

Basic idea of DCT & Watson's DCT-Based Visual Mode

展示人: 罗文杰

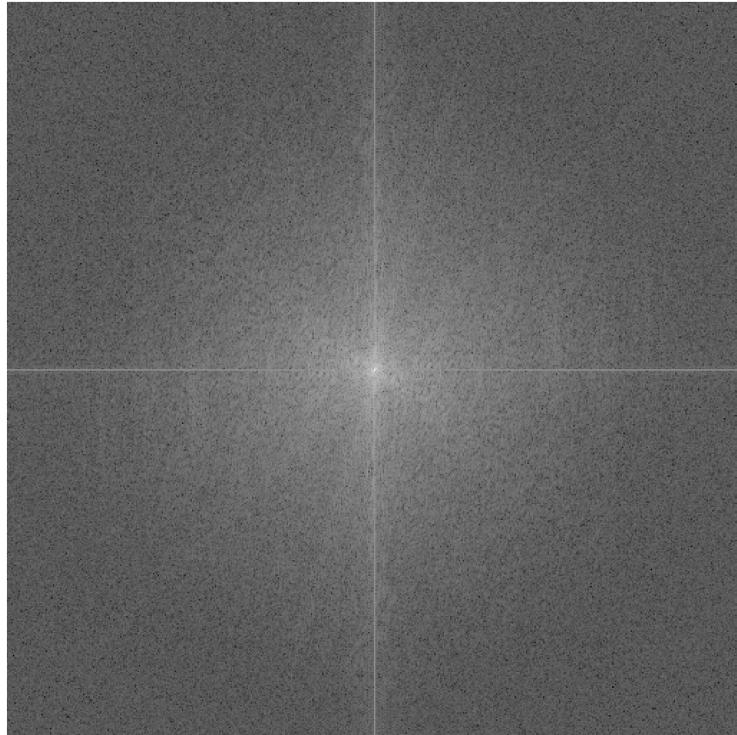
为什么需要 DCT?

对于一张图片，我们有两种理解视角

Spatial Domain



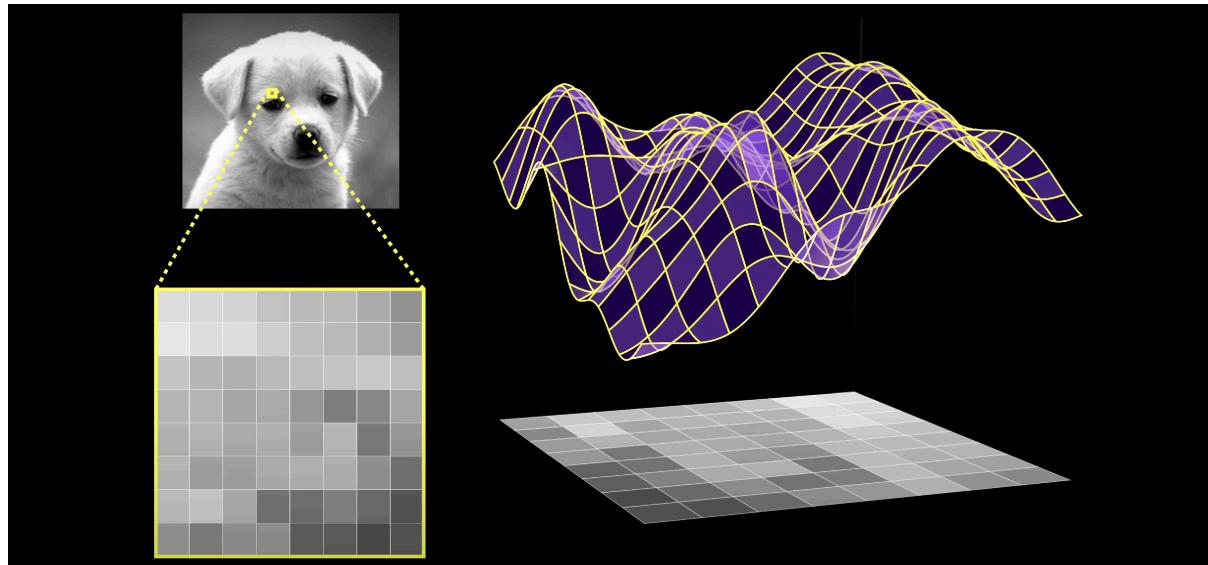
Frequency Domain



为什么需要 DCT?

