



Computer Culture Advanced

信息表示与处理3

感知型数字化

连续物理信号经过采样、量化、编码、转化为二进制

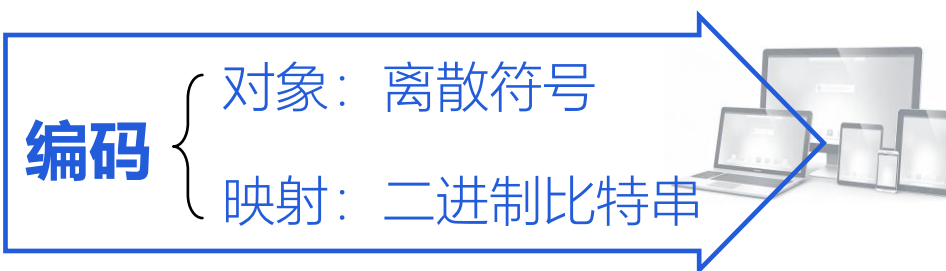
数字化：感知型 与 逻辑型

所有信息进入计算机前，都必须被**数字化**，不同的信息，有不同的**转换**策略。



$$y = f(x)$$

符号变换规则



逻辑型数字化（核心动作编码）

输入：离散符号

映射：人为约定的精确规则

特点：无近似、无误差

例子：数值、文本、矢量图形、程序代码



（基于规则绘制的图形）可无限缩放

数字化：感知型 与 逻辑型

所有信息进入计算机前，都必须被**数字化**，不同的信息，有不同的**转换**策略。



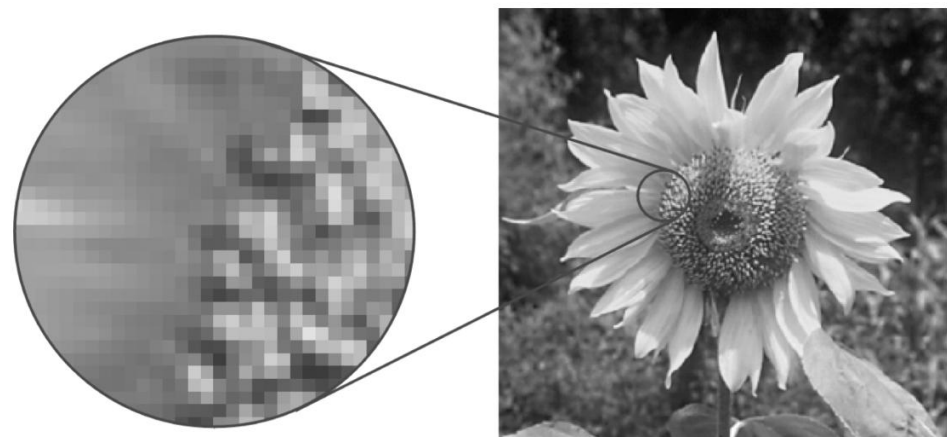
感知型数字化

输入：物理世界**连续**信号

映射：**采样 + 量化 + 编码**

特点：**近似记录，存在误差与权衡**

例子：图像、声音、视频



(基于有限记录还原的**图像**) 缩放失真

本次课的主题：感知型数字化

感知型数字化

连续物理信号 → 采样 + 量化 完成 离散化 → 编码映射成 0/1 →
计算机处理和保存 f → 解码还原感知信息

$$y = f(x)$$

- 音频数字化（一维时域信号）

声波连续幅值 → 采样（时间取样） + 量化（幅值分级） → 编码（二进制映射）

要点：奈奎斯特原则、采样率、频率混叠、量化误差等

- 图像数字化（二维空域信号）

光影连续亮度 → 采样（空间取样 = 像素） + 量化（亮度 / 色彩分级） → 编码（二进制映射）

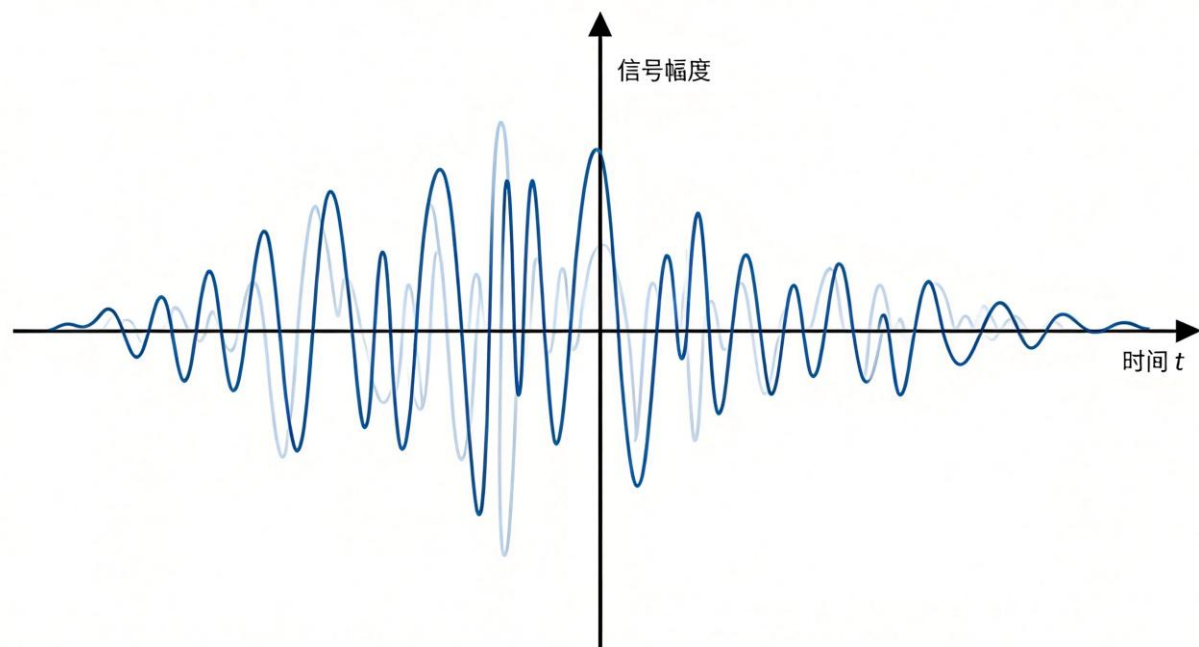
要点：空间混叠、色彩模型、色深等

- 数据压缩（海量的多媒体数据实用化的前提）

压缩：借助编码实现更紧凑的映射，分为无损格式和有损格式

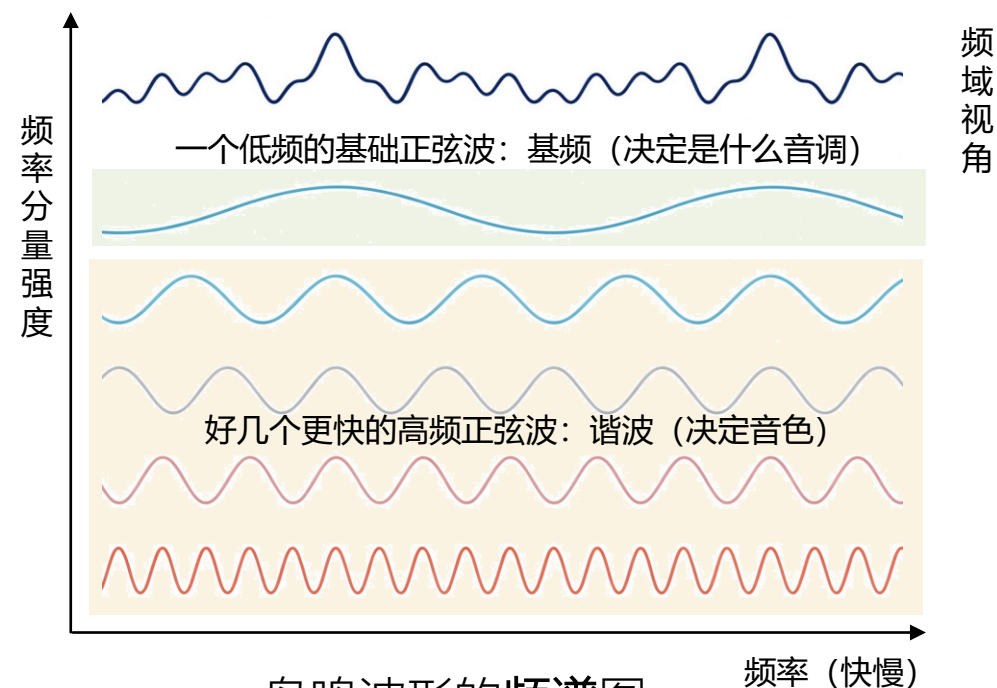


认识连续的声音波形信号



一段鸟鸣的（时域）波形图

真实声音波形（一条不规则的振动曲线）
很难说清楚有哪些成分、高音多还是低音多



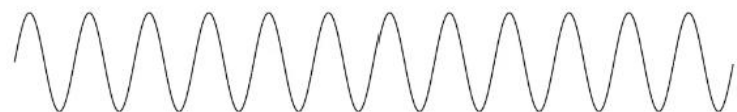
鸟鸣波形的频谱图

把这个声音“拆开”，列出它里面包含的所有
振动频率，以及每一种振动有多强（频谱图）

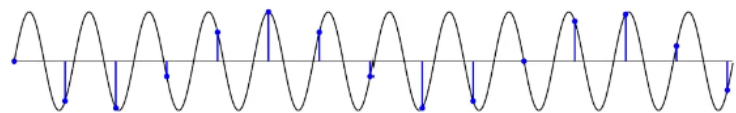


声音波形数字化（采样频率与混叠现象）

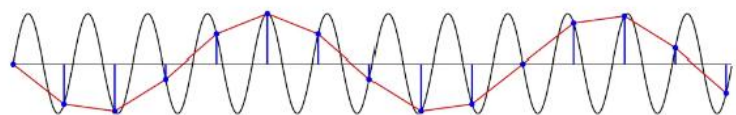
