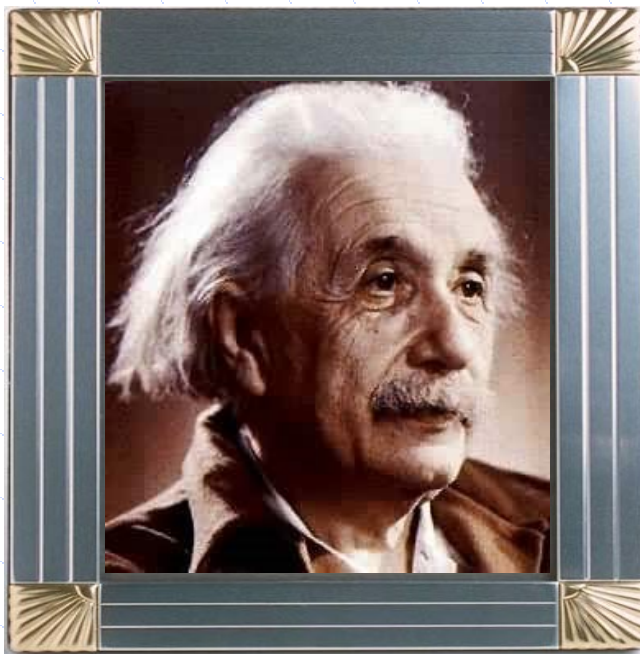
颜色：256

大小：61.4 KB

压缩比：1.3



Comparison:

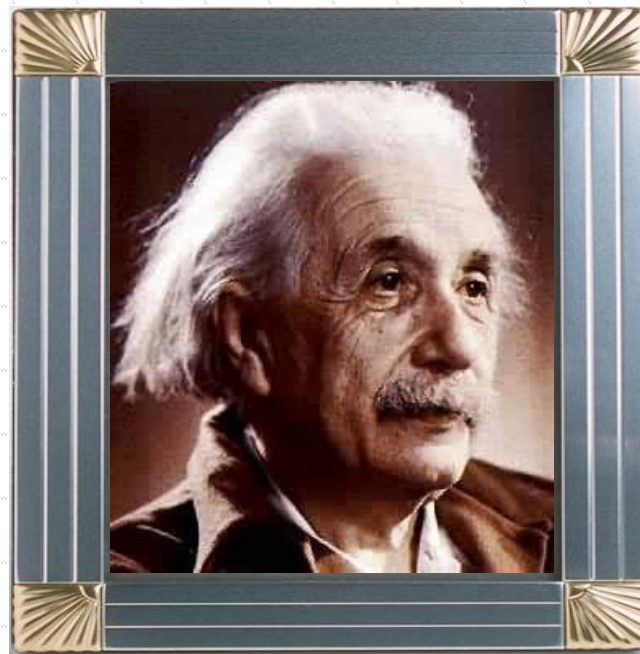


Einstein.jpg

颜色：真彩色

大小：11.1 KB

压缩比：21.4



Einstein.tif

颜色：真彩色

大小：238.4 KB

压缩比：1.0



(4) 图像打印密度/pixel density

- ❖ 图像密度pixel density, 指每英寸所打印的点数, 单位为ppi (pixels per Inch), 又是也称resolution
- ❖ 扫描图像时, 通常会选择150/200/300/600dpi, 从而将原始图像扫描为数字图像的ppi

A4纸大小: 210mm x 297mm 按 300dpi
= 3.27in x 11.7 in
= 2480 x 3508 像素

- ❖ 因此, 在打印图像时, 需要设置resolution的值, 否则图像会变形
- ❖ 注意, 像素密度与像素分辨率的区别, 通常根据前后文决定

(5) 图像的参数

- ❖ 分辨率：如1024 x 768
- ❖ 通道/位深：单色/彩色3通道/24位/32位
- ❖ 文件格式：jpg/bmp/png/tif/gif
- ❖ 压缩率
- ❖ Dpi值
- ❖ 相机照片的Exif信息
 - 包含相机型号/照片的日期/gps位置等

