



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ,

ΤΜΗΜΑ ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΒΕΣ 04: ΣΥΜΠΙΕΣΗ ΚΑΙ ΜΕΤΑΔΟΣΗ ΠΟΛΥΜΕΣΩΝ

Ακαδημαϊκό Έτος 2005 - 2006, Χειμερινό Εξάμηνο

Θεωρητικές Ασκήσεις (# 3): Ήχος, Εικόνα και Βίντεο

1. Ο διακριτός μετασχηματισμός σνημίτονου (Discrete Cosine Transform) για ένα 8×8 πίνακα ορίζεται από τη σχέση:

$$F(u, v) = C(u, v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right)$$

Όπου $F(u, v)$ είναι ο DCT συντελεστής για τη χωρική συχνότητα (u, v)

$f(x, y)$ είναι η αρχική ένταση φωτεινότητας στη θέση (x, y)

και $C(u, v)$ είναι ο παράγοντας κανονικοποίησης που ορίζεται από:

$$C(0, 0) = \frac{1}{8},$$

$$C(0, v) = C(u, 0) = \frac{1}{4\sqrt{2}}$$

$$C(u, v) = \frac{1}{4} \text{ διαφορετικά}$$

- (a) Υπολογίστε τους συντελεστές DCT για τα πιο κάτω 8×8 τμήματα δεδομένων (data blocks), τα οποία είναι αποσπάσματα μιας εικόνας, και εξηγήστε τις απαντήσεις σας. Μπορείτε να γράψετε ένα μικρό πρόγραμμα (script) για να το κάνει αυτό ή να χρησιμοποιήσετε την εντολή `dct2` στη Matlab. - (4 μονάδες)

(b) Στην πάνω αριστερή εικόνα έχουμε μια κύρια μεταβολή φωτεινότητας κατά την κάθετη κατεύθυνση (λευκό - μαύρο). Επομένως εκτός από το συντελεστή DC ($F(0,0)$) σημαντικός θα είναι και ο συντελεστής $F(1,0)$ (μη μηδενικοί θα είναι και άλλοι συντελεστές $F(u,0)$).

Στη πάνω δεξιά εικόνα έχουμε δύο ουσιαστικές μεταβολές κατά την διαγώνια κατεύθυνση από πάνω αριστερά προς κάτω δεξιά (λευκό - μαύρο - λευκό) άρα εκτός από το συντελεστή DC σημαντικοί θα είναι και οι συντελεστές $F(1,1)$ και $F(2,2)$ (μη μηδενικοί θα είναι και άλλοι συντελεστές $F(u,v)$).

Στην κάτω αριστερή εικόνα έχουμε μια κύρια μεταβολή φωτεινότητας κατά την κάθετη κατεύθυνση (μαύρο - λευκό) και μία κύρια μεταβολή κατά την οριζόντια κατεύθυνση (λευκό - μαύρο). Επομένως εκτός από το συντελεστή DC σημαντικοί θα είναι και οι συντελεστές $F(1,0)$, $F(0,1)$, $F(1,1)$ (μη μηδενικοί θα είναι και άλλοι συντελεστές $F(u,v)$).

Στη κάτω δεξιά εικόνα έχουμε δύο ουσιαστικές μεταβολές κατά την κάθετη κατεύθυνση (μαύρο - λευκό - μαύρο) και την οριζόντια κατεύθυνση (μαύρο - λευκό - μαύρο). Κατά συνέπεια έχουμε και τις αντίστοιχες μεταβολές κατά τη διαγώνια κατεύθυνση. Επομένως εκτός από το συντελεστή DC σημαντικοί θα είναι και οι συντελεστές $F(2,0)$, $F(0,2)$ και $F(2,2)$ (μη μηδενικοί θα είναι και άλλοι συντελεστές $F(u,v)$).

Είναι προφανές ότι το zig-zag scanning δεν είναι υποχρεωτικά η καλύτερη σάρωση των συντελεστών για τη δημιουργία της μεγαλύτερης ακολουθίας μηδενικών συντελεστών όπως φαίνεται στις περιπτώσεις που έχουμε μεταβολές μόνο κατά την κάθετη ή οριζόντια κατεύθυνση.