单独拿出一根波纹（单一频率的正弦波），做采样实验，观察采样太慢的时候，混叠是怎么发生的



假设要把这个**模拟**波形转化为**数字**信号



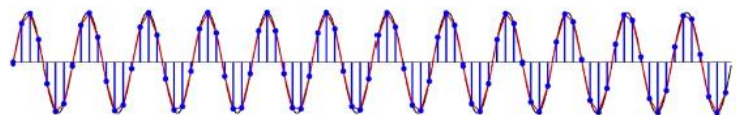
采样速度不足（**采样频率过低**），只能采集到**稀疏的蓝色采样点**



当系统根据这些稀疏采样点还原波形时，可能会把它误判为一个**频率更低**的红色波形



采样点太少时，高频信号会被误判成一个更低频率的波形，这个**虚假低频**就是高频信号的**别名**。**混入**原始声音之后，真假低频叠加在一起，我们无法区分，这个现象就是**混叠**（aliasing）



为了保证不会认错波形、准确还原原始信号，采样频率不能无限制降低

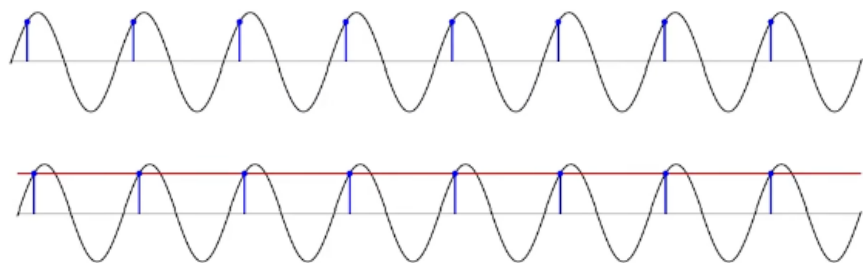
采样频率至少要多高，才能避免混叠



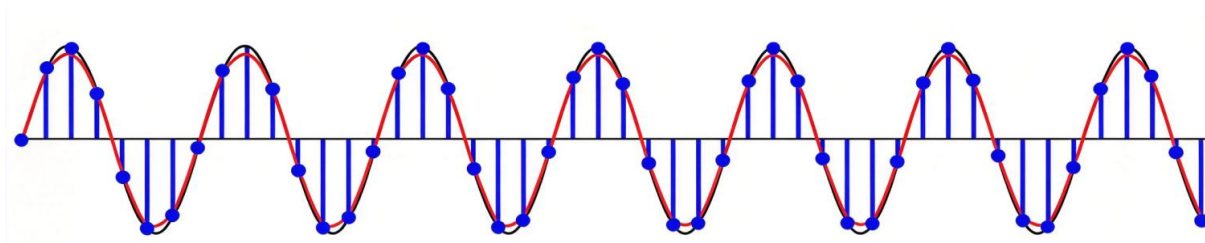


采样下限准则：奈奎斯特（Nyquist）准则

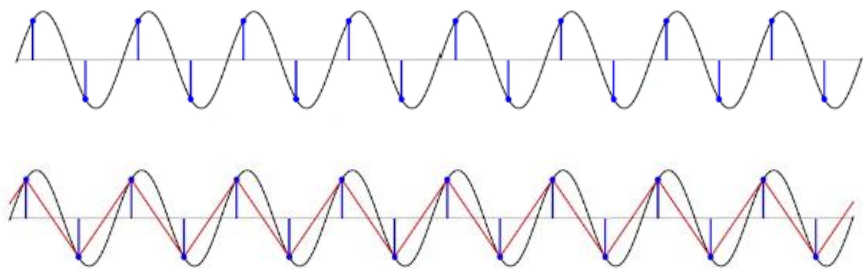
奈奎斯特准则：为了避免混叠，**采样频率必须大于信号中最高频率分量的 2 倍**。
这个两倍值就是采样频率的下限，也叫**奈奎斯特频率**。



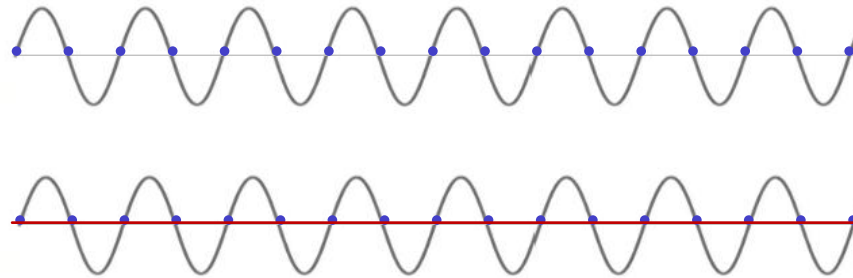
❶ 每周期仅 1 个采样点 → 翻译成了 0Hz（混叠特例）



❸ 每个周期采样多于 2 个点：采样频率 > 2 倍信号中最高频率分量，能够稳定识别频率，避免混叠



采样点落在波峰、波谷 → 碰巧可以读出原始频率 采样点落在零点 → 波动信息可能直接丢失，误判为无变化信号

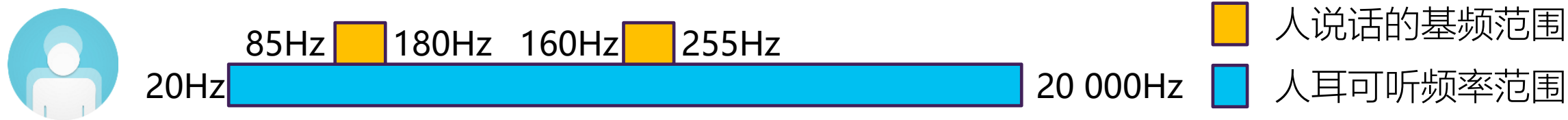


❷ 每周期 2 个采样点（临界，不稳定）



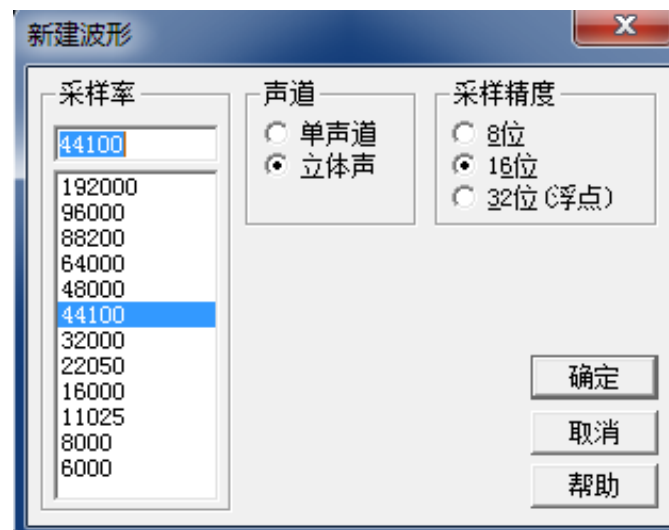
音频采样：适配人耳听觉的采样频率

奈奎斯特准则：为了避免混叠，**采样频率，必须大于信号当中最高频率分量的 2 倍。**这个两倍值就是采样频率的下限，也叫**奈奎斯特频率**。



采样频率决定音频能够稳定记录的最高信号频率
(**最高信号频率 < 采样率 / 2**)：

- 8,000 Hz：电话音质，仅保留语音核心频段
- 11,025 Hz：低音质音频采样率
- 22,050 Hz：中音质 / FM 收音机级音质
- 44,100 Hz：高保真音频，完整覆盖人耳20kHz听觉上限
- 48,000 Hz及更高：数字电视、DVD等专业影音场景

