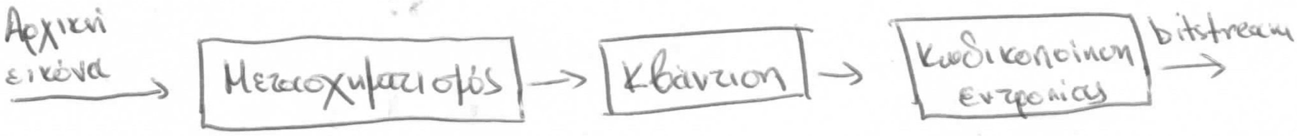
Πολυμέσα

24/1/2015

1) Δώστε το (block) διάγραμμα συμπίεσης εικόνας (ή αλλιώς από ποιά μέρη αποτελείται μια συμπίεση εικόνας ή τι κάνει ο κωδικοποιητής εικόνας).

Λύση:



Μετασχηματισμός: Είναι συνήθως DCT (ή DWT) και μετασχηματίζει την αρχική εικόνα $f(x,y)$ σε $F(u,v)$ (δεκαδικές τιμές)

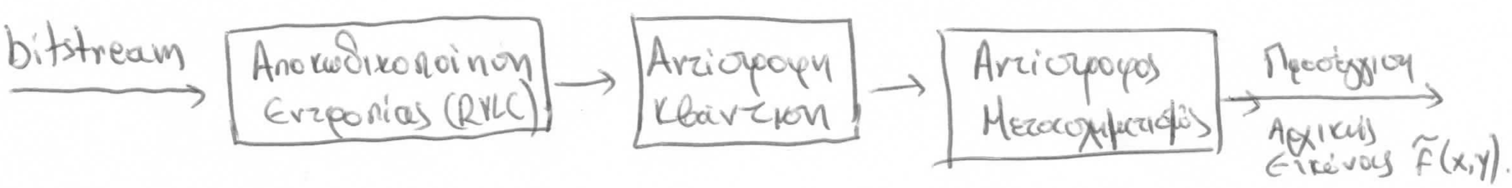
Κβάντιση: Διαιρεί τη μετασχηματισμένη εικόνα, δηλαδή: $\hat{F}(u,v) = \left[\frac{F(u,v)}{Q(u,v)} \right]$ ακέραιες τιμές

όπου $Q(u,v)$ είναι πίνακας κβάντισης.

Κωδικοποίηση εντροπίας (VLC): Αντιστοιχίζει σε κάθε ακέραιο $\hat{F}(u,v)$ μια ακολουθία από bits 0,1.

2) Ποιό είναι το (block) διάγραμμα της αποσυμπίεσης εικόνας. (τι κάνει ο αποκωδικοποιητής)

Λύση:



Αποκωδικοποίηση εντροπίας: από τα bits, επιστρέφει ακεραίους $\hat{F}(u,v)$

Αντίστροφη κβάντιση: κάνει πολλαπλασιασμό $\tilde{F}(u,v) = \hat{F}(u,v) \cdot Q(u,v)$.

Αντίστροφος Μετασχηματισμός: μετατρέπει το $\tilde{F}(u,v)$ σε τιμές φωτεινότητας για τη νέα εικόνα $\tilde{F}(x,y)$.

24/1/2015

②

Πολυμέσα.

3) Σφάλμα μεταξύ των εικόνων $f(x,y)$ και $\hat{f}(x,y)$: $MSE = \frac{1}{N\cdot M}\sum_x\sum_y\left[f(x,y)-\hat{f}(x,y)\right]^2$

$\downarrow$
$N\cdot M$

Ποιότητα $PSNR = 10\cdot\log_{10}\left(\frac{255^2}{MSE}\right)$

Μικρό σφάλμα MSE σημαίνει μεγάλο PSNR, άρα καλή ποιότητα.

(Σ/Λ): Το PSNR αντιστοιχεί <u>πάντα</u> στην ποιότητα της εικόνας.

(Λ): γιατί αν πάρω μια εικόνα και την μετατοπίσω 1 pixel π.χ. δεξιά τότε θα έχει σχετικά μεγάλο MSE, άρα μικρό PSNR ενώ η ποιότητα είναι καλή.