2. Θεωρείστε ένα σήμα βίντεο που έχει ληφθεί από μια βιντεοκάμερα πεπλεγμένη σάρωσης (interlaced camera) στις 30 εικόνες ανά δευτερόλεπτο (frames per second), σχήμα δειγματοληψίας χρώματος 4:2:0, 10 bits ανά εικονοστοιχείο (pixel) για κάθε κανάλι χρώματος, 480 εικονοστοιχεία ανά γραμμή, 360 γραμμές ανά εικόνα.
- (a) Ποιο είναι το bit rate του σήματος αυτού; - (3 μονάδες)
 - (b) Υποθέστε ότι το πρώτο πεδίο (field) της κάθε εικόνας (frame) του βίντεο έχει αφαιρεθεί. Το βίντεο που προκύπτει είναι σε πεπλεγμένη μορφή (interlaced) ή σε προοδευτική μορφή (progressive); - (4 μονάδες)
 - (c) Ποιο είναι το νέο bit rate; - (3 μονάδες)

Λύση

(a) Η χρωματική δειγματοληψία 4:2:0 έχει τη μορφή που δίνεται στο επόμενο σχήμα. Για κάθε 4 τιμές φωτεινότητας έχουμε από μία τιμή για τις χρωματικές συνιστώσες C_r , C_b . Οι τιμές χρώματος εμφανίζονται στις περιττές γραμμές τις εικόνες. Επειδή το σήμα είναι interlaced τα δεύτερα πεδία αποτελούνται μόνο από τιμές φωτεινότητας (εξ ου και η ονομασία 4:2:0 αντί της 4:1:1) όπως φαίνεται και στο σχήμα.

Οι διαστάσεις των εικόνων των διαφόρων καναλιών είναι:

$Y=360 \times 480$, $C_r=180 \times 240$, $C_b=180 \times 240$. Επομένως έχουμε τον ακόλουθο υπολογισμό bit rate:

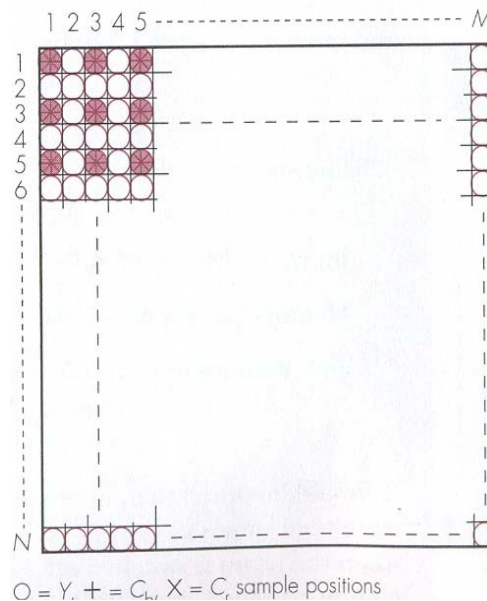
Περιττά fields:

$$15 \text{ fields /sec} \times (180 \times 480 \text{ pixels/ field} + 180 \times 240 \text{ pixels/ field} + 180 \times 240 \text{ pixels/ field}) \times 10 \text{ bits /pixel} = 25\,920\,000 \text{ bps}$$

Άρτια fields:

$$15 \text{ fields /sec} \times (180 \times 480 \text{ pixels/ field}) \times 10 \text{ bits /pixel} = 12\,960\,000 \text{ bps}$$

Άρα συνολικά έχουμε συνολικό bitrate 38 880 000 bps



- (b) Αφαιρώντας το πρώτο πεδίο (field) από κάθε frame του βίντεο τότε το βίντεο που προκύπτει είναι σε προοδευτική μορφή αλλά με μισή χωρική ανάλυση. Θεωρητικά αν αφαιρεθεί το πρώτο πεδίο το σήμα βίντεο θα περιέχει μόνο το κανάλι φωτεινότητας. Αυτό δεν γίνεται στη πράξη. Αντίθετα αφαιρείται μόνο το κανάλι φωτεινότητας του συγκεκριμένου πεδίου (στην ουσία είναι σαν να αφαιρείται το δεύτερο πεδίο). Επομένως το σήμα βίντεο που προκύπτει με την αφαίρεση του ενός πεδίου είναι τύπου 4:2:2 με ανάλυση καναλιών $Y=180 \times 280$, $Cr=180 \times 240$, $Cb=180 \times 240$, και 15 frames. Το αντίστοιχο bit-rate θα είναι:

$$15 \text{ frames/sec} \times (180 \times 480 \text{ pixels} + 180 \times 240 \text{ pixels} + 180 \times 240 \text{ pixels}) \times 10 \text{ bits / pixel} = 25\,920\,000 \text{ bps}$$

