

第二章 多媒体基础

周满

15271802577

zhouman@hust.edu.cn

华中科技大学网络空间安全学院

图像多媒体基础

- 一、图像基础知识
- 二、为什么压缩
- 三、JPEG编码压缩

图像基础

什么是图像

- 人类由外界获得的信息约有70%来自于视觉，图像是人类视觉的基础
- 图像会以各种各样的形式出现：可视的和非可视的、抽象的和物理的
- 可视图像：由可见光形成，能被人类的视觉系统所感受，如人们通常见到的自然景物等
- 非可视图像：不能为人眼直接感受，是某种物理量的平面或空间分布形成的，如X射线、红外线、微波等。这类图像通过换能与显示装置变换后，仍能为视觉所感受

图像基础

什么是图像

- 图像可被划分为称作像素点 (pixel) 的小区域，在每个像素点，表示亮度或色彩的值被采样和量化，得到像素值。该过程称为数字化，得到的图像称为**数字图像**
- 数字图像以其存储方式的不同又可分为不同的文件格式，如位图 (BMP)、矢量图 (EPS)、动态图 (GIF)、无损压缩 (PNG)、有损压缩 (JPG) 等等

图像基础

◆ 数字图像的基本表示单元-像素



3:1



24:1

图像基础

图像的两大类型

➤ 位图

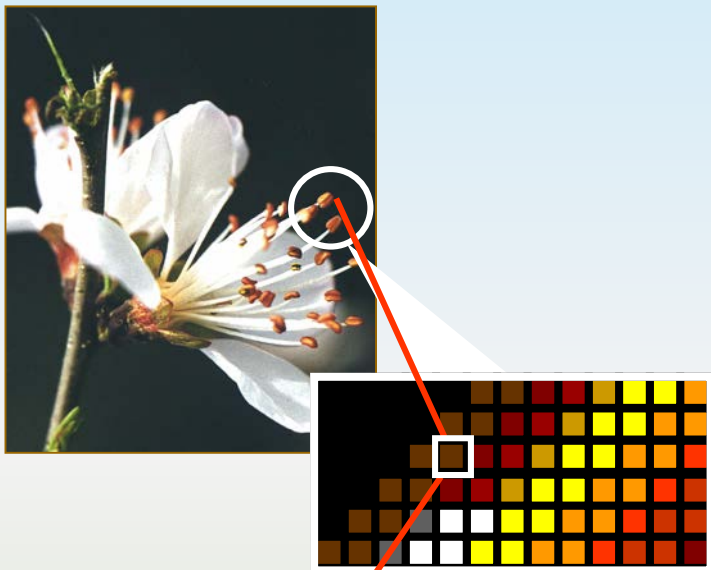
- 采用像素矩阵方法来显示和存储图像
- 真实细腻，反映图像层次色彩，体积较大
- 更适合描述照片

➤ 矢量图

- 采用数学方法描述由几何元素组成的图像
- 文件小，能够随意缩放而不改变图像质量
- 更适合描述图形

图像基础

数字图像的色彩



图像像点

- 图像由基本显示单元“像素”构成

- 像素值由若干个二进制位描述

- 二进制位数代表图像颜色的数量

8位图像

16位图像

24位图像

8 bit ($2^8 = 256$ 色)

16 bit ($2^{16} = 65536$ 色)

24 bit ($2^{24} = 16M$ 色)

图像基础

图像类型

- 二值图像
- 灰度图像
- RGB图像

图像基础

◆ 数字图像中表示单位-像素

◆ 二值图像-0或1-1位

- ◆ 图像可以看做一个二维矩阵，每个元素只存在0，1两种取值，0为黑，1为白

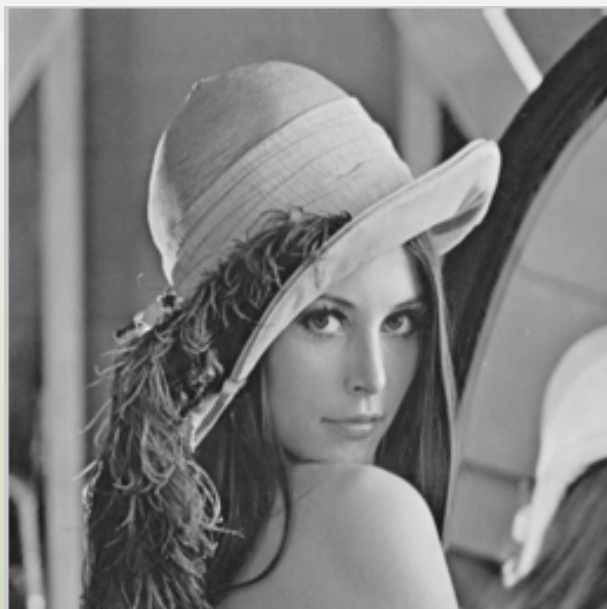


256x256 [logical](#)

	1	2	3	4	5	6	7	8	
37	1	1	1	1	1	1	1	1	0
38	1	1	1	1	1	1	1	1	0
39	1	1	1	1	1	1	1	0	0
40	1	1	1	1	1	1	0	0	0
41	1	1	1	1	1	1	0	0	0
42	1	1	1	1	1	0	0	0	0
43	1	1	1	1	1	0	0	0	0
44	1	1	1	1	0	0	0	0	0
45	1	1	1	0	0	0	0	0	0
46	1	1	1	0	0	0	0	0	0
47	1	1	0	0	0	0	0	0	0
48	1	1	0	0	0	0	0	0	0
49	1	0	0	0	0	0	0	0	0

图像基础

- ◆ 数字图像表示单位-像素
 - ◆ 灰度图像-0~255-8位
 - ◆ 图像也可以看做一个二维矩阵，每个元素取值为0到255，0为黑，255为白



256x256 uint8							
	1	2	3	4	5	6	7
1	162	162	160	162	163	160	159
2	162	162	160	162	163	160	159
3	163	160	160	160	161	158	157
4	160	158	159	158	160	159	155
5	155	157	157	157	159	158	157
6	156	158	156	153	157	157	156
7	157	157	157	156	157	156	155
8	158	157	157	157	156	156	157
9	157	157	156	154	157	156	157
10	156	156	158	155	157	156	157
11	156	156	156	156	156	157	157
12	158	155	153	157	157	157	158

图像基础

- ◆ 数字图像中表示单位-像素
 - ◆ RGB彩色图像-- $3 \times (0 \sim 255)$ -24位



	R	G	B
256x256 uint8			
	1	2	3
1	137	137	134
2	137	137	134
3	138	135	134
4	134	133	133
5	130	133	131
6	131	133	130

	R	G	B
256x256 uint8			
	1	2	3
1	226	223	226
2	226	223	226
3	227	224	226
4	226	225	226
5	224	223	225
6	224	224	227

	R	G	B
256x256 uint8			
	1	2	3
1	126	130	119
2	126	130	119
3	124	122	118
4	118	113	119
5	106	111	113
6	107	110	106

图像基础

常见标准颜色的RGB值

颜色	R	G	B
红	255	0	0
蓝	0	255	0
绿	0	0	255
黄	255	255	0
紫	255	0	255
青	0	255	255
白	255	255	255
黑	0	0	0
灰	128	128	128

数字视频

- ◆ 数字图像（帧）序列
 - ◆ 视觉暂留现象
 - ◆ 一副图像会在脑海中停留一瞬
 - ◆ 当图像切换速度足够快，就会形成连续的画面
 - ◆ 通常需要大于25帧每秒



图像多媒体基础

- 一、图像基础知识
- 二、为什么压缩
- 三、JPEG编码压缩

为什么要压缩？

◆ 数据过于庞大

◆ 一个视频例子 (DVD)

◆ 每个像素 (RGB) : 24比特

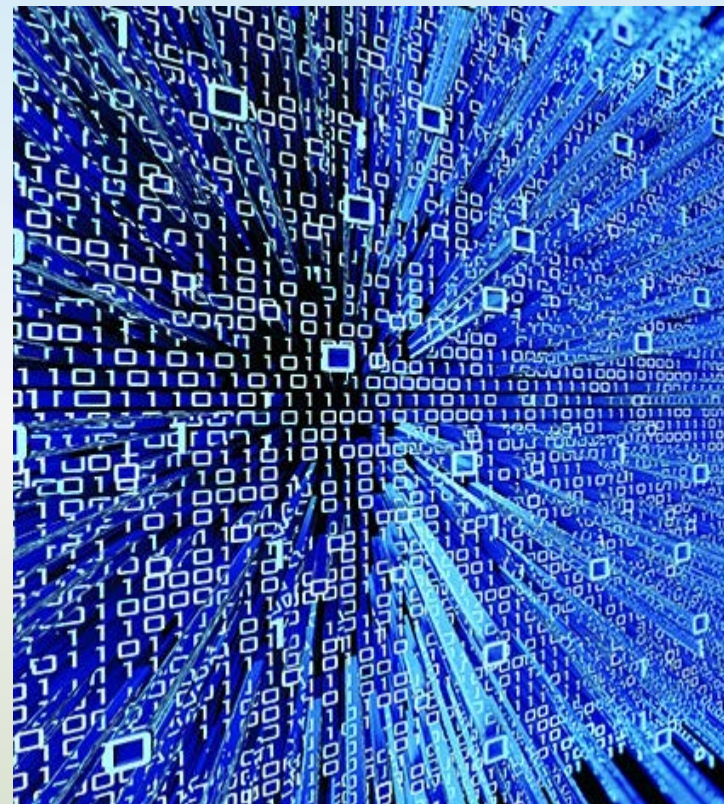
◆ 分辨率: 1920×1080

◆ 帧率: 30帧/秒

◆ 比特率 = $1920 \times 1080 \times 24 \times 30 \approx 1493 \text{ Mbps}$

◆ 一部两小时电影的数据量?

