

数字媒体技术实验二

刘绍辉

shliu@hit.edu.cn

哈尔滨工业大学计算机科学与技术学院

2023春

■ 实验二（了解原理，可以直接调用函数实现）

- 能对音频文件进行DFT, DCT和DWT变换
 - 例如，读入音频文件，以1024长度对音频分窗处理，然后对其进行一维的DFT, DCT, DWT处理，然后画出原始音频，以及处理后音频的图形
- 能对图像进行二维的DFT, DCT, DWT变换
 - 理解FFT的优势：用一维的DFT变换对图像块进行变换，分别显示其幅度图和相位图（注意如何可视化结果，例如，将系数区间归一化到[0,255]，或者系数取对数等）；用FFT变换来对图像块进行变换，看看其速度；
 - 理解DCT变换的能量聚集特性，8*8变换后，保留左上角k个系数后（做Zigzag扫描，将8*8转化为64维向量），再做逆DCT变换，恢复原始图像，比较原始图像与恢复图像的PSNR值和SSIM值(需要查阅PSNR和SSIM的公式并实现)
 - 理解DWT变换的频率特性
- 有以上知识的同学尝试下列任务（选做）
 - 理解MP3, JPEG, H.264, AVS, H.265里面DCT算法的应用流程，理解从浮点数变换到整数变换的变换思想

思考题：一般做目标检测、人脸识别等任务，都是从图像空域考虑，能否从DCT变换考虑？直接将JPEG的DCT系数输入分类器来进行检测和识别？

- 设 $f(x)$ 是离散变量 x 的函数，一维离散函数 $f(x)$ 的傅里叶变换 $F(u)$ 定义为

$$F(u) = \sum_{x=-\infty}^{\infty} f(x)e^{-j2\pi ux}$$

- 这里 u 为连续频率变量， $F(u)$ 是周期为1的复值、连续函数。显然，离散信号傅里叶变换的频率范围限定在 $[-1/2, 1/2]$ 或等价的 $[0, 1]$ 上，在这个区间之外的任何频率与这个区间之内频率的傅里叶变换是相等的
- 对应的，一维离散函数的傅里叶逆变换定义为

$$f(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} F(u)e^{j2\pi ux} du,$$

- 注意：离散函数 $f(x)$ 是绝对可和的，即 $\sum_{x=-\infty}^{\infty} |f(x)| < \infty$ 是离散函数傅里叶变换存在的充分条件

1-D DFT



- 离散函数傅里叶变换具有连续谱

- 计算机只能处理离散数据
- 傅里叶变换的计算是作用于全局的，具有很大的计算量，而离散化则有利于计算推导快速算法
- 因此对离散函数的傅里叶变换的连续谱进行采样而获得离散谱

- 通过对连续谱在 $[0,1]$ 区间抽取 N 个等间距采样样本，采样间隔 $\Delta u = \frac{1}{N}$ ， N 为频率采样的样本数，计算连续谱的采样值，则其一维离散函数的离散傅里叶变换定义为：

$$F(u) = \sum_{x=0}^{N-1} f(x) e^{-j \frac{2\pi u x}{N}}, u = 0, 1, \dots, N-1$$

- 这里 u 为离散的频率变量，对应的一维离散傅里叶变换的逆变换为

$$f(x) = \frac{1}{N} \sum_{u=0}^{N-1} F(u) e^{j \frac{2\pi u x}{N}}, x = 0, 1, \dots, N-1$$

$$f(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} F(u) e^{j 2\pi u x} du$$

- 从计算公式可以得到
 - 每次计算 $F(u)$,需要 N 次复数乘法和 $N - 1$ 次复数加法, 等价于 $4N$ 次实数乘法和 $4N - 2$ 次实数加法, 因此计算长度为 N 的离散傅里叶变换, 需要 N^2 次复数乘法和 $N^2 - N$ 次复数加法
- $F(u) = |F(u)|e^{j\theta(u)}$
- 其中 $|F(u)| = \sqrt{R^2(u) + I^2(u)}$,称为傅里叶幅度谱或傅里叶谱, 而 $\theta(u) = \arctan\left(\frac{I(u)}{R(u)}\right)$,则称为傅里叶相位谱, 而功率谱一般定义为 $P(u) = |F(u)|^2$,反映信号中能量的分布情况
- 二维DFT是一维DFT的直接推广

- 设一幅图像 $f(x, y)$ 的大小为 $M \times N$, 则其二维离散傅里叶变换 $F(u, v)$ 定义为:

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)}$$

- 这里 $u = 0, 1, \dots, M-1, v = 0, 1, \dots, N-1$
- 其二维离散傅里叶逆变换定义为:

$$f(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)}$$

- 这里 $x = 0, 1, \dots, M-1, y = 0, 1, \dots, N-1$
- 注意: $\frac{1}{MN}$ 的位置不固定, 既可以在正变换中, 也可以在逆变换中, 只是确保正变换再逆变换后恒等

- 类似一维情况，可以定义二维离散傅里叶变换的傅里叶谱 $|F(u, v)|$ ，相位谱 $\theta(u, v)$ 和功率谱 $P(u, v)$ 为

$$|F(u, v)| = \sqrt{R^2(u, v) + I^2(u, v)}$$

$$\theta(u, v) = \arctan\left(\frac{I(u, v)}{R(u, v)}\right)$$

$$P(u, v) = |F(u, v)|^2$$

- $F(u, v) = R(u, v) + iI(u, v) = |F(u, v)|e^{i\theta(u, v)}$
- 空域与频域采样间隔之间的关系如下
- $\Delta u = \frac{1}{M\Delta x}$, $\Delta v = \frac{1}{N\Delta y}$, 分别表示空域和频域中采样间隔的关系
- 实验中要求将图像的幅度谱、相位谱、功率谱都计算并显示出来，体会幅度谱、相位谱和功率谱的重要性

傅里叶变换的卷积定理



- 两个支撑大小为 $M \times N$ 的离散函数 $f(x, y)$ 与 $h(x, y)$ 的卷积 $f(x, y) * h(x, y)$ 定义为

$$f(x, y) * h(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(x - m, y - n)$$

- 卷积定理表明空域卷积等价于频域乘积

$$f(x, y) * h(x, y) \leftrightarrow F(u, v) H(u, v)$$

- 证明:** $\mathcal{F}[f(x, y) * h(x, y)] = \sum_x \sum_y f(x, y) * h(x, y) e^{-j2\pi(ux+vy)}$

$$\begin{aligned} &= \sum_x \sum_y \left[\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(x - m, y - n) \right] e^{-j2\pi(ux+vy)} \\ &= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) \left[\sum_x \sum_y h(x - m, y - n) e^{-j2\pi(u(x-m)+v(y-n))} \right] e^{-j2\pi(um+vn)} \\ &= F(u, v) H(u, v) \end{aligned}$$

- 两个支撑尺寸大小为 $M \times N$ 的离散函数 $f(x, y)$, $h(x, y)$ 的相关 $f(x, y) \circ h(x, y)$ 定义为

傅里叶变换的卷积定理



- 两个支撑尺寸大小为 $M \times N$ 的离散函数 $f(x, y), h(x, y)$ 的相关 $f(x, y) \circ h(x, y)$ 定义为

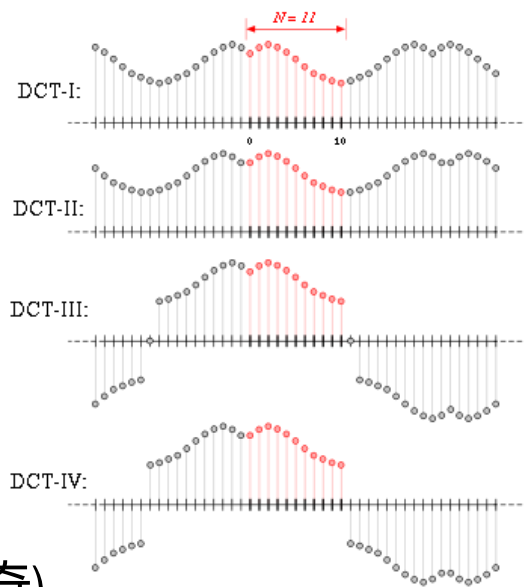
$$f(x, y) \circ h(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(m - x, n - y)$$

- 对比卷积与相关运算
- $f(x, y) \circ h(-x, -y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(-(m - x), -(n - y))$
- $= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(x - m, y - n) = f(x, y) * h(x, y)$
- 从而可知, 卷积计算包括反褶, 平移, 相乘再相加。而相关计算没有反褶。如果将函数 $f(x, y)$ 与反褶函数 $h(-x, -y)$ 作为输入, 则可以直接使用计算卷积的程序来计算相关
- 类似, $\mathcal{F}[f(x, y) \circ h(x, y)] = \sum_x \sum_y [f(x, y) \circ h(x, y)] e^{-j2\pi(ux+vy)}$
- $= \sum_x \sum_y [\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(m - x, n - y)] e^{-j2\pi(ux+vy)}$
- $= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) [\sum_x \sum_y h(m - x, n - y) e^{-j2\pi(u(m-x)+v(n-y))}]^* e^{-j2\pi(um+vn)} = \mathcal{F}[f(x, y)] \mathcal{F}[h(x, y)]^* = F(u, v) H^*(u, v)$

