

Συστήματα και Αλγόριθμοι Πολυμέσων

Ιωάννης Χαρ. Κατσαβουνίδης

Ομιλία #3: Αρχές Επεξεργασίας
Σημάτων Πολυμέσων

10 Οκτωβρίου 2005

Πανεπιστήμιο Θεσσαλίας
Τμήμα Μηχ. Η/Υ, Τηλεπ. & Δικτύων

Επανάληψη (1)

- Δειγματοληψία – επανα-δειγματοληψία
 - Τεχνικές φίλτρων (συνέλιξη)
 - $\text{sync}(x) = \sin(x)/x$
 - Παράθυρα για περιορισμό αλληλοκάλυψης φάσματος (aliasing) και κυματισμών (ripples)
 - Τεχνικές παρεμβολής
 - Πρώτου βαθμού (γραμμική)
 - Δευτέρου ή ανώτερου βαθμού
 - Splines

Επανάληψη (2)

- Κβαντοποίηση
 - Quantization. Μη αντιστρεπτή διαδικασία – απώλεια ποιότητας κατά τη φάση της κωδικοποίησης
 - Inverse quantization. Ανάκτηση του αρχικού σήματος κατά τη φάση της αποκωδικοποίησης
 - Ομοιόμορφη κβαντοποίηση: $PSNR = 6N + 11$ (dB)
 - Μη-ομοιόμορφη κβαντοποίηση: Lloyd-Max quantizer, companding
- YUV420 video format

Βασικές τεχνικές (αρχές) επεξεργασίας πολυμέσων

- *Δειγματοληψία*
- *Κβαντοποίηση*
- Μετασχηματισμοί
- Θεωρία Πληροφορίας
- Προ- και Μετά-επεξεργασία πολυμεσικών σημάτων

Μετασχηματισμοί (1)

- Σκοπός είναι η αποσυσχέτιση (de-correlation) του σήματος
- Μετασχηματίζουμε το σήμα από το πεδίο του χρόνου (ήχος) ή το πεδίο των pixels (εικόνα/βίντεο) στο πεδίο των «συχνοτήτων», παράγοντας για κάθε N δείγματα του αρχικού σήματος N συντελεστές μετασχηματισμού
- Πλήρως αντιστρεπτή διαδικασία

Μετασχηματισμοί (2)

- Διακριτοί μετασχηματισμοί :
πολλαπλασιασμός πίνακα επί διάνυσμα
Ευθύς μετασχηματισμός

$$y = Tx$$

Αντίστροφος μετασχηματισμός

$$x = T^{-1}y$$

Μετασχηματισμοί (3)

- Ο πίνακας του μετασχηματισμού T έχει διαστάσεις $N \times N$ και (συνήθως) τις ακόλουθες ιδιότητες

$$\|T\| = 1$$

$$T^{-1} = T^T$$

Μετασχηματισμοί (4)

- Ισχύει η αρχή διατήρησης της ενέργειας (Parseval's Theorem)

$$\|x\| = \|y\|$$

- Αλλά η κατανομή ενέργειας είναι πολύ διαφορετική μεταξύ του αρχικού σήματος και των συντελεστών του μετασχηματισμού: «κέρδος μετασχηματισμού»

Μετασχηματισμοί (5)

- Διακριτός μετασχηματισμός συνημιτόνου (DCT) 8 σημείων:

$$t_u = \frac{C_u}{2} \sum_{x=0}^7 s_x \cos \frac{(2x+1)\pi u}{16}, u = 0..7$$

$$C_0 = \frac{1}{\sqrt{2}}, C_n = 1 \text{ for } n = 1..7$$

Μετασχηματισμοί (6)

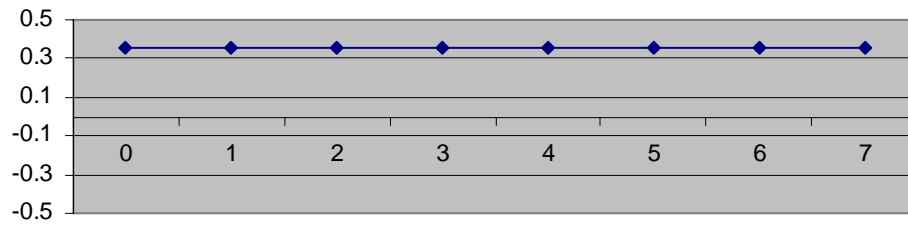
- Αντίστροφος διακριτός μετασχηματισμός συνημιτόνου (DCT) 8 σημείων:

$$s_x = \frac{1}{2} \sum_{u=0}^7 t_u C_u \cos \frac{(2x+1)\pi u}{16}, u = 0..7$$

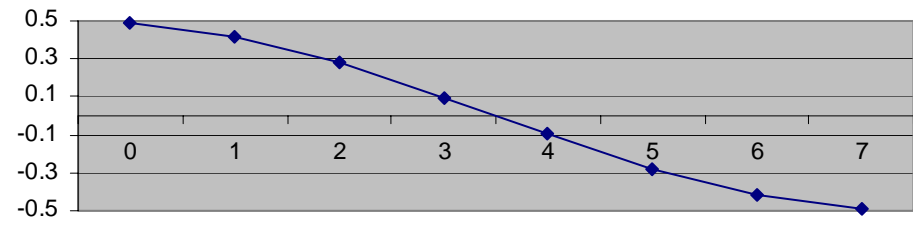
$$C_0 = \frac{1}{\sqrt{2}}, C_n = 1 \text{ for } n = 1..7$$

Μετασχηματισμός DCT 8 σημείων

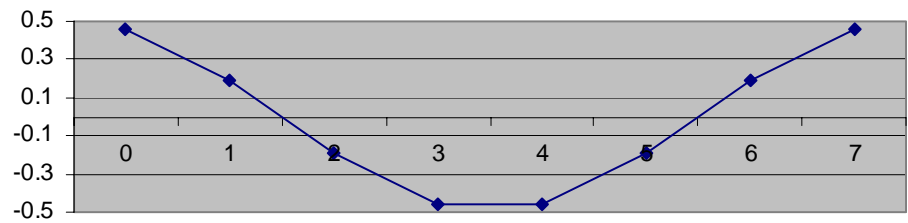
DC-0



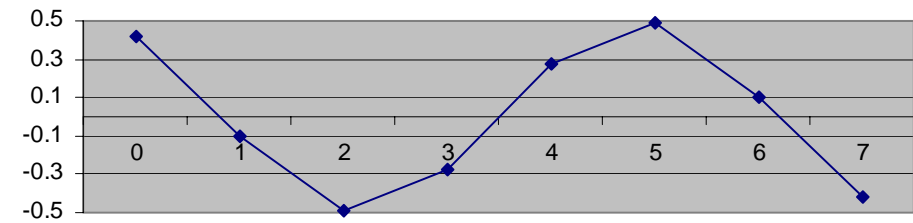
AC-1



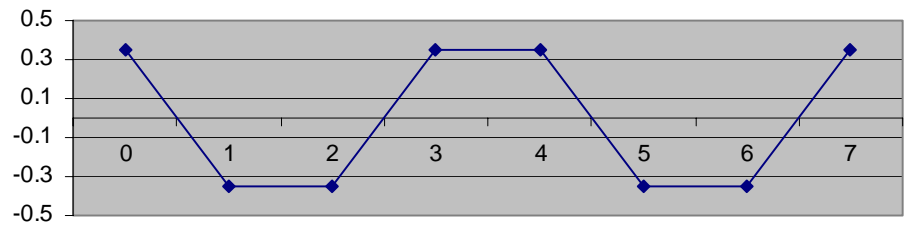
AC-2



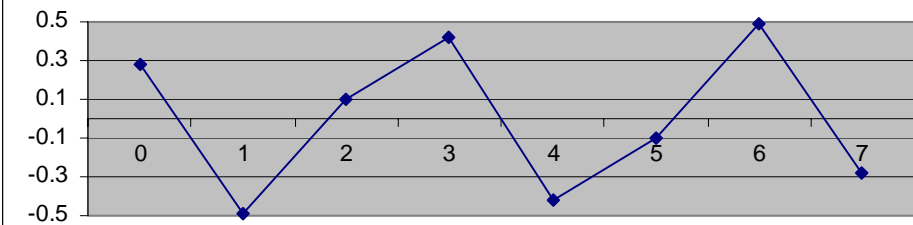
AC-3



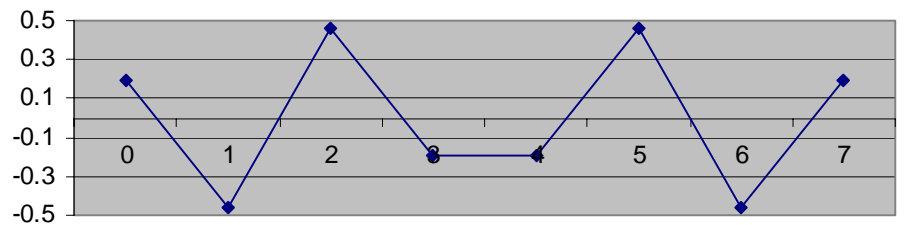
AC-4



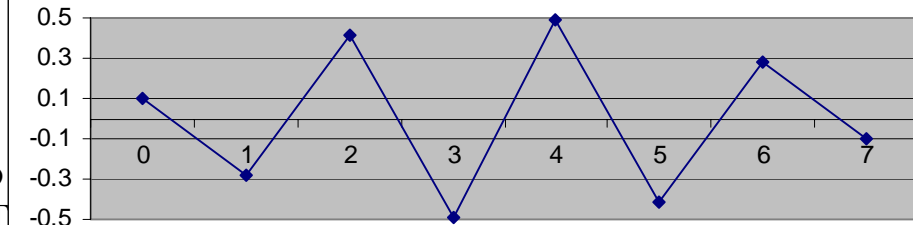
AC-5



AC-6



AC-7



μο
, T

«Κέρδος Μετασχηματισμού» (1)

- Υποθέτοντας ομοιόμορφη κβαντοποίηση N bits τυχαίας μεταβλητής με διασπορά σ , η διασπορά του σφάλματος κβαντοποίησης σ_q δίνεται από τη σχέση

$$\sigma_q = \sigma \cdot 2^{-N}$$

«Κέρδος Μετασχηματισμού» (2)

- Εάν έχουμε 2 τυχαίες μεταβλητές με διασπορά σ_1 και σ_2 τις οποίες κβαντίζουμε με χρήση N_1 και N_2 bits αντίστοιχα, τότε η διασπορά του σφάλματος κβαντοποίησης δίνεται από τις σχέσεις

$$\sigma_{q1} = \sigma_1 \cdot 2^{-N_1} \quad \text{και} \quad \sigma_{q2} = \sigma_2 \cdot 2^{-N_2}$$

«Κέρδος Μετασχηματισμού» (3)

- Συνεπώς, εάν επιθυμούμε να ελαχιστοποιήσουμε τη συνολική ενέργεια του σφάλματος κβαντοποίησης, υπό τη συνθήκη ότι ο συνολικός αριθμός bits είναι N , τότε έχουμε το πρόβλημα βελτιστοποίησης υπό συνθήκη

$$\min(\sigma_{q1}^2 + \sigma_{q2}^2) = \min(\sigma_1^2 \cdot 2^{-2N_1} + \sigma_2^2 \cdot 2^{-2N_2})$$