3. (a) Απαριθμείστε δύο λόγους που να εξηγούν γιατί χρειαζόμαστε να μεταφέρουμε τα χρώματα από τον RGB χώρο σε άλλους χώρους χρωμάτων (π.χ. YUV). - (6 μονάδες)
- (b) Ο μετασχηματισμός από τον (R, G, B) χώρο χρωμάτων στον (Y, U, V) πραγματοποιείται με τη χρήση της παρακάτω σχέσης πολλαπλασιασμού πινάκων:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.437 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Υπολογίστε την (Y, U, V) αναπαράσταση (0-255) για τις (R, G, B) τιμές {200, 100, 255}. - (3 μονάδες)

Λύση

- (a) Λόγοι που χρειαζόμαστε να μεταφέρουμε τα χρώματα από τον RGB χώρο σε άλλους χώρους χρωμάτων είναι:

- Η ανθρώπινη όραση είναι πιο ευαίσθητη στη φωτεινότητα παρά στα χρώματα με αποτέλεσμα τη μεταφορά των χρωμάτων από τον RGB χώρο σε άλλους χώρους χρωμάτων όπως ο χώρος YCrCb έτσι ώστε να υπάρχει δυνατότητα διαφορετικού χειρισμού της φωτεινότητας

από τα χρωματικά κανάλια (π.χ. διαφορετική ανάλυση, διατήρηση συμβατότητας μεταξύ έγχρωμης και ασπρόμαυρης τηλεόρασης)

- Ο χώρος RGB είναι ευνοϊκός όταν απαιτείται σύνθεση χρωμάτων όπως π.χ. απεικόνιση στην οθόνη υπολογιστή. Αντίθετα σε άλλες περιπτώσεις όπως π.χ. εκτύπωση έχουμε αφαιρετική από το λευκό δημιουργία άρα χρήση του μοντέλου CMYK. Σε εφαρμογές υπολογιστικής όρασης η ανάλυση των χρωμάτων με βάση τη χροιά, κορεσμό, λαμπρότητα (μοντέλο HSB) είναι καταλληλότερη.

$$(b) \begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.437 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \cdot \begin{bmatrix} 200 \\ 100 \\ 255 \end{bmatrix} \Rightarrow \{Y,U,V\}=\{148, 53, 46\} \text{ (Οι τιμές πρέπει να είναι ακέραιες! Άρα το } Y=147.6 \text{ θα γίνει } Y=148)$$

(c) Απαιτούμενη μνήμη = 64 000 bits / sec * 120 sec = 7680000 bits = 960000 bytes = 937,5 Kbytes (1 Kbyte = 1024 bytes)

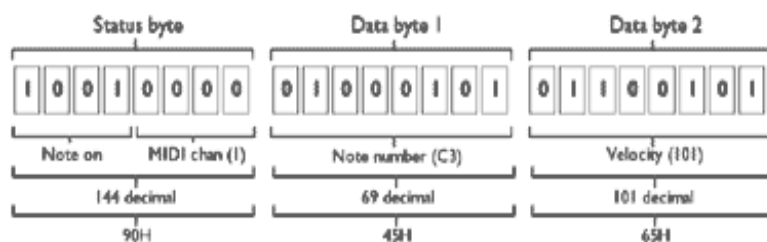
4. Τι είναι το MIDI (γράψτε ένα σύντομο ορισμό); Δώστε ένα MIDI μήνυμα και εξηγήστε τη σημασία του κάθε byte. Ποια είναι τα πλεονεκτήματα και τα μειονεκτήματα του MIDI έναντι του ψηφιακού ήχου (Digital Audio); - (10 μονάδες)

Λύση

MIDI (Musical Instrument Digital Interface):

Είναι ένα πρότυπο για τη μουσική , ανάλογο των διανυσματικών γραφικών. Καθορίζει πως κωδικοποιούνται τα διάφορα στοιχεία μιας παρτιτούρας καθώς και τα όργανα που συμμετέχουν. Περιέχει και πρότυπα για την επικοινωνία μουσικών οργάνων με υπολογιστή.