4) <u>ΘΕΜΑ</u>

Ο μετασχηματισμός DCT μιας διάστασης είναι:

$$F(u) = w(u)\cdot\sum_{x=0}^{N-1} f(x)\cdot\cos\left[\frac{(2x+1)\pi u}{N}\right],\quad u=0,1,\dots,N-1 \quad \text{όπου } w(u)=\begin{cases}\frac{1}{\sqrt{N}}, & \text{αν } u=0\\ \sqrt{2/N}, & \text{αν } u=1,\dots,N-1\end{cases}$$

για το $f(x)$, $x=0,1,\dots,N-1$. Αν θεωρήσετε τα διανύσματα:

$f=[f(0)\ f(1)\dots f(N-1)]^T$

$F=[F(0)\ F(1)\dots F(N-1)]^T$.

Δώστε τον πίνακα $N\times N$ G ώστε $F=G\cdot f$ για $N=4$.

<u>Λύση:</u>

$$F(u) \overset{N=4}{=} w(u)\cdot\sum_{x=0}^{3} f(x)\cdot\cos\left(\frac{(2x+1)\pi u}{4}\right)$$

<u>Για $u=0$</u> : $F(0)=w(0)\cdot\left[F(0)\cdot\cos\left(\frac{\pi\cdot 0}{4}\right)+F(1)\cdot\cos\left(\frac{3\pi\cdot 0}{4}\right)+F(2)\cdot\cos\left(\frac{5\pi\cdot 0}{4}\right)+F(3)\cdot\cos\left(\frac{7\pi\cdot 0}{4}\right)\right]$

θα αλλάξει μόνο αυτό

Ομοίως για $u=1,2,3$

Άρα:
$$G=\begin{bmatrix} w(0)\cdot\cos\left(\frac{\pi\cdot 0}{4}\right) & w(0)\cos\left(\frac{3\pi\cdot 0}{4}\right) & w(0)\cdot\cos\left(\frac{5\pi\cdot 0}{4}\right) & w(0)\cos\left(\frac{7\pi\cdot 0}{4}\right)\\ w(1)\cdot\cos\left(\frac{\pi\cdot 1}{4}\right) & w(1)\cdot\cos\left(\frac{3\pi\cdot 1}{4}\right) & w(1)\cdot\cos\left(\frac{5\pi\cdot 1}{4}\right) & w(1)\cos\left(\frac{7\pi\cdot 1}{4}\right)\\ w(2)\cdot\cos\left(\frac{\pi\cdot 2}{4}\right) & w(2)\cdot\cos\left(\frac{3\pi\cdot 2}{4}\right) & w(2)\cdot\cos\left(\frac{5\pi\cdot 2}{4}\right) & w(2)\cos\left(\frac{7\pi\cdot 2}{4}\right)\\ w(3)\cdot\cos\left(\frac{\pi\cdot 3}{4}\right) & w(3)\cdot\cos\left(\frac{3\pi\cdot 3}{4}\right) & w(3)\cdot\cos\left(\frac{5\pi\cdot 3}{4}\right) & w(3)\cos\left(\frac{7\pi\cdot 3}{4}\right)\end{bmatrix}$$

Πολυμέσα

24/1/2015

5) ΘΕΜΑ

Ο μονοδιάστατος DCT υλοποιείται στο πρότυπο H.264 με τον πίνακα $C = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & -1 & 1 & 2 \\ 3 & 2 & 0 & 1 \\ 2 & 1 & 1 & 2 \end{bmatrix}$. Βρείτε πίνακα μετασχηματισμού T που να υλοποιεί δισδιάστατο μετασχηματισμό DCT* σε πίνακα P που να έχει τις τιμές φωτεινότητας

(π.χ. $P = \begin{bmatrix} 123 & 150 & 132 & 100 \\ 100 & 200 & 300 & 400 \\ 0 & 1 & 2 & 3 \\ 4 & 5 & 6 & 7 \end{bmatrix}$)

* με row-column decomposition

Λύση:

Αρκεί να εφαρμόσω το γινόμενο πινάκων: $T = C \cdot P \cdot C^T$

ανάστροφος (γραμμές τις κάνω στήλες)

όπου $C^T = \begin{bmatrix} 1 & 1 & 3 & 2 \\ 1 & -1 & 2 & 1 \\ 2 & 1 & 0 & 1 \\ 2 & 2 & 1 & 2 \end{bmatrix}$

$$T = C \cdot P \cdot C^T =$$

$C \cdot P = \begin{bmatrix} 231 & 362 & \cdots & \\ 31 & -36 & \cdots & \\ \vdots & \vdots & & \\ & & & \end{bmatrix}$

6) (Σ/Λ): Αν δεν υπήρχε κβάντιση τότε θα ανακτούσαμε την αρχική εικόνα χωρίς σφάλματα. (εννοεί: $\tilde{f}(x,y) = f(x,y)$).