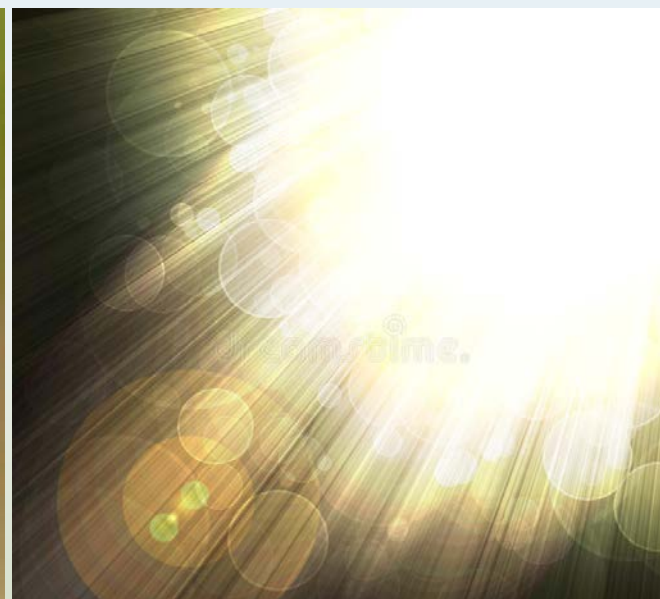
◆ 数据量 = $1493 \times 2 \times 60 \times 60 \approx 10749 \text{ G}$



为什么可以压缩？

- 存在冗余

- ◇ 视觉冗余-掩蔽效应



视觉掩蔽效应：在图像呈现后，在人眼前马上呈现一个其他刺激信息，该图像就会被抹去。

为什么可以压缩？

• 存在冗余

◇ 视觉冗余-掩蔽效应

◇ 亮度掩蔽

- ◆ 物理学将人眼最小可觉察的亮度差值随背景亮度变化而变化的现象称为人眼的亮度掩蔽特性
- ◆ 背景越亮，能不被感知到的亮度差异，也就是视觉能容忍的差异越大

为什么可以压缩？

◆ 亮度掩蔽

$$\frac{\Delta I}{I} \approx \text{const}$$

- ◆ const是一个常数
- ◆ 分子 ΔI 是亮度探测阈值（即最小可觉差），指人眼恰可察觉的目标亮度与背景亮度之间的最小亮度差异
- ◆ 分母 I 是背景亮度值
- ◆ 这个公式意味着背景越亮，能不被感知到的亮度差异（即视觉能容忍的差异）越大

