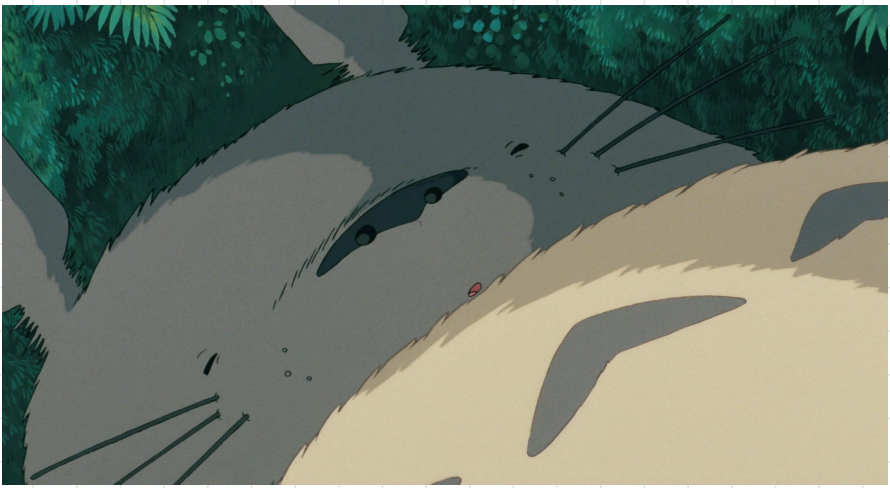
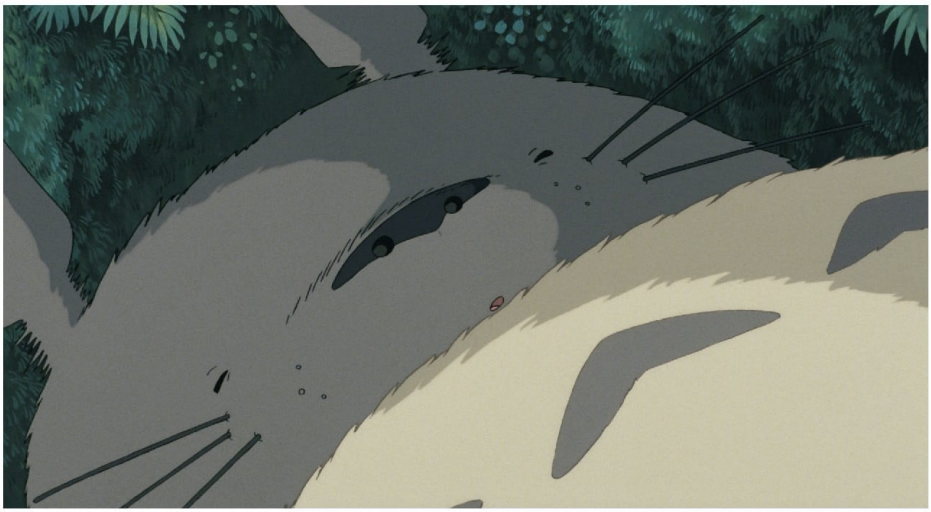
(1) Write a Matlab or Python code for the 4:2:0 image compression technique.

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)

Original image



Reconstructed image



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

Because the cutoff frequencies of windows match the characteristic of hearing. ⇒ The sensitivity of the human ear to frequency is viewed in multiples rather than numerical values.

(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)

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

$X(z)=\sum_{n=-\infty}^{+\infty} x[n]\cdot z^{-n} \Rightarrow \hat{X}(z)=\sum_{n=-\infty}^{\infty}\hat{x}[n]\cdot z^{-n}=\hat{x}[2]\cdot z^{-2}$

$=1\cdot z^{-2}=z^{-2}$

z transform of $\hat{x}[n]$

$\hat{X}(z)=\ln X(z) \Rightarrow z^{-2}=\ln X(z)$

$X(z)=e^{z^{-2}} \Rightarrow e^{X}=e^{z^{-2}}$

The Maclaurin series of e^x is the exponential function

$X(z)=\sum_{n=0}^{\infty}\frac{x^n}{n!}=\sum_{n=0}^{\infty}\frac{z^{-2n}}{n!}=1+z^{-2}+\frac{z^{-4}}{2}+\frac{z^{-6}}{6}$ $\quad x=z^{-2}$

⇒ Let $z^{-2n}=z^{-k}$, $n=\frac{k}{2}$,

$\Rightarrow \sum_{k=0,2,4\cdots}^{\infty}\frac{1}{(\frac{k}{2})!}z^{-k}$

⇒ inverse z transform

$\Rightarrow x[n]=\begin{cases}\frac{1}{\frac{n}{2}!}, & \text{if } n=0,2,4\cdots \\ 0, & \text{if } n=1,3,5\cdots\end{cases}$

(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)