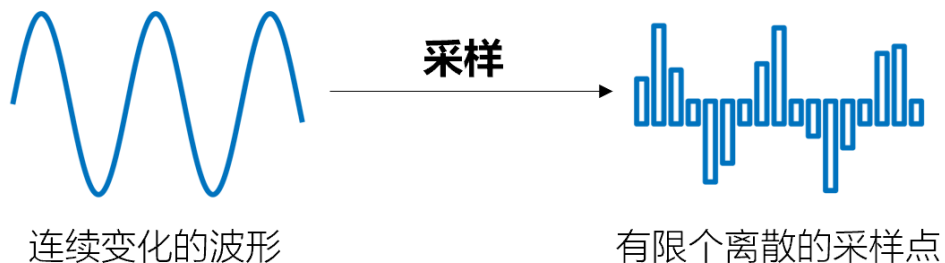


为什么数字录音时经常看到44100Hz的采样频率

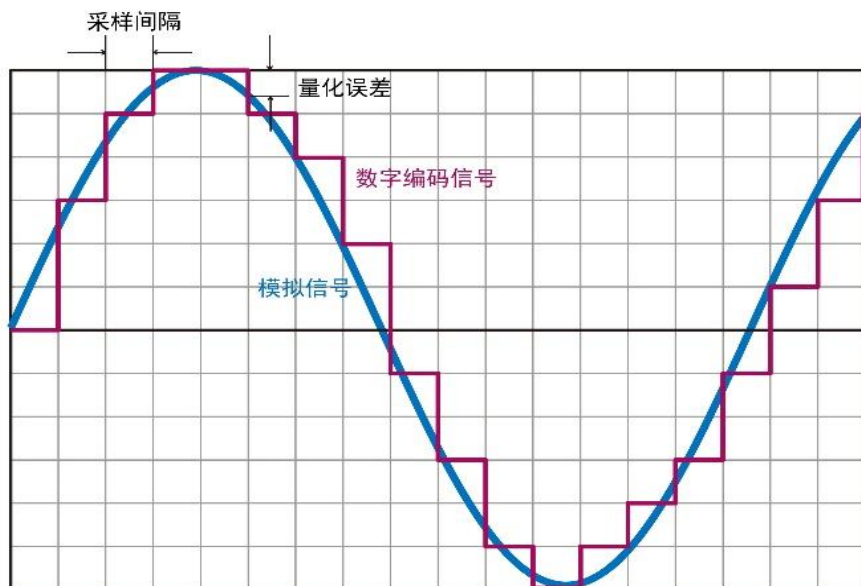




离散化 = 采样 + 量化



声音波形 ↔ 数字音频



采样 = 时间尺子

每隔多久采集一次信号,一串带有天然时序的样本符号序列。满足奈奎斯特原则, 保证信号里的频率成分不会认错。

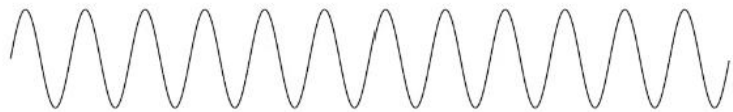
量化 = 幅值尺子

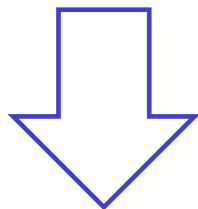
给连续的信号幅值划分有限刻度, 把采样点的幅值就近归到刻度上。它决定波形起伏能被复刻到多精细。

有限刻度 = 幅值台阶



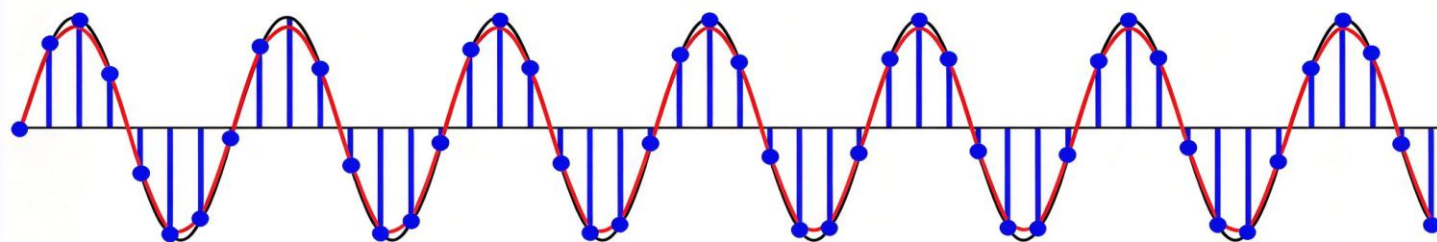
音频编码：音频符号样本的幅值编码

现实世界**连续**的物理信号 



离散化 (采样 + 量化)

数字化 = **离散符号** + **编码** = 符号与比特数串的一一映射



音频样本是一串**带有天然时序的样本符号序列**：样本的先后顺序直接代表时间位置，顺序本身承载时序信息，不需要额外编码，**每个样本符号里需要编码的有效信息，是它对应的幅值台阶。**

选用多长的二进制串，来给**幅值台阶**做**编码**？

- 256级： 8bit
- 65536级： 16bit



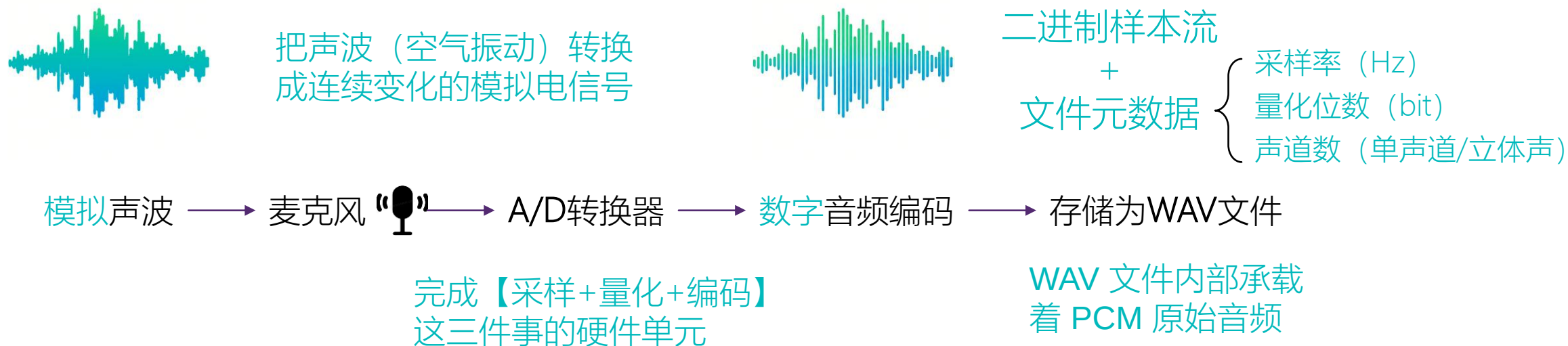


从声音波形物理量到数字音频

概念流程

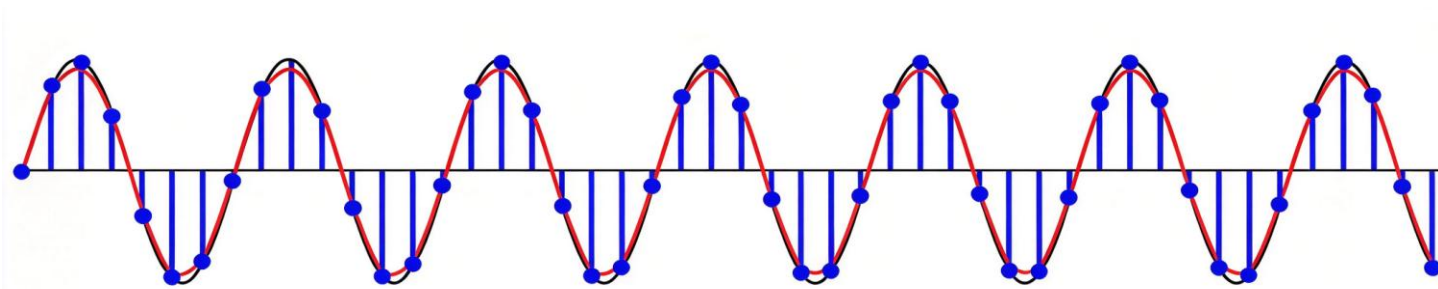


工程实现流程





音频文件



$$\text{数字音频数据量} = \text{采样频率 (每秒样本点数量)} \times \text{每个样本的量化位数 (bit)} \times \text{时长} \times \text{声道数}$$