为什么可以压缩？

• 存在冗余

◇ 视觉冗余-掩蔽效应

◇ 对比度掩蔽

- ◆ 邻域掩蔽：由于空间相邻系数导致的掩盖效应，人眼视觉对平坦区域失真敏感性强于复杂纹理区域
- ◆ 自对比度掩蔽：具有相同空间频率、取向和位置的信号掩蔽效应

对比度掩蔽（Contrast Masking）：反映在存在背景信号的前提下，人眼视觉对一个信号的感知掩蔽特性

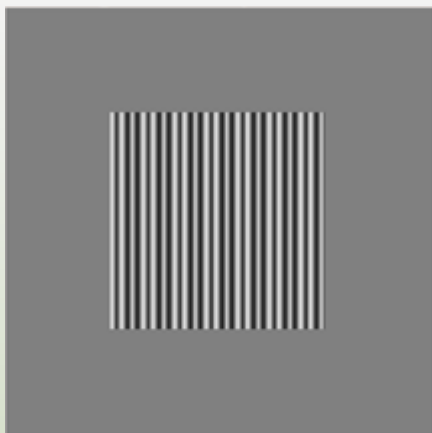
为什么可以压缩？



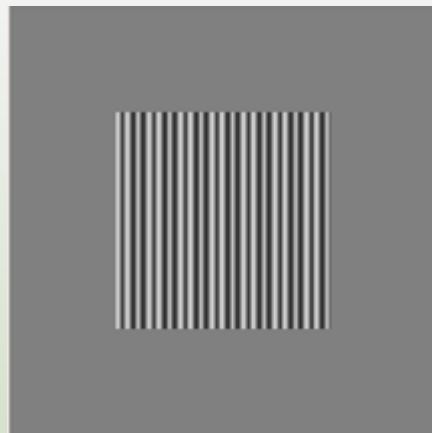
-



$= S_1$



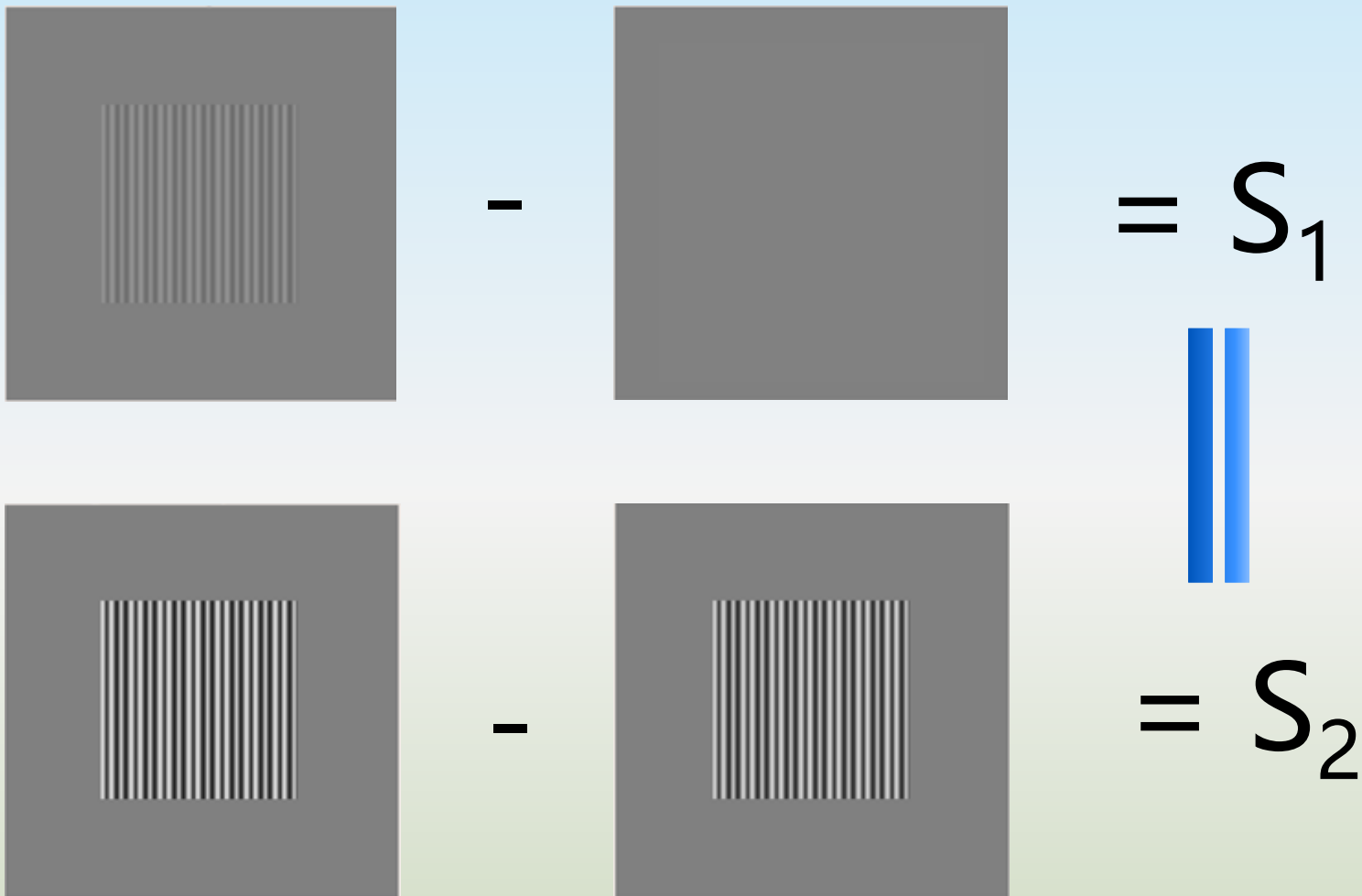
-



$= S_2$

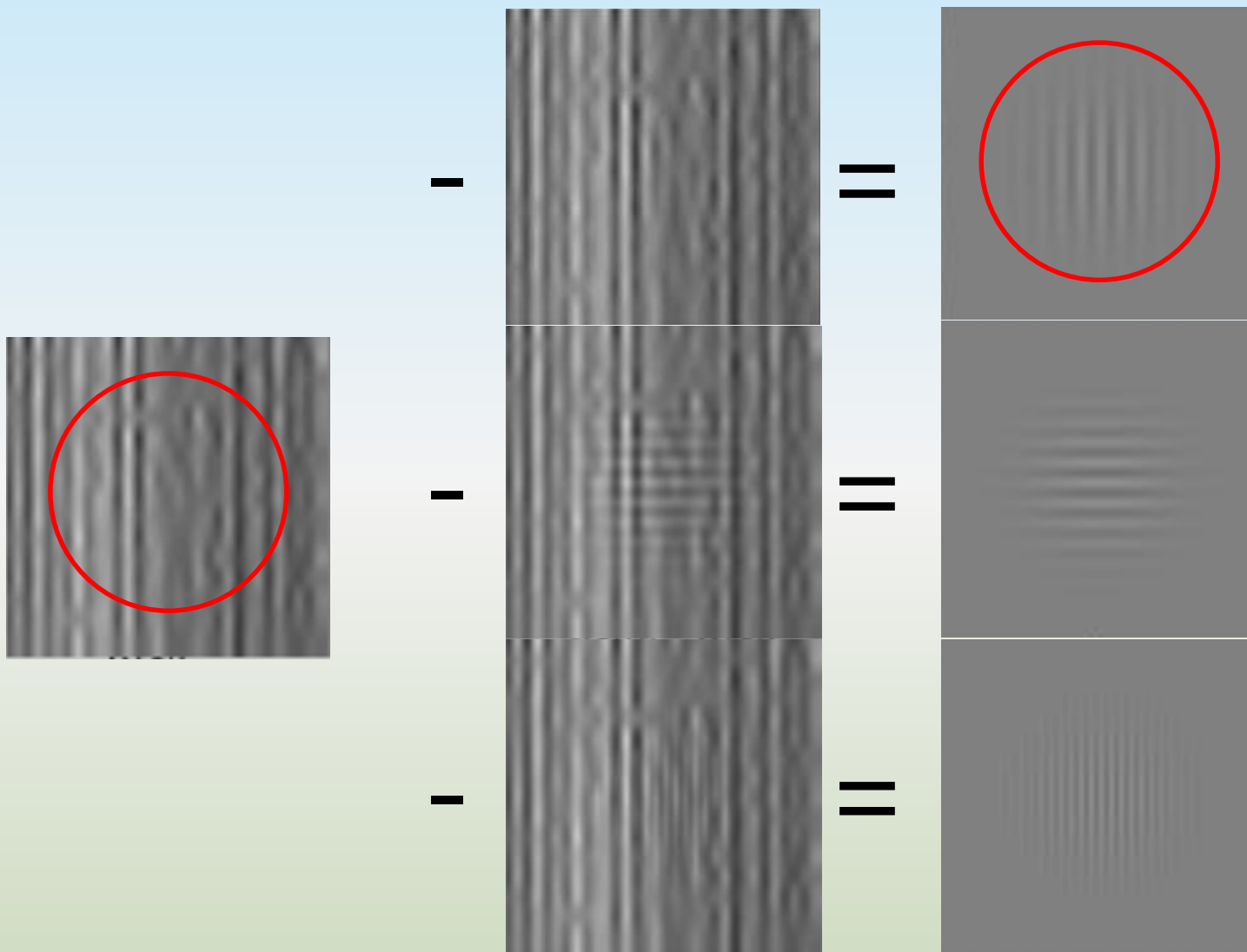
哪幅图修改更明显？

为什么可以压缩？