Exercise 3.2

- ❖ 安装acdsee、photoshop软件
- 3 用mspaint/画图工具，打开给定的图片，转换为bmp/png/jpg格式
- 4 用acdsee，打开给定的图片，查看相机型号/照片日期信息，然后缩小，并按70%的压缩率转换为jpg图像
- 5 用photoshop软件，修改图片的resolution

目录/Agenda

3.1 数字声音/Digital Sound

3.2 图像/Images

3.3 矢量和三维图形/Vector and 3-D Graphics

3.4 数字视频/Digital Video

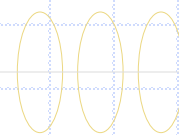
3.2 矢量和三维图形

Vector and 3-D Graphics

- ❖ 矢量图形的概念
- ❖ 矢量图形转换为图像/光栅化

图形

- 图形 (Graphic) 一般指用计算机绘制的画面, 如直线、圆、圆弧、矩形、任意曲线和图表等。
- 图形的格式是一组描述点、线、面等几何图形的大小、形状及其位置、维数的指令集合。
- 在图形文件中只记录生成图的算法和图上的某些特征点, 因此也称**矢量图**。
- 矢量图形的文件容量较小, 可以在进行放大、缩小或旋转等操作时保持图像的清晰度, 并且与分辨率无关。它们适用于图形设计、文字设计、标志设计、版式设计等领域, 但难以表现色彩层次丰富的逼真图像效果。
- 常见的矢量图形格式包括**Adobe Illustrator**的*.ai、*.eps和.svg、**Auto CAD**的*.dwg和*.dxf、**Corel DRAW**的*.cdr等。



(1) 矢量图形

- ❖ 包含重建图片的指令，包括绘制对象的形状、尺寸、位置、颜色等指令
- ❖ 这些对象均可修改

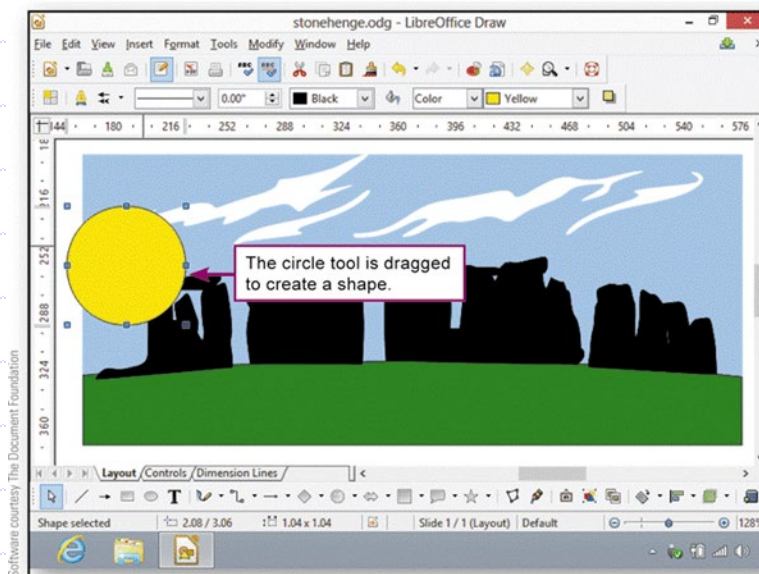


FIGURE 8-32

To draw a circle, select the filled circle tool, and then drag the mouse pointer to indicate the circle's location and size. A color palette allows you to select the circle color. After you create the circle object, you can move it and change its size or color. You can also create irregular shapes for objects, such as clouds, by connecting short line segments. Learn the basics of drawing vector images by accessing this figure in your interactive eBook.



(2) 光栅化/Rasterization

- ❖ 矢量图形可通过光栅化，转换为图像
- ❖ 转换时，需要指定分辨率
- ❖ 转换后不能恢复

FIGURE 8-34

When vector images are rasterized, they become bitmaps and can't be enlarged without becoming pixelated.



