Lösningsförslag till tentamen i

Kompression av ljud och bild

TSBK35

23:e mars 2023

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 **I** Kodas oberoende av andra bilder. Bilden delas upp i block om 8×8 bildpunkter och transformeras (DCT). Transformkomponenterna kvantiseras likformigt och ordnas i zigzag-ordning. Nollorna skurlängdskodas och par (skurlängd, nollskild komponent) kodas med fixa trädkoder.

P Rörelsekompenserad prediktion från närmast tidigare I- eller P-bild på block om 16×16 bildpunkter (makroblock). För varje makroblock skickas en rörelsevektor som talar om positionen på det block man predikterar från. Prediktionsfelet (skillnaden mellan riktiga data och prediktionen) kodas sen med en transformkodare av samma sort som används för I-bilder.

B Rörelsekompenserad prediktion från närmast tidigare och/eller närmast efterföljande I- eller P-bild på block om 16×16 bildpunkter (makroblock). För varje makroblock skickas en eller två rörelsevektorer som talar om positionerna på de block man predikterar från. Prediktionsfelet (skillnaden mellan riktiga data och prediktionen) kodas sen med en transformkodare av samma sort som används för I- pch P-bilder.

I-bilder ger högst datatakt och B-bilder lägst datatakt (givet att man håller kvaliteten konstant). För att enkelt kunna hoppa in mitt i en videoström (t.ex. när man byter kanal på sin digitalbox) så måste man med jämna mellanrum skicka I-bilder. Typiskt låter man var 10:e till var 20:e bild vara en I-bild.

3 a) Se kurslitteraturen.

b) Se kurslitteraturen.

4 Se kurslitteraturen.

5 Se kurslitteraturen.

6 a) Den teoretiskt lägsta datatakten ges av källans entropi:

$$H = -0.5 \cdot \log_2 0.5 - 0.4 \cdot \log_2 0.4 - 0.1 \cdot \log_2 0.1 \approx 1.3610 \text{ bitar/symbol}$$

b) Huffmankoden får en kodordsmedellängd som är 2.78 bitar/kodord, vilket ger datatakten 1.39 bitar/symbol. Jämför med den teoretiska gränsen (entropin).

7 Den avkodade sekvensen är

$$poppoppomomomopopnomopom\ldots$$

och ordboken ser i detta läge ut som

index	sekvens	index	sekvens	index	sekvens
0	m	6	pp	12	$omop$
1	n	7	pop	13	$popn$
2	o	8	ppo	14	no
3	p	9	om	15	$omopo$
4	po	10	mo	16	$om*$
5	op	11	omo		

8 Prediktionen ser ut som
$p_i = a_1 \cdot \hat{X}_{i-1} + a_2 \cdot \hat{X}_{i-2} \approx a_1 \cdot X_{i-1} + a_2 \cdot X_{i-2}$.
Prediktionsfelets varians σ_d^2 ges av

$$\begin{aligned}
\sigma_d^2 &= E\{(X_i - p_i)^2\} \approx \\
&\approx E\{(X_i - a_1 \cdot X_{i-1} - a_2 \cdot X_{i-2})^2\} = \\
&= (1 + a_1^2 + a_2^2) R_{XX}(0) - 2a_1 \cdot R_{XX}(1) - 2a_2 \cdot R_{XX}(2) + 2a_1 a_2 \cdot R_{XX}(1)
\end{aligned}$$

Derivera med avseende på a_1 respektive a_2 och sätt lika med 0, vilket ger oss lösningen