- 表示互为傅里叶变换对
 - $f(x, y) * h(x, y) \leftrightarrow F(u, v)H(u, v)$
 - $\mathcal{F}[f(x, y) \circ h(x, y)] \leftrightarrow F(u, v)H^*(u, v)$
- 在图像处理中，空域卷积的主要作用是空域滤波，而空域相关的主要作用是图像匹配
 - 利用 $h(x, y)$ 表示待检测目标或感兴趣区域，通常称为模板(template)，因此图像匹配也成为模板匹配
 - 若输入图像 $f(x, y)$ 中包含匹配的模板，则在 $f(x, y)$ 和 $h(x, y)$ 的完全匹配位置上这两个函数的相关系数达到最大值
 - 由于模板 $h(x, y)$ 和输入图像 $f(x, y)$ 中的目标大小未必一样，通常利用高斯金字塔方法对模板进行逐级下采样，分别用不同分辨率的模板与输入图像做模板匹配

- DCT (Discrete Cosine Transform)
 - N 维的离散输入序列 $x(n), n = 0, 1, \dots, N - 1$
 - DFT为复数的线性组合, 即使 $x(n)$ 是实数
- 假设我们找到一个 $N \times N$ 的正交变换, 将实数序列表达为余弦序列的线性组合
 - $X(k) = \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2\pi kn}{N}\right), 0 \leq k \leq N - 1$
 - $x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \cos\left(\frac{2\pi kn}{N}\right), 0 \leq n \leq N - 1$
- 问题: 这是否可能?
 - 如何 N 点的序列 $x(n)$ 是实偶序列, 这是可能的, $x(n) = x(N - n), 0 \leq n \leq N - 1$, 结果其DFT就是实偶的
 - 因此, 对任意的 N 点序列 $x(n)$, 可能存在该序列偶扩展序列的 $2N$ 点的DFT来获得其离散余弦表示
 - 通过对 n 点离散输入序列 $x(n)$ 做偶对称扩展, 然后对该实偶序列做DFT, 后可获得其离散余弦表示

- DCT和DFT变换的区别在于原始信号是在正弦与余弦上展开，还是仅在余弦函数上展开
 - 其本质在于对DFT变换的不同边界条件的处理
 - DFT为周期扩展，DCT为偶对称扩展
- 假设输入序列为 x_0, x_1, x_2, x_3 , 或者 $abcd$, 这时候在序列左右两端是偶对称还是奇对称，对于偶的左边界对称又可分为是整数点对称还是半整数点对称
 - $abcd \rightarrow \cdots dcbabcdcba \cdots$
 - $abcd \rightarrow \cdots dcbaabcd dcba \cdots$
 - $abcd \rightarrow \cdots (-b)(-c)(-d)0dcbabcd0(-d)(-c) \cdots$
 - $abcd \rightarrow \cdots (-a)(-b)(-c)(-d)dcbaabcd(-d)(-c) \cdots$
- 根据这种讨论，边界：左右，边界对称有
- 两种情况：奇偶，对称有两种情况：
- 整点和半整点一共导致16种情况，其中8种
- 为余弦（左边界为偶），8种为正弦(左边界为奇)



■ 最常见的DCT变换为DCTII型变换

- $X(k) = \sum_{n=0}^{N-1} x(n) \cos \left[\frac{\pi(2n+1)k}{2N} \right], k = 0, \dots, N-1$
- 其边界条件对应为: $x(n)$ 关于 $n = -\frac{1}{2}$ 偶对称, $n = N - \frac{1}{2}$ 偶对称; $X(k)$ 在 $k = 0$ 偶对称, $k = N$ 奇对称

