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

B = C420(A), where 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)

Before:

Original Image

After:

reconstructed image

(2) Why the Mel-frequency cepstrum is more suitable for dealing with the acoustic signal than the original cepstrum? (10 scores)

Sol:

(1)Always real.

(2)$\sum|X[k]|^2 B_m[k]$ has much less probability to be 0.

(3)The cut off frequencies of windows math the characteristic of hearing.

(4) DCT is applied instead of the IFT.

(3) Suppose that the cepstrum of $x[n]$ is

$\hat{x}[2]=1 \qquad \hat{x}[n]=0$ otherwise

Please determine $x[n]$. (10 scores)

$\hat{x}[n] \xrightarrow{z\text{ transform}} \hat{X}(z)=\sum_n \hat{x}[n]z^{-n}=1z^{-2}$

$\begin{cases}\hat{x}[2]=1\\ \hat{x}[n]=0\end{cases}$ $\quad \begin{array}{c|c} 1 & \\ \hline z & \end{array}$ $\quad =\dfrac{A(1-1z^{-1})}{(1-2z^{-1})}$

$\Rightarrow \begin{cases}\log A, & n=0\\ -\sum_{k=1}^{m_i}\frac{a_k^n}{n}+\sum_{k=1}^{P_i}\frac{c_k^n}{n}, & n>0\end{cases}$

$\downarrow \exp$

$X(z)=\exp(\hat{X}(z))$

$\exp(x)=\sum_{k=0}^{\infty}\frac{x^k}{k!}$

$=\exp(A)+\exp(1-z^{-1})-\exp(1-2z^{-1})$

$\downarrow z^{-1}$

$x[n]=\begin{cases}\overset{e}{\underset{\|}{\exp(A)}}, & n=0\\ 0, & n<0\\ -\frac{1^n}{n}+\frac{2^n}{n}, & n>0\end{cases}$ #

(4) Suppose that, for a stringed instrument, the frequency of Do is 250Hz. (a) Determine the string length corresponding to Do if the speed of sound at 15°C is considered. (b) What is the string length corresponding to La? (10 scores)

Sol:

a. 假設在 15 度 c，速度為 340 m/s

$$\lambda \times f = 340\ m/s$$

$$\downarrow$$

$$250$$

$$\lambda = \frac{340}{250} = 1.36\ m = 136\ cm$$

b.

Do ~ La

↓ ↓

262 | 440

$$9 \quad 250 \times 2^{\frac{9}{12}} = 420.45$$

$$\lambda \times 420.45 = 340$$

$$\lambda = 0.8087 = 80.87\ cm$$

(5) (a) Why a music signal is easier to compress than other vocal signals? (Write at least 3 reasons) (b) Why a cartoon / mark image is easier to compress than other images? (Write at least 2 reasons) (10 scores)

Sol:

a. (1) Move concentration f0,2f0,3f0,⋯, at the frequencies.

(2) For each note, the frequency is fixed.

(3) Fundamental frequencies are $f_0 \cdot 2^{\frac{k}{12}}$

(4) Beat intervals are $T \cdot 2^k$,k=-1,0,1,2,⋯

(5) Repeated melody

b.

(1) The color/intensity is fixed within a region.

(2) Edges can be approximated by lines or arcs.

(6) Suppose that there are three vocal signals (i) -cos(1200πt); (ii) sin(5400πt); (iii) cos(20000πt). (a) Which one sounds the loudest? (b) Which one is the most suitable to sound? (10 scores)

Sol:

a.(ii)

b.(i)

(7) (a) Why we always use the DCT instead of the DFT and the KLT to image compression? (Write two reasons). (b) Why we apply the 8x8 DCT instead of performing the DCT on the whole image in the JPEG process? (Write three reasons). (10 scores)

Sol:

a.

(1) DCT near optimal.

(2) DCT independent of input compare to KLT.

(3) Real output.

b.

(1)The characteristics of an image vary with the location.

(2)The memory requirement is reduced.

(3)Low complexity $\Theta(MN\log MN)$ reduce $\rightarrow \Theta(MN)$

(8) Suppose that $P(x = \text{'a'}) = 0.45$, $P(x = \text{'b'}) = 0.3$, $P(x = \text{'c'}) = 0.16$, $P(x = \text{'d'}) = 0.06$, $P(x = \text{'e'}) = 0.03$.

(a) What is the entropy of x?

(b) Determine the coding tree of x when using the Huffman code in the binary (二進位) system.

(c) What is the average coding length for each input when using the Huffman code to encode x? (15 scores)

Sol:

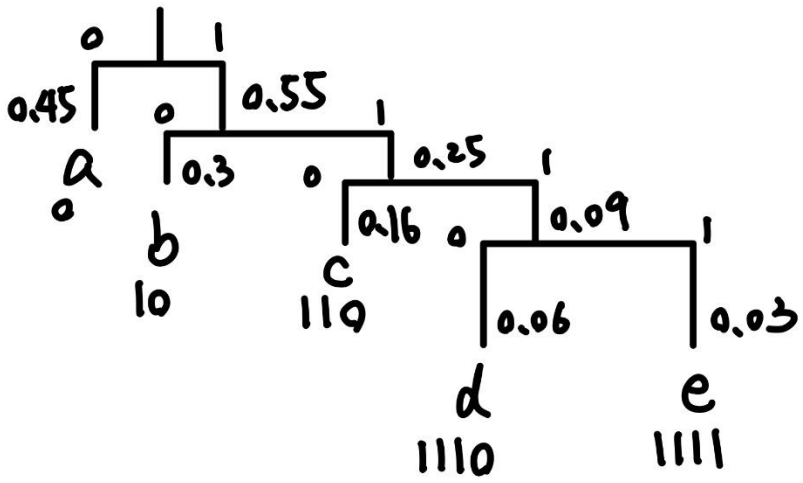
(a)

$X = 0.45 \ln\frac{20}{9} + 0.3 \ln\frac{10}{3} + 0.16 \ln\frac{25}{4} + 0.06 \ln\frac{50}{3} + 0.03 \ln\frac{100}{3}$

=0.3593+0.3612+0.2932+0.1688+0.1052

=1.2877

(b)



(c)假設總共是 100，45*1+30*2+16*3+6*4+3*4=189

189/100=1.89

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

人的耳朵可以聽到的頻率為 20Hz-20000Hz 是以幾分貝為標準呢?

Sol:80dB