我们来理解两个视角是如何转换的

1. 如果将图片的某行切出来，以坐标为 x 轴，以该行所有像素点的灰度值为 y 轴作图，我们可以得到图片在该行的像素信号。进一步，将整张图片的所有像素点的灰度值画在一张图上，我们得到了图片的信号描述
2. 回顾傅里叶变换的基本思想：将一个函数分解为一组正弦或者余弦函数的加权和。通过这样的手段对图片信号进行分解，我们就成功地将图片信号从空间域转换到了频域



为什么需要 DCT?

频域变换能够帮助我们理解图像，它带来的好处有很多

- **压缩**: 大量实践表明，人眼更容易感知图片的低频信息，对高频部分则不太敏感。依靠此事实，我们可以选择去除图片的一部分高频信息，进而压缩图片大小 (JPEG)
- **去噪**: 在频域中，图像的噪声往往集中在高频部分。通过频域滤波，可以有效地去除噪而保留图像的主要信息 (高斯滤波器、中值滤波器)

...

作为一种频域变换的方法，DCT 则具有如下优点：

- **能量集中**: DCT 可以将信号的大部分能量集中在少量的系数上，这意味着我们可以用较少的系数来近似表示原始信号
- **易于计算**: 与傅里叶变换相比，DCT 的计算复杂度更低。DCT 中只涉及余弦函数的计算，而不涉及正弦函数，这使得计算更加简单和高效
- **频率敏感**: DCT 对低频信号更加敏感，这意味着在 DCT 域中，较低频率的成分通常会有较大的幅度，这对于许多信号处理和图像处理应用是有益的

如何进行 DCT

来看 DCT 算法是怎么做的

DCT 算法的全称是 Discrete Cosine Transform，核心思想是用不同频率的**余弦函数**之和来表示离散的**数据序列**

具体来说，针对 N 个实数 x_0, \dots, x_{N-1} 组成的序列，通过

$$X_k = \sum_{n=0}^{N-1} x_n \cos \left[\frac{(2n+1)\pi k}{2N} \right]$$

得到一组新序列 X_0, \dots, X_{N-1} ，其中 X_k 代表了最终的频域表示中 $\cos(kx)$ 的系数

最终的分解结果为

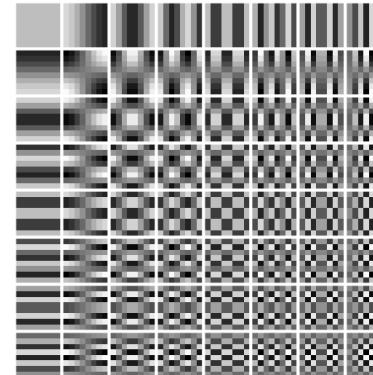
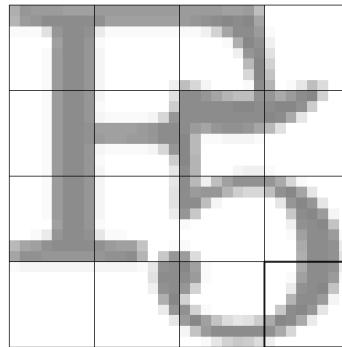
$$\sum_{k=0}^{N-1} X_k \cos(kx)$$