计算题

一首时长 3 分 25 秒的立体声（双声道）歌曲，采样频率 44.1kHz，每个样本量化位数 16bit，不做任何压缩，求原始 PCM 音频的数据量（单位：KB）



PCM（脉冲编码调制）：直接存储采样、量化得到的原始样本编码，无压缩，是音频最基础的原始存储形式。

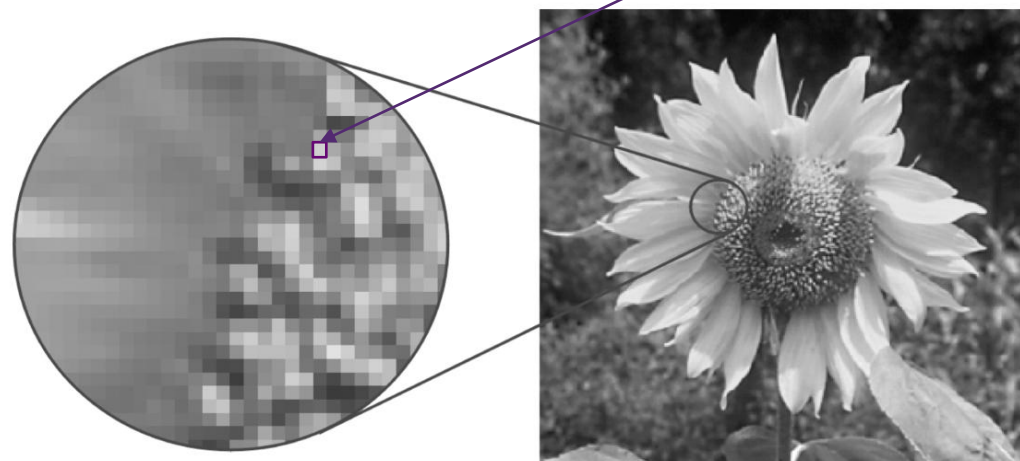
数字图像（光影连续亮度的离散化和编码表示）

图像（Raster Graphics）是用**像素**（Pixel，能被独立赋予颜色和亮度的最小单位）阵列记录视觉画面的光影信息，通过数码相机拍摄、扫描仪扫描、屏幕截图等方式获取。

图像的特点：

- **细节丰富**：每个像素都可独立设置颜色和亮度，能精准还原自然场景的色彩层次、明暗变化和细节纹理，适用于呈现照片、实景画面等复杂视觉内容
- **缩放失真**：放大图像时，像素矩阵会被拉伸，导致画面细节模糊；缩小图像时，会丢失部分像素，可能导致细节损失
- **文件量大**：需存储每个像素的颜色、亮度等信息，分辨率越高、色彩越丰富，文件体积越大

通过空间维度的采样，自带平面位置信息的光影离散样本，也就是**像素阵列**



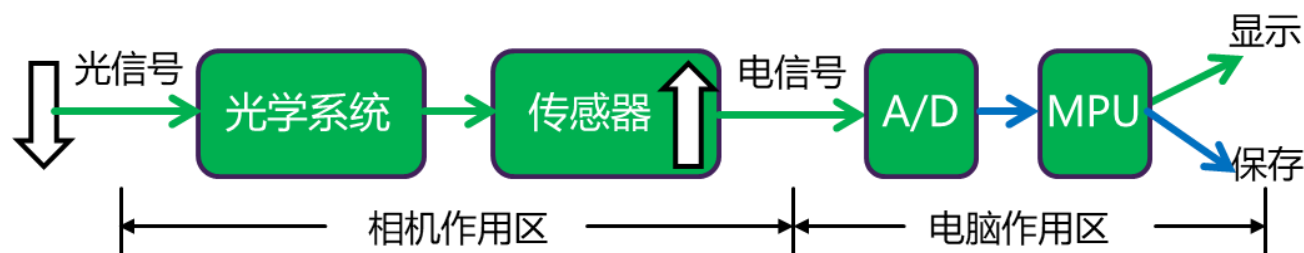
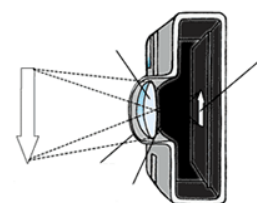
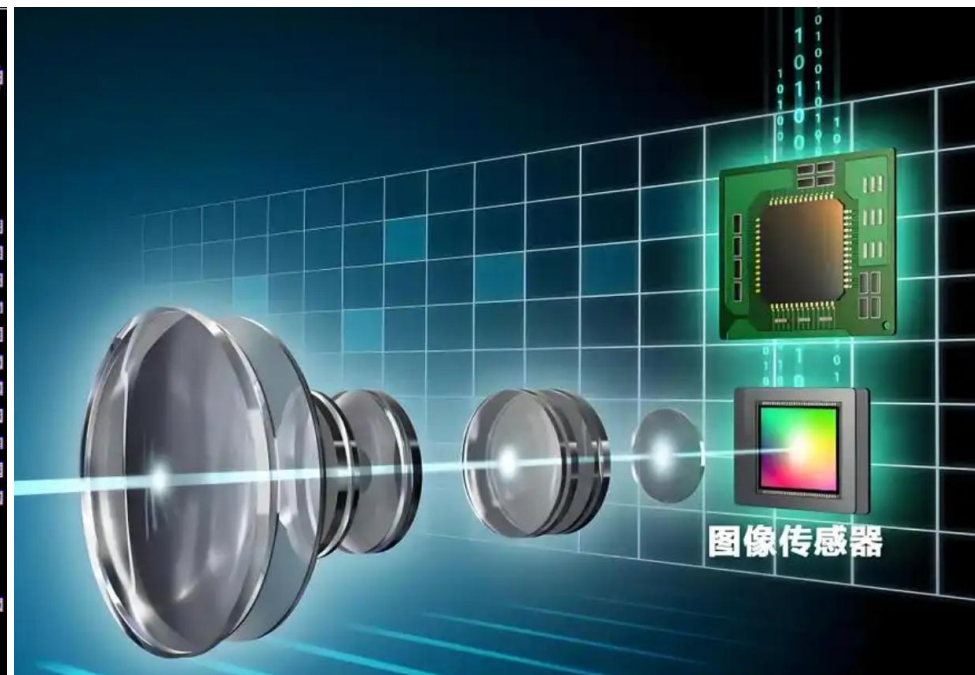
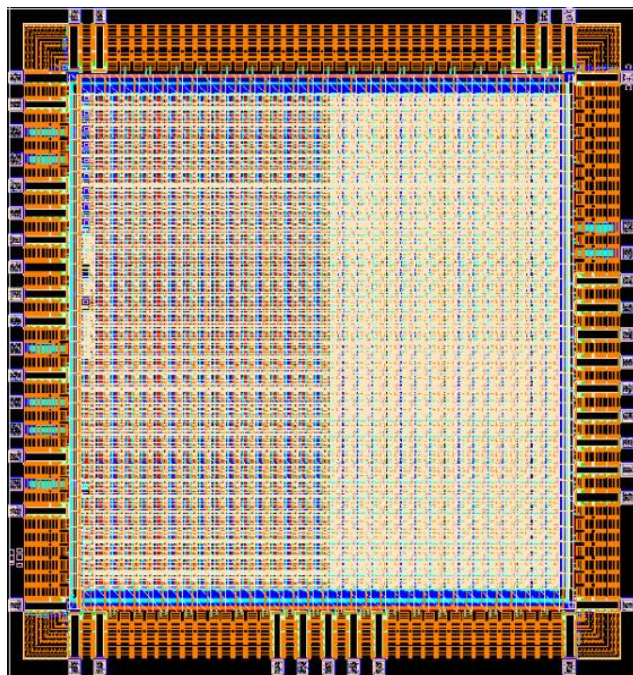
从数码相机视角理解连续光影的采样