Περιγραφή δομής MIDI μηνύματος:



Ένα μήνυμα MIDI αποτελείται συνήθως από 3 bytes. Το πρώτο είναι ένα μήνυμα ελέγχου (εντολή) ενώ τα επόμενα δύο bytes είναι στην ουσία δεδομένα (παράμετροι) της εντολής. Τα μηνύματα MIDI γράφονται συνήθως σε δεκαεξαδική μορφή. Το μήνυμα που εικονίζεται παραπάνω είναι: 90 45 65

Αυτό δηλώνει ενεργοποίηση νότας στο κανάλι 1 (πρώτο byte), ο καθορισμός της νότας που θα εκτελεστεί δίνεται στο δεύτερο byte ενώ η ταχύτητα εκτέλεσης δίνεται στο τρίτο byte.

Πλεονεκτήματα έναντι ψηφιοποιημένου ήχου:

- Υπάρχει ευελιξία στην επεξεργασία της μουσικής MIDI.

- Απαιτείται σημαντικά μικρότερος χώρος αποθήκευσης.
- Σε ορισμένες περιπτώσεις η ποιότητα της μουσικής MIDI είναι πολύ υψηλή εξαιτίας της πηγής εισόδου.

Μειονεκτήματα έναντι ψηφιοποιημένου ήχου:

- Το υπολογιστικό κόστος μετατροπής MIDI σε ακουστική κυματομορφή.
 - Το ακουστικό αποτέλεσμα εξαρτάται από τη συσκευή εξόδου και συνήθως είναι υποδεέστερο από τη ψηφιοποιημένη μουσική.
 - Αδυναμία αναπαράστασης ομιλίας σε αποδεκτό βαθμό.
5. Ποιος είναι ο πιο γνωστός μηχανισμός που χρησιμοποιείται για να δείξουμε true color εικόνες (24 bits) όταν χρησιμοποιούμε λιγότερα χρώματα (256); - (6 μονάδες)

Λύση

Ο πιο γνωστός μηχανισμός που χρησιμοποιείται για να δείξουμε true color εικόνες (24 bits) όταν χρησιμοποιούμε λιγότερα χρώματα (256) είναι η χρήση της τεχνικής dithering κατά την οποία κάθε τριάδα συνιστωσών χρωμάτων RGB απεικονίζεται από το πλησιέστερο από τα χρώματα που υπάρχουν σε μια παλέτα χρωμάτων (CLUT=>Color Look Up Table). Η τεχνική αυτή είναι ιδιαίτερα διαδεδομένη στις εικόνες τύπου GIF.

6. Θεωρείστε μια γκριζα εικόνα η οποία δίνεται στο (a) πιο κάτω, όπου η τιμή στο κάθε κουτί δίνει την ένταση φωτεινότητας. Ζωγραφίστε το ιστόγραμμα αυτής της εικόνας. Επίσης, υπολογίστε την ομοιότητα μεταξύ των εικόνων (a) και (b). (Χρησιμοποιώντας ιστόγραμμα, Ευκλείδεια απόσταση και μετρικές διασταύρωσης (intersection) ιστογράμματος) - (10 μονάδες)

1	4	5	8	10	5
9	10	8	7	1	2
3	3	5	4	1	3
3	1	3	6	7	3
3	2	3	10	9	3
1	2	3	4	6	3

(a)

7	4	5	8	10	5
9	9	3	7	1	2
3	3	5	4	1	3
3	5	3	8	7	9
4	5	3	9	9	3
1	5	2	3	7	4

(b)

Λύση

Τα ιστογράμματα των εικόνων A και B φαίνονται στα σχήματα (a) και (b) αντίστοιχα, ενώ τα κανονικοποιημένα ιστογράμματα (συχνότητα εμφάνισης / συνολικός αριθμός pixel) δίνονται στα σχήματα (c) και (d).

Οι τιμές των κανονικοποιημένων ιστογραμμάτων είναι $H_A = \{0.139, 0.083, 0.305, 0.083, 0.083, 0.056, 0.056, 0.056, 0.056, 0.083\}$ και $H_B = \{0.083, 0.056, 0.25, 0.111, 0.167, 0, 0.111, 0.056, 0.139, 0.028\}$

Η Ευκλείδεια απόσταση των ιστογραμμάτων δίνεται από τη σχέση:

$$D(H_A, H_B) = \sqrt{\sum_{i=1}^{NumberOfBins} (H_A(i) - H_B(i))^2} = \sqrt{\sum_{i=1}^{10} (H_A(i) - H_B(i))^2} = \sqrt{(0.139 - 0.083)^2 + \dots + (0.083 - 0.028)^2}$$