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} R_{XX}(0) & R_{XX}(1) \\ R_{XX}(1) & R_{XX}(0) \end{pmatrix}^{-1} \begin{pmatrix} R_{XX}(1) \\ R_{XX}(2) \end{pmatrix} \approx \begin{pmatrix} 1.2377 \\ -0.2623 \end{pmatrix}$$

$$\sigma_d^2 \approx 0.001107$$

Antag att den aritmetisk kodaren är perfekt, dvs att datatakten är lika med entropin för det kvantiserade prediktionsfelet. Prediktionsfelet kommer approximativt att vara gaussiskt. Detta ger oss distorsionen

$$D \approx \frac{\pi e}{6} \cdot \sigma_d^2 \cdot 2^{-2 \cdot 2.9} \approx 2.8266 \cdot 10^{-5}$$

och signal-brus-förhållandet

$$\text{SNR} = 10 \cdot \log_{10} \frac{0.0307}{D} \approx 30.36 \text{ [dB]}$$

9 En tvåpunkters hadamardtransform ges av transformmatrisen

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

Ett block av bildpunkter ser ut som

$$\mathbf{X} = \begin{pmatrix} X_{00} & X_{01} \\ X_{10} & X_{11} \end{pmatrix}$$

Motsvarande transformerade block blir

$$\mathbf{\Theta} = \begin{pmatrix} \theta_{00} & \theta_{01} \\ \theta_{10} & \theta_{11} \end{pmatrix} = \mathbf{A}\mathbf{X}\mathbf{A}^T =$$

$$= \frac{1}{2} \begin{pmatrix} X_{00} + X_{01} + X_{10} + X_{11} & X_{00} - X_{01} + X_{10} - X_{11} \\ X_{00} + X_{01} - X_{10} - X_{11} & X_{00} - X_{01} - X_{10} + X_{11} \end{pmatrix}$$

Varianserna för de fyra transformkomponenterna är

$$
\begin{aligned}
\sigma_{00}^2 &= E\{\theta_{00}^2\} = \frac{1}{4}E\{(X_{00}+X_{01}+X_{10}+X_{11})^2\} \\
&= R_{XX}(0,0)+R_{XX}(0,1)+R_{XX}(1,0)+\frac{1}{2}R_{XX}(1,1)+\frac{1}{2}R_{XX}(1,-1) = 3785 \\
\sigma_{01}^2 &= E\{\theta_{01}^2\} = \frac{1}{4}E\{(X_{00}-X_{01}+X_{10}-X_{11})^2\} \\
&= R_{XX}(0,0)-R_{XX}(0,1)+R_{XX}(1,0)-\frac{1}{2}R_{XX}(1,1)-\frac{1}{2}R_{XX}(1,-1) = 75 \\
\sigma_{10}^2 &= E\{\theta_{10}^2\} = \frac{1}{4}E\{(X_{00}+X_{01}-X_{10}-X_{11})^2\} \\
&= R_{XX}(0,0)+R_{XX}(0,1)-R_{XX}(1,0)-\frac{1}{2}R_{XX}(1,1)-\frac{1}{2}R_{XX}(1,-1) = 115 \\
\sigma_{11}^2 &= E\{\theta_{11}^2\} = \frac{1}{4}E\{(X_{00}-X_{01}-X_{10}+X_{11})^2\} = \\
&= R_{XX}(0,0)-R_{XX}(0,1)-R_{XX}(1,0)+\frac{1}{2}R_{XX}(1,1)+\frac{1}{2}R_{XX}(1,-1) = 25
\end{aligned}
$$

Kontroll: $(\sigma_{00}^2+\sigma_{01}^2+\sigma_{10}^2+\sigma_{11}^2)/4 = \sigma_X^2 = R_{XX}(0,0) = 1000$

Vi har total $4 \cdot 1.5 = 6$ bitar att fördela på de fyra transformkomponenterna. Medeldistorsionen minimeras om vi kvantiserar θ_{00} med 4 bitar, θ_{01} och θ_{10} med 1 bit vardera och θ_{11} med 0 bitar.

Den resulterande distorsionen blir då

$$
D \approx \frac{1}{4}(0.009497 \cdot \sigma_{00}^2 + 0.3634 \cdot \sigma_{01}^2 + 0.3634 \cdot \sigma_{10}^2 + \sigma_{11}^2) \approx 32.498
$$

och motsvarande SNR är

$$
10 \cdot \log_{10} \frac{1000}{D} \approx 14.88 \text{ [dB]}
$$