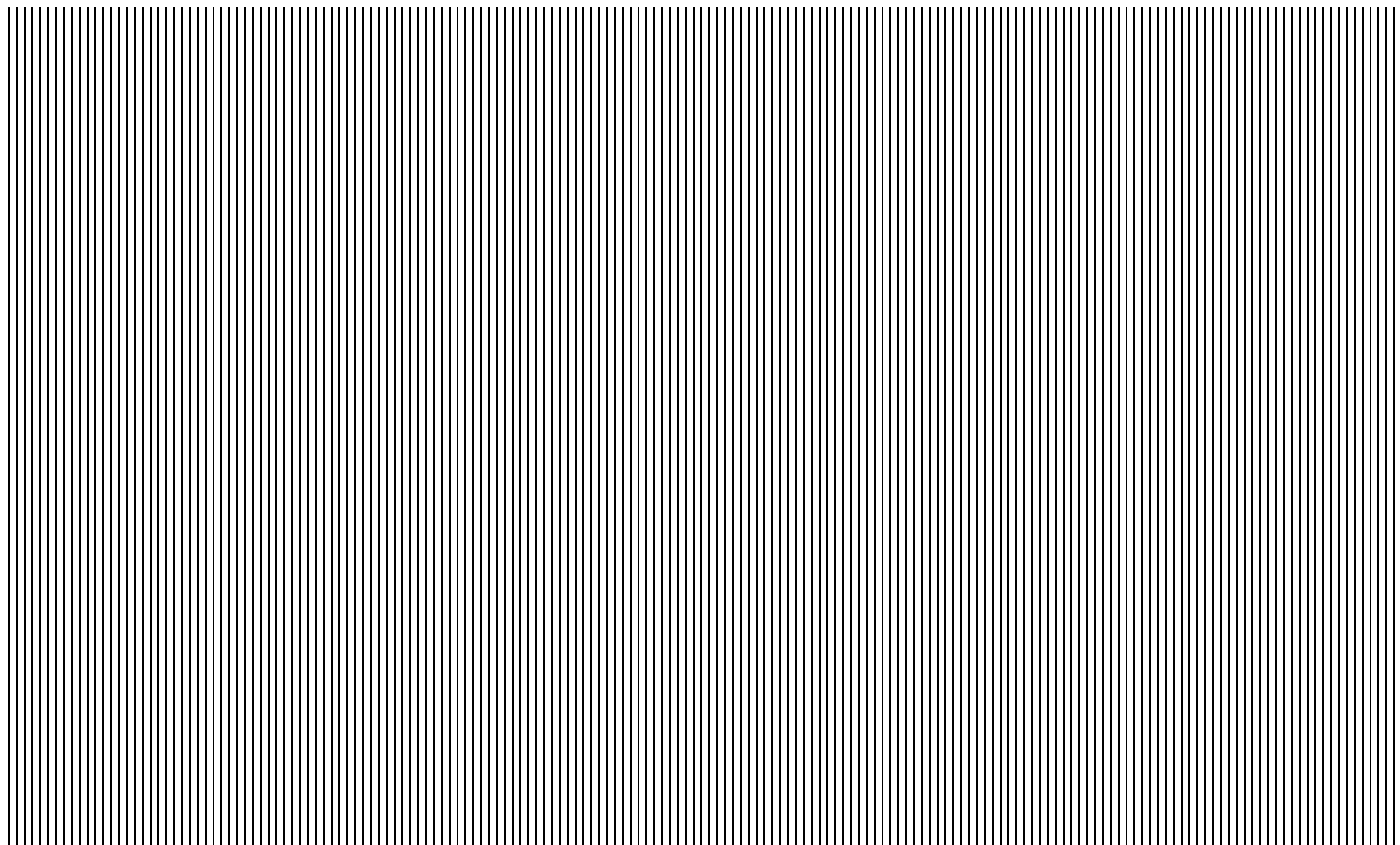
采样（Sampling）把空间上连续变化的光影画面，转换成离散的像素点阵，用有限个采样点记录画面信息。

图像传感器CMOS上排布着数百万甚至上亿个微小的**光敏单元**，每个光敏单元是一个独立的**采样点**，快门打开的瞬间，每个光敏单元捕捉对应位置的光线强弱与色彩信息，完成一次空间采样。

光敏单元接收到光线后，将光信号转换成电信号（光照越强，输出的电信号幅值越大）。

相机芯片按顺序读取全部光敏单元的电信号，再经过模/数转换，最终生成由一个个**像素**构成的图像阵列。





动手试一试：

拿出手机打开相机，对准屏幕上的细密竖条纹图案；缓慢改变手机和屏幕之间的距离，仔细观察画面变化。



想一想：

出现的这种摩尔纹，是条纹本身变了吗？它和图像的空间采样率有什么样的关系？

连续光影的采样也要遵循奈奎斯特原则吗

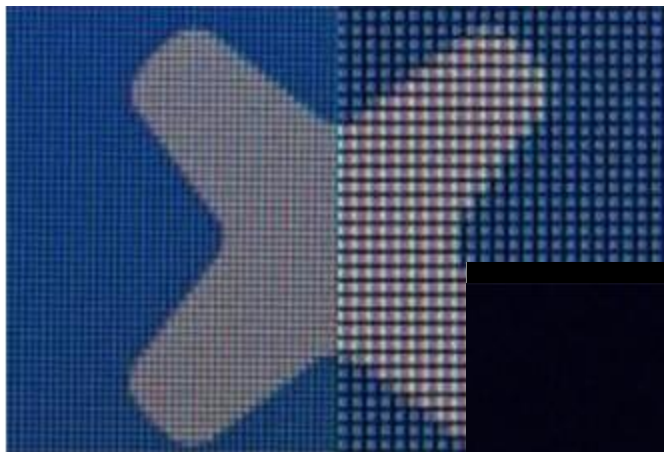




采样密度与分辨率：找到合适的平衡点

4K

1080P



标清 (720P)

小屏清晰，大屏略欠细腻

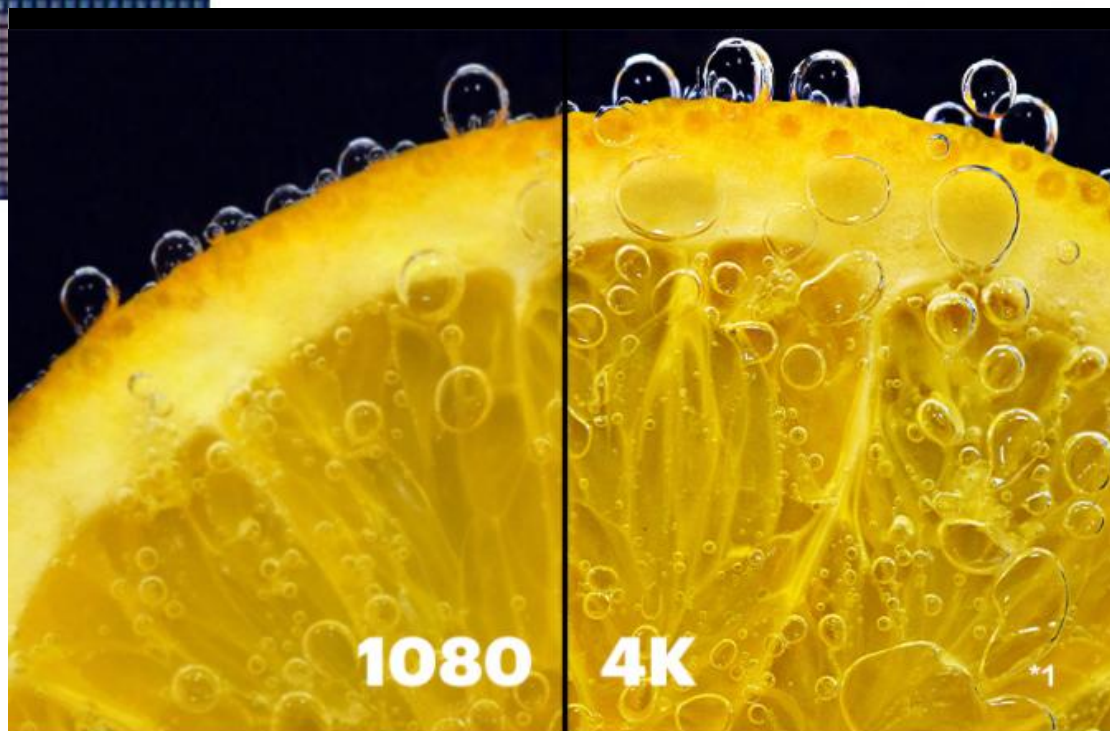
全高清 (1080P)