(Σ), γιατί κατά την κβάντιση κάνουμε διαίρεση και αποκοπή του κλασματικού μέρους, άρα χάνεται πληροφορία.

Παρατήρηση:

$\hat{F}(u,v) = \left\lfloor \frac{F(u,v)}{Q(u,v)} \right\rfloor$. Όσο μεγαλύτερη η κβάντιση ($Q(u,v)$), τόσο μεγαλύτερη πληροφορία χάνω.

Πολυμέσα

24/1/2015.

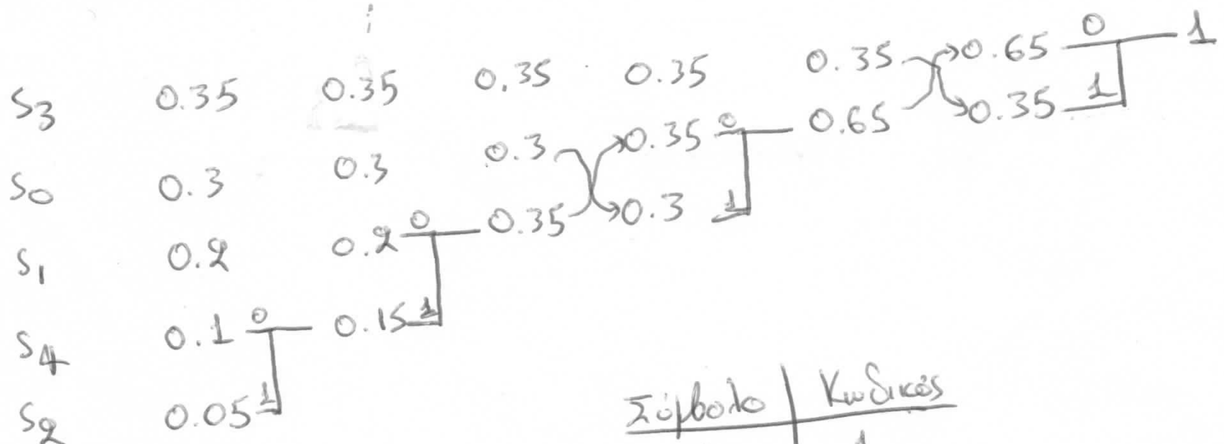
4

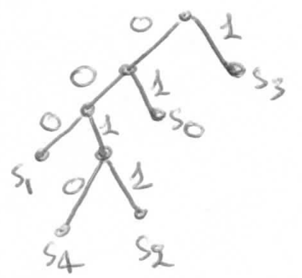
7) Άσκηση (Κωδικοποίηση Huffman)

Έστω τα σύμβολα s_0, s_1, s_2, s_3, s_4 που παράγει μια πηγή πληροφορίας με αντίστοιχες πιθανότητες 0.3, 0.2, 0.05, 0.35, 0.1.
Δώστε τον κώδικα Huffman και επιβεβαιώστε τη γνωστή σχέση μεταξύ εντροπίας και μέσου μήκους κωδικής λέξης.

Λύση:

Διατάσσω τα σύμβολα κατά φθίνουσα σειρά πιθανότητας:





Σύμβολο	Κωδικός
s_3	1
s_0	01
s_1	000
s_4	0010
s_2	0011

Γνωστή σχέση εντροπίας - μέσου μήκους:

$H(X) \le \bar{L} \le H(X)+1$ (1), $H(X)$: εντροπία, $\bar{L}$: μέσο μήκος

$$H(X) = -\sum_i p_i \log_2(p_i) = -\big(0.35\log(0.35) + 0.3\log(0.3) + 0.2\log(0.2) + 0.1\log(0.1) + 0.05\log(0.05)\big) \simeq 1.9$$

$$\bar{L} = \sum p_i l_i = 0.35\cdot 1 + 0.3\cdot 2 + 0.2\cdot 3 + 0.1\cdot 4 + 0.05\cdot 4 = 0.35+0.6+0.6+0.4+0.2 = 2.15 \text{ bits/σύμβολο.}$$

l_i: μήκος i-συμβόλου (πόσα bit έχει)