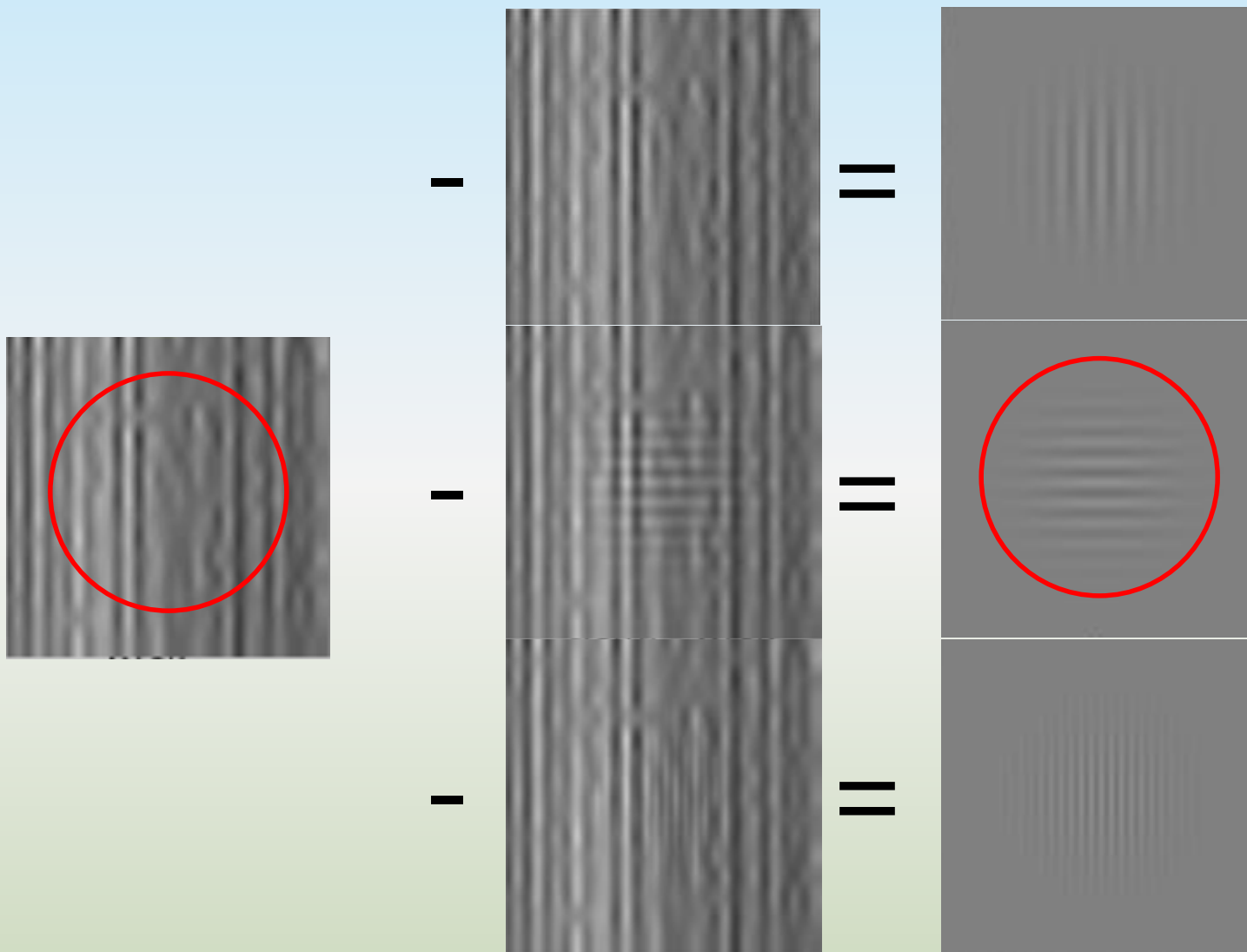
邻域掩蔽：人眼视觉对平坦区域失真敏感性强于复杂纹理区域

为什么可以压缩？



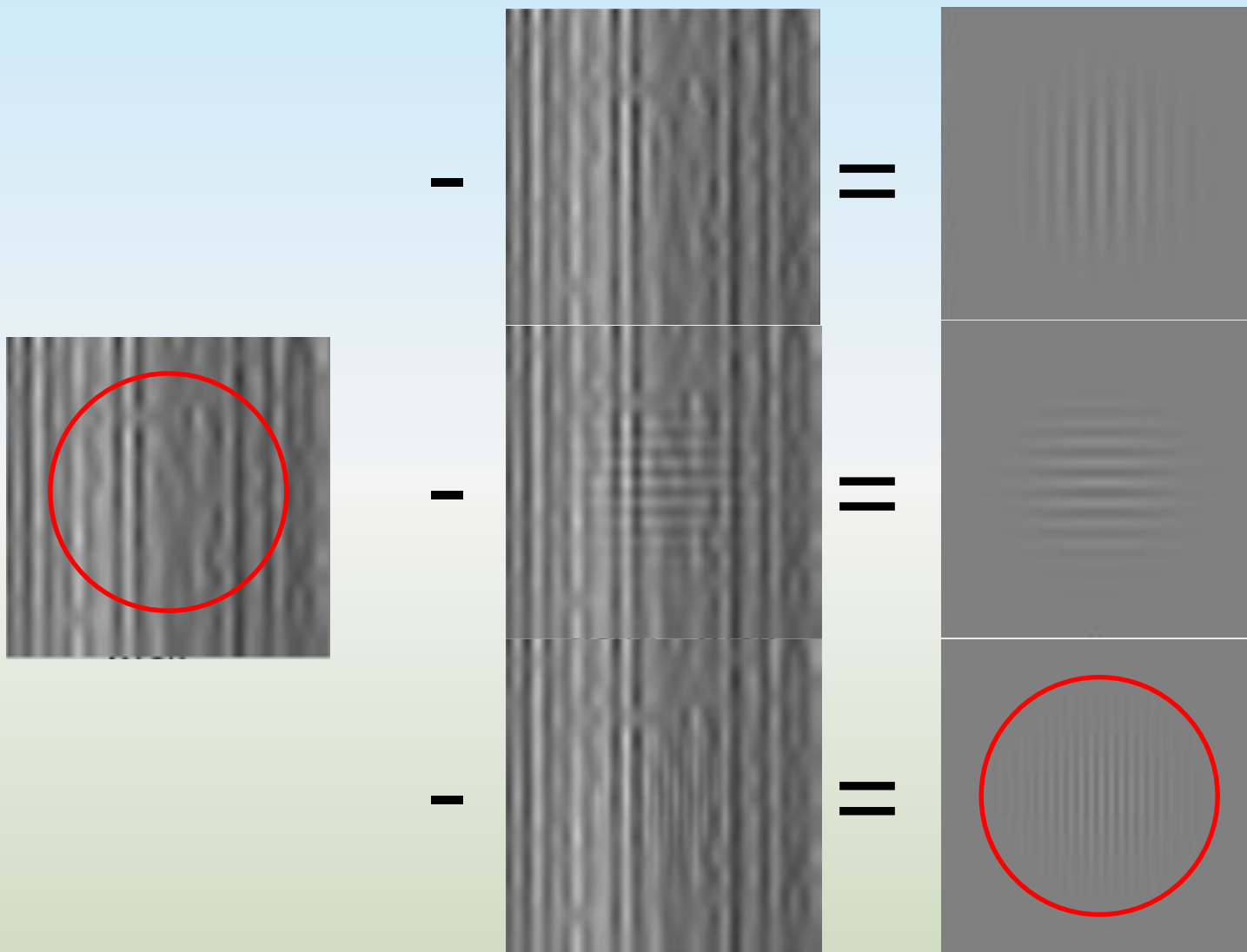
自对比度掩蔽：
具有相同空间频率、取向和位置的信号掩蔽效应

为什么可以压缩？



自对比度掩蔽：
具有相同空间频率、取向和位置的信号掩蔽效应

为什么可以压缩？



自对比度掩蔽：
具有相同空间频
率、取向和位置
的信号掩蔽效应

为什么可以压缩？

• 存在冗余

◇ 视觉冗余-掩蔽效应

◇ 色度掩蔽

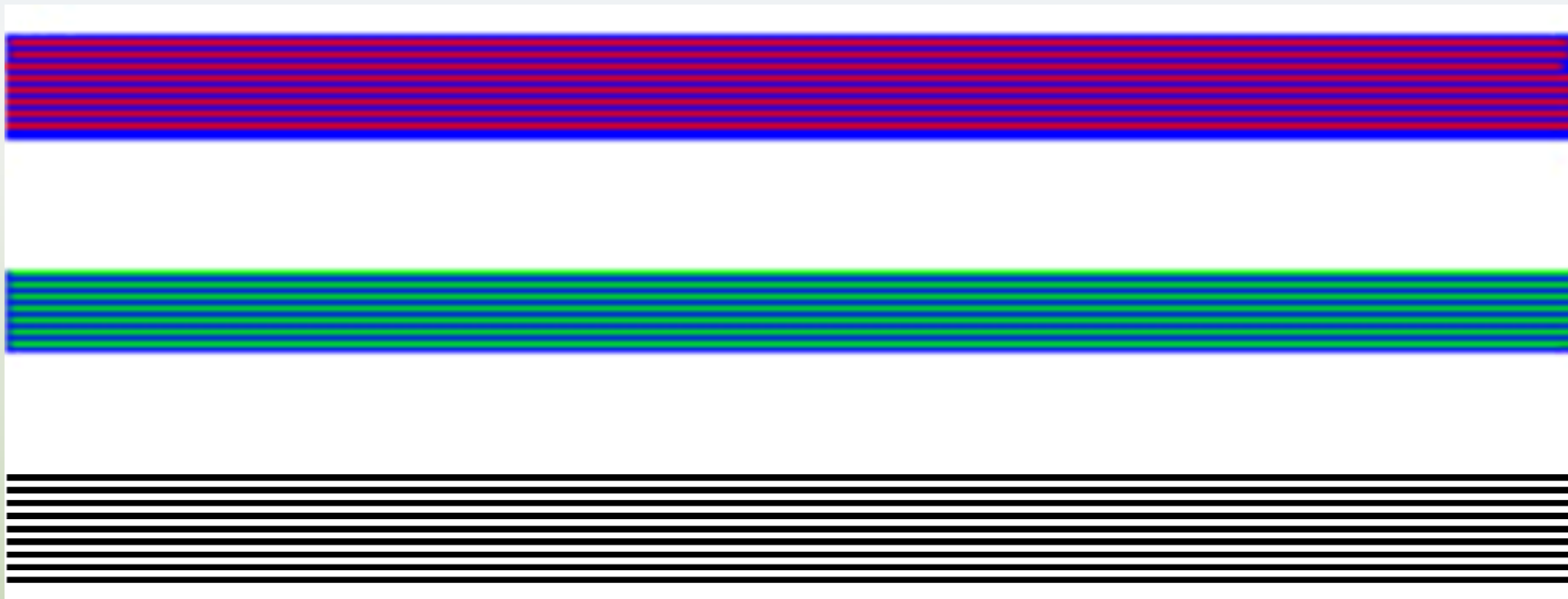
- ◆ 人眼对亮度的分辨力要明显比彩色的高，对间隔较密的黑白正弦光栅我们可能分辨清楚，而同样间距的彩色光栅，则可能分辨不清
- ◆ 另外，在亮度变化剧烈的背景上，人眼对色彩变化的敏感度明显降低

