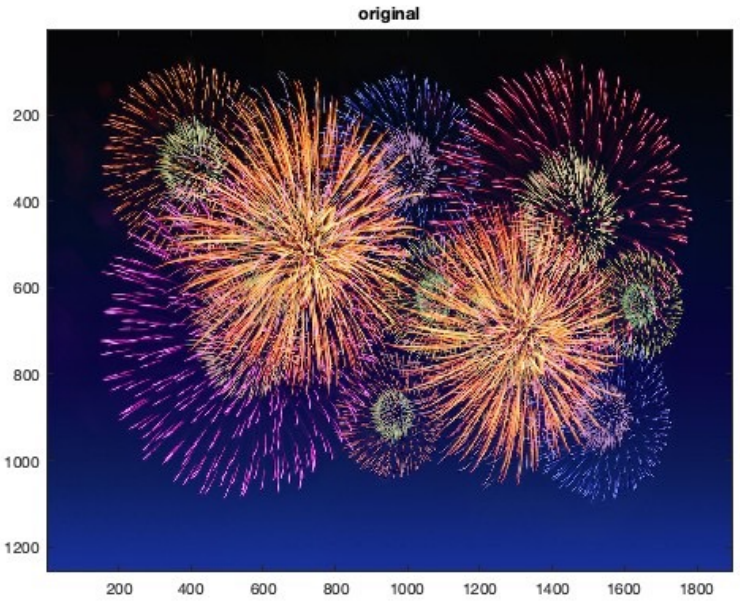
Homework 3 (Due: May 3rd)

(1) Write a Matlab or Python code for the 4:2:0 image compression technique.

B = C420(A), A is the input color image and B is the reconstructed image. Just use the interpolation method for reconstruction. The code should be handed out by NTUCool. (Note: The command rgb2ycbcr cannot be used.)

(25 scores)



(2) Suppose that there is a multipath system $y[n] = x[n] + 0.3x[n\text{-}15] + 0.2x[n\text{-}25]$. (a) Find $p[n]$ such that $y[n] = x[n] * p[n]$. (b) Design the lifter to remove the effect of $p[n]$ and try to not destroy $x[n]$ as possible. (10 scores)

(a) 根據講義 p.193，$y[n] = x[n] + \alpha x[n\text{-}N_p] = x[n] * p[n]$ ， $p[n] = \delta[n] + \alpha\delta[n\text{-}N_p]$

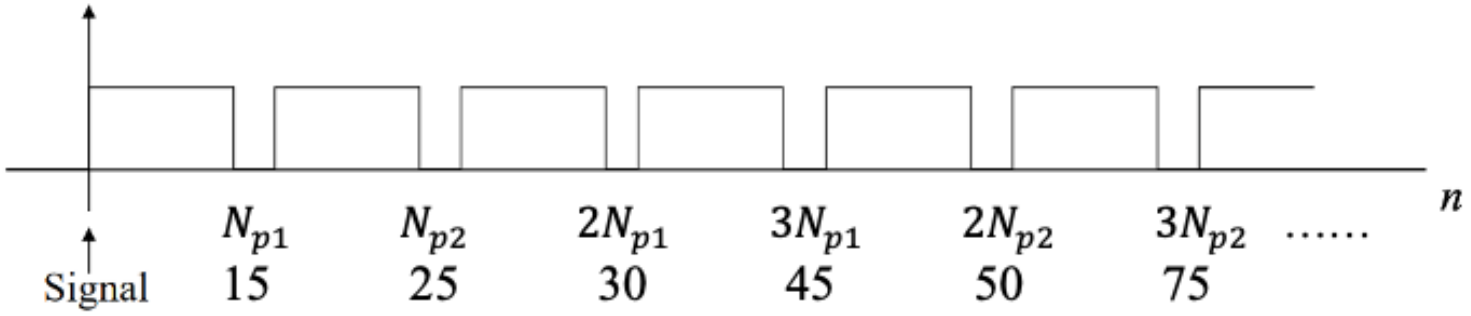
$$\Rightarrow p[n] = \delta[n] + 0.3\delta[n-15] + 0.2\delta[n-25]$$

(b) z transform $\Rightarrow P(Z) = 1 + 0.3Z^{-15} + 0.2Z^{-25}$

$$\hat{P}(Z) = \log(1 + 0.3Z^{-15} + 0.2Z^{-25}) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{0.3^k}{k}Z^{-15k} + \sum_{k=1}^{\infty}(-1)^{k+1}\frac{0.2^k}{k}Z^{-25k}$$

$$Z^{-1} \Rightarrow \hat{p}[n] = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{0.3^k}{k}\delta(n-k\cdot 15) + \sum_{k=1}^{\infty}(-1)^{k+1}\frac{0.2^k}{k}\delta(n-k\cdot 25)$$

Filtering out the echo by the following "lifter":



(3) Suppose that there are three vocal signals: (i) cos(300πt), (ii) -sin(1200πt), (iii) sin(6000πt). (a) Which voice sounds louder? (b) Which voice signal can be propagated to a longest distance? (10 scores)

頻率：

(i) $\cos(300\pi t):150$, (ii) $-\sin(1200\pi t):600$, (iii) $\sin(6000\pi t):3000$