细节饱满，无明显锯齿

超高清 (4K)

近距离细节拉满，色彩过渡自然

在相同物理尺寸下，
分辨率越高，像素
的物理尺寸越小



分辨率（采样密度）的选择，不只看技术，而是受「人眼感知能力 + 文件体量 + 使用场景」共同约束

采样密度与分辨率：找到合适的平衡点

- 采样太少：低于奈奎斯特 → 混叠失真（摩尔纹）
- 采样过多：超出感知极限 → 产生感知冗余、浪费存储空间

用途	所需空间频率（细节程度）	推荐分辨率（打印/显示）	说明
电视、投影大屏	中等（看脸、看图）	72–150 PPI	远距离观看，无需极高分辨率
网页图片	低–中	72–96 PPI	优先控制文件大小，加快加载
高清印刷（画册、海报）	高（看纹理、文字）	300 PPI	满足奈奎斯特对印刷网点的要求
医学影像（X光、病理切片）	极高（细胞级细节）	>500 PPI	不允许任何混叠，关乎诊断
卫星遥感、显微成像	超高	数千 PPI	科研级需求

分辨率（采样密度）的选择，不只看技术，而是受「人眼感知能力 + 文件体量 + 使用场景」共同约束



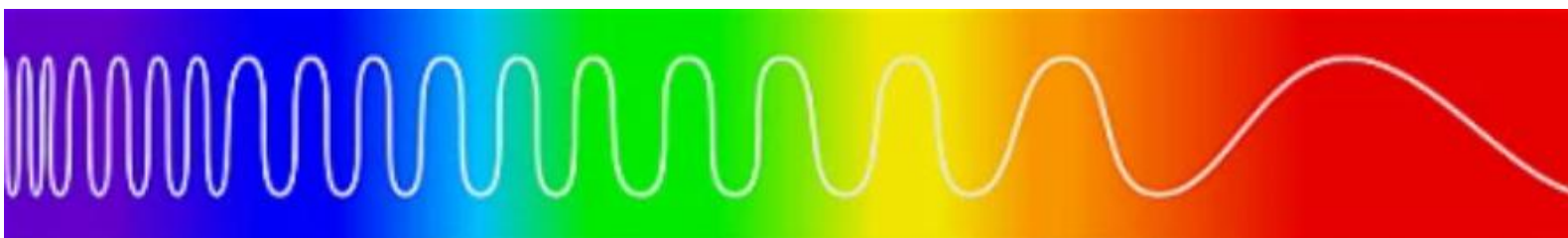
亮度

暗



明

色彩



不同色彩的光，本质是不同频率的电磁波

阵列中的像素点承载着哪些**量化**信息





量化：亮度信息

图像亮度信息的量化和音频幅值的量化逻辑是相通的：

光的明暗强弱是一个连续变化的量，拿一把有限刻度的尺子，把无穷多的亮度，切成有限的等级，这就是**灰度的量化**

亮度量化模型：**连续幅值 + 均匀刻度尺**

最暗 (0) **亮度量化** (用8位编码, 00000000~11111111映射到离散的亮度等级) 最亮 (255)



一共256个台阶

对于人类来说亮度的**量化等级**多少合适





量化编码位 8bit



量化编码位 5bit



量化编码位 4bit



量化编码位 3bit

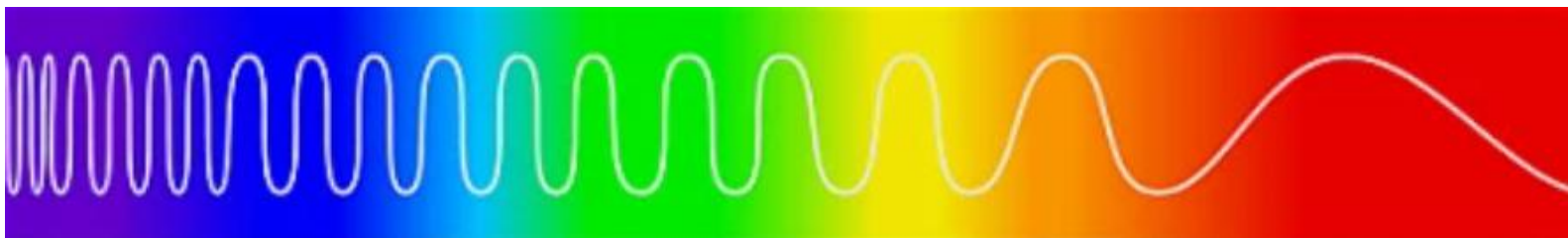


量化编码位 2bit



量化编码位 1bit

灰度图像：不同量化等级带来的视觉效果 —— 人眼适合多少级亮度台阶？



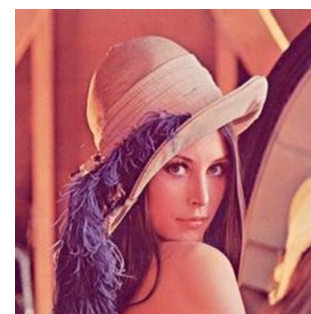
不同色彩的光，本质是不同频率的电磁波

从记录和处理



灰度图

到能够记录和处理



彩色图

音频用一把幅值尺子记录声波强弱，**灰度图**用一把亮度尺子记录光影明暗。**彩色**，需要一把什么样的量化尺子？





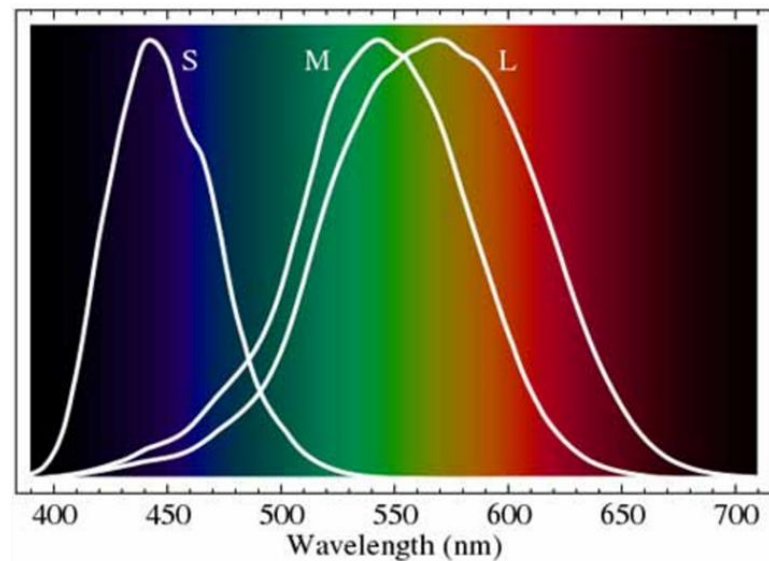
色彩的量化：一套多维度的尺子组合

我们的眼睛并不会直接分辨光谱上成千上万种波长。视网膜上有三类感光细胞，分别对蓝、绿、红三个波段的光最敏感；大脑靠这三路信号的强弱组合，就合成出我们感知到的全部色彩。



