Homework 3 (Due: 5/21st)

(1) Write a Matlab or Python program 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 submitted to ceiba. (Note: The command rgb2ycbcr cannot be used.)

(25 scores)

original image

reconstructed image
RGB:PSNR
39.6419

ruselt

(2) Write two concepts you learned from the oral presentation on 4/30. (10 scores)

1. 第三場的 Hidden Voice Attack：

我覺得這個想法很有趣，的確人聽到雜音跟一般對話能區別的出來，但那是對於我們來說，但對機器來說，這些聲音訊號被處理過後，得到的 result 可能是一樣的，會讓我們以為是誤判，透過這類的方式可攻擊某些 API，相反的說不定可以在 AI model (不一定是聲音) 找出 bug，並修護。

2. 第一場：Laplacian Pyramid Networks

Laplacian Pyramid 蠻常做為圖像的特徵資料，像是 sift、surf 之類的，而這邊類似於當成骨架的概念，可以拿到不同 scale 的資料，來做為高解晰度的資料。

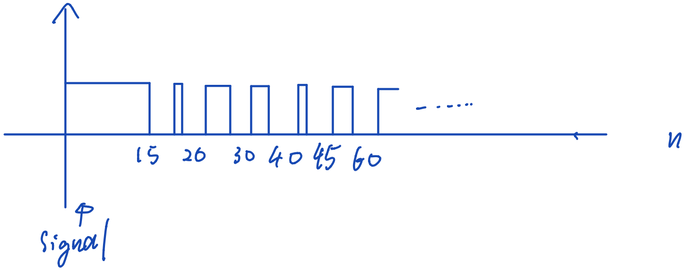
(3) Suppose that $y[n] = x[n] + a_1x[n\text{-}15] + a_2x[n\text{-}20]$. How do we use the lifter to remove the multipath problem and reconstruct $x[n]$ from $y[n]$? (10 scores)

defind $N_{p_1} = 15$, $N_{p_2} = 20$

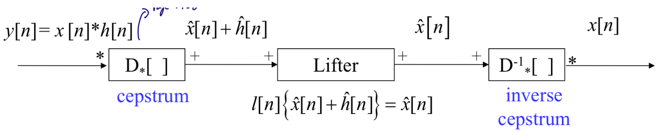
$$P(z) = 1 + a_1 z^{-N_{p_1}} + a_2 z^{-N_{p_2}} \Rightarrow \hat{P}(z) = \log(1 + a_1 z^{-N_{p_1}} + a_2 z^{-N_{p_2}})$$

$$\Rightarrow \sum_{k=1}^{\infty} (-1)^{k+1} \frac{a_1^k}{k} z^{-kN_{p_1}} + \sum_{i=1}^{\infty} (-1)^{i+1} \frac{a_2^i}{i} z^{-iN_{p_2}} \xRightarrow{z^{-1}} \hat{p}[n] = \sum_{k=1}^{\infty} (-1)^{k+1} \left[\frac{a_1^k}{k}\delta[n-k\cdot N_{p_1}] + \frac{a_2^k}{k}\delta[n-kN_{p_2}]\right]$$

$\Rightarrow$ by filter



$\Rightarrow$ 透過下列濾波，上述已計算到Lifter，最後只需將result進 inverse cepstrum 可得到 $x[n]$



(4) Suppose that one of the string of the guitar has the length of 65cm. (a) What are the frequencies of the vocal signal that the string will generate at 15°C? (b) Suppose that the vocal signal generated by the string is Do. What are the string lengths to generate Mi and So? 基頻 261.63 (10 scores)

(a) $15°C$, $V = 340 (m/s)$, $d = 0.65\,m$

以共振觀點出發 $d = \frac{k}{2}\lambda \Rightarrow \lambda = \frac{2}{k}d$

$\because V = f \times \lambda \Rightarrow f = \frac{V}{\lambda} = \overset{170}{\cancel{340}} \cdot \frac{k}{\cancel{2}} \cdot \frac{1}{d} = \frac{170 \cdot k}{0.65} = 261.5385 \cdot k\,(Hz)$

k 為任意整數 $\Rightarrow 261.5385 \cdot k\,(Hz)$, $k = 1, 2, 3, 4 \cdots$

(b) $Mi \rightarrow 2^{\frac{4}{12}}$

$So \rightarrow 2^{7/12}$ $\Rightarrow \because \lambda = \frac{2}{k}d$, $f = \frac{V}{\lambda}$, V 不變