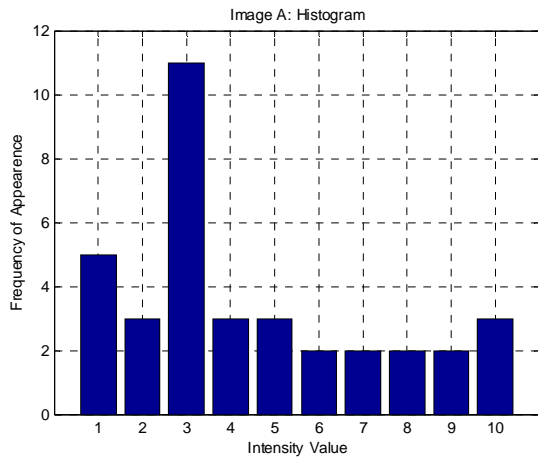
$$= 0.176$$

Η ομοιότητα με βάση την Ευκλείδεια απόσταση δίνεται από τη σχέση $S_{Euclidian} = 1 - D = 0.874$.

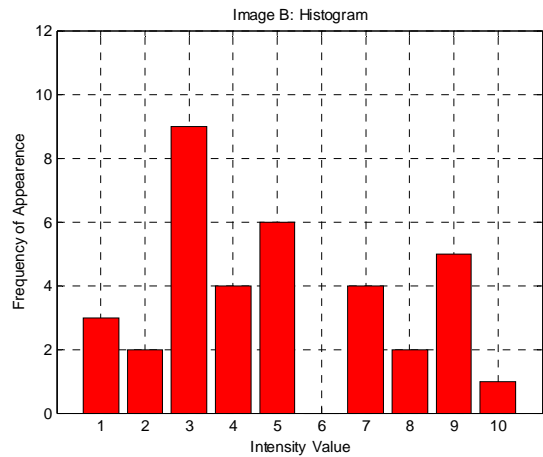
Η διασταύρωση ιστογράμματος (Histogram Intersection) δίνεται από τη σχέση:

$$I(H_A, H_B) = \sum_{i=1}^{NumberOfBins} \min(H_A(i), H_B(i)) = \min(0.139, 0.083) + \dots + \min(0.083, 0.028) = 0.75$$

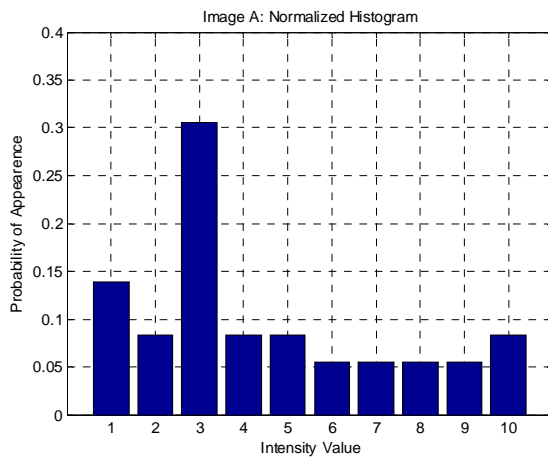
Η ομοιότητα με βάση τη διασταύρωση ιστογράμματος είναι η ίδια η τιμή της διασταύρωσης ιστογράμματος άρα $S_{Intersection} = I = 0.75$



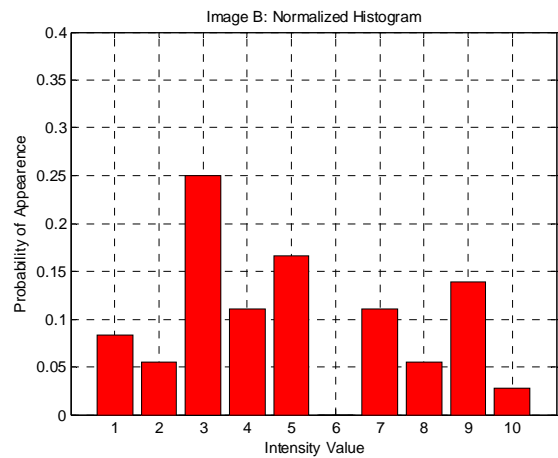
(a)



(b)



(c)



(d)

7. Το JPEG χρησιμοποιεί Διακριτό Μετασχηματισμό Σνημίτονου (DCT) για συμπίεση εικόνας.
 - (a). Ποια είναι η τιμή $F(0,0)$ εάν η εικόνα $f(x, y)$ είναι όπως φαίνεται πιο κάτω; - (3 μονάδες)
 - (b). Ποιος AC συντελεστής $|F(u, v)|$ είναι ο μεγαλύτερος για αυτή την $f(x, y)$; Γιατί; Είναι το $F(u, v)$ θετικό ή αρνητικό; - (9 μονάδες)

20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
80	80	80	80	80	80	80	80
80	80	80	80	80	80	80	80
140	140	140	140	140	140	140	140
140	140	140	140	140	140	140	140
200	200	200	200	200	200	200	200
200	200	200	200	200	200	200	200

Λύση

$$(a) F(u, v) = C(u, v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left(f(x, y) \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right) \right) =$$

$$F(0,0) = C(0,0) \sum_{x=0}^7 \sum_{y=0}^7 \left(f(x, y) \cos\left(\frac{(2x+1) \cdot 0 \cdot \pi}{16}\right) \cos\left(\frac{(2y+1) \cdot 0 \cdot \pi}{16}\right) \right) =$$

$$= C(0,0) \sum_{x=0}^7 \sum_{y=0}^7 (f(x, y))$$

Από την τελευταία σχέση προκύπτει ότι ο συντελεστής $F(0,0)$ αντιστοιχεί στο άθροισμα των τιμών φωτεινότητας της εικόνας διαιρεμένο με το συντελεστή $C(0,0) = 8$. Επομένως με βάση τις τιμές τις εικόνας προκύπτει $F(0,0) = 880$.

(b) Όπως φαίνεται από τη μορφή της εικόνας έχουμε μεταβολές της φωτεινότητας μόνο κατά την κάθετη κατεύθυνση άρα αναμένουμε όλες τους συντελεστές DCT να είναι μηδενική με εξαίρεση τους συντελεστές $F(u,0)$.

Ο συντελεστής $F(1,0)$ αθροίζει θετικά τις πρώτες μισές γραμμές της εικόνας και αρνητικά τις τελευταίες μισές (++++----).

Ο συντελεστής $F(2,0)$ αθροίζει θετικά το πρώτο και το τελευταίο ένα τέταρτο των γραμμών και αρνητικά τις ενδιάμεσες (+-+--+). Με βάση τη μορφή του συγκεκριμένου πίνακα ο συντελεστής $F(2,0) = 0$.

Ο συντελεστής $F(3,0)$ αθροίζει τις γραμμές της εικόνας με βάση την παρακάτω σειρά προσήμων (+--++-+). Οπότε οι πρώτες δύο και οι τελευταίες δύο γραμμές αλληλοαναιρούνται. Η 3^η και 4^η γραμμή προστίθενται αρνητικά και η 5^η και 6^η γραμμή θετικά.

Ο συντελεστής $F(4,0)$ αθροίζει τις γραμμές της εικόνας με βάση την παρακάτω σειρά προσήμων (+---++-+). Με βάση τη μορφή του συγκεκριμένου πίνακα ο συντελεστής $F(4,0) = 0$.

Ο συντελεστής $F(5,0)$ αθροίζει τις γραμμές της εικόνας με βάση την παρακάτω σειρά προσήμων (+---++-+). Οπότε οι πρώτες δύο και οι τελευταίες δύο γραμμές αλληλοαναιρούνται. Η 3^η και 4^η γραμμή προστίθενται θετικά και η 5^η και 6^η γραμμή αρνητικά.

Ο συντελεστής $F(6,0)$ αθροίζει τις γραμμές της εικόνας με βάση την παρακάτω σειρά προσήμων (+---++-+). Με βάση τη μορφή του συγκεκριμένου πίνακα ο συντελεστής $F(6,0) = 0$.

Ο συντελεστής $F(7,0)$ αθροίζει τις γραμμές της εικόνας με βάση την παρακάτω σειρά προσήμων (+---++-+). Με βάση τη μορφή του συγκεκριμένου πίνακα ο συντελεστής $F(7,0) = 0$ σχεδόν μηδενίζεται (οι συντελεστές των προσήμων δεν είναι οδηγούν σε πλήρη αναίρεση).

Από την παραπάνω ανάλυση προκύπτει ότι ο συντελεστής $F(1,0)$ θα έχει τη μέγιστη (κατά απόλυτη τιμή) τιμή επειδή το κάτω μέρος της εικόνας έχει σαφώς μεγαλύτερες τιμές από το άνω μέρος. Για τον ίδιο λόγο ο συντελεστής $F(1,0)$ έχει αρνητική τιμή.

8. Το MPEG ορίζει τα I, B και P frames για να πετύχει συμπίεση. - (20 μονάδες)
- Εξηγήστε τους ορισμούς και τις διαφορές των I, B και P frames.
 - Μπορούμε να προχωρήσουμε γρήγορα μπροστά (fast forward) ή πίσω (backward) για να προσπελάσουμε οποιοδήποτε τυχαίο MPEG frame; Γιατί;
 - Όταν κωδικοποιούμε ένα βίντεο, πως επιλέγουμε να κωδικοποιήσουμε το τρέχον frame σαν I, B ή P frame; Δώστε μερικούς γενικούς κανόνες σχετικά με αυτό.
 - Τι είναι το MPEG-I GoP;
 - Ένα GoP μπορεί να είναι «ανοικτό» ή «κλειστό». Ποια η διαφορά μεταξύ των δύο αυτών τύπων GoP;

Λύση

(a) I frame

- Το είδος των πλαισίων αυτών κάνει χρήση του intra frame - coding. Τα πλαίσια τύπου I είναι τα μόνα που είναι κωδικοποιημένα στο σύνολο τους και η αποκωδικοποίηση μπορεί να γίνει χωρίς αναφορά σε κάποιο άλλο.
- Αποτελούν σημεία αναφοράς κατά την τυχαία προσπέλαση ενός σήματος. Επειδή η παρουσία τους είναι απαραίτητη σαν σημείο χρονικής αναφοράς επιβάλλεται να μεταδίδονται ανά τακτά χρονικά πλαίσια (περίπου κάθε 15 πλαίσια).
- Η διαδικασία της κωδικοποίησης ενός I πλαισίου βασίζεται στην λογική του προτύπου JPEG.

P frame

- Τα πλαίσια τύπου P είναι βασισμένα σε ένα προηγούμενο I ή P πλαίσιο.
- Με την βοήθεια του motion estimation προβλέπουν τη νέα θέση όποιων macroblocks έχουν απλά μετακινηθεί και κωδικοποιούν τον αριθμό του macroblock, ένα διάνυσμα κίνησης και τη διαφορά των μπλοκ (σφάλμα πρόβλεψης).
- Με την σειρά τους μπορούν να αποτελέσουν και αυτά σημείο αναφοράς για επόμενα πλαίσια (P ή B frames) και αυτός είναι ο λόγος που συμβάλλουν στην εισαγωγή και διάδοση σφαλμάτων, αφού η διαδικασία της πρόβλεψης κίνησης δεν μπορεί να είναι απόλυτα ακριβής.
- Συμπιέζονται σε μεγαλύτερο βαθμό από ότι τα I πλαίσια (αν η εκτίμηση κίνησης δεν οδηγεί σε μεγαλύτερη συμπίεση τα P πλαίσια κωδικοποιούνται ως I πλαίσια)

B frame

- Τα πλαίσια τύπου B είναι πλαίσια που δημιουργούνται με εκτίμηση κίνησης από το αμέσως προηγούμενο και επόμενο I ή P πλαίσιο.
- Δεν συντελούν τόσο πολύ στην διάδοση των σφαλμάτων γιατί δεν χρησιμοποιούνται ως σημεία αναφοράς και επιπλέον μειώνουν σημαντικά το σφάλμα πρόβλεψης παίρνοντας τον μέσο όρο από την εκτίμηση κίνησης σε δύο πλαίσια.
- Ο κύκλος ζωής τους περιορίζεται μόνο σε αυτά και δεν επεκτείνεται με το να κληροδοτούν πληροφορίες σε άλλα πλαίσια, κάτι που πολλές φορές σε συνδυασμό και με την υπολογιστική

πολυπλοκότητα που απαιτούν για την κωδικοποίηση και αποκωδικοποίηση τα κάνει μη επιθυμητά από τους κατασκευαστές.

- (b) Τυχαία προσπέλαση σε επίπεδο frame δεν είναι εφικτή γιατί για την αποκωδικοποίηση των B και P frames χρειαζόμαστε κάποιο / κάποια άλλα frames λόγω της αντιστάθμισης κίνησης. Άρα τυχαία προσπέλαση επιτυγχάνεται μόνο με βάση το πλησιέστερο I frame.
- (c) I frames χρησιμοποιούμε σε αλλαγές σκηνών ή σε περιπτώσεις μικρής ομοιότητας ανάμεσα σε διαδοχικά frames. Σε κάθε περίπτωση πρέπει να έχουμε κωδικοποίηση ενός I frame τουλάχιστον κάθε 132 πλαίσια ώστε να δίνεται η δυνατότητα επανασυγχρονισμού του δέκτη.
- P frames χρησιμοποιούμε όταν υπάρχει ομοιότητα ανάμεσα σε διαδοχικά frames αλλά ταυτόχρονα υπάρχει και σημαντική κίνηση επομένως η πολυπλοκότητα της αντιστάθμισης κίνησης είναι μεγάλη για να επεκταθεί σε δύο frames.
- B frames χρησιμοποιούμε όταν υπάρχει μικρή κίνηση ανάμεσα σε διαδοχικά frames ώστε η αντιστάθμιση κίνησης σε δύο πλαίσια να μην απαιτεί υπερβολικό υπολογιστικό φόρτο.
- (d) Το MPEG - I GoP (Group of Pictures) είναι μια αλληλουχία από I, P και B-frames. Αρχίζουν πάντα με I- frame και καταλήγουν σε B ή P frame μετά από τα οποία ακολουθεί υποχρεωτικά I- frame.
- (e) Ανάλογα με το αν ένα GoP καταλήγει σε P ή B frame είναι κλειστό ή ανοικτό αντίστοιχα. Ένα κλειστό GoP μπορεί να αποκωδικοποιηθεί ολόκληρο χωρίς να γίνεται αναφορά σε πλαίσιο από άλλο GoP. Αντίθετα ένα ανοικτό GoP χρειάζεται για την αποκωδικοποίηση του το πρώτο I frame του επόμενου GoP.

9. Υποθέστε ότι κωδικοποιούμε μια σειρά από 18 frames ($F_1, F_2, F_3, \dots, F_{18}$) χρησιμοποιώντας MPEG. Κάθε frame στην σειρά κωδικοποιείται ως I, B ή P σύμφωνα με την ακόλουθη σειρά:

I B B P B I I P P B B P P B I I P I

- (a). Δώστε τη σωστή σειρά μεταφοράς, αποκωδικοποίησης και παρουσίασης. - (4 μονάδες)
- (b). Πόσα frames τουλάχιστον χρειάζεται να αποθηκευτούν σε buffer για να αποκωδικοποιηθεί το MPEG-I βίντεο; Εξηγήστε την απάντησή σας. - (4 μονάδες)

Λύση

- (a) Η σειρά μεταφοράς είναι η σειρά κωδικοποίησης δηλαδή $F_1, F_2, F_3, \dots, F_{18}$. Για την αποκωδικοποίηση πρέπει να είμαστε βέβαιοι ότι για κάθε frame υπάρχουν ήδη αποκωδικοποιημένα όλα τα frame από τα οποία εξαρτάται. Επομένως για την αποκωδικοποίηση ενός P frame χρειαζόμαστε το αμέσως προηγούμενο P ή I frame ενώ για την αποκωδικοποίηση ενός B frame χρειαζόμαστε το αμέσως προηγούμενο P ή I frame και το αμέσως επόμενο P ή I frame. Επομένως η σειρά αποκωδικοποίησης είναι:

I P B B I B I P P P B B P I B I P I

δηλαδή $F_1, F_4, F_2, F_3, F_6, F_5, F_7, F_8, F_9, F_{12}, F_{10}, F_{11}, F_{13}, F_{15}, F_{14}, F_{16}, F_{17}, F_{18}$

Η σειρά παρουσίασης είναι ίδια με τη αρχική σειρά κωδικοποίησης δηλαδή $F_1, F_2, F_3, \dots, F_{18}$

- (b) Για την αποκωδικοποίηση MPEG-1 βίντεο χρειαζόμαστε τουλάχιστον την αποθήκευση δύο frames σε buffer γιατί στη δυσμενέστερη περίπτωση (B frames) είναι πιθανό η κωδικοποίηση ενός frame να εξαρτάται από άλλα δύο (το αμέσως προηγούμενο P ή I frame και το αμέσως επόμενο P ή I frame)