

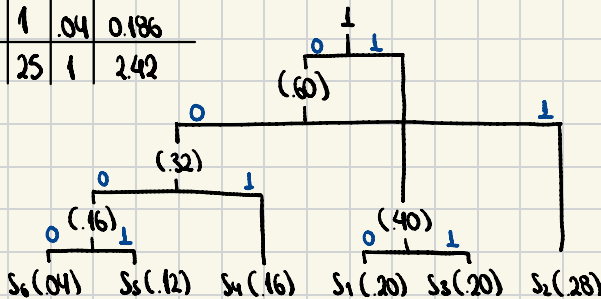
Θέμα 1 (2.6 μονάδες)

1.1 (1.8) Να θεωρήσετε την ακόλουθη εικόνα 5 × 5 με τις αντίστοιχες τιμές pixel. Δώστε τον πίνακα συχνοτήτων για κάθε σύμβολο. Ποιος είναι ο ελάχιστος αριθμός bits ανά σύμβολο για αυτό το μήνυμα; Σχεδιάστε τον κωδικό Huffman για κάθε σύμβολο. Υπολογίστε τον μέσο αριθμό bits ανά σύμβολο.

#	τιμή	f	p	$-p \log_2 p$
1	110	5	.20	0.464
2	120	7	.28	0.516
3	140	5	.20	0.464
4	160	4	.16	0.423
5	170	3	.12	0.367
6	180	1	.04	0.186
Σ		25	1	2.42

180	160	160	140	120
110	110	120	140	120
110	140	120	120	140
120	160	160	170	170
170	120	110	140	110

$H = -\sum p_i \log_2(p_i) = 2.42 \text{ bits/symbol}$
ελάχιστος αριθμός bits ανά σύμβολο = 2.42



τιμή	p	Huffman	N	N*p
180	.04	0000	4	0.16
170	.12	0001	4	0.48
160	.16	001	3	0.48
120	.28	01	2	0.56
110	.20	10	2	0.4
140	.20	11	2	0.4
				2.48

μέσος αριθμός bit ανά σύμβολο: $\bar{L} = 2.48 \text{ bits/symbol}$
ελάχιστος θεωρητικός αριθμός bits: $H = 2.42 \text{ bits/symbol}$ } πολύ αποδοτική συμπίεση

1.2 (0.8) Γιατί οι υπηρεσίες βίντεο χρησιμοποιούν συνήθως υποδειγματοληψία χρώματος 4:2:0;

Οι υπηρεσίες βίντεο χρησιμοποιούν ευρέως την υποδειγματοληψία χρώματος 4:2:0 διότι επιτυγχάνει δραστητική μείωση του όγκου των δεδομένων με ελάχιστη αντιληπτή απώλεια ποιότητας. Αυτό βασίζεται σε μια θεμελιώδη ιδιότητα του ανθρώπινου οπτικού συστήματος: η αντίληψη της φωτεινότητας (luminance - Y) είναι πολύ εντονότερη από αυτή του χρώματος (chrominance - Cb, Cr). Στο σχήμα 4:2:0:

- Η φωτεινότητα δειγματοποιείται πλήρως (ένα δείγμα ανά pixel),
- ενώ οι χρωματικές συνιστώσες υποδειγματοποιούνται με λόγο 1:4 (δηλαδή μία φορά ανά 2x2 pixels).

Αυτό σημαίνει ότι, με τη μισή ή και λιγότερη πληροφορία χρώματος, διατηρείται οπτικά αποδεκτή ποιότητα, ειδικά σε κινούμενη εικόνα όπου η αντίληψη λεπτομερειών είναι περιορισμένη.

Το 4:2:0 προσφέρει ιδανική ισορροπία μεταξύ συμπίεσης και ποιότητας και χρησιμοποιείται σε τυπικά πρότυπα βίντεο όπως MPEG-2, H.264, HEVC, καθώς και σε υπηρεσίες streaming (π.χ. YouTube, Netflix) και ψηφιακή τηλεόραση.

Σχήμα	Δειγματοληψία Y (φωτεινότητα)	Δειγματοληψία Cb/Cr (χρώμα)	Ποιότητα χρώματος	Χρήση	Όγκος δεδομένων
4:4:4	1:1	1:1	Άριστη (χωρίς απώλεια)	Επαγγελματικό βίντεο, post-production	Υψηλός
4:2:2	1:1	1:2 (οριζόντια)	Πολύ καλή	Τηλεοπτική παραγωγή, broadcast	Μέτριος
4:2:0	1:1	1:4 (σε 2x2 μπλοκ)	Καλή (οπτικά αποδεκτή)	Streaming, Blu-ray, video conferencing	Χαμηλός

Θέμα 2 (2.7 μονάδες)

2.1 (0.9) Να θεωρήσετε τον μ/μο DFT των δύο εικόνων που φαίνονται παρακάτω. Ποια είναι η σχέση της φάσης των μετασχηματισμών DFT των δύο εικόνων; Να εξηγήσετε σύντομα.

D.I.P

D.I.P

I1: $f_1(x,y) = f(x,y)$ με διάσταση $M \times N$

I2: $f_2(x,y) = f(M-x, N-y)$ με διάσταση $M \times N$

$$\text{DFT}_1: F_1(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cdot \exp\left(-2\pi j \left(\frac{ux}{M} + \frac{vy}{N}\right)\right) = F(u,v)$$

$$\begin{aligned} \text{DFT}_2: F_2(u,v) &= \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(M-x, N-y) \cdot \exp\left(-2\pi j \left(\frac{ux}{M} + \frac{vy}{N}\right)\right) \stackrel{\substack{x'=M-x \\ x=0 \Rightarrow x'=M \\ x=M \Rightarrow x'=0}}{=} \sum_{x'=0}^{M-1} \sum_{y'=0}^{N-1} f(x',y') \cdot \exp\left(-2\pi j \left(\frac{u(M-x')}{M} + \frac{v(N-y')}{N}\right)\right) \\ &= \sum_{x'=0}^{M-1} \sum_{y'=0}^{N-1} f(x',y') \cdot \underbrace{e^{-2\pi j u}}_{=1 \text{ λόγω περιόδικοτητας}} \cdot \exp\left(-2\pi j \left(\frac{-ux'}{M} + \frac{-vy'}{N}\right)\right) = \sum_{x'=0}^{M-1} \sum_{y'=0}^{N-1} f(x',y') \cdot \exp\left(2\pi j \left(\frac{ux'}{M} + \frac{vy'}{N}\right)\right) = \overline{F(u,v)} \\ &= 1 \text{ λόγω περιόδικοτητας} \end{aligned}$$

$F_2(u,v) = \overline{F_1(u,v)} \Rightarrow \Phi_2(u,v) = -\Phi_1(u,v) \Rightarrow$ ο DFT2 έχει αναθεωρημένη φάση από DFT1.

2.2 (0.9) Ένας μαθητής έγραψε την ακόλουθη πρόταση: "Το φίλτρο ενδιάμεσης τιμής μπορεί να μειώσει τον αριθμό των κατειλημμένων τμημάτων (bins) σε ένα ιστογράμμο. Δεν θα αυξήσει ποτέ τον αριθμό τους." Συμφωνείτε, ή διαφωνείτε, γιατί;

Συμφωνώ με τη δήλωση.

Το φίλτρο ενδιάμεσης τιμής επιλέγει την κεντρική τιμή από τις υπάρχουσες στη γειτονιά, χωρίς να δημιουργεί νέες. Συνεπώς, μπορεί να μειώσει τα bins στο ιστογράμμο, εφόσον κάποιες τιμές εξαφανίζονται, αλλά δεν μπορεί να αυξήσει τον αριθμό τους.

2.3 (0.9) Γιατί τα πρότυπα MPEG-1 και MPEG-2 χρησιμοποιούν τη δομή GOP με περιοδικά πλαίσια I; Για βιντεοδιάσκεψη ή εφαρμογές τηλεφώνων βίντεο, μπορεί ο κωδικοποιητής να εισάγει περιοδικά I-frames;

Τα πρότυπα MPEG-1 και MPEG-2 χρησιμοποιούν τη δομή GOP (Group of Pictures) με περιοδικά I-frames για να επιτυγχάνεται ισορροπία μεταξύ συμπίεσης και προσβασιμότητας. Τα I-frames είναι πλήρως ανεξάρτητα και επιτρέπουν:

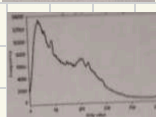
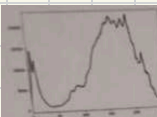
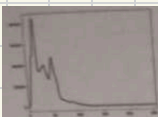
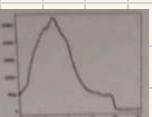
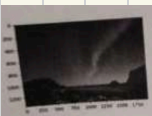
- Τυχαία προσπέλαση (π.χ. χρήση μετακίνησης σε ένα σημείο του βίντεο),
- Επαναφορά σε περίπτωση απώλειας δεδομένων (π.χ. σε μετάδοση).

Σε εφαρμογές όπως βιντεοδιάσκεψη ή τηλεφωνία βίντεο, η συχνή χρήση I-frames δεν είναι ιδανική, επειδή αυξάνει σημαντικά τον ρυθμό μετάδοσης. Ωστόσο, ο κωδικοποιητής μπορεί να εισάγει περιοδικά I-frames, ειδικά μετά από σκληρή αλλαγή, ή για συγχρονισμό σε περίπτωση απώλειας πακέτων.

Η συχνότητα εξαρτάται από τις απαιτήσεις ποιότητας και καθυστέρησης της εφαρμογής.

Θέμα 3 (2 μονάδες)

Να αντιστοιχίσετε τις εικόνες (1)-(4) με τα ιστογράμματα (α)-(δ) δίδοντας σύντομη εξήγηση.



2. πολύ σκοτεινή: ισοχρώμια προς αριστερά

3. πιο φωτεινή: ισοχρώμια προς δεξιά

1. πιο smooth μεταβολές χρωμάτων → πιο smooth ισοχρώμια

4. εικόνα με έντονες μεταβολές → έντονη αντίθεση, δύο κορυφές

Θέμα 4 (2.7 μονάδες)

4.1 (0.9) Η είσοδος σε ένα νευρωνικό δίκτυο (όχι συνελκτικό) είναι μια έγχρωμη εικόνα 300×300 . Το πρώτο κρυμμένο επίπεδο του δικτύου έχει 100 νευρώνες, πλήρως συνδεδεμένους στην είσοδο. Πόσες άγνωστες παραμέτρους έχει το επίπεδο αυτό;

input: RGB image 300×300 $W \times H \times C = 300 \times 300 \times 3 \Rightarrow$ αριθμός εισόδων: $D = W \times H \times C = 270.000$

νευρώνων: $N = 100$, κάθε νευρώνας: $\left\{ \begin{array}{l} D \text{ βάρη} \\ 1 \text{ bias} \end{array} \right\} \Rightarrow$ παράμετροι / νευρώνα = $D + 1$

συνολικό # παραμέτρων = $N \times (D + 1) = 100 \times (270.000 + 1) = 27.000.100$

4.2 (0.9) Στα συνελκτικά δίκτυα η χρήση parameter sharing επιτρέπει τα επόμενα:

- Μείωση του συνολικού αριθμού άγνωστων παραμέτρων
 - Εφαρμογή ανίχνευσης features σε πολλαπλές θέσεις των εικόνων εισόδου
 - Το sharing των παραμέτρων που υπολογίστηκαν για μία κατηγορία εικόνων και σε άλλες κατηγορίες (transfer learning)
- Συμφωνείτε ή όχι; Γιατί;

Συμφωνώ με όλα.

1 **Μείωση παραμέτρων:** Στα CNNs, αντί κάθε φίλτρο να έχει ξεχωριστό βάρος για κάθε pixel εισόδου, χρησιμοποιούμε το ίδιο σύνολο βαρών (kernel) σε όλη την εικόνα. Αυτό μειώνει δραματικά τον αριθμό των παραμέτρων, π.χ. από εκατομμύρια (σε πλήρως συνδεδεμένα) σε λίγες χιλιάδες.

2 **Ανίχνευση features σε πολλές θέσεις:** Επειδή το kernel "σαρώνει" την εικόνα, μπορεί να ανιχνεύσει το ίδιο pattern (π.χ. μια ακμή) ανεξάρτητα από το πού βρίσκεται — κάτι κρίσιμο για εικόνες όπου τα αντικείμενα εμφανίζονται σε διαφορετικές θέσεις.

3 **Μεταφορά μάθησης (Transfer Learning):** Εφόσον τα φίλτρα μαθαίνουν γενικά patterns (π.χ. γραμμές, υφές), μπορούν να επαναχρησιμοποιηθούν και σε άλλες εφαρμογές — ακόμη και σε διαφορετικά dataset. Αυτό κάνει το transfer learning πολύ αποτελεσματικό.

4.3 (0.9) Ένα συνελκτικό δίκτυο εφαρμόζεται σε εικόνες προσώπων και βγάξει στην έξοδο N χαρακτηριστικά σημεία (landmarks) για το εν λόγω πρόσωπο. Πόσες εξόδους έχει το δίκτυο: N^2 , N , $2N$, $3N$;

N landmarks από πρόσωπο
 $\forall$ landmark χρειαζόμαστε (x, y) } $\Rightarrow$ output = $2N$ τιμές