第一张彩色相片 James Clerk Maxwell (1861)
这正是「分色 + 幅度量化」思路的最早实践

色彩是外界物体反射或发射的光线进入人眼后，经视觉系统处理后产生的一种知觉体验。人类拥有三色视觉，采用 RGB 三基色量化方案



数字化时，我们仿照人眼，准备三把独立的亮度尺子，分别记录红、绿、蓝分量（三路信号）的强弱



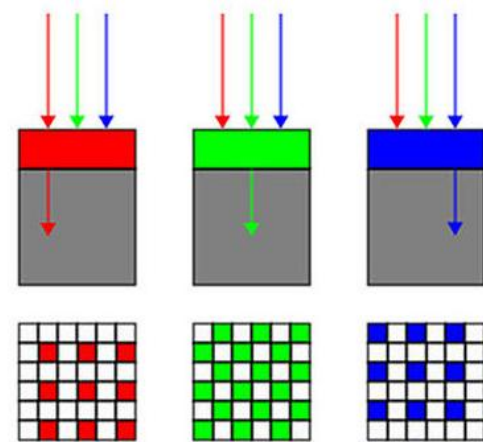
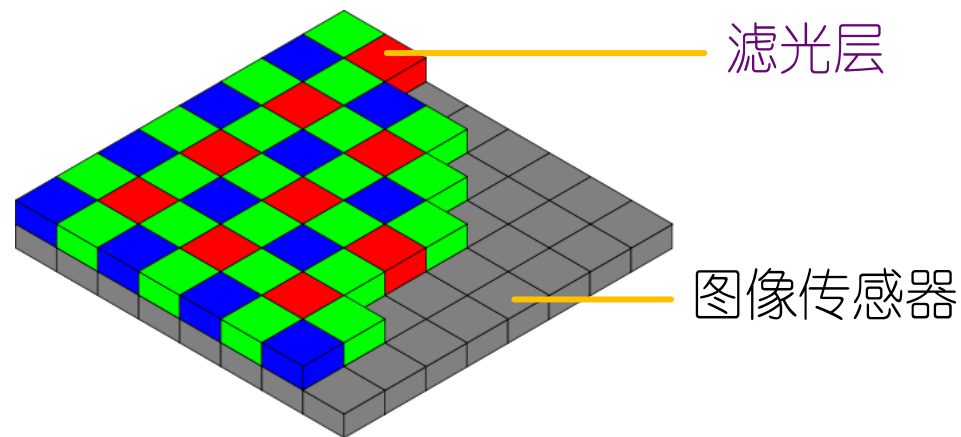
用数码相机实例理解这套量化尺子

如何让图像传感器收集到色彩？

工程上通过增加一层滤光层，配合算法实现色彩收集：

- **分工采样：**
每个感光点只采集单一基色的光强电信号。
- **量化单一颜色：**
和前面灰度亮度量化逻辑一样，每个感光点得到的单色电信号，被量化成 0~255 的数值。
- **插值补全颜色：**
根据相邻感光点的基色数据，用插值算法估算出这个像素缺失的另外两种基色量化值，补齐成一个拥有 R、G、B 三个分量的彩色感光点。

依靠滤光层分色采样分工和插值补齐色彩，是传感器在成本、体积、画质之间做出的经典工程权衡。

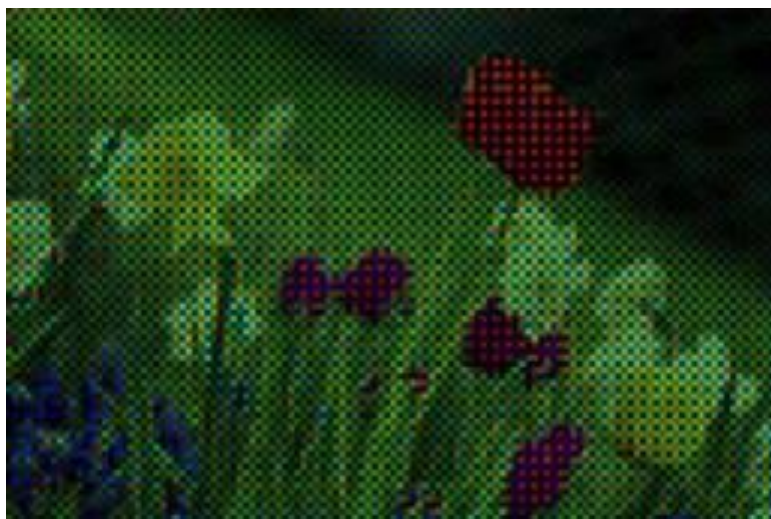


滤光层采用拜尔阵列排布（2绿:1红:1蓝）这样的分配，匹配人眼对绿光亮度、细节更敏感的特点



数码相机

RAW格式



按下快门获得一张数字底片



JPG格式



数码相机处理得到的拍摄照片

RAW（原始采样数据：拜尔单色采样点阵）→插值计算，为每个感光点补齐 R/G/B 三分量→每个感光点补齐 R/G/B 三分量 →再经过相机内部处理与压缩，得到适配使用场景的图像（比如 **JPG**）



彩色图像：不同量化等级带来的视觉效果

色彩量化的两步：**分色**（RGB / 调色板方案） + **幅度量化**（比特深度决定视觉效果）

4



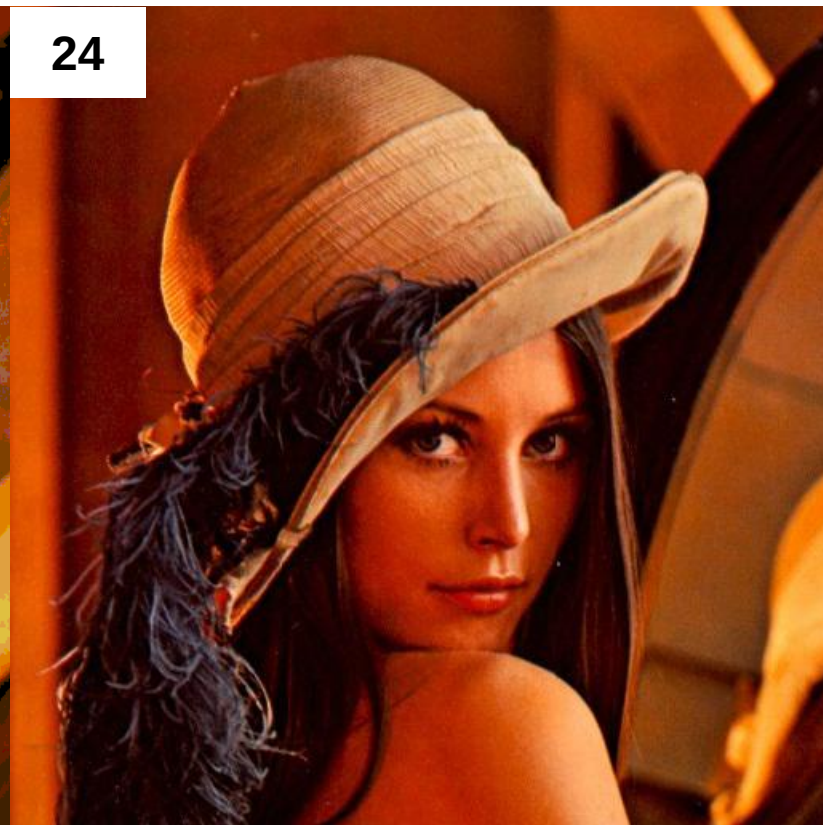
调色板模式，总位深4bit

8



调色板模式，总位深8bit

24



RGB色彩模式，单通道位深=8bit

所有量化方案，都是在有限比特下分配色彩台阶：比特越多，台阶越细密，画面过渡越顺滑



图像颜色深度

- **黑白图**，图像颜色深度为 **1**
- **灰度图**，图像颜色深度为 **8**
- **RGB 8色图**，图像颜色深度为 **3**
- **索引16色图**，图像颜色深度为 **4**，引入调色板机制
- **索引256色图**，图像颜色深度为 **8**
- **真彩色图**，图像颜色深度为 **24**



Black & White (黑白图)



Gray Scale (灰度图)



RGB Color 8色图



Indexed 16-Color (索引16色图)



Indexed 256-Color (索引256色图)



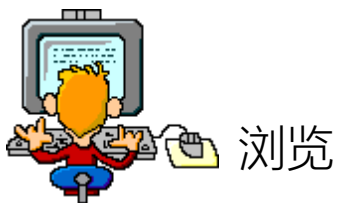
RGB True Color (真彩色图)

调色板本身：调色板表里存的条目，每一条依然是 RGB，是完整的色彩描述。

工作逻辑：针对具体一张图像统计全部出现的颜色，按出现频次排序，挑最重要的一批（比如 16 色、256 色）存入调色板；图像本体不再存每个像素完整 RGB，只存一个索引编号（序号），去查表找对应的 RGB。



连续光影物理信号数字化： 采样密度越高、量化位越高、细节越丰富、文件体量也越大



勃朗峰全景照片
(3650亿像素)