(1) $\Rightarrow$ Κάνω αντικατάσταση στην (1): $1.9 \le 2.15 < 2.9$

Πολυμέσα

24/1/2015

8) -(Σ/Λ): Ο κώδικας Huffman έχει την προθεματική ιδιότητα (ή αλλιώς είναι στιγμιαία αποκωδικοποιήσιμος).

(Σ), γιατί κάθε κωδικός που δίνει στα σύμβολα δεν είναι πρόθεμα κανενός άλλου κωδικού.

-(Σ/Λ): Μπορεί να βρεθεί κώδικας που να ελαχιστοποιεί το $\bar{L}$ και να έχει την προθεματική ιδιότητα.

(Σ), γιατί η κωδικοποίηση Huffman βρίσκει τέτοιο βέλτιστο κώδικα και έχει την προθεματική ιδιότητα γιατί κάθε κωδικός που δίνει στα σύμβολα δεν είναι πρόθεμα κανενός άλλου κωδικού.

-: Πότε μπορώ να πετύχω $H(x) = \bar{L}$;

ΜΟΝΟ όταν οι πιθανότητες p_i είναι αρνητικές δυνάμεις του 2 π.χ. $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{10}, \ldots$

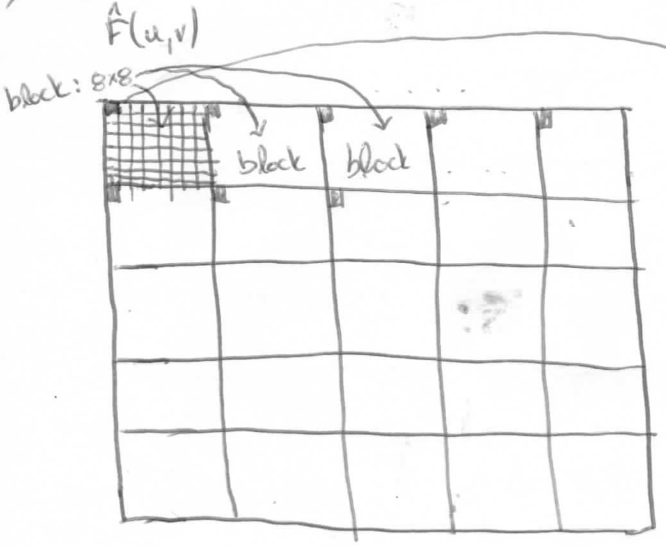
-(Σ/Λ): Αν οι πιθανότητες των συμβόλων είναι ίσες, τότε βέλτιστος κώδικας είναι αυτός που δίνει ίσο αριθμό bits ανά κωδική λέξη.

(Σ) Τότε: $\bar{L} = \lceil \log_2(N) \rceil$, N: πλήθος συμβόλων.

s_0 000
s_1 001
⋮ 010
011
100
101
110
s_7 111

Πολυμέσα

9) Κωδικοποίηση στο JPEG



- λέγεται dc συντελεστής
  είναι σημαντικός γιατί η τιμή του ισούται με τη μέση τιμή φωτεινότητας του block του.

- λέγονται ac συντελεστές
  έχουν πολύ μικρές τιμές (κοντά στο 0) άρα δεν είναι τόσο σημαντικοί.

Κωδικοποίηση DC συντελεστών

π.χ. 125 130 122 118 127

κωδικοποιώ τις διαφορές ανά δύο των dc συντελεστών, δηλαδή:

5 -8 -4 9.

Στην αρχή βάζω το 125

Τελικά: 125 5 -8 -4 9.

Κωδικοποίηση AC συντελεστών:

π.χ.

120	5	4	1	0	3	0	1
3	2	0	1	5	-3	0	0
0	0	0	1	2	-2	0	0
0	1	0	2	-1	0	0	0
1	2	0	-3	0	1	0	0
-2	-1	-2	2	0	1	-1	0
-4	0	-1	0	0	0	0	-2
1	0	1	0	0	1	0	0

block (8x8) δίνεται

*** Σάρωση zig-zag.

Κωδικοποιώ μόνο τους μη-μηδενικούς συντελεστές ως ζεύγη (LEVEL, RUN) (RunLength Encoding)

5 3 0 2 4 1 0 0 0 0 1 1 0 1 0 3 5 1 0
2 -2 . . . . .