如何进行 DCT

针对图像，我们需要使用 DCT 的二维版本

首先**按行**对图片信号进行一次一维 DCT 变换，得到一组 DCT 系数矩阵，接着对该矩阵**按列**再进行一次 DCT 变换，得到最终的 DCT 系数

大多数时候，我们会将图片分成若干 8×8 大小的块，以块为单位作 DCT 变换



$$\begin{matrix} \text{Block} \\ = c_1 \cdot \text{Block}_1 + c_2 \cdot \text{Block}_2 + \dots + c_{64} \cdot \text{Block}_{64} \end{matrix}$$

Watson's Visual Model

借助 DCT 算法，我们可以评估图片添加水印前后的视觉保真度（visual fidelity）

一种朴素的做法是，直接求两张图片的 DCT 系数之差作为感知距离。这种做法的缺点在于，并没有考虑人眼对图片的感知特点，计算的结果不能很好地反映 JND (Just Noticeable Difference) 的大小

具体来说，人眼的视觉感知有以下特点：

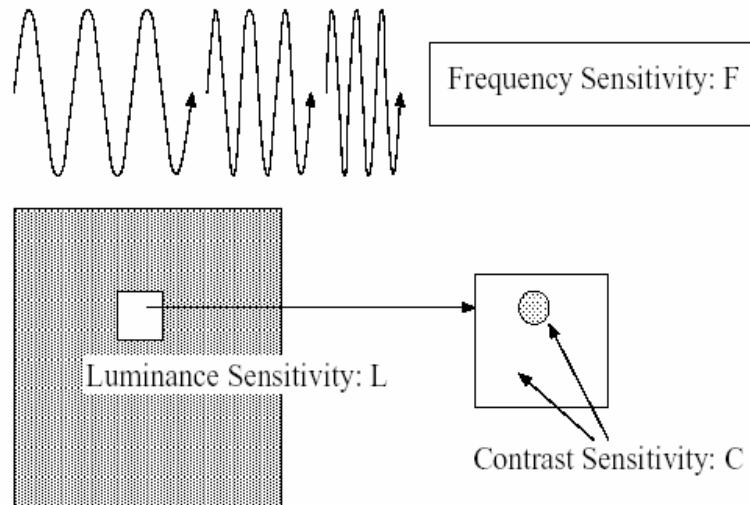
- 对高频信息不敏感，对低频信息敏感
- 如果亮度已经很高，人眼几乎感觉不到任何亮度增加
- 如果两个色调之间的差异太小，人眼几乎感觉不到任何差异

Watson's Visual Model 利用了如上特点，能够很好地进行 JND 估计

Watson's Visual Model

Watson's Visual Model 的流程如下：

给定原始图像 c_o 和被破坏的图像 c_w ，将图像分割成若干大小为 8×8 的块，将第 k 个块上第 i 行 j 列的像素表示为 $c[i, j, k]$ 。接着对所有块进行 DCT 变换，得到图片的 DCT 表示 C ，其中第 k 个块的系数矩阵为 $C[i, j, k]$ ， $C[0, 0, k]$ 代表第 k 个块的 DC (Direct Current) 分量，也即块中像素的平均能量。接着，通过 Sensitivity, Luminance Masking 和 Contrast Masking 对每个块上两张图片的 DCT 差作缩放，再做一次池化处理，得到两张图片的差异 $D_{wat}(c_o, c_w)$



Watson's Visual Model

Sensitivity

Watson's Visual Model 定义了一组频率敏感度表 t . 表中的数据 $t[i, j]$ 衡量了人眼对某种频率的敏感程度，值越低，代表人眼对其越敏感

Table 8.2 DCT frequency sensitivity table.