■ 一般情况

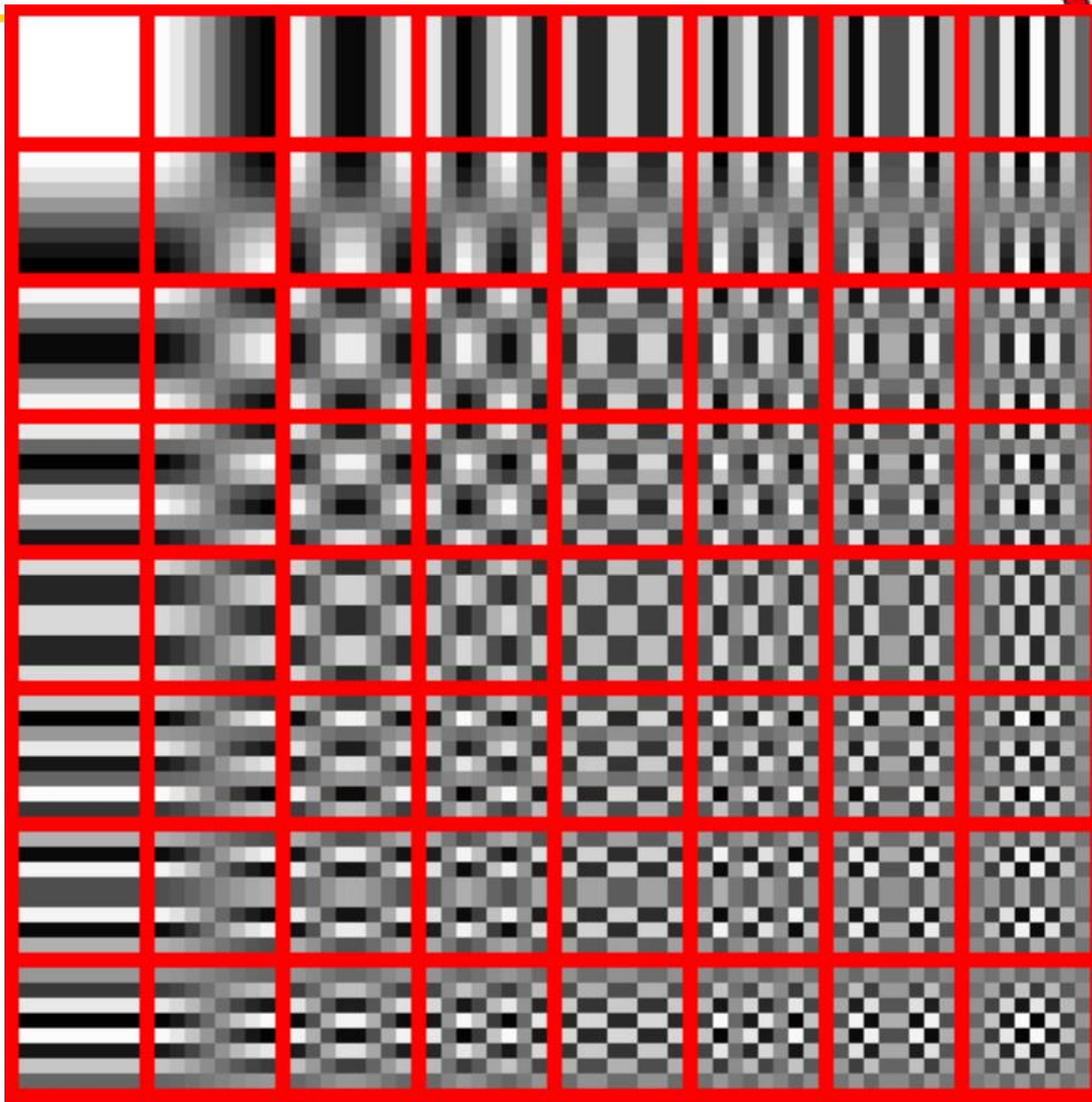
- $k = 0, X(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n)$
- $k \in [1, N-1], X(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos \left[\frac{(2n+1)k\pi}{2N} \right]$
- $x(n) = \frac{1}{\sqrt{N}} X(0) + \sqrt{\frac{2}{N}} \sum_{k=1}^{N-1} X(k) \cos \left[\frac{(2n+1)k\pi}{2N} \right], n \in [0, N-1]$

■ 矩阵形式

- $\mathbf{X} = \mathbf{C}\mathbf{x}, \mathbf{C} = [c(k, n)], c(k, n) = \begin{cases} \frac{1}{\sqrt{N}}, k = 0, n \in [0, N-1] \\ \sqrt{2/N} \cos \left[\frac{(2n+1)k\pi}{2N} \right], k \in [1, N-1], n \in [0, N-1] \end{cases}$, 且 $\mathbf{x} = \mathbf{C}^{-1}\mathbf{X} = \mathbf{C}^t\mathbf{X}, \mathbf{C} = \mathbf{C}^*, \mathbf{C}^{-1} = \mathbf{C}^t$