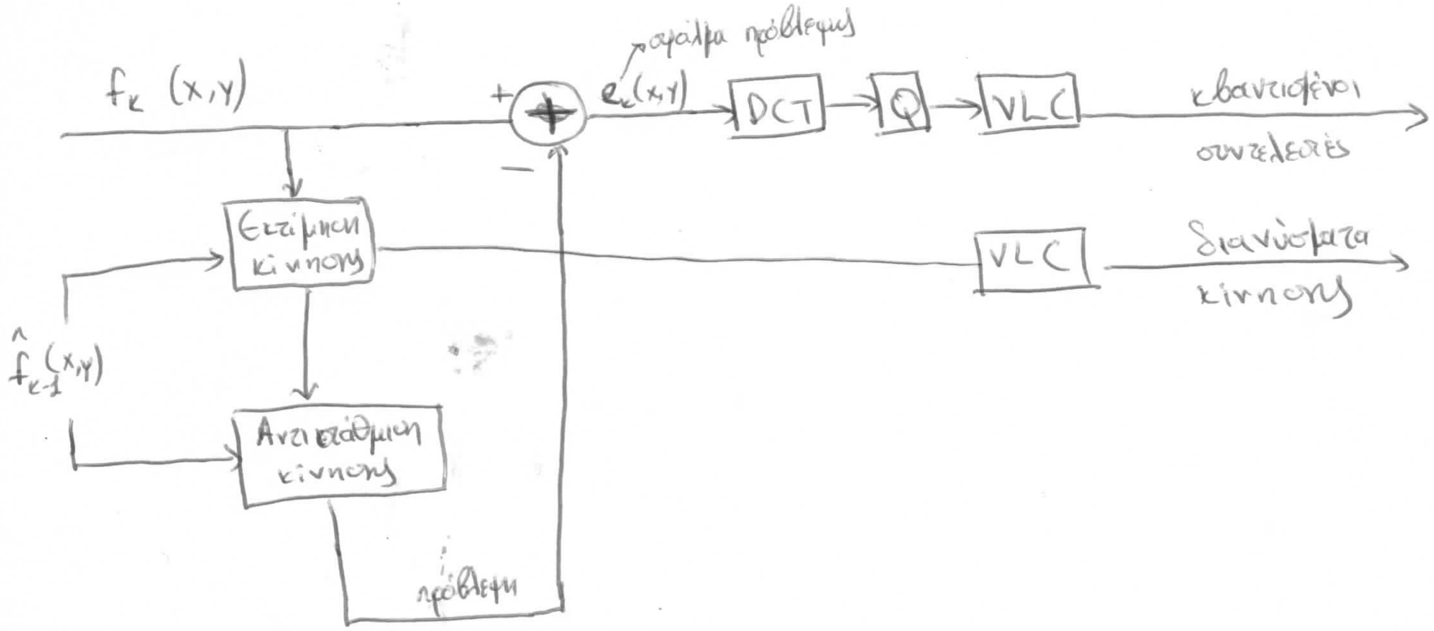
(5, 0) (3, 0) (2, 1) (4, 0) (1, 0) (1, 3)

αριθμός — πόσα μηδενικά προηγούνται 'μεταξύ'

(1, 0) (1, 0) (3, 1) (5, 0) (1, 0)

(2, 1) (-2, 0) . . .