1.40	1.01	1.16	1.66	2.40	3.43	4.79	6.56
1.01	1.45	1.32	1.52	2.00	2.71	3.67	4.93
1.16	1.32	2.24	2.59	2.98	3.64	4.60	5.88
1.66	1.52	2.59	3.77	4.55	5.30	6.28	7.60
2.40	2.00	2.98	4.55	6.15	7.46	8.71	10.17
3.43	2.71	3.64	5.30	7.46	9.62	11.58	13.51
4.79	3.67	4.60	6.28	8.71	11.58	14.50	17.29
6.56	4.93	5.88	7.60	10.17	13.51	17.29	21.15

Watson's Visual Model

Luminance Masking

利用图像的亮度信息，Watson's Visual Model 对敏感度表进行修正，修正公式为

$$t_L[i, j, k] = t[i, j] (C_o[0, 0, k]/C_{0,0})^{a_T}$$

其中， $C_{0,0}$ 代表整张图片的直流分量之和， a_T 为常数，建议为 0.649

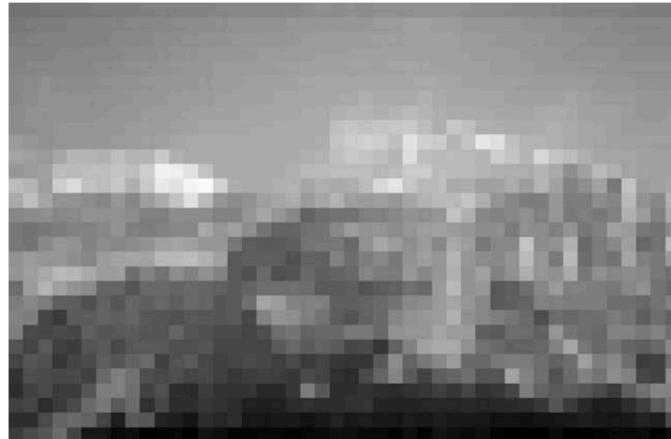


FIGURE 8.9

Relative luminance-masking thresholds for Figure 8.8(a). Bright areas indicate blocks with high luminance-masking values.

Watson's Visual Model

Contrast Masking

利用对比度信息，Watson's Visual Model 对敏感度表做进一步修正，修正公式为

$$s[i, j, k] = \max\{t_L[i, j, k], |C_o[i, j, k]|^{w[i, j]} t_L[i, j, k]^{1-w[i, j]}\}$$

其中 $w[i, j]$ 为常数，建议为 0.7

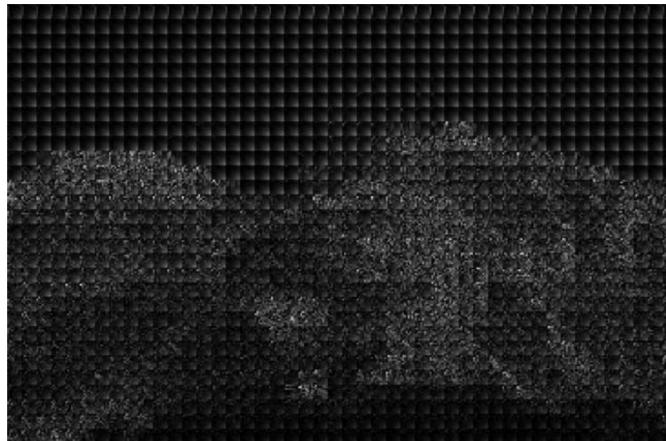


FIGURE 8.10

Relative masking thresholds, or slacks, for Figure 8.8(a). The brightness of each pixel here is proportional to the slack of the corresponding block-DCT coefficient.

Watson's Visual Model

Pooling

借助上一步得到的约束项，我们可以计算两个对应块的感知距离

$$d[i, j, k] = \frac{e[i, j, k]}{s[i, j, k]}$$

其中 $e[i, j, k] = C_w[i, j, k] - C_o[i, j, k]$, 代表了 DCT 系数差

最后，将每个块中不同频率的误差组合在一起得到最终的感知距离

$$D_{wat}(c_o, c_w) = \left(\sum_{i,j,k} |d[i, j, k]|^p \right)^{\frac{1}{p}}$$

其中 p 为常数，建议为 4

谢谢大家