64个8x8的2维DCT基函数

ILAB



DCT(Discrete)变换的例子

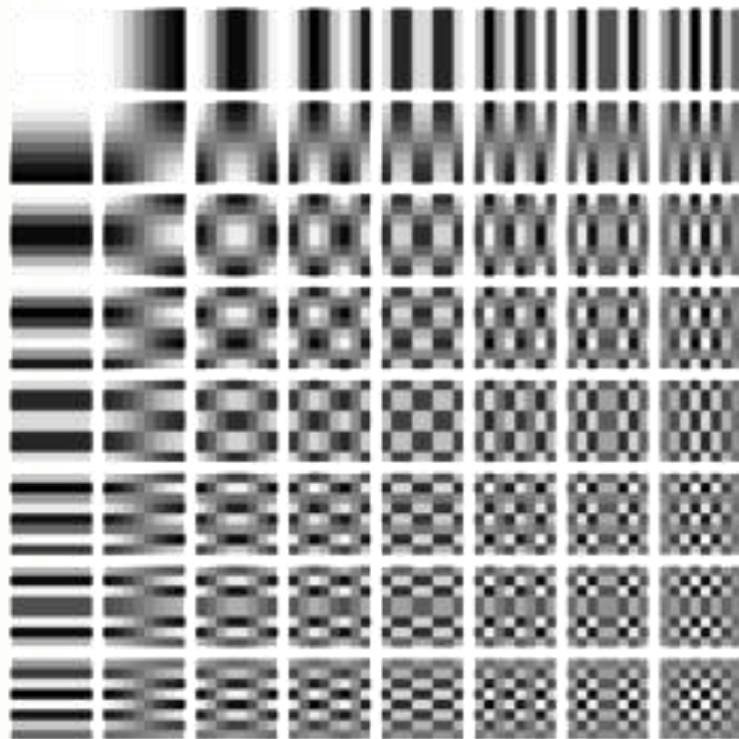


- 假设图像大小为 8×8 的字符图像



原始大小图像, 扩大10倍(最近邻方法), 扩大10倍(双线性差值).

- DCT基函数



5.1917	-0.3411	1.2418	0.1492	0.1583	0.2742	-0.0724	0.0561
0.2205	0.0214	0.4503	0.3947	-0.7845	-0.4391	0.1001	-0.2554
1.0423	0.2214	-1.0017	-0.2720	0.0789	-0.1952	0.2801	0.4719
-0.2340	-0.0392	-0.2517	-0.2665	0.6351	0.3501	-0.1433	0.3550
0.2750	0.0225	0.1229	0.2183	-0.2583	-0.0742	-0.2042	-0.5905
0.0558	0.0428	-0.4721	-0.2905	0.4745	0.2875	-0.0284	-0.1611
0.3169	0.0541	-0.1033	-0.0225	-0.0055	0.1017	-0.1650	-0.1500
-0.2970	-0.0627	0.1960	0.0544	-0.1136	-0.1031	0.1887	0.1444

DCT变换的例子



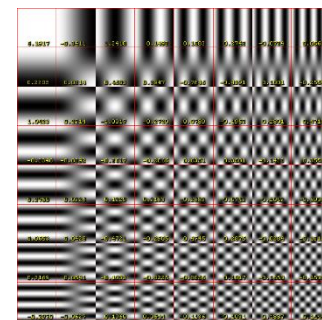
■ DCT系数

6.1917	-0.3411	1.2418	0.1492	0.1583	0.2742	-0.0724	0.0561
0.2205	0.0214	0.4503	0.3947	-0.7846	-0.4391	0.1001	-0.2554
1.0423	0.2214	-1.0017	-0.2720	0.0789	-0.1952	0.2801	0.4713
-0.2340	-0.0392	-0.2617	-0.2866	0.6351	0.3501	-0.1433	0.3550
0.2750	0.0226	0.1229	0.2183	-0.2583	-0.0742	-0.2042	-0.5906
0.0653	0.0428	-0.4721	-0.2905	0.4745	0.2875	-0.0284	-0.1311
0.3169	0.0541	-0.1033	-0.0225	-0.0056	0.1017	-0.1650	-0.1500
-0.2970	-0.0627	0.1960	0.0644	-0.1136	-0.1031	0.1887	0.1444

■ 基函数和变换系数之间的关系?

+

6.192 ×



左边为最后图像. 中间是加权函数 (基×系数) 加入到左边图像后更新显示. 右边是当前函数和相应的系数. 图像使用双线性差值扩大10倍后用于显示.