10) Ποιό είναι το block διάγραμμα συμπίεσης video;



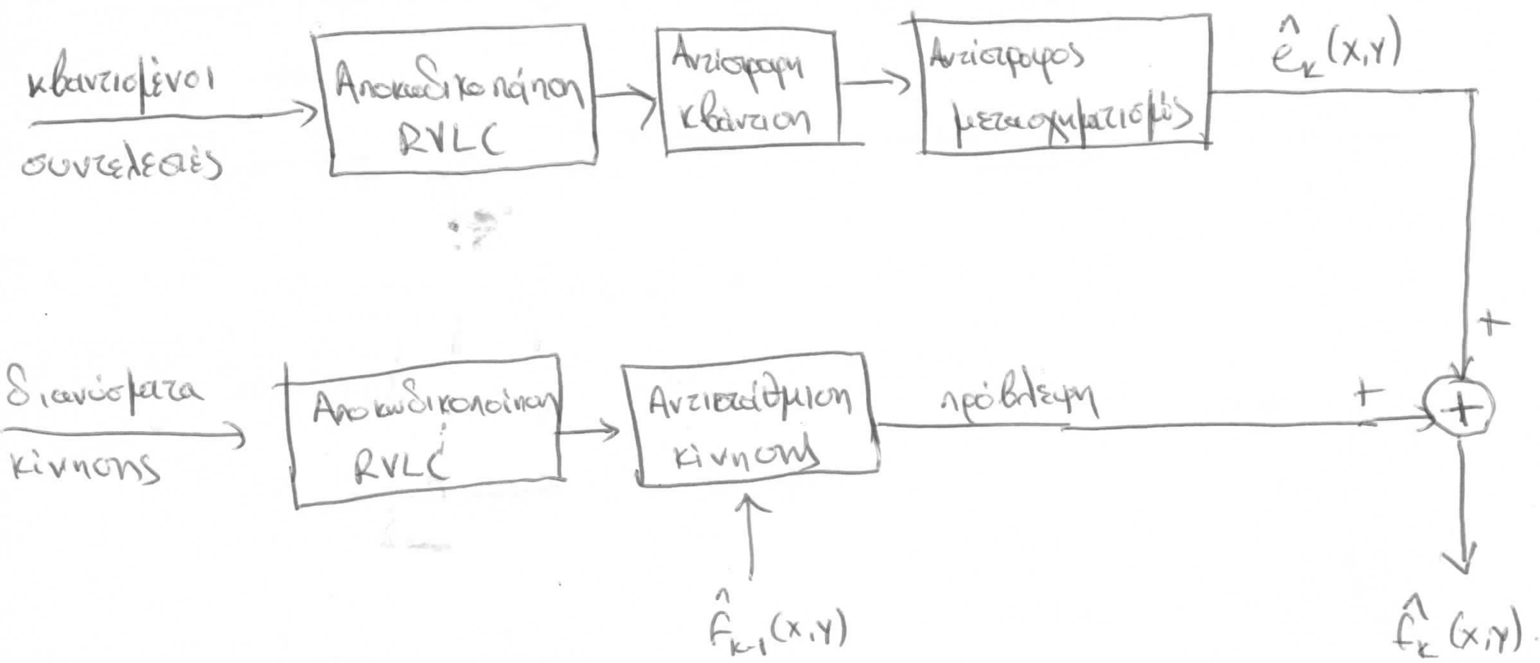
Εκτίμηση κίνησης: υπολογίζει για κάθε macroblock 16x16 το καλύτερο διάνυσμα κίνησης, συγκρίνοντας την εικόνα $f_k(x,y)$ με την προηγούμενή της $\hat{f}_{k-1}(x,y)$

Αντιστάθμιση κίνησης: χρησιμοποιεί τα διανύσματα κίνησης που βρήκε η εκτίμηση κίνησης και την προηγούμενη εικόνα $\hat{f}_{k-1}(x,y)$ για να δημιουργήσει μια πρόβλεψη της εικόνας $f_k(x,y)$.

DCT
Q
VLC

} ίδια με εικόνας

11) Ποιό είναι το block διάγραμμα αποσυμπίεσης video;
(ή αλλιώς αποκωδικοποιητή;)



12) (Σ/Λ) Η κωδικοποίηση απαιτεί περισσότερο χρόνο από την αποκωδικοποίηση.

(Σ) γιατί η κωδικοποίηση απαιτεί πολύ χρόνο λόγω του υπολογισμού των διανυσμάτων κίνησης σε κάθε macroblock. Αντίθετα, η αποκωδικοποίηση τα παίρνει ως είσοδο, άρα δεν τα ξαναυπολογίζει.

(Σ/Λ) Ο κωδικοποιητής περιέχει αποκωδικοποιητή.