为什么可以压缩？

◇ 视觉冗余-掩蔽效应

◇ 色度掩蔽

- ◆ 人眼对亮度的分辨力要明显比彩色的高，对间隔较密的黑白正弦光栅我们可能分辨清楚，而同样间距的彩色光栅，则可能分辨不清

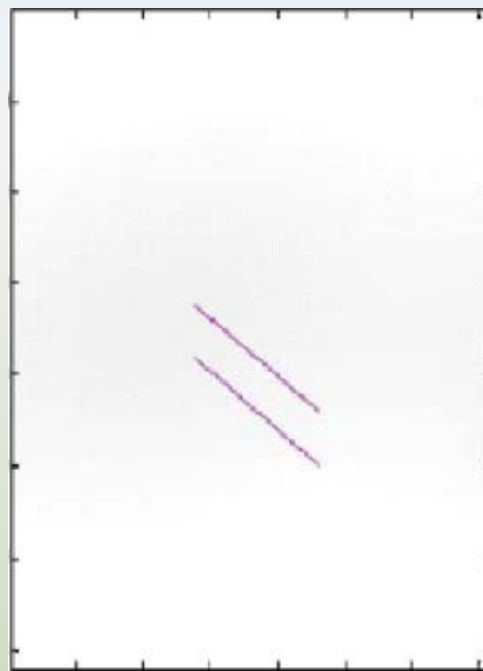


为什么可以压缩？

◇ 视觉冗余-掩蔽效应

◇ 色度掩蔽

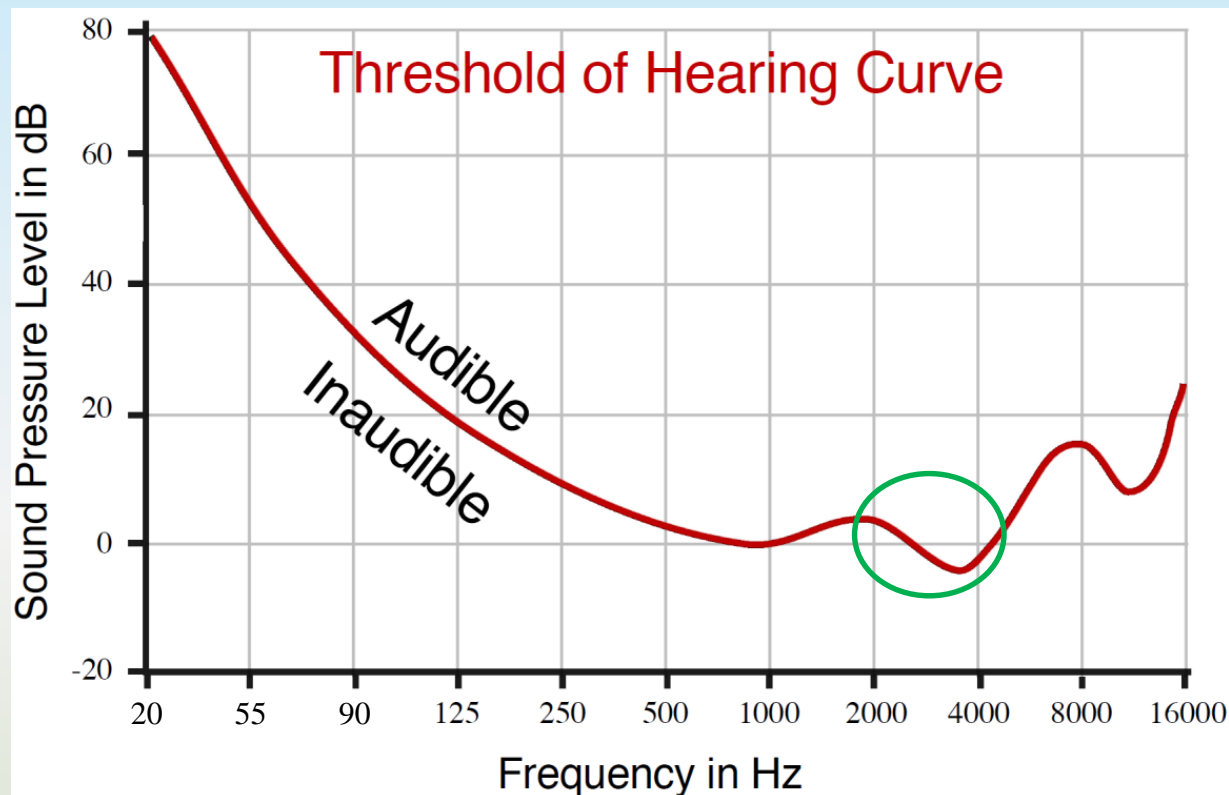
- ◆ 在亮度变化剧烈的背景上，人眼对色彩变化的敏感度明显降低



为什么可以压缩？

- 存在冗余

- ◆ 听觉冗余



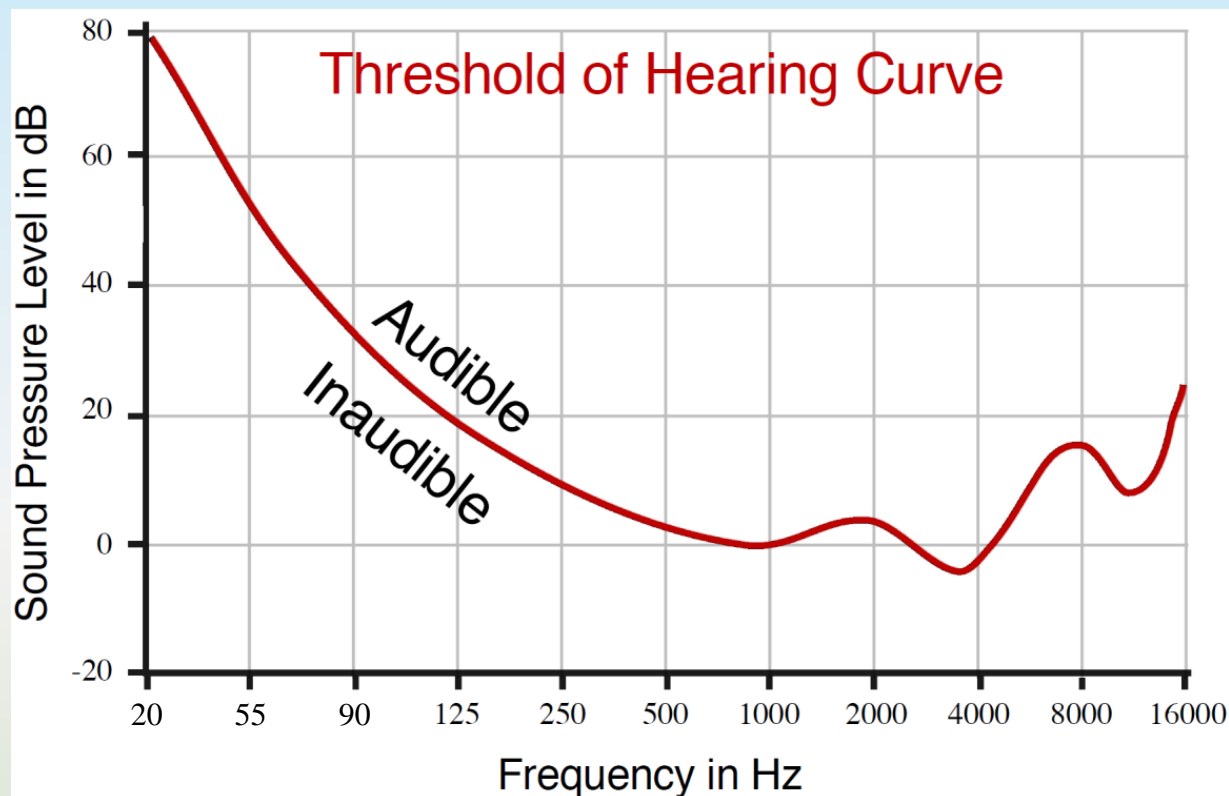
- ◆ 听觉阈值

- ◆ 听觉阈值的大小随声音频率的改变而改变，每个人的听觉阈值也不同。大多数人的听觉系统对2 ~ 5 kHz之间的声音最敏感

为什么可以压缩？

- 存在冗余

- ◆ 听觉冗余



- ◆ 听觉阈值

- ◆ 一个人是否能听到声音取决于声音的频率，以及声音的幅度是否高于这种频率下的听觉阈值

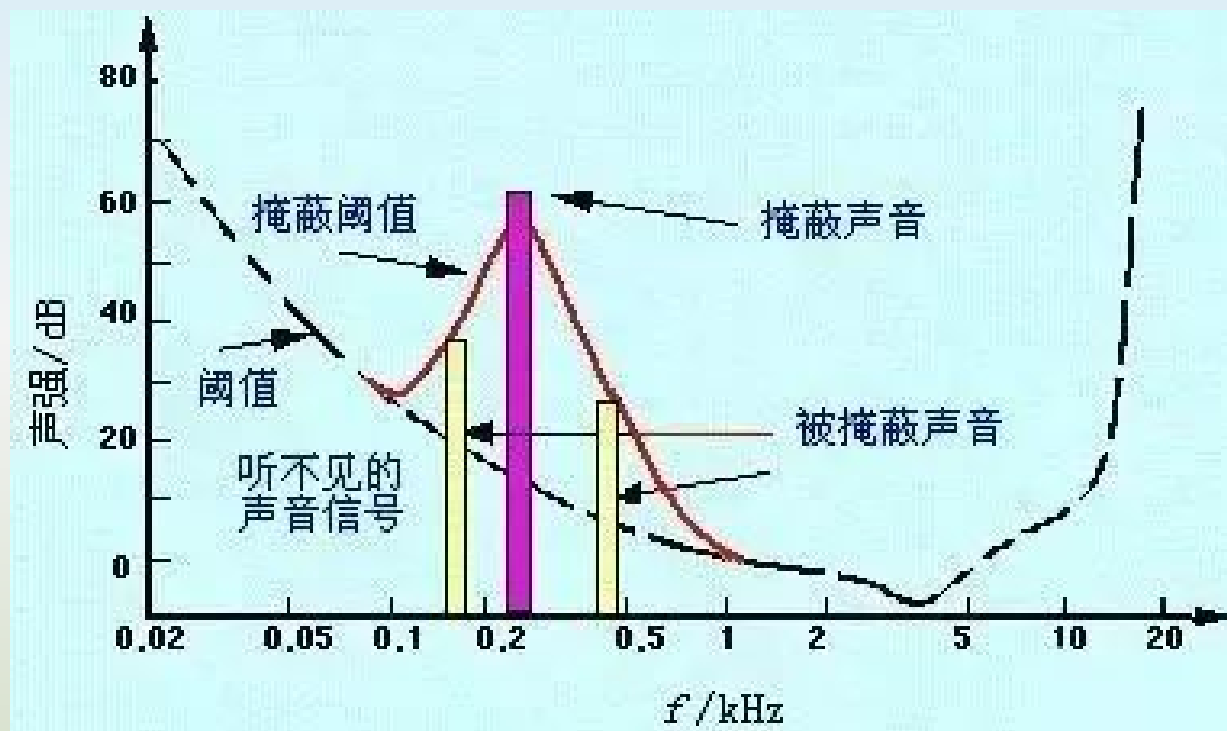
为什么可以压缩？

• 存在冗余

◇ 听觉冗余



食堂噪声掩蔽



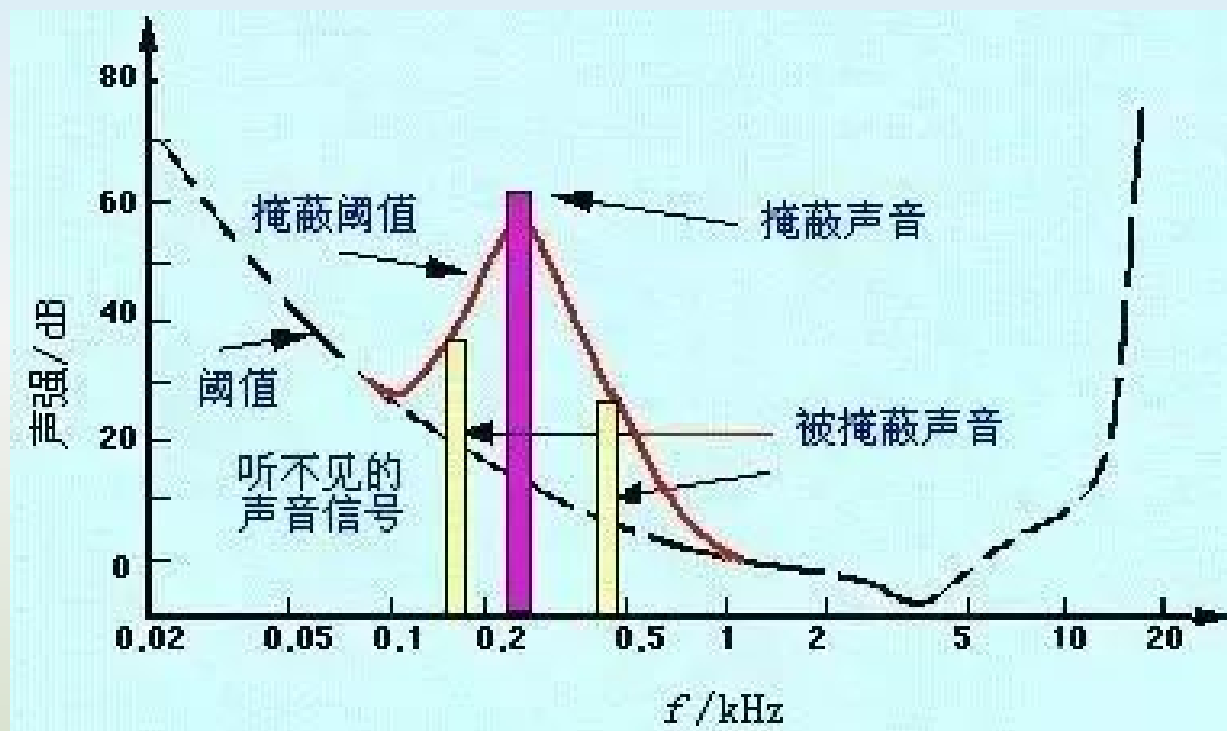
◇ 听觉掩蔽

- ◇ 同时掩蔽：几个强弱不同的声音同时存在时，强声使弱声难以听见

为什么可以压缩？

• 存在冗余

◇ 听觉冗余



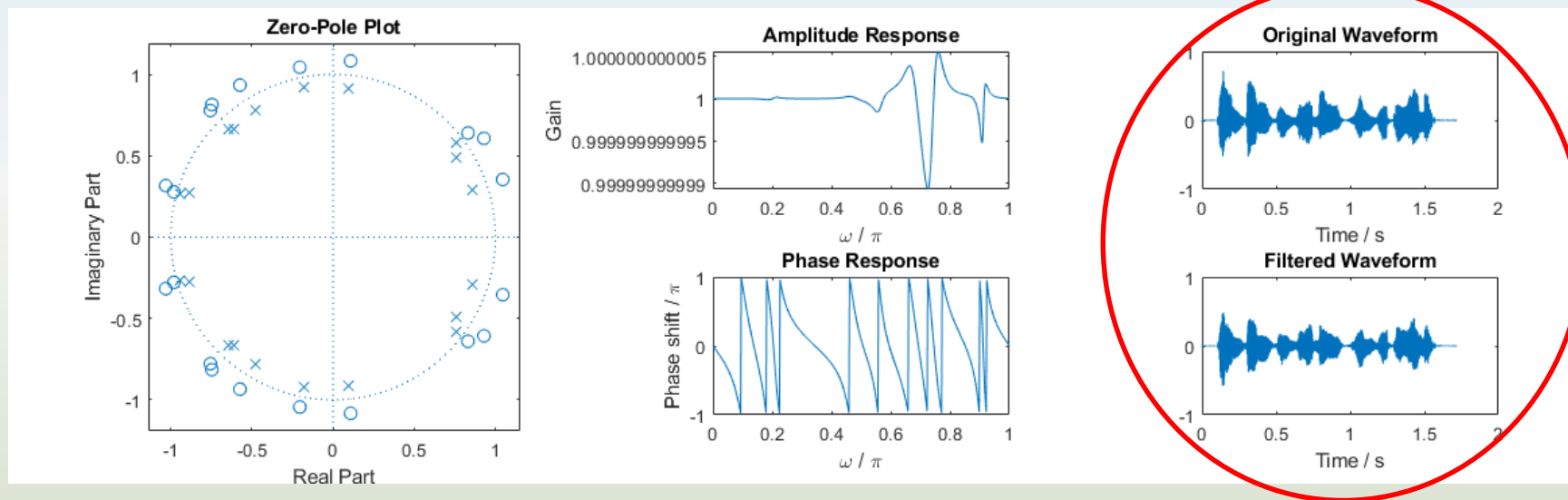
◇ 听觉掩蔽

- ◇ 异时掩蔽：声音在不同时间先后发生时，强声使其周围的弱声难以听见

为什么可以压缩？

• 存在冗余

◇ 听觉冗余



◇ 相位不敏感

- ◇ 人耳对振幅、频率的变化较为敏感，而对相位变化的敏感程度则要较弱

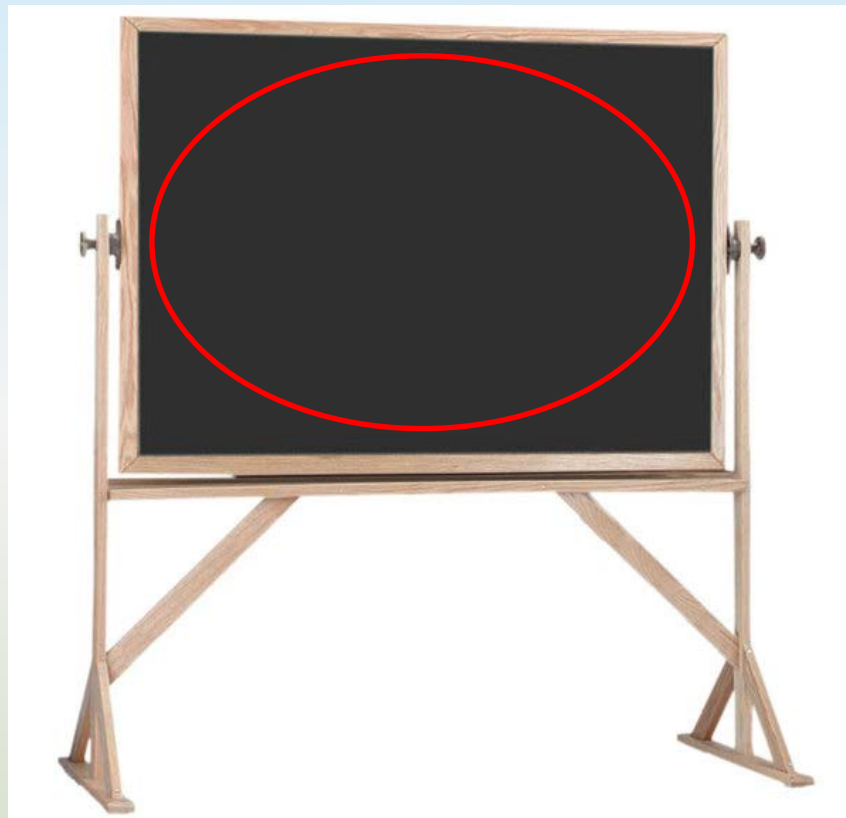
为什么可以压缩？

- 存在冗余

- ◆ 统计冗余

- ◆ 空域冗余

- ◆ 相邻空间的像素值是非常相似的，即空间相关性很强



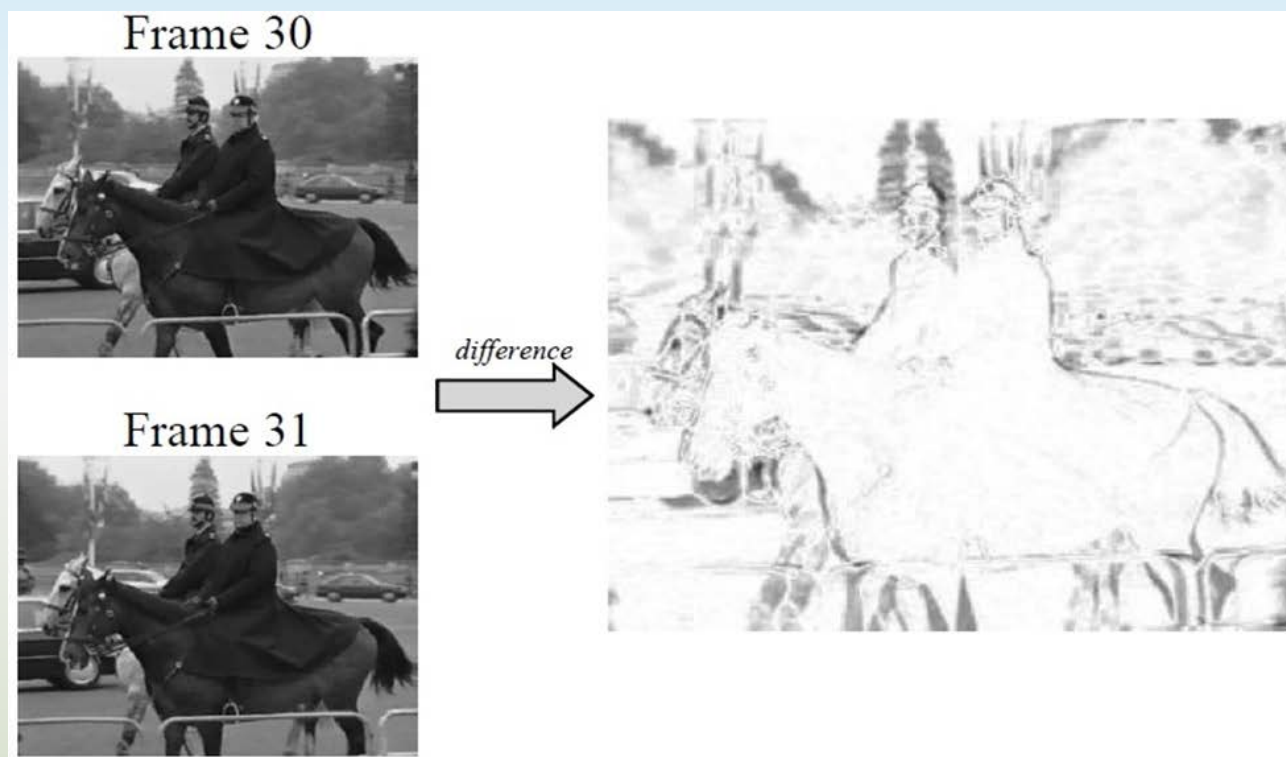
为什么可以压缩？

• 存在冗余

◆ 统计冗余

◆ 时域冗余

◆ 视频图像在时间上存在很大相关性，相邻帧的像素值重复或非常接近