© MediaTechnics

(3) 3D图形

- ❖ 3D图形也是一组指令
- ❖ 在计算机中，三维图形的绘制通常需要使用特定的图形处理软件和算法来实现。这些算法可以通过在二维平面上使用数学模型来模拟光线追踪、阴影、纹理等三维效果，从而生成逼真的三维图形。
- ❖ 三维图形可以用于各种应用，包括游戏制作、电影特效、虚拟现实、科学可视化等领域。

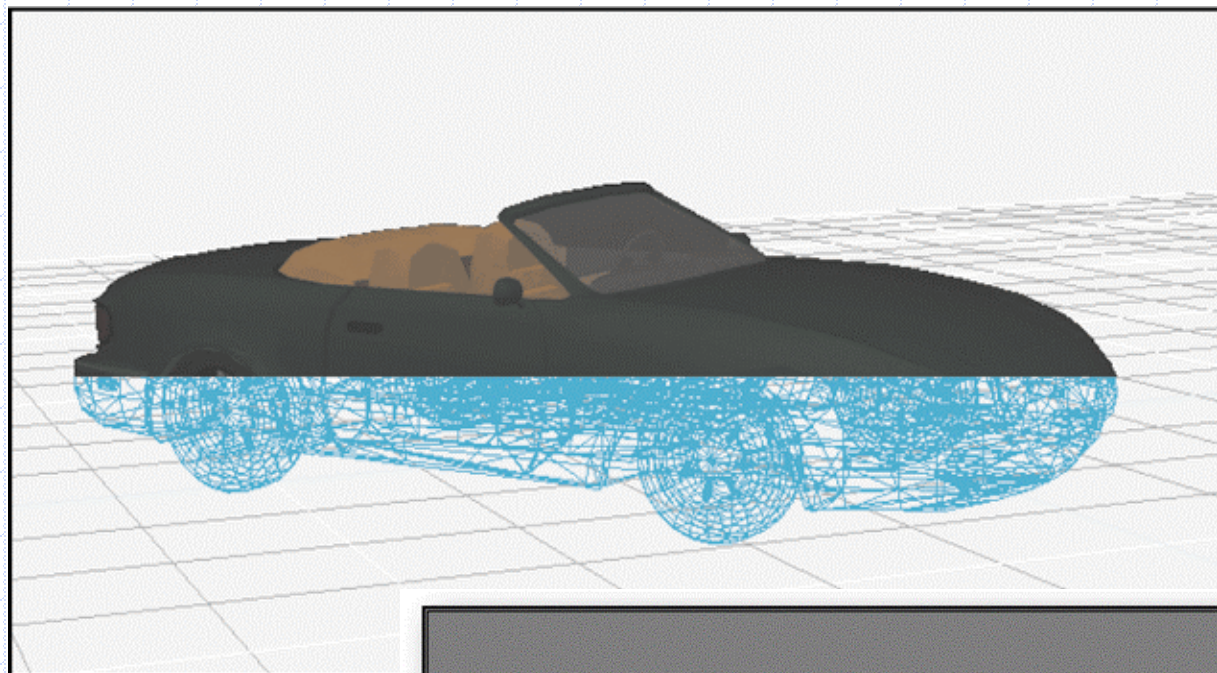


FIGURE 8-36

3-D graphics are based on a wireframe, which can be rendered into a bitmap image that looks three-dimensional.