(Σ) γιατί πρέπει και οι δύο να έχουν την ίδια είσοδο $\hat{f}_{k-1}(x,y)$ για να μπορεί να γίνει σωστά η αντιστάθμιση κίνησης

Πολυμέσα

24/1/2015

13) Είδη πλαισίων

I: Intra (χωρίς πρόβλεψη)

P: Productive (πρόβλεψη από 1 πλαίσιο)

B: Bidirectional (πρόβλεψη από 2 πλαίσια).

Άσκηση:

Έστω η σειρά κωδικοποίησης των πλαισίων είναι:

0, 2, 1, 4, 3, 6, 5, 8, 7, ...

Τι είδους πλαίσια είναι;

Λύση:

0: I (χωρίς πρόβλεψη γιατί είναι πρώτο).

2: P (με πρόβλεψη από το 0).

1: B (με πρόβλεψη από το 0,2) → επειδή είναι ανάμεσα από 0,2.

4: P (με πρόβλεψη από το 2)

3: B (πρόβλεψη από το 2,4) → επειδή ανάμεσα από 2,4.

Τα άρτια είναι P και τα περιττά B.

14) Ποιά είναι η διαφορά του I με το IDR;

Το IDR δεν επιτρέπει σε πλαίσια που είναι μετά από αυτό να κάνουν πρόβλεψη από πλαίσια που είναι πριν από αυτό. Ενώ το I δεν θέτει περιορισμό.

15) Ποιά είναι η διαφορά του B και B_R; R --> Reference

Το B_R μπορεί να είναι πρόβλεψη για άλλο πλαίσιο, ενώ το B δεν το επιτρέπει.

16) Είδη Πλαισίων

Intra (χωρίς πρόβλεψη): I, IDR

Inter (με πρόβλεψη): P, B, B_R

17) Ανθεκτικότητα σε σφάλματα (error resilience):

Κλιμακωτή κωδικοποίηση
- Υπάρχουν τρία επίπεδα σε ιεραρχία
- Όσο ανεβαίνουμε επίπεδο τόσο μεγαλώνει η προστασία από σφάλματα.
- Χρησιμοποιείται σε ετερογενή δίκτυα
  ↓ διαφορετικές ταχύτητες μετάδοσης.

Κωδικοποίηση πολλαπλών περιγραφών
- Τα επίπεδα δεν έχουν ιεραρχία
- Υπάρχουν πολλές δυνατές διαφορές κωδικοποίησης σε περίπτωση σφάλματος σε κάποια.

Κλιμακωτή (κατηγορίες) → SNR, χρονική, χωρική

Πολυμέσα

18) (Σ/Λ) Ο μετασχηματισμός DWT χρησιμοποιεί ~~...~~ 2 κανάλια (L, H) όταν είναι μονοδιάστατος και 4 κανάλια
Low High

(LL, LH, HL, HH) όταν είναι διδιάστατος. (π.χ. σε εικόνα)

(Σ) γιατί κάνει "wavelet" μετασχηματισμό κατά γραμμή και στο αποτέλεσμα κάνει "wavelet" κατά στήλη.

(Σ/Λ) ~~...~~ Ένα κανάλι που μεταδίδει αθλητικό αγώνα έχει μικρότερο bit~~...~~ rate (ή bitstream) από ένα κανάλι που μεταδίδει ειδήσεις.

(Λ) γιατί ο αγώνας έχει περισσότερη κίνηση, άρα δημιουργείται περισσότερο σφάλμα πρόβλεψης, που σημαίνει ότι απαιτούνται πολλά περισσότερα bits.

(Σ/Λ) Η συχνότητα δειγματοληψίας f_s σε μια τηλεφωνική γραμμή είναι 8000 Hz (ή 8 KHz).

(Σ) Η μέγιστη συχνότητα φωνής είναι $f_{max} = 4000$ Hz και από Θ. Nyquist πρέπει $f_s \geq 2 \cdot f_{max}$ για να γίνει σωστά η δειγματοληψία. Εδώ ισχύει όντως: $8000 \geq 2 \cdot 4000$.

Παρατήρηση:

- Γιατί στα CD η $f_s = 44100$ Hz (ή 44,1 KHz);
  Είναι σωστό γιατί η μέγιστη ακουστική συχνότητα είναι $\boxed{f_{max} = 20.000 \text{ Hz}}$
  Πράγματι, ισχύει ~~...~~ το Θ. Nyquist: $44.100 \geq 2 \cdot 20.000$