(iii)>(ii)>(i)

波長：

(i)>(ii)>(iii)

(a) 在頻率小於 3000HZ 的情況下，頻率越高人耳聽到聲音所需要的分貝數就越低，(iii)的頻率最大，所以人耳聽到 (iii) 的聲音是最大的。

(b) 波長越長 => 傳播距離較遠，(i)的波長最大，所以 (i) 傳播的距離最遠。

(4) Suppose that for a stringed instrument the frequency of Do is 240 Hz. (a) Determine the frequencies of Mi and So for the instrument. (b) Suppose that the rate of wave propagation is 340m/sec. Determine the lengths of the strings to generate Mi and So for the stringed instrument. (10 scores)

(a) 根據講義 p.240，

$$f_0 \cdot 2^{\frac{k}{12}} HZ,\ f_0 = frequency\ of\ Do, k = 和\ Do\ 差多少個半音$$

Mi 和 Do 差 4 個半音，frequency of Mi : $240 \cdot 2^{\frac{4}{12}} = 302.381052\ HZ$

So 和 Do 差 7 個半音，frequency of So : $240 \cdot 2^{\frac{7}{12}} = 359.593698\ HZ$

(b) $f = \frac{340}{\lambda},\ \lambda = 2L \Rightarrow f = \frac{340}{2L} \Rightarrow L = \frac{340}{2f}$

$$L\ to\ generate\ Mi : \frac{340}{2 \cdot 302.381052} \Rightarrow \frac{0.5622045392}{n}\ m$$

$$L\ to\ generate\ \text{So} : \frac{340}{2 \cdot 359.593698} \Rightarrow \frac{0.4727557823}{n}\ m$$

$$for\ n = 1, 2, 3, \ldots$$

(5) In addition to the DCT, which is adopted by MP3, write at least three possible ways that can compress a music signal more efficiently. (10 scores)

1. 能量只集中在某些特定區域，例如：f_0、$2f_0$、$3f_0$ HZ…，只要記錄這些地方，其他地方可以精簡。
2. 在同一個音當中頻率是穩定的。
3. Repeated melody，大部分的音樂旋律都會重複。

(6) In the JPEG process, (a) why the DCT is used instead of the DFT for transformation? Write at least two reasons. (b) Why the input image is separated into several 8x8 blocks before using the DCT? Write at least two reasons. (c) Why the DC difference is encoded instead of the original DC value? (d) Why zigzag is beneficial for AC term encoding? (20 scores)

(a) 1. near optimal：跟 fourier transform 相比，DCT 能量集中度更好，可以讓能量更集中。
2. real output： DCT 是實數的，而 fourier transform 是複數，要記錄實數與虛數的結果，不利於做壓縮，需要的記憶體量會比較多。

(b) 1. 頻率分布通常會隨著空間改變，而某些部分很有可能只有低頻的成分，該方格會接近常數，經過 DCT 就會變成一個點，所以取整張圖不如取局部的，只記錄該點的值，其他可以忽略。
2. 整張圖一起存會耗費很多暫存記憶體，增加硬體成本，8x8的話需要的暫存記憶體比較少。

(c) 在大部分的情況下 DC difference 會近似於 0，有集中性就利於做壓縮。

(d) 信號大部分都集中在低頻，越高頻為 0 的機率越高，利用 zigzag 的順序再配合 EOB 的話，就可以知道某個地方它後面的數值都為 0 不需要做編碼。

(7) Suppose that $P(x = n) = e^{-\lambda}\lambda^n/(n!)$ for $n = 0, 1, 2, 3\ldots..,$ 40 where $\lambda = 0.97$. Also suppose that length(x) = 50000. Estimate the range of the total coding lengths in the binary system when using (i) the Huffman code and (ii) the arithmetic code. (15 scores)

$$entropy = \sum_{j=1}^{J} P(S_j) \log \frac{1}{P(S_j)} \qquad P\text{: probability}$$

$$= \sum_{n=0}^{\infty} e^{-0.97} \frac{0.97^n}{n!} \ln \left(\frac{1}{e^{-0.97} \frac{0.97^n}{n!}} \right) = 1.2873$$

(i)

$$ceil\left(N \frac{entropy}{\log k} \right) \le b \le floor\left(N \frac{entropy}{\log k} + N \right)$$

$$\Rightarrow ceil\left(50000 \frac{1.2873}{log2}\right) \le b \le floor\left(50000 \frac{1.2873}{log2} + 50000\right)$$

$$\Rightarrow 92860 \le b \le 142859$$

(ii)

$$ceil\left(N \cdot \frac{entropy}{\log k} \right) \le b \le floor\left(N \cdot \frac{entropy}{\log k} + \log_k 2 + 1 \right)$$

$$\Rightarrow ceil\left(50000 \frac{1.2873}{log2}\right) \le b \le floor\left(50000 \frac{1.2873}{log2} + 1 + 1\right)$$

$$\Rightarrow 92860 \le b \le 92861$$

(Extra): Answer the questions according to your student ID number.
(ended with 0, 1, 3, 4, 5, 6, 8, 9)

Q：cepstrum 要取前面幾個 coefficient ?

A：13個， $c_x[1], c_x[2], c_x[3], \ldots\ldots, c_x[13]$