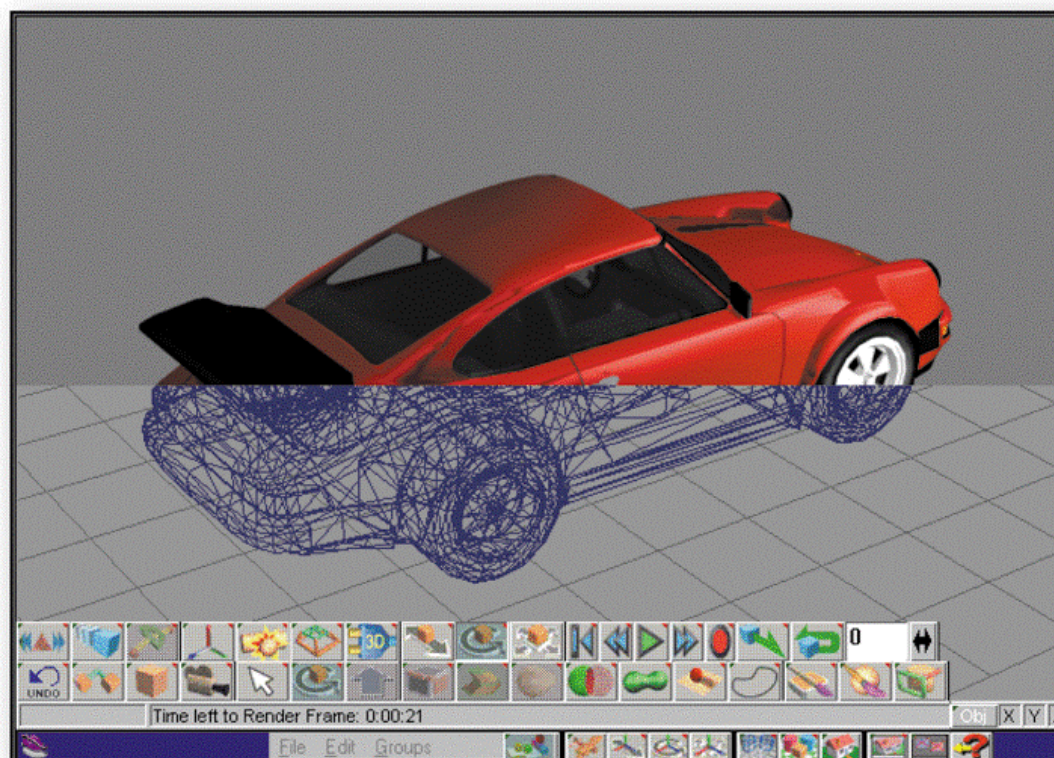


FIGURE 8-38

3-D graphics software provides tools for drawing a wireframe and then specifying colors and textures for rendering.

▶ Watch a wireframe being rendered and animated in your interactive eBook.

Exercise 3.3

❖ 下载“微易CAD图形绘制设计器”软件

6 用两个三角形绘制一个六角形，设置成透明线性

目录/Agenda

3.1 数字声音/Digital Sound

3.2 图像/Images

3.3 矢量和三维图形/Vector and 3-D Graphics

3.4 数字视频/Digital Video

3.2 数字视频

Digital Video

- ❖ 视频文件格式
- ❖ 视频文件格式转换/编码

视频

视频是由一幅幅单独的画面序列（帧frame）组成，这些画面以一定的速率（fps）连续地投射在屏幕上，使观察者具有图像连续运动的感觉。

视频文件的存储格式有AVI、MPG MOV等。



(1) 视频文件格式

- **AVI**。由微软公司发布的视频格式。
- **WMV**。独立于编码方式的在Internet上实时传播多媒体的技术标准。
- **MPEG**。此格式包括了MPEG-1、MPEG-2和MPEG-4在内的多种视频格式。
- **MP4**。此种格式是一种适合于在移动设备上播放的视频文件格式。
- **MOV**。此种格式是Apple公司开发的一种视频文件格式，被广泛使用。
- **MKV**。此种格式是一种多媒体封装格式，它主要的特性为能容纳多种不同类型编码的视频、音频及字幕流。主要优势在于其能在一个文件中容纳无限数量的视频、音频、图片或字幕轨道，因此常被用作保存电影、电视节目等多媒体内容。
- **ASF**。全称是Advanced Streaming Format，是微软开发的一种流媒体格式。
- **FLV**。此种格式是Adobe Flash Player使用的视频文件格式，被广泛用于网页上的在线视频播放。
- **F4V**。此种格式也是一种流媒体格式，被广泛用于网络视频播放。
- **RMVB**。全称是RealMedia Variable Bitrate，是RealNetworks公司开发的一种视频文件格式，主要被用于在线视频播放。
- **RM**。全称是RealMedia，是RealNetworks公司开发的一种流媒体格式，主要用于在线视频播放。
- **3GP**。此种格式主要被用于移动电话上的视频播放。

