LINKÖPINGS UNIVERSITET

Lösningsförslag till tentamen i
Kompression av ljud och bild
TSBK35

23:e augusti 2019

1 Vissa symboler (eller sekvenser av symboler) kommer att vara vanligare än andra (i en stokastisk modell motsvaras detta av olika sannolikheter). Genom att ha korta kodord för vanliga symboler och långa kodord för ovanliga symboler kan man få en lägre datatakt än om man använt kodord av samma längd för alla symboler.

De flesta källor har minne (beroende mellan symboler i sekvensen) vilket kan utnyttjas för att få en lägre datatakt än om man inte tar hänsyn till minnet.

2 Se kurslitteraturen

3 Se kurslitteraturen.

4 Se kurslitteraturen

5 Frekvensmaskeringen i hörselsinnet innebär att starka ljud vid en viss frekvens gör att svaga ljud vid närliggande frekvenser inte uppfattas (maskeras).

Vid ljudkodning kan detta utnyttjas genom att man försöker ta bort ljud ur signalen som ändå inte kommer att uppfattas av lyssnaren. På så sätt man kan man få en lägre datatakt utan att den subjektiva ljudkvaliteten påverkas. P.g.a. detta använder de flesta ljudkodare någon form av frekvensbeskrivning av signalen (transform- eller delbandskodning).

6 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 nollskurar använder man helt enkelt flera par. Huffmankoden optimeras för paren (skurlängd, kategori) och extra bitar skickas för att precisera värdena på de nollskilda komponenterna.

7 a) Entropin för en första ordningens markovkälla X_i ges av $H(X_i|X_{i-1})$.

Den stationära fördelningen ges av ekvationssystemet

$$\begin{cases} p(a) = 0.9\cdot p(a) + 0.3\cdot p(b) \\ p(a)+p(b) = 1 \end{cases}$$

med lösningen

$$p(a) = \frac{3}{4}, \quad p(b) = \frac{1}{4}$$

Källans entropi blir då

$$\begin{aligned} H(X_i|X_{i-1}) &= \frac{3}{4}\cdot(-0.9\cdot\log_2 0.9 - 0.1\cdot\log_2 0.1) + \\ & \quad \frac{1}{4}\cdot(-0.7\cdot\log_2 0.7 - 0.3\cdot\log_2 0.3) \\ &\approx 0.5721 \end{aligned}$$

b) För att klara kravet måste vi koda tre symboler i taget. Sannolikheterna för tripplar är

$$p(aaa) = \tfrac{243}{400},\quad p(aab) = \tfrac{27}{400},\quad p(aba) = \tfrac{9}{400},\quad p(abb) = \tfrac{21}{400}$$

$$p(baa) = \tfrac{27}{400},\quad p(bab) = \tfrac{3}{400},\quad p(bba) = \tfrac{21}{400},\quad p(bbb) = \tfrac{49}{400}$$

En huffmankod för denna fördelning kan se ut som

symboler	kodord	symboler	kodord
aaa	0	aab	100
aba	10100	abb	1011
baa	1100	bab	10101
bba	1101	bbb	111

Kodordsmedellängden för denna kod är $\bar{l} = 2.0175$ bitar/kodord, vilket ger datatakten $R = \frac{\bar{l}}{3} = 0.6725$ bitar/symbol.

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

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

Korrelationsmatris för signalen

$$\mathbf{R}_X = \begin{pmatrix} 0.307 & 0.264 & 0.223 & 0.253 \\ 0.264 & 0.307 & 0.264 & 0.223 \\ 0.223 & 0.264 & 0.307 & 0.264 \\ 0.253 & 0.223 & 0.264 & 0.307 \end{pmatrix}$$

Korrelationsmatris för transformkomponenterna. Varianser för transformkomponenterna i huvuddiagonalen

$$\mathbf{R}_\theta = \mathbf{A}\mathbf{R}_X\mathbf{A}^T = \begin{pmatrix} 1.0525 & 0 & -0.0055 & 0 \\ 0 & 0.0895 & 0 & 0.0055 \\ -0.0055 & 0 & 0.0785 & 0 \\ 0 & 0.0055 & 0 & 0.0075 \end{pmatrix}$$

Varianserna kan förstås också beräknas direkt från definitionen.

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

$$D = \frac{0.009497 \cdot 1.0525 + 0.1175 \cdot 0.0895 + 0.1175 \cdot 0.0785 + 0.0075)}{4} \approx 0.009309$$

vilket ger SNR

$$\text{SNR} = 10 \cdot \log_{10} \frac{0.307}{D} \approx 15.18 \text{ [dB]}$$

För att uppnå minst samma SNR utan transform med Lloyd-Max-kvantisering måste vi använda 4 bitars kvantisering, vilket ger SNR$\approx 10 \cdot \log_{10} \frac{0.307}{0.009497 \cdot 0.307} \approx 20.22$ dB. Tre bitars kvantisering kommer inte att räcka, för det ger ett SNR som är 14.62 dB.

9 Prediktorkoefficenterna fås som lösning till ekvationssystemet

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

$$\begin{pmatrix} 121 & 102 \\ 102 & 121 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 109 \\ 112 \end{pmatrix}$$

dvs

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \approx \begin{pmatrix} 0.4166 \\ 0.5745 \end{pmatrix}$$

Prediktionsfelets varians:

$$\sigma_e^2 = 121 - \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} 109 \\ 112 \end{pmatrix} \approx 11.254$$

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

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

och signalbrusförhållandet

$$10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{121}{D}$$

För att få ett SNR som är minst 36 dB, så måste alltså distorsionen vara högst $121/10^{3.6} \approx 0.03039$. För att detta ska vara uppfyllt måste vi välja R enligt

$$R > \frac{1}{2} \log_2 \frac{\pi e \cdot \sigma_e^2 \cdot 10^{3.6}}{6 \cdot 121} \approx 4.52 \text{ [bitar/pixel]}$$