$$N_1 + N_2 = N$$

«Κέρδος Μετασχηματισμού» (4)

- Μπορούμε να δείξουμε ότι το ελάχιστο σφάλμα επιτυγχάνεται όταν

$$\frac{\sigma_1}{\sigma_2} = \frac{2^{N_1}}{2^{N_2}} \Rightarrow 2^N \frac{\sigma_1}{\sigma_2} = 2^{2N_1} \Rightarrow N_1 = \frac{N}{2} + \log_2 \frac{\sigma_1}{\sqrt{\sigma_1 \sigma_2}}$$

- Όταν $\sigma_1 = \sigma_2$, $N_1 = N_2 = N/2$
- Όταν $\sigma_1 = 0$, $N_1 = 0$, $N_2 = N$
- Όταν $\sigma_1 = 4^* \sigma_2$, $N_1 = N/2 + 1$, $N_2 = N/2 - 1$

«Κέρδος Μετασχηματισμού» (5)

- Έτσι λοιπόν, το ζητούμενο του μετασχηματισμού είναι να αποσυσχετίσει τις τυχαίες μεταβλητές, ούτως ώστε να «μετατοπίσει» ενέργεια σε ορισμένους μόνο συντελεστές
- Π.χ. Εάν $\sigma_1 = \sigma_2 = 7.07$ πριν από το μετασχηματισμό, και $N = 10$ bits, $E(q) = 3.125$
Εάν μετά το μετασχηματισμό $\sigma_1 = 9$ και $\sigma_2 = 3$,
τότε $N_1 = 5.79$, $N_2 = 4.21$ και $E'(q) = 1.95$
Το «κέρδος μετασχηματισμού» είναι 2 dB

Πολυδιάστατοι Μετασχηματισμοί

- Διαχωρίσιμοι μετασχηματισμοί
 - Εφαρμόζονται πρώτα κατά τη μία διάσταση και εν συνεχεία κατά τη δεύτερη κ.ο.κ.
- Ο πίνακας ενός πολυδιάστατου μετασχηματισμού μπορεί να εκφραστεί ως καρτεσιανό γινόμενο των αντίστοιχων πινάκων μονοδιάστατου πίνακα
- Τα ενδιάμεσα αποτελέσματα – το σήμα εξόδου από τον μετασχηματισμό, το οποίο αποτελεί το σήμα εισόδου για το δεύτερο μετασχηματισμό – μπορούν να αποθηκευτούν στον ίδιο χώρο με το αρχικό σήμα εισόδου (in-place computation)