$\hookrightarrow$ mi

$f \propto \frac{1}{\lambda}$

① $f \cdot 2^{4/12} \propto \frac{1}{\lambda} \cdot 2^{4/12} \Rightarrow \frac{1}{1.2599} \times 0.65 \Rightarrow \frac{0.5159}{k}\,m$, $k = 1, 2, 3, 4 \ldots$ #

So

② $f \cdot 2^{7/12} \propto \frac{1}{\lambda} \cdot 2^{7/12} \Rightarrow \frac{1}{1.4983} \times 0.65 \Rightarrow \frac{0.4338}{k}\,m$, $k = 1, 2, 3, 4 \ldots$ #

(5) (a) Discuss <u>three possible ways</u> to compress a music signal more effectively.

(b) Discuss <u>two possible ways</u> to compress a cartoon image more effectively. (10 scores)

(a) ① repeated melody.

② frequency is stable within a beat.

③ interval of a note is a multiple of b/2 (b: beat length)

(b) ① edge information + color in each region.

② edge can be encode using few parameters.

(6) Why it is better to apply the DCT instead of the DFT and the KLT for image compression? (10 scores)

DFT 計算量大還會計算出多餘的高頻成份，而KLT缺點為 dependent on image，且還需一併記錄 transform matrix，而 DCT 在大部分影像而言，DCT 能夠近似KLT，尤其在

$\mathrm{corr}\{f[m,n], f[m+\tau, n+\eta]\} = \rho^{|\tau|}\rho^{|\eta|}$，$\rho \to 1$ 時有 fast Algorithm，

且 ① independent of the input ② near optimal ③ real output，DCT 會比其它兩個 algorithm 好。

(7) What are the advantages to apply the techniques of (a) differential coding for DC terms and (b) zigzag for AC terms in the JPEG process? (10 scores)

(a) differential coding:

由於每個 8x8 區塊的DC值基本上相等，所以前後兩項值求出中間值，就可不必多存一個值，這樣就可達到資料壓縮的效果。

(b) zigzag:

可將低頻參數和高頻參數分別集中在一起，隨著頻率的增加，高頻參數基本為 0，可讓變動長度編碼達到最高的壓縮率。

(8) Suppose that $x[k]$ is a series of natural number and the length of $x[k]$ is 50000. Also suppose that $P(x[k]=n) = (1-e^{-\lambda})\boxed{e^{-\lambda n}}$ for $n = 0, 1, 2$......... where $\lambda = 0.005$. Estimate the range of the total coding lengths in the binary system when using (i) the Huffman code and (ii) the arithmetic code to encode $x[k]$. (15 scores)

$N = 50000$, $K=2$ (Binary),

收斂

$$\because \text{Entropy} = \sum_{j=1}^{J} p(s_j) \ln \frac{1}{p(s_j)} = \sum_{j=1}^{\infty} (1-e^{-0.005}) e^{-0.005n} \cdot \ln\left(\frac{1}{(1-e^{-0.005})e^{-0.005n}}\right)$$

$$\Rightarrow \frac{a}{1-r} \cong 6.2983$$

(i)

$$\left\lceil N \cdot \frac{\text{Entropy}}{\log k} \right\rceil \le b \le \left\lfloor N \cdot \frac{\text{Entropy}}{\log k} + N \right\rfloor$$

$$\Rightarrow \left\lceil 50000 \cdot \frac{6.2983}{\log 2} \right\rceil \le b \le \left\lfloor 50000 \cdot \frac{6.2983}{\log 2} + 50000 \right\rfloor$$

$$\Rightarrow 454328 \le b \le 504327$$

#

(ii)

$$\left\lceil N \cdot \frac{\text{Entropy}}{\log K} \right\rceil \le b \le \left\lfloor N \cdot \frac{\text{Entropy}}{\log K} + \log_K 2 + 1 \right\rfloor$$

$$\Rightarrow \left\lceil 50000 \cdot \frac{6.2983}{\log 2} \right\rceil \le b \le \left\lfloor 50000 \cdot \frac{6.2983}{\log 2} + \log_2 2 + 1 \right\rfloor$$

$$\Rightarrow 454328 \le b \le 454329$$

#