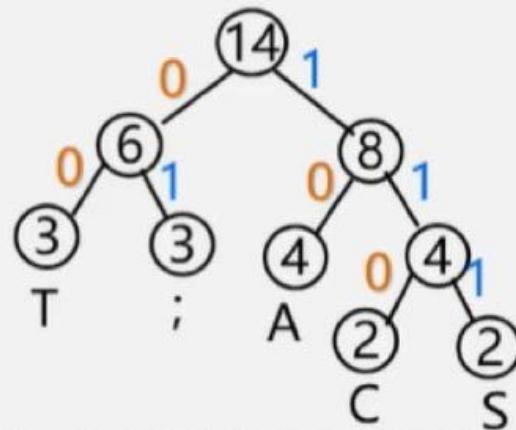


为什么可以压缩？

• 存在冗余

◇ 统计冗余

【例】要传输的字符集 $D = \{C, A, S, T, ;\}$
字符出现频率 $w = \{2, 4, 2, 3, 3\}$



T: 00
;: 01
A: 10
C: 110
S: 111

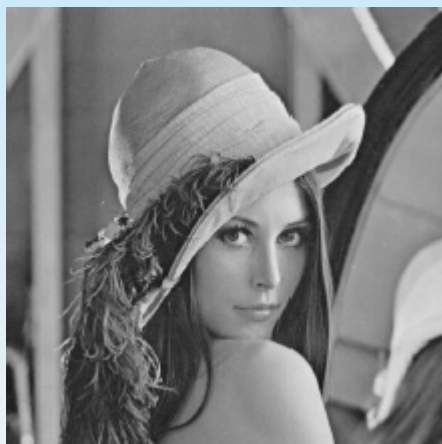
例 电文是 {CAS;CAT;SAT;AT}

其编码是: 11010111011101000011111000011000

◇ 编码冗余

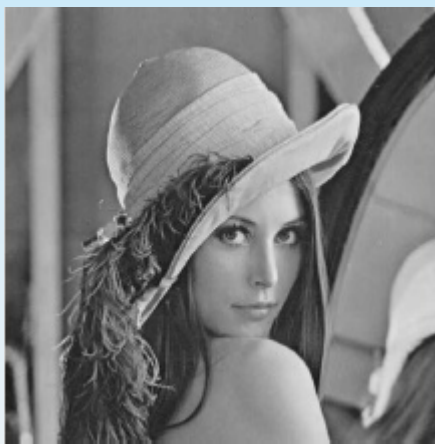
- 不同码字出现的频率不同，我们可以根据码字出现的概率，提高编码效率

压缩结果



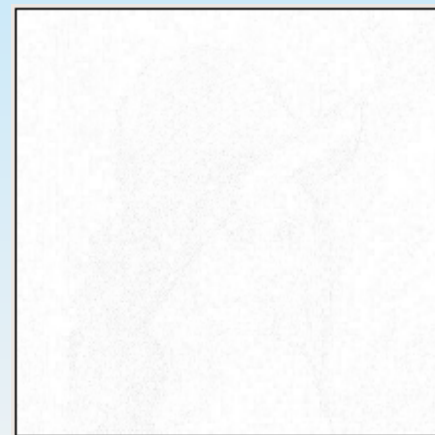
原图

-



12: 1

=



原图

-



43: 1

=



如何评价压缩效果，从哪些方面来进行评价？

图像多媒体基础

- 一、图像基础知识
- 二、为什么压缩
- 三、JPEG编码压缩

JPEG是什么

- ◆ 联合图像专家组（Joint Photographic Experts Group）的缩写
 - ◆ 由ISO和IEC两个组织机构联合组成的专家组，负责制定静态的数字图像数据压缩编码标准
- ◆ 可用于压缩灰度图像和彩色图像，包括两种基本压缩算法
 - ◆ 有损压缩算法：以离散余弦变换(DCT)为基础，在压缩比为25:1的情况下，压缩后还原得到的图像与原始图像相比，非图像专家难于找出它们之间的区别
 - ◆ 无损压缩算法：以预测技术为基础

JPEG是什么

◆ JPEG格式

- ◆ 存放使用JPEG压缩的图像文件交换格式，后缀为 .JPG或 .JFF，大多数浏览器支持此格式

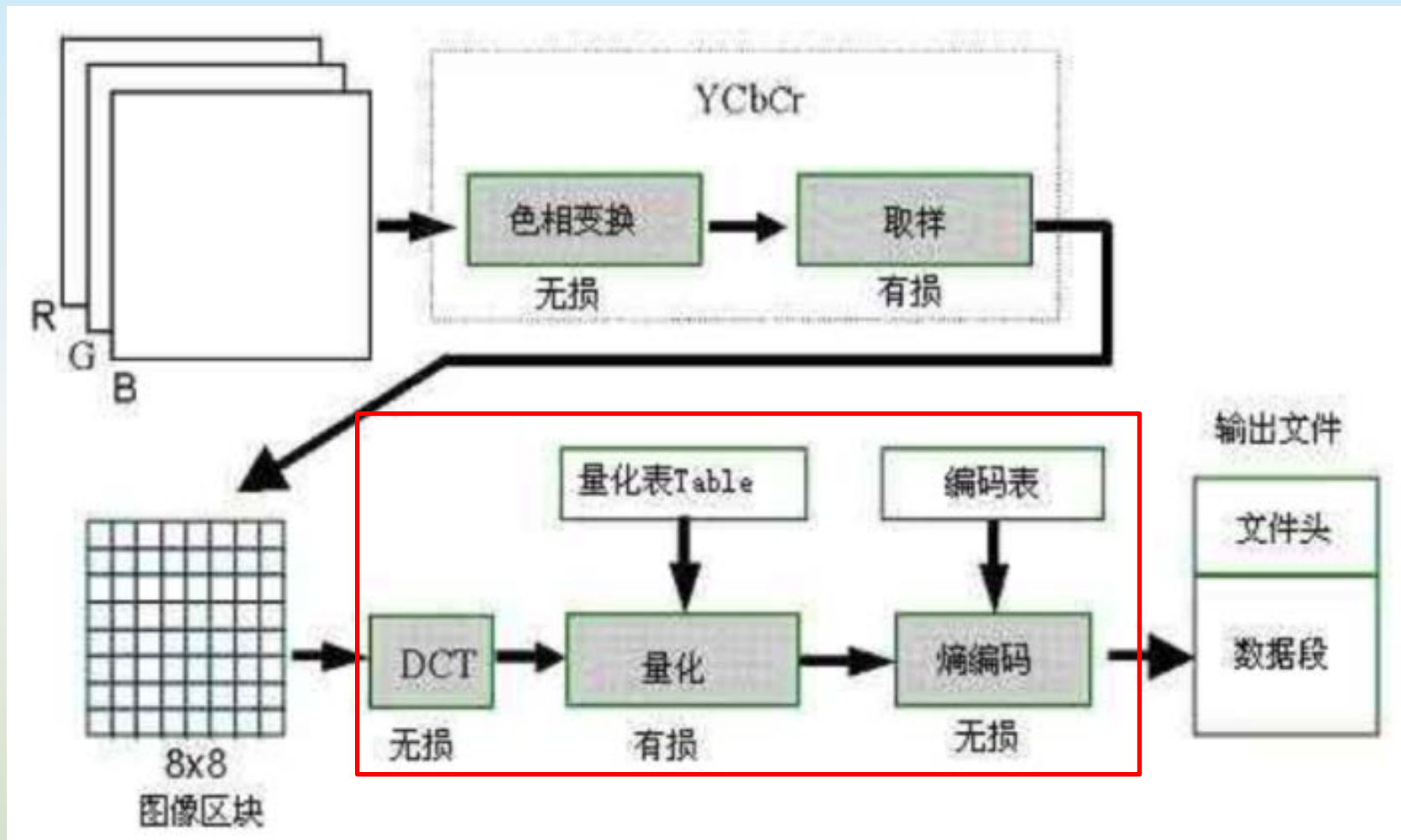
◆ JPEG算法

- ◆ 利用视觉系统特性，使用变换、量化和熵编码相结合的方法，以去掉或减少视觉的冗余信息和数据本身的冗余信息

JPEG是什么

- ◆ JPEG标准的压缩算法大致可以分成以下三个步骤:
 - ◆ 使用正向离散余弦变换(FDCT)将空间域表示的图像变换成频率域表示的图像
 - ◆ 使用加权函数对DCT系数进行量化, 加权函数对人的视觉系统是最佳的
 - ◆ 使用霍夫曼编码器对量化系数进行编码

