

Lösningförslag till tentamen i
Kompression av ljud och bild
TSBK35

7:e juni 2022

- 1 Se kurslitteraturen
- 2 En Lloyd-Max-kvantiserare är en kvantiserare som minimerar den resulterande medeldistorsionen, givet fördelningen på signalen som ska kvantiseras och antalet rekonstruktionspunkter i kvantiseraren.
- 3 Se kurslitteraturen.

4 Ett bra svar bör innehålla det mesta av nedanstående detaljer:

Bilden delas in i block om 8×8 bildpunkter. Blocken transformeras med en separabel DCT.

Transformkomponenterna kvantiseras likformigt (steglängden kan väljas fritt för var och en av de 64 komponenterna).

Skillnaden mellan DC-komponentens värde och DC-komponentens värde i det föregående blocket kodas med en huffmankod. För att få ett litet mindre alfabet skapar man en kategori genom att beräkna $\lceil \log_2(|d| + 1) \rceil$ för skillnaden d och gör huffmankodningen på denna signal. Extra bitar skickas för att precisera vilket d som kodas.

Övriga 63 transformkomponenter ordnas i zig-zag-ordning, för att få dem i en approximativ frekvensordning. Nollorna skurlängdskodas. Par (skurlängd, nollskild komponent) kodas med en huffmankod. Återigen, för att få ner alfabetsstorleken beräknar man en kategori för de nollskilda komponenterna på samma sätt som görs för DC-komponenten och skurlängderna maximeras till 16. För att koda längre nollskurur använder man helt enkelt flera par. Huffmankoden optimeras för paren (skurlängd, kategori) och extra bitar skickas för att precisera värdena på de nollskilda komponenterna.

5 Se kurslitteraturen.

6 a)

$$-0.8 \cdot \log 0.8 - 0.1 \cdot \log 0.1 - 0.1 \cdot \log 0.1 \approx 0.9219$$

b) En huffmankod för par av symboler ger en kodordsmedellängd på 1.92 bitar/kodord och medeldataakt 0.96 bitar/symbol.

c) Intervallet som hör till sekvensen är $[0.5696 \ 0.57472)$, av storlek 0.00512.

Antalet bitar i kodordet måste vara minst

$$\lceil -\log_2 0.00512 \rceil = 8$$

eventuellt krävs 9 bitar.

Alternativ 1: Det minsta talet med 8 bitar i intervallet är $(0.10010010)_2 = 0.5703125$. Alla tal som börjar med dessa 8 bitar ligger också i intervallet, så det räcker med 8 bitar. Kodordet blir **10010010**.

Alternativ 2: Mittpunkten i intervallet är $0.57216 = (0.10010010011110\dots)_2$. Ta de första 9 bitarna som kodord: **100100100**.

- 7 Transformmatris med sorterade basfunktioner (sorteringen är inte nödvändig för att lösa uppgiften)

$$A = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$$

Korrelationsmatris för signalen

$$R_X = \begin{pmatrix} 0.600 & 0.576 & 0.546 & 0.522 \\ 0.576 & 0.600 & 0.576 & 0.546 \\ 0.546 & 0.576 & 0.600 & 0.576 \\ 0.522 & 0.546 & 0.576 & 0.600 \end{pmatrix}$$

Korrelationsmatris för transformkomponenterna. Varianser för transformkomponenterna i huvuddiagonalen

$$R_\theta = AR_X A^T = \begin{pmatrix} 2.271 & 0 & -0.027 & 0 \\ 0 & 0.081 & 0 & 0.027 \\ -0.027 & 0 & 0.027 & 0 \\ 0 & 0.027 & 0 & 0.021 \end{pmatrix}$$

Varianserna kan förstås också beräknas för en transformkomponent i taget.

Den bittilldelning som minimerar distorsionen är fyra bitar till den första komponenten, en bit till komponent två och inga bitar till komponent tre och fyra. Den resulterande medeldistorsionen blir

$$D \approx \frac{0.009497 \cdot 2.271 + 0.3634 \cdot 0.081 + 0.027 + 0.021}{4} \approx 0.02475$$

vilket ger SNR

$$\text{SNR} = 10 \cdot \log_{10} \frac{0.600}{D} \approx 13.85 \text{ [dB]}$$

- 8 Prediktorkoefficienterna fås som lösning till ekvationssystemet

$$\begin{pmatrix} E\{X_{i-1,j}^2\} & E\{X_{i,j-1} \cdot X_{i-1,j}\} \\ E\{X_{i,j-1} \cdot X_{i-1,j}\} & E\{X_{i,j-1}^2\} \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} E\{X_{i,j} \cdot X_{i-1,j}\} \\ E\{X_{i,j} \cdot X_{i,j-1}\} \end{pmatrix}$$

$$\begin{pmatrix} 6.25 & 5.15 \\ 5.15 & 6.25 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 5.70 \\ 5.50 \end{pmatrix}$$

dvs

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \approx \begin{pmatrix} 0.5821 \\ 0.4003 \end{pmatrix}$$

Prediktionsfelets varians:

$$\sigma_d^2 = 6.25 - (a_1 \ a_2) \begin{pmatrix} 5.70 \\ 5.50 \end{pmatrix} \approx 0.7301$$

Givet att kvantiseringen är fin så kommer prediktionsfelet approximativt att vara normalfördelat. Om vi gör likformig kvantisering följt av perfekt entropikodning blir den resulterande distorsionen

$$D \approx \sigma_d^2 \cdot \frac{\pi e}{6} \cdot 2^{-2R}$$

Givet datatakten $R = 5$ får vi alltså distorsionen

$$D \approx 0.001015$$

och signal-brusförhållandet

$$10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{6.25}{D} \approx 37.90 \text{ [dB]}$$