原图网址: <http://www.in2white.com/>



算一算：

这张 3650 亿像素的全景图，如果是 24bit RGB 无压缩 BMP 格式，文件有多大？



多媒体技术实用化的前提：压缩编码

信息的作用是消除不确定性

香农熵公式

$$I = D - du$$

I —— 信息量

D —— 数据量

Du —— 冗余量

$$H(X) = - \sum_{i=1}^n p_i \log_2 p_i$$

信息熵
(单位：比特)

香农熵给出了**信息的最小理论比特下限**：一段信号真正携带的**信息量**是**I**，我们通过采样+量化直接得到的**数据量D**通常更大，多出信息量的部分就是**冗余du**。

压缩编码的目标：设计**更紧凑的二进制映射**，剔除冗余，让编码后的数据量尽量尽可能贴近理论信息量下限。

名称	大小
 陈奕迅-十年_无损.flac	21,270 KB
 陈奕迅-十年_超品.mp3	7,998 KB
 陈奕迅-十年_高品.mp3	3,200 KB
 陈奕迅-十年_流畅.wma	2,422 KB

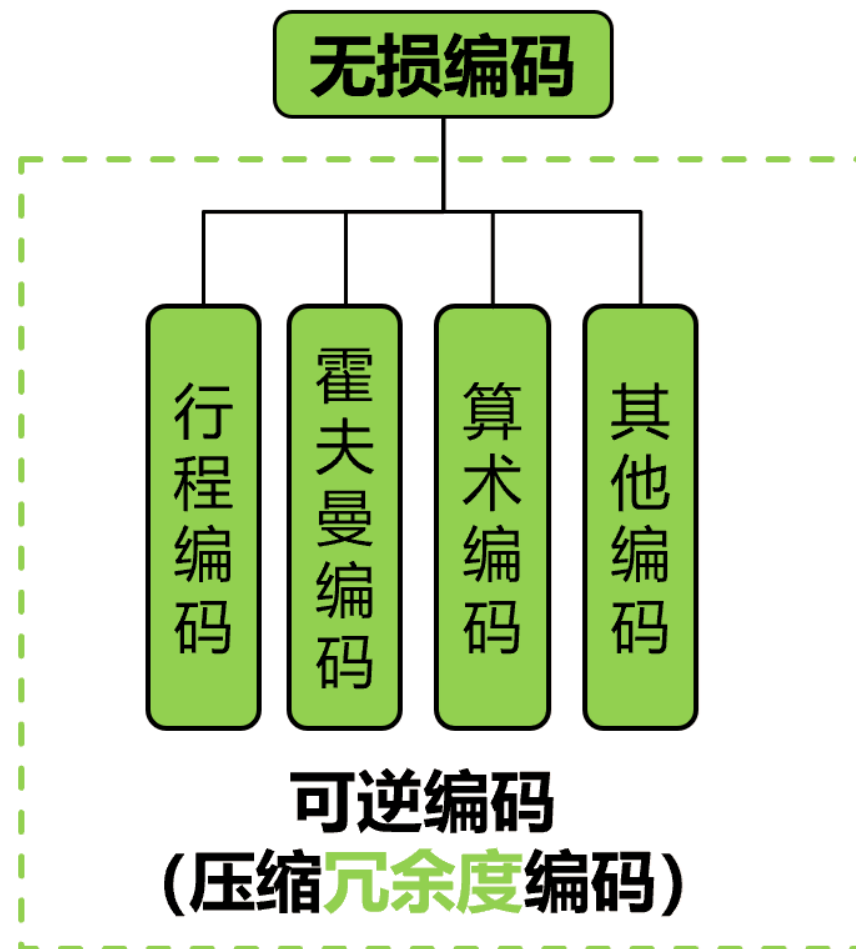


无损压缩编码

I = D - du — 无损压缩 (编码)

无损压缩 (编码)

又称**可逆编码** (压缩冗余度编码) :
由于不会产生失真, 多用于文本、
数据以及应用程序的压缩, 多媒体
中常见PNG图片、FLAC音频等。它
能**保证完全地恢复原始数据**。但压
缩比较低, 一般在2:1~5:1之间。





有损压缩编码

$I = D - du$ — 有损压缩 (编码)

有损压缩 (编码)

又称**不可逆编码** (压缩熵编码) :
除了去除统计冗余, 还会依据人眼、人耳的感知特性, 舍弃人类感知无法分辨的细节信息。解压后无法 100% 还原原始采样数据, 但在合适参数下, 人主观感受差别很小, 可以实现很高的压缩比。

感知型数字化里两类不同来源的损失:

- 量化误差
- 有损压缩的主动损失



名称	大小
陈奕迅-十年_无损.flac	21,270 KB
陈奕迅-十年_超品.mp3	7,998 KB
陈奕迅-十年_高品.mp3	3,200 KB
陈奕迅-十年_流畅.wma	2,422 KB

有损编码

预测编码

变换编码

基于模型编码

直接映射编码

其他编码

不可逆编码
(压缩熵编码)



压缩编码与常见的文件格式

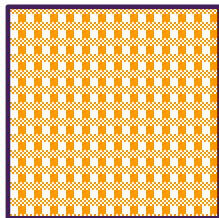
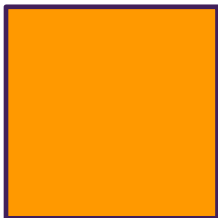
$$I = D - Du$$

I —— 信息量

D —— 数据量

Du —— 冗余量

无损

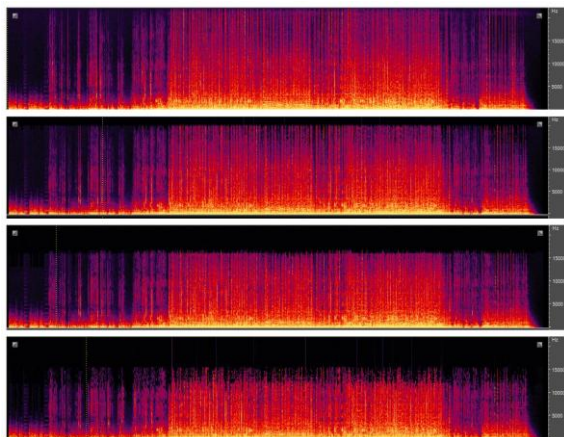


文本、程序、需要反复编辑的图像、追求原汁原味的音频，不允许丢失任何采样信息，比如PNG、FLAC、WAV、ZIP等

压缩依靠

统计冗余（数据内部重复、规律）

有损



照片分享、在线音频、网络视频。优先控制文件体积，接受感知层面可控的信息取舍，比如JPG、MP3、WMA、MP4等

名称	大小
陈奕迅-十年_无损.flac	21,270 KB
陈奕迅-十年_超品.mp3	7,998 KB
陈奕迅-十年_高品.mp3	3,200 KB
陈奕迅-十年_流畅.wma	2,422 KB

压缩依靠

统计冗余 + 感知冗余（人眼 / 人耳分辨极限）

从选择采样频率、量化比特，再到压缩方案，全部都是面向人感知的工程权衡。

数字化：感知型 与 逻辑型

所有信息进入计算机前，都必须被**数字化**，不同的信息，有不同的**转换**策略。

逻辑型数字化是基于人预先定义的**规则**完成**编码**表示。计算机硬件仅处理 0/1 符号，**含义、语义、结构**是人通过**编码规则**赋予的。

典型应用：数值编码、文本编码、矢量图形、程序代码、数学公式。

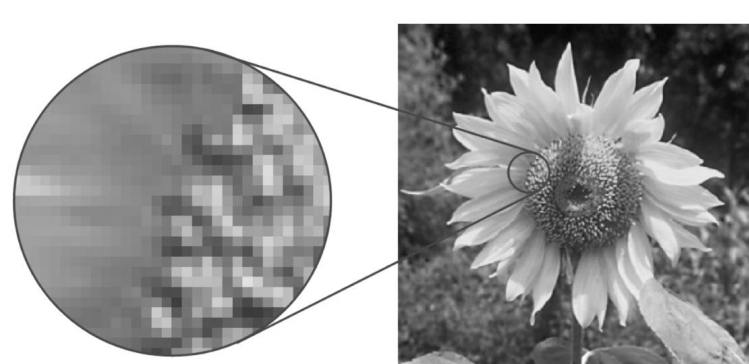
感知型数字化是对物理世界中**连续**变化的信号，通过**采样、量化**，得到**有限离散**样本，是对真实世界的**近似记录**。典型应用：图像、声音和视频的数字化记录。



信息加工处理机器



(基于规则绘制的**图形**) 可无限缩放



(基于有限记录还原的**图像**) 缩放失真