JPEG算法概要



JPEG压缩编码流程图

DCT变换

- 离散余弦变换（Discrete Cosine Transform, DCT）是一种实数域变换，其变换核为实数余弦函数
- **特点：**对一幅图像进行离散余弦变换后，许多有关图像的重要可视信息都集中在DCT变换的一小部分系数中
- 因此，离散余弦变换（DCT）是有损图像压缩JPEG的核心，同时也是所谓“变换域信息隐藏算法”的主要“变换域”之一

DCT变换

二维离散余弦变换 (2DCT)

$$F(u, v) = \frac{1}{\sqrt{MN}} c(u)c(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cos\left[\frac{\pi u(2x+1)}{2M}\right] \cos\left[\frac{\pi v(2y+1)}{2N}\right]$$

$$u = 0, 1, 2, \dots, M-1 \quad v = 0, 1, 2, \dots, N-1$$

逆变换

$$f(x, y) = \frac{1}{\sqrt{MN}} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} c(u)c(v) F(u, v) \cos\left[\frac{\pi u(2x+1)}{2M}\right] \cos\left[\frac{\pi v(2y+1)}{2N}\right]$$

$$x = 0, 1, 2, \dots, M-1 \quad y = 0, 1, 2, \dots, N-1$$

$$c(u) = \begin{cases} 1 & u = 0 \\ \sqrt{2} & u \neq 0 \end{cases} \quad c(v) = \begin{cases} 1 & v = 0 \\ \sqrt{2} & v \neq 0 \end{cases}$$

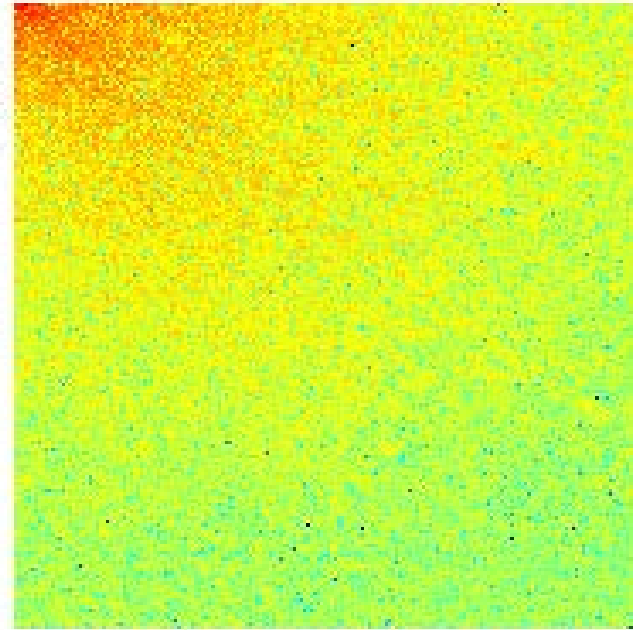
DCT变换

DCT系数的性质

原始图像

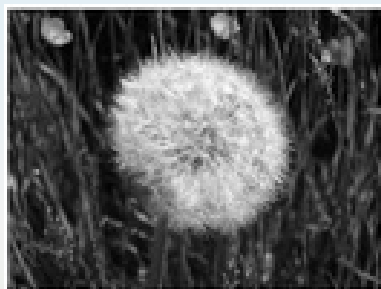


图像DCT系数的光谱表示

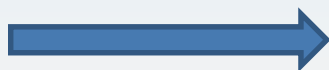


DCT变换

将时空域能量转换为频域



转换



DC系数

-415	-30	-61	27	56	-20	-2	0
4	-22	-61	10	13	-7	-9	5
-47	7	77	-25	-29	10	5	-6
-49	12	34	-15	10	6	2	2
12	-7	-13	-4	2	2	-3	3
-8	3	2	-6	-2	1	4	2
-1	0	0	-2	-1	-3	4	-1
0	0	-1	-4	-1	0	1	2

AC系数

特点：

- 左上角为直流DC系数，其余为交流AC系数
- 左上角到右下角数值的绝对值大体趋势逐渐变小
- 能量主要集中在左上角DC系数附近的一小部分

DCT变换

分块DCT (8×8)

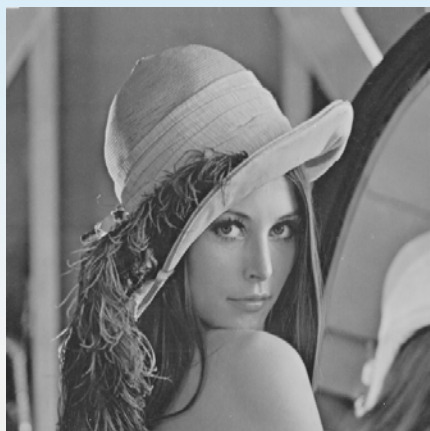
Array Editor: blocDCTmatrix

File Edit View Web Window Help

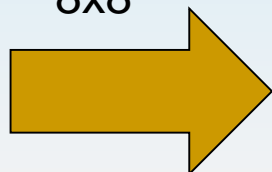
Numeric format: shortG Size: 256 by 768

	1	2	3	4	5	6	7	8	9
1	5.5294	0.02495	-0.00090576	0.030216	-0.00098039	0.0018297	0.00037518	0.0035344	5.4333
2	-0.011966	0.022077	-0.028234	1.9082e-005	-0.00027049	-0.0012524	0.00010351	-0.0018048	-0.058609
3	0.033176	-0.0018085	-0.0021638	-0.0011417	-0.00053058	0.0021862	0.00089629	-0.0015537	0.03152
4	0.0031053	2.8559e-005	0.00071165	0.0017299	0.00077029	0.0023427	-0.00063014	0.002811	0.00090639
5	-0.0019608	-0.0017719	-0.00090576	-0.00073024	0.00098039	-0.0017682	0.00037518	-0.0042651	-0.0029412
6	-0.0015462	0.00024911	-0.0010651	-0.00031136	-0.0011528	0.002003	-0.00031156	0.0007218	-0.0022236
7	0.0032368	0.0007228	0.0018767	-0.0017207	0.0012809	7.1779e-005	-0.00077735	-0.0003508	0.0018005
8	0.0086864	-0.00053729	0.0019961	0.00055914	0.0013598	-0.0010803	-0.00052039	-0.00032021	0.0057345
9	5.5005	-0.020722	-0.00064047	0.00083538	0.0004902	0.0019396	-0.00026529	0.0027038	5.5608
10	0.031344	-0.0015836	-0.00059558	0.0011489	-0.0023325	0.0014451	0.0011039	0.0030323	-0.037566
11	0.0083261	0.00066075	-0.00063682	0.0011943	0.00064047	0.00079799	-0.00034662	-0.0013987	0.032956
12	-0.0042148	-0.0025179	0.0022593	-0.00041249	0.00028363	-0.00084992	0.0017009	-0.0033385	-0.0015045
13	0.002451	-0.00071797	-0.00064047	-0.00021247	0.0004902	0.00083484	-0.00026529	-0.0049693	-0.0019608
14	0.0017817	-0.0016824	0.0015096	-0.00027562	0.001857	-0.0005679	-0.0032209	-0.0022307	0.0038542
15	0.0034488	0.0016243	-0.00034662	0.0012597	0.00026529	-0.0035157	-0.00014358	0.00032309	0.00014436
16	-0.0096583	-0.00031499	-0.0016486	0.00022853	-0.0018776	0.00028745	0.00021958	0.00060316	-0.0075936
17	5.6348	-0.018868	0.014822	0.0049372	-0.0004902	-0.0017036	-0.0013642	0.00048792	5.6848

DCT变换



8x8



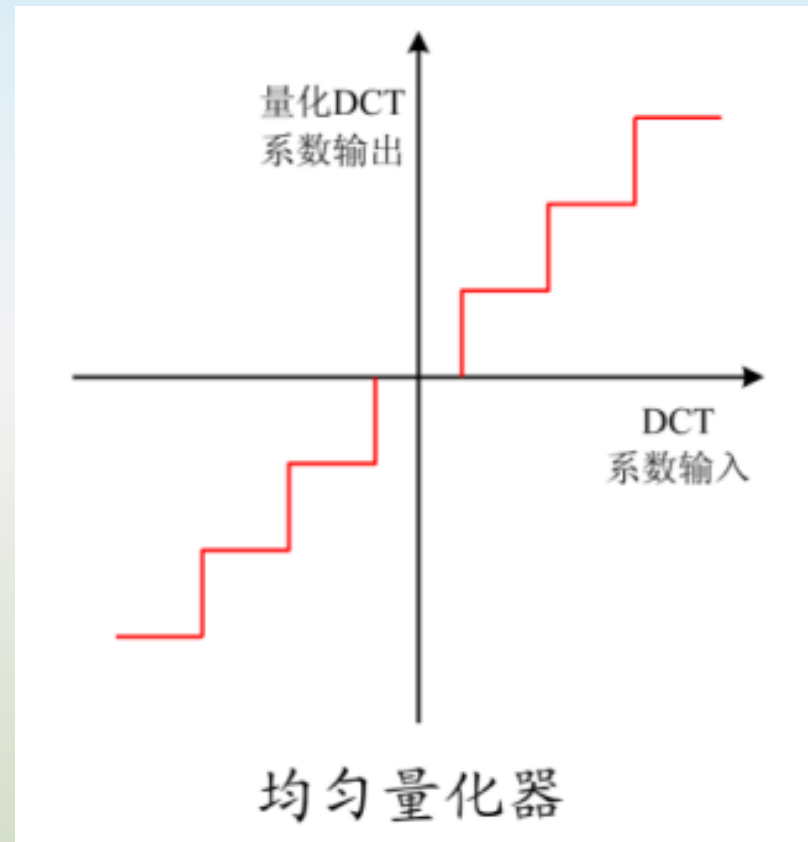
-415	-30	-61	27	56	-20	-2	0
4	-22	-61	10	13	-7	-9	5
-47	7	77	-25	-29	10	5	-6
-49	12	34	-15	-10	6	2	2
12	-7	-13	-4	-2	2	-3	3
-8	3	2	-6	-2	1	4	2
-1	0	0	-2	-1	-3	4	-1
0	0	-1	-4	-1	0	1	2

特点:

1. 变换后包涵低频，中频，高频系数；
2. 低频：集中了主要能量；
3. 中频：适宜做嵌入，既不引起视觉变化，也不会被轻易破坏；
4. 高频：噪声部分，容易被压缩过滤。