340 m/s

K: 分頻半音

(a) $\lambda = \frac{2L}{n}$, $n \in \mathbb{N}$

$f = \frac{340}{\frac{2L}{n}} = 250\,\text{Hz} \Rightarrow \frac{340n}{2L} = 250 \Rightarrow 2L = \frac{340n}{250}$

$\Rightarrow L = 0.68n$ m

$= 68n$ cm #

, $n = 1, 2, 3 \cdots$

(b) $f_{La} = 250 \times 2^{\frac{9}{12}} = 420.448307\ldots \doteq 420.448$

$f = 420.448 = \frac{340n}{2L} \Rightarrow 2L = \frac{340n}{420.448}$

$L = 0.404n$ m

$= 40.4n$ cm #

$n = 1, 2, 3 \cdots$

(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) 一般化 (10 scores)

(a)

① Concentration f_0, $2f_0$, $3f_0$ .... 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 $T \cdot 2^k$, $k = -1, 0, 1, 2 \cdots$

⑤ repeated melody

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

② edges can be approximated by lines or arcs

600Hz 2700Hz

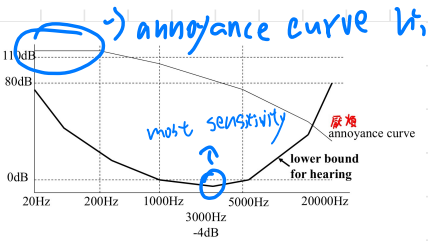
(6) Suppose that there are three vocal signals (i) -cos(1200πt); (ii) sin(5400πt); (iii) cos(20000πt). (a) Which one sounds the loudest? (b) Which one is the most suitable to sound? (10 scores)

10000Hz

1Hz → cos(2πt)

(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)

(a) DCT have:
① Independent of the input
② real output

(b) ① The characteristics of an image vary with the location.
② The memory requirement is reduced
③ Low complexity $O(MN\log MN) \xrightarrow{\text{reduce}} O(MN)$

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

0.16
0.175

(a) What is the entropy of x?

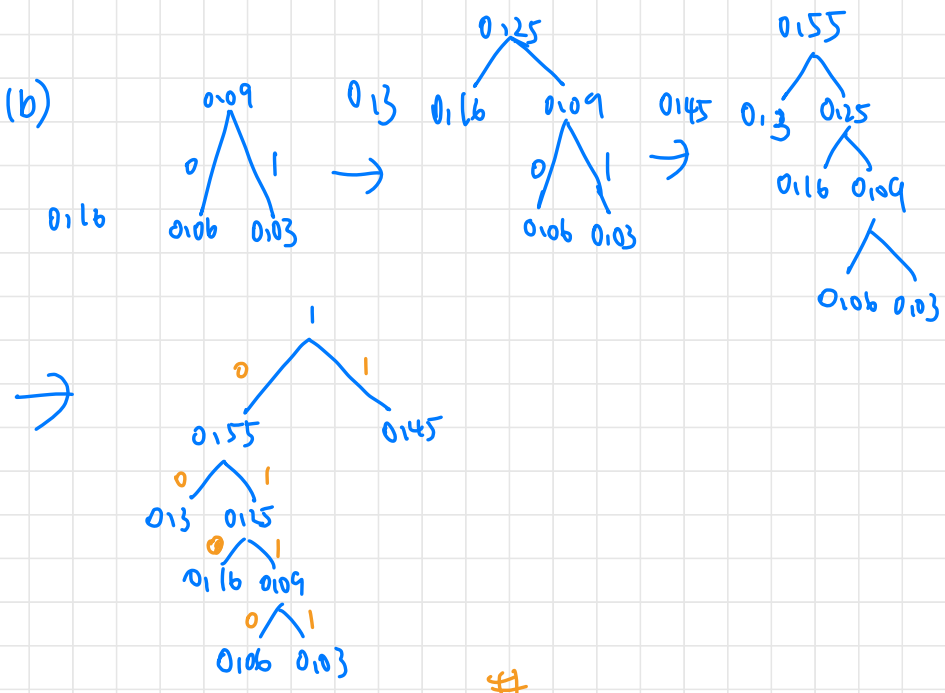
(b) Determine the coding tree of x when using the Huffman code in the binary (二進位) system.

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

entropy $= \sum_{j=1}^{J} P(S_j) \ln \frac{1}{P(S_j)}$

(a) $0.45 \ln \frac{1}{0.45} + 0.3 \ln \frac{1}{0.3} + 0.16 \ln \frac{1}{0.16} + 0.06 \ln \frac{1}{0.06} + 0.03 \ln \frac{1}{0.03} = 1.2877$ #

(b)



(c) mean(L) $= 1 \times 0.145 + 2 \times 0.13 + 3 \times 0.16 + 4 \times (0.06 + 0.03)$

$= 1.89$ #

加分題：學號尾數9

用前面 cx[1], cx[2] ~ cx[13] coefficient 當作語音辨識的特徵