❖ 每一种视频文件，必须有相应的解码器才能播放

视频和动画的文件格式

动画文件的格式主要有两种：

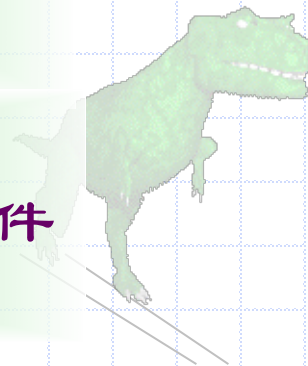
FLIC 格式和MMM 格式。

1. FLIC动画

早期版本的FLIC文件只支持 $320 \times 200 \times 256$ 色模式，文件的扩展名为“.FIY”。较新版本支持的分辨率和颜色数都有所提高，文件的扩展名也改为“.FLC”。它使用了无损压缩方法，画面效果十分清晰，但本身不能存储同步声音。

2. MMM动画

MMM格式是微软多媒体动画的文件格式。



视频文件大小计算

- ❖ 完整的视频文件是由**音频流与视频流**2个部分组成的，音频和视频分别使用的是不同的编码率，因此一个视频文件的最终技术大小的编码率是音频编码率+视频编码率。例如一个音频编码率为128Kbps，视频编码率为800Kbps的文件，其总编码率为928Kbps，意思是经过编码后的数据每秒钟需要用928K比特来表示。
- ❖ **视频输出文件大小公式：**
$$\left(\text{音频编码率 (Kbit为单位)} / 8 + \text{视频编码率 (Kbit为单位)} / 8 \right) \times \text{影片总长度 (秒为单位)} = \text{文件大小 (MB为单位)}$$
- ❖ 一部MKV高清电影（30Mbps码率）一分钟影片容量：
 - $30000\text{kbps} \times 60 / 8 = 225000\text{KB} = 219.7265625\text{MB}$
 - 一部90分钟电影约**19.3G**

(2) 视频文件转换/编码器

- ❖ 视频文件转换时，需要
 - 解码器：对原格式解码
 - 编码器：对新格式进行编码
- ❖ 许多编解码器是有专利保护的
 - 如mpeg2/dvd

多媒体图像数据压缩技术

❖ 运动图像压缩编码的国际标准：

- **MPEG-I**：针对运动图像的数据压缩技术，其图像质量是家用录像机的质量
- **MPEG-II**：运动图像及其伴音的通用压缩技术标准
- **MPEG-IV**：高清晰度电视
- **MPEG-VI**：低数据率的电视节目标准，主要用于交互式多媒体场合

多媒体音频数据压缩技术

❖ **音频压缩编码**的国际标准：高保真音频数据压缩的国际标准（MPEG）。该标准提供三个独立的压缩层次：

❖ **第一层**：用于数字录像机，压缩后的数据速率为384kbps

❖ **第二层**：用于数字广播的音频编码、CD-ROM的音频信号以及VCD的音频编码，压缩后的数据速率为192kbps

❖ **第三层**：MP3和Microsoft公司的Windows Media 文件格式，压缩后的数据速率为64kbps

(3) 相关软件

- ❖ K-Lite Mega Codec Pack
 - 经典的编码/解码器包
 - 支持几乎所有的视频格式
- ❖ VirtualDub
 - 经典的视频编辑软件

Exercise 3.4

❖ 下载klitecodec和virtualdub

7 打开virtualdub, 将给定视频文件的格式转换为mp4