量化

- ◆ 对DCT变换后的频率系数进行量化
- ◆ 量化目的是降低非“0”系数的幅度以及增加“0”值系数的个数
- ◆ 量化是造成图像质量下降的最主要原因

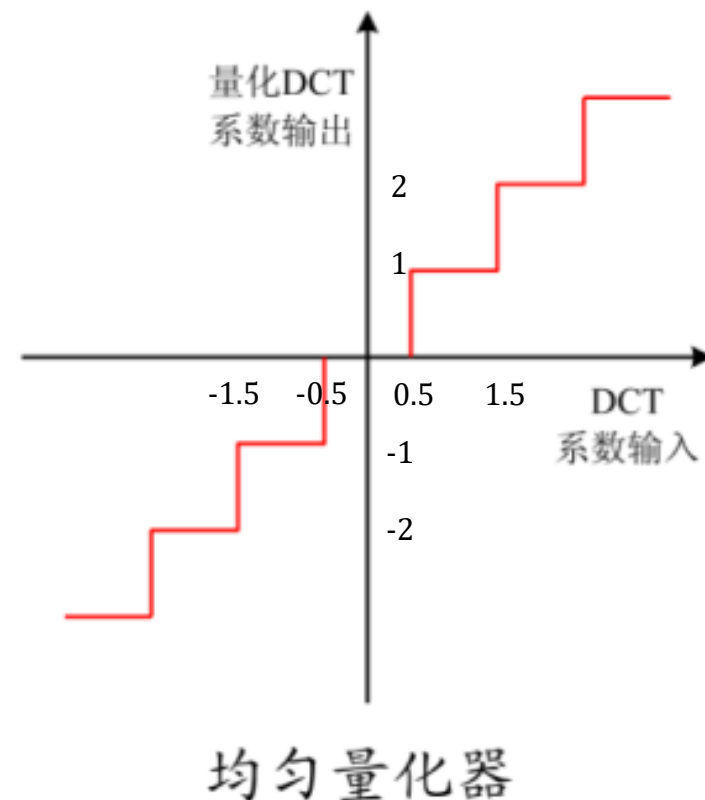


量化

◆ 用于量化的均匀量化器如右图所示，简单来说就是做除法，取整

◆ 量化公式如下：

$$\hat{F}(u, v) = \text{round}\left(\frac{F(u, v)}{Q(u, v)}\right)$$



量化

◆ 量化步距

- ◆ 按照系数所在的位置和每种颜色分量的色调值来确定的，亮度和色度量化步长表如下所示
- ◆ 因为人眼对亮度信号比对色差信号更敏感，亮度量化步长一般小于色差量化步长

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

亮度量化步长表

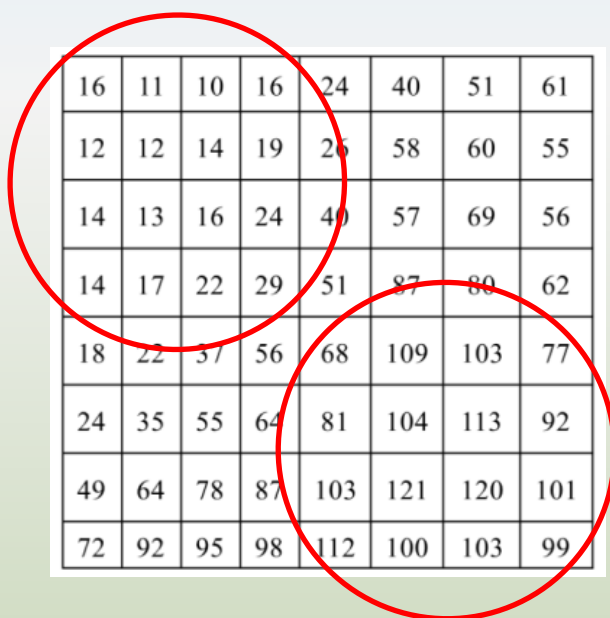
17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

色差量化步长表

量化

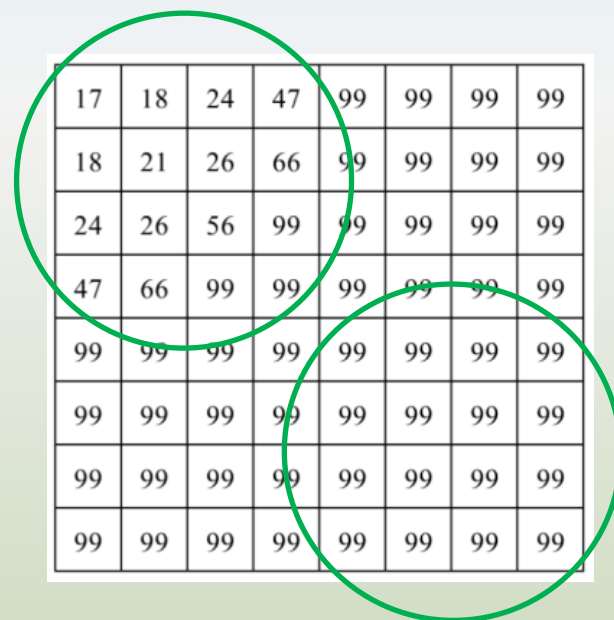
◆ 量化步距

- ◆ 由于人眼对低频分量的图像比对高频分量的图像更敏感，因此表中的左上角的量化步距要比右下角的量化步距小



16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

亮度量化步长表



17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

色差量化步长表

量化

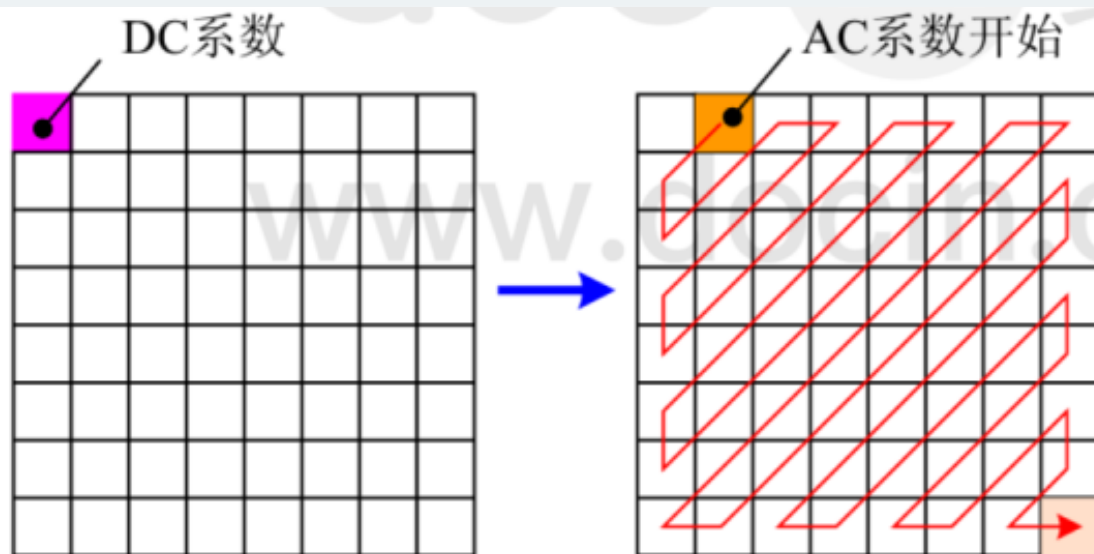
- ◇ 信息损失的来源
- ◇ 是多媒体数据压缩的一个关键步骤
- ◇ 为了对这个量化的效果进行客观的量化的评价，量化误差的公式如下所示：

$$e_q = x - Q^{-1}(Q(x))$$

量化

◆ Z字形编排

- ◆ 为增加连续的“0”值系数的个数，即“0”的游程长度，DCT系数的序号如图所示，序号小的位置表示频率较低，通过Z字形编排将一个8 x 8的矩阵变成一个1 x 64的矢量



0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

量化DCT系数的编排

DCT系数序号

游程编码

- ◇ 接下来，需要对“之”字形扫描后的量化DCT系数进行游程编码
- ◇ 例如游程编码前DCT AC系数为57,45,0,0,0,0,23, 0,-30,-16,0,0,1,0,0,0 ,..., 0,0,0；游程编码后DCT AC系数为(0,57), (0,45) , (4,23), (1,-30), (0,-16), (2,1), (0,0)

结束标志，后面都是“0”

游程编码

- ◆ 由于Huffman编码要求，每组前一个表示系数零数量的数字大小为4 bit，只能在0~15范围内取值

- ◆ 例如游程编码前DCT系数57, $\overbrace{0,0,0,\dots,0,0,0}^{18\text{个“0”}}$, 3, 0, 0, 0, 0, -2, 0, 0, \dots , 0, 0, 0; 进行编码，得到的游程编码 结果为(0,57), (15,0), (2,3), (4,-2), (0,0)
- 16个“0” 2个“0”, 3

熵编码

- ❖ 为了提高存储效率，JPEG 压缩并不是直接存储游程编码值，而是根据量化 DCT 系数的统计特性，对游程编码后得到的系数值再进行 Huffman 编码压缩
- ❖ Huffman 编码是一种基于概率统计思想的编码方式，它对概率较大的事件分配较小的空间存储，对概率较小的事件分配较多的空间存储
- ❖ 熵编码可以分为 AC 系数和 DC 系数的 Huffman 编码两部分

Huffman 编码

编码长度	实际数值	编码Bits
1	-1,1	0,1
2	-3,-2,2,3	00,01,10,11
3	-7,-6,-5,-4,4,5,6,7	000,001,010,011,100,101,110,111
4	-15,...,-8,8,...,15	0000,...,0111,1000,...,1111
5	-31,...,-16,16,...,31	00000,...,01111,10000,...,11111
6	-63,...,-32,32,...,63	000000,...,011111,100000,...,111111
7	-127,...,-64,64,...,127	0000000,...,0111111,1000000,...,1111111
8	-255,...,-128,128,...,255	...
9	-511,...,-256,256,...,511	...
10	-1023,...,-512,512,...,1023	...
11	-2047,...,-1024,1024,...,2047	...

AC系数Huffman 编码

Huffman 编码

● 示例

DCT变换后63位 AC系数的游程编码如下：

(0,57) ; (0,45) ; (4,23) ; (1,-30) ; (0,-8) ; (2,1) ; (0,0)

对应的Huffman 编码如下：

(0,6,111001) ; (0,6,101101) ; (4,5,10111);
(1,5,00001) ; (0,4,0111) ; (2,1,1) ; (0,0)

练习

DCT变换后63位 AC系数为98, $\overbrace{0, 0, \dots, 0}^{19\text{个“0”}}$, 9, 0, 0, 0, 0, 7, $\underbrace{0, 0, \dots, 0}_{33\text{个“0”}}$, 2, 0, 0, 0, 写出游程编码结果。