Γρήγοροι μετασχηματισμοί (1)

- Βασίζονται στα υπολογιστικά σχήματα των Cooley & Tukey για τον γρήγορο διακριτό μετασχηματισμό Fourier (FFT)
- Η ιδέα είναι η εύρεση κοινών παραγόντων στους συντελεστές του πίνακα του μετασχηματισμού κατά τρόπο που να μειώνεται η υπολογιστική πολυπλοκότητα με τη χρήση ενδιάμεσων μεταβλητών.

Γρήγοροι μετασχηματισμοί

- Πολυπλοκότητα πολλαπλασιασμού πίνακα επί διάνυσμα (N): N^2 MACs.
- Πολυπλοκότητα γρήγορου μετασχηματισμού N σημείων: $N \log N$
- FDCT: Χρησιμοποιείται από τους codecs JPEG, MPEG1/2/4 Video (8x8), όπως επίσης και σε τροποποιημένη μορφή από τον codec MPEG1-Audio (“MP3”) – 12 ή 36 σημείων

Μετασχηματισμός DCT

- Χρησιμοποιεί συντελεστές που δεν είναι ρητοί, και συνεπώς εισάγονται υπολογιστικά σφάλματα κατά την υλοποίησή του σε αριθμητική σταθερής υποδιαστολής.
- Ο αντίστροφος μετασχηματισμός (IDCT) υπόκειται σε συγκεκριμένους περιορισμούς σχετικά με τον αριθμό των bits που χρησιμοποιούνται για την αποθήκευση των συντελεστών και των ενδιάμεσων αποτελεσμάτων, σύμφωνα με το διεθνές στάνταρ IEEE-P1180/D2 (1990)

Άλλοι μετασχηματισμοί σημάτων πολυμέσων (1)

- Wavelets (JPEG2000)
 - Επέκταση της θεωρίας πολυφασικών φίλτρων
 - Υπάρχουν πολλές «οικογένειες» μετασχηματισμών Wavelets (Haar, Daubechies, bi-orthogonal, κλπ.)
 - Χαρακτηρίζονται από 2 ορθογώνια μεταξύ τους φίλτρα
 - ένα χαμηλοπερατό και ένα υψηλοπερατό.
 - Έχουν πεπερασμένο αριθμό μη-μηδενικών σημείων
 - Εφαρμόζονται σε μορφή πυραμίδας, και γι' αυτό το λόγο έχουν πολυπλοκότητα $O(N)$

Άλλοι μετασχηματισμοί σημάτων πολυμέσων (2)

- Hadamard (H.264)
 - Αναδρομική κατασκευή με τη βοήθεια της σχέσης
$$\begin{bmatrix} H & H \\ H & -H \end{bmatrix}$$

Όπου

$$H_1 = [1] \quad H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Μετασχηματισμός Hadamard

- Δεν χρησιμοποιεί πολλαπλασιασμούς αλλά μόνο προσθέσεις, αφαιρέσεις και δυαδικές ολισθήσεις
- Έχει πολύ καλές ιδιότητες συγκέντρωσης ενέργειας για σήματα υψηλής συσχέτισης, όπως τα σήματα βίντεο και εικόνες
- Το μέγεθος του μετασχηματισμού που χρησιμοποιεί το στάνταρ συμπίεσης είναι 4×4 και 8×8 (μόνο το υψηλό προφίλ)

Μετασχηματισμοί σημάτων πολυμέσων

- Εφαρμόζονται (συνήθως) σε κάποιο υποσύνολο του σήματος, όπως για παράδειγμα κομμάτια 8x8 της εικόνας (JPEG, MPEG1/2/4) ή 12/36 δείγματα ήχου (MP3) ή 4x4 κομμάτια της εικόνας (H.264)
- Για να μειωθεί το φαινόμενο του σφάλματος στα όρια των κομματιών, εφαρμόζονται τεχνικές είτε αλληλοκάλυψης (MP3) είτε μετα-επεξεργασίας φιλτραρίσματος των ορίων (de-blocking filter – H.264)

Άσκηση

- Να χωρίσετε το πρώτο καρέ του αρχείου src13 σε υποπίνακες 8×8 . Στη συνέχεια να εφαρμόσετε το διακριτό μετασχηματισμό DCT και να υπολογίσετε την ενέργεια των 64 συντελεστών για